New Directions in Advanced RFID Systems

DISSERTATION SUBMITTED TO

THE SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

OF THE UNIVERSITY OF ADELAIDE

BY

Damith Chinthana Ranasinghe

IN FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

January 2007

Declaration of Originality

I declare that this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge it does not contain any material previously published or written by any other person except where due reference is made in the text.

I give consent to this copy of my thesis, when deposited in The University of Adelaide's Library, being available in all forms of media, now or hereafter known.

Damith Chinthana Ranasinghe
Department of Electrical and Electronic Engineering
The University of Adelaide
1 January 2007.

Acknowledgments

The author wishes to express his sincere appreciation and deepest felt gratitude to Prof. Peter H. Cole for his kindness, continued support, encouragement, and foremost for giving me the opportunity to work with him. I have been touched by his tireless enthusiasm from the first time I visited the early Adelaide Auto-ID Centre. He has been an inspiration to me both academically and personally. Over the years Prof. Cole has provided me with tremendous freedom and flexibility to pursue various disciplines of research and I have accumulated an enormous amount of knowledge as a result.

I would like to thank Dr. Braden Phillips and Prof. Daniel Engels, for their many suggestions, guidance and encouragement provided to direct my research in the areas of privacy and security. Discussions with them were always insightful and stimulating.

I also had the great pleasure of collaborating with Mr. David Hall, Mr. Daihyn Lim and Prof. Srini Devadas on a number of projects. I would like to thank them for their collaboration. I will always remember the many stimulating conversations I shared with David. My brief encounters with Leigh Turner and his insights have also been very valuable to my research.

I would like to thank the following people at the Auto-ID lab for their support, their valuable assistance and for making the Auto-ID lab an enjoyable place to work: Mr. Alfio Grasso, Mr. Kin Seong Leong, Ms. Ng Mun Leng, Mr. Raja Goshal, Mr. Zheng Zhu and Dr. Behnam Jamali. I must especially thank Mr. Alfio Grasso for reading my thesis and providing valuable comments and feedback.

I must also thank the following people at the school of electrical and electronic engineering for their invaluable assistance: Ms. Colleen Greenwood, Mrs. Rose-Marie Descalzi, Mr. David Bowler, Mr. Geoffrey Pook, Mr. Ian Linke and Mr. Stephen Guest.

In addition, a special note of gratitude to Jivaka Jayasundera and Ava Hayman for fastidiously labouring through my thesis, and Antonietta Sisto for her love and support, while I live so far away from my family and friends. I must especially thank my friend Jivaka Jayasundera for his steadfast encouragement and support through some difficult times. I would also like to thank all my friends who have stood by my side like a rock throughout my life.

Most importantly, I am forever indebt to my father and my mother for their love, encouragement, patience and above all, the sacrifices they have made on my behalf to help me reach my goals. I am forever in your debt. Finally, I must not forget my little sister and her love and support through many difficult times.

January 2007.

Abstract

A combination of Radio Frequency Identification technology and ubiquitous computing are revolutionising the manner in which we look at simple objects. Radio Frequency Identification (RFID) allows RFID labeled objects to be identified at a distance without physical contact, and ubiquitous computing provides a virtually connected environment for the objects. RFID labels are frequently referred to as the next generation barcodes.

RFID Systems provide increased productivity, efficiency, convenience and many advantages over bar codes for numerous applications, especially global supply chain management.

RFID labeling has a number of advantages over conventional bar code systems. The optics based bar code systems could be rendered useless by common everyday environments containing dirt, dust, smoke, grease, condensation and by misorientation and misalignment. Furthermore bar codes are subject to fraudulent duplication and counterfeiting with minimal effort.

However, there are limitations and constraints inherent to RFID technology: semiconductor thresholds, limits on transmitted power, costs, antenna and coupling inefficiencies. Thus it is important for RFID designers to understand these limitations and constraints in order to optimise system designs and overcome inefficiencies where possible. Therefore the work presented in this dissertation seeks to improve the performance of advanced RFID systems by overcoming a number of these limitations.

Prior to a discussion of improving performance, the author's interpretation of a modern RFID system along its evolutionary path as a ubiquitous RFID network and its application to supply chain management is described. Performance improvements are achieved by: the development of electromagnetic theory for RFID system analysis and optimisation; design and development of interrogator antennas; analysis of electrically small and tiny antennas for RFID labels; and development and utilisation of a design methodology for creating high performance label antennas and antennas for tagging metallic objects.

Implementations of RFID systems have raised concerns regarding information security and possible violations of end-user privacy. The most profound concerns are raised against low cost RFID technology because of its potential for mass scale deployment, its pervasive nature, and the resource limitations preventing the provision of strong cryptographic solutions. There is a growing need in the RFID community to discover and develop techniques and methods to overcome various hurdles posed by the above-mentioned concerns.

Thus, the thesis also considers the vulnerabilities of low cost RFID systems and associated insecurities and privacy concerns resulting from the latter. Prior to addressing such concerns impeding the deployment of low cost RFID technology, a framework within

which to provide security services is also detailed. It has become important to both define and identify a framework based around low cost RFID systems since RFID has become a "catch all" phrase for various other forms of technology.

Addressing security and privacy of low cost RFID systems requires novel thinking. The later parts of the thesis outline design considerations for security mechanisms and a number of practicable solutions for providing the features of: mutual authentication; confidentiality; message content security; product authentication; anonymity and untraceability, that are necessary for low cost RFID systems to overcome the weaknesses identified in this dissertation. Implementing these security mechanisms requires the generation of true random tag parameters and true random numbers. Achieving these objectives using a hardware based true random number generator is also described and analysed.

A final part of the thesis focuses on active RFID labels and improving their performance. The primary concern with active labels is the life of the onboard battery. Turn-on circuits provide a method of turning "on" and "off" an active label remotely to conserve valuable battery power. Analysis, development and testing of a turn-on circuit concept, based on interrogator field sensing, have provided a means of remotely activating and deactivating active RFID labels and conserving battery power. The final chapter of this thesis provides a detailed analysis, based on coupling relations between electromechanical systems, for evaluating the feasibility of a theft detection sensor, based on a turn-on circuit for an active RFID label, for preventing the theft of high value items.

While low cost RFID needs to overcome certain security and privacy related barriers, RFID technology does provide novel and valid approaches to such security related applications as product authentication, anti-counterfeiting and theft detection. It is believed that the contributions from this thesis will extend and elaborate on the existing knowledge base, paving the way forward to allow further significant deployment of advanced RFID technology.

Dedicated to my mother
Who taught me
Never to forget my humble beginnings
And that
Education is the only path to true freedom

CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Backward channel	The term refers to the communication channel from the label to the reader
Ciphertext	The result of encrypting a string of plaintext
CMOS	Complementary Metal Oxide Semiconductor
Collision	The term is defined as an event where multiple tags simultaneously reply to a stimulus from a reader, where information is lost in the process. A classic example of a collision would be a packet collision
CRP	Challenge-response pair
DOS	Denial of Service
ECC	Elliptic Curve Cryptosystems
EM	Electromagnetic
EMC	Electromagnetic Compatibility
EPC	Electronic Product Code. This is a naming convention developed for all physical objects. It is a scheme proposed by the former Auto-ID Centre group
FCC	Federal Communication Commission. It is the organisation responsible for EMC standards and regulations in the United States
Forward channel	The term refers to the communication channel from the reader to the label
HF	High Frequency. Refers to a band with a centre frequency at 13.56 MHz
IC	Integrated Circuit
ITU	International Telecommunication Union. An international organisation involved in worldwide radio frequency spectrum management
MAC	Message Authentication Code
Nonce	Number used only once

NP	NP stands for non-deterministic polynomial time. In the context of complexity theory it classifies a set of problems that can be solved in polynomial time on a nondeterministic Turing machine.
NP-Hard	Here NP stands for non-deterministic polynomial time. NP-Hard problems are a class of problems where there is no known algorithm for solving the problem in polynomial time
NRZ	Data coding method Non-Return to Zero.
PPM	Data coding method Pulse Pause Modulation.
PRNG	Pseudorandom number generator
PTM,	Data coding method Pulse Time Modulation.
PWM	Data coding method Pulse Width Modulation
RF	Radio Frequency
RFID	Radio Frequency Identification
RZ	Data coding method Return to Zero.
Spoofing	The term is taken to refer to an attack on the privacy or integrity of an RFID system in which the attacker creates a misleading context, by way of creating a fake label, to trick the readers into making an inappropriate security-relevant decision, or fooling the readers into believing that the label legitimately belongs to the labeled article.
Synchronous stream ciphers	A stream cipher is said to be synchronous if the key stream is generated independently of the plaintext and of the ciphertext
UHF	Ultra High Frequency
UPC	Universal Product Code

Chapter 1

INTRODUCTION

The purpose of this chapter is to present an overview of this dissertation and describe the organisation of the material presented. A summary of problems and performance issues related to modern RFID systems that are addressed in this thesis is provided along with contributions made towards addressing the outlined problems and performance issues. A list of the author's publications are also provided.

1.1 Overview

An evolving set of identification technologies continues to play an important role in the provision of identification and other related information about entities both animate and inanimate in various contexts, ranging from Allied air planes in World War II [1] to items stacked on supermarket shelves. The technologies used range from biometric data recognition, optical character recognition, magnetically coded ink (MICR) used on cheques, magnetic strips used on credit cards, Wiegand wire and barium ferrite inserts used in access control badges, speech recognition, smart cards and barcodes [2]. The term "Auto-ID" is used to refer to systems based on the above technologies because automatic identification is almost always a primary function performed by such systems [2].

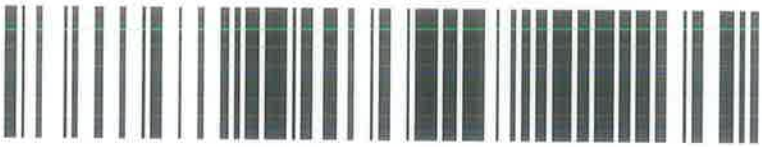

Figure 1.1 A conventional barcode with 10 encoded characters.

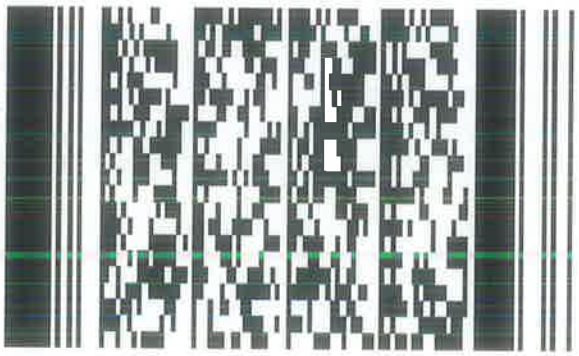

Figure 1.2 An example of a 2D barcode with 62 encoded characters.

Amongst the identification technologies, barcodes, of which a standardised form (Universal Product Code) developed by a consortium of inter-industry trade associations called the Uniform Code Council (UCC), recently renamed as GS1 [3], are available for consumer items [4], and have continued to proliferate throughout the world as a solution to both logistic and inventory control by providing a product identification code. The proliferation of Universal Product Code (UPC) technology was aided by the continued growth in the commerce of goods and the need for efficient logistical support for global trade and transport of goods. A research report compiled by PriceWaterhouseCoopers in 1999 entitled "17 Billion Reasons to Say Thanks" [4] depicts the impact of the UPC on international economies and industry and on the consuming public. Since the inception of the idea by Norman Joseph Woodland in 1949, various forms of the barcode have been developed over

the years [5]. One such example is the 2D bar code which can carry a considerable amount of data on a smaller surface area than compared with its predecessor [3]. The latter fact is made clear by comparing Figure 1.1 and Figure 1.2 which show that the space required to encode 62 characters using 2D barcodes is much less than that possible with conventional barcodes.

UPC codes are a contactless technology where scanners work in the range of a few centimetres but require a clear vision of the bar code in the scanning field. Bar code labels have to be printed carefully and exactly since line thickness and spaces signify the alphanumeric characters forming the UPC, while the objects labeled with barcodes must be physically manoeuvred to align the barcodes with the scanners. In addition, barcodes are not suitable for hostile environments where optical recognition of the barcode may not be possible. Another common failure of barcodes is that they are prone to physical damage as the printed symbols may be smudged or concealed, making the process of scanning difficult or impossible. These issues limit the usefulness, possible application and performance of barcode based "Auto-ID" technologies.

An "Auto-ID" technology devoid of many of the imperfections of barcodes has emerged to rival the simple barcode. This technology is commonly referred to as Radio Frequency Identification (RFID). Radio Frequency Identification (RFID) systems allow automatic identification using a unique identifier associated with an object or person. RFID technology has been in existence for more than sixty years with the earliest application of RFID commonly cited as the "Identify Friend or Foe" system used in Allied aircraft in the Second World War [1]. Current applications of RFID are a diverse collection of identification and data capture applications. Some typical applications are automatic toll collection, animal identification, proximity cards for secure access, authentication, theft detection, tamper proofing containers and last but not least, supply chain logistics [2 and 6].

The work presented in this thesis investigates advanced RFID systems. What soon becomes apparent with study of advanced RFID technology is the large number of issues leading to disappointing performance, and the imminent threats posed by vulnerabilities of low cost RFID systems. The following section summarises the issues plaguing the advanced RFID systems addressed in this dissertation.

1.2 Problem Statements

Modern RFID systems are almost always supported by a network infrastructure and thus have come to be called networked RFID systems. The architecture to build an ubiquitous item identification network originated at the former Auto-ID Center, now the Auto-ID Labs [7] while the process of standardisation issues and future developments is currently managed by a host of working groups, primarily the Software Action Group (SAG) and the Hardware Action Group (HAG), established by EPCglobal Inc. The Auto-ID Center's vision was to create a "Networked Physical World" by building an intelligent infrastructure linking objects, information, and people through computer networks oblivious to the users [31]. Such a system can be realised with a combination of automatic identification technology and a ubiquitous computer network that will glue the physical

world together. The ability to form a ubiquitous item identification network has a wide range of applications including manufacturing automation and supply chain management. A clear understanding of this network architecture, the data flows within, and the usefulness of the architecture, is not widely available in literature.

One avenue for addressing the inadequate performance of RFID systems is by understanding the coupling link between a tag and a reader, and the coupling elements: antennas [10, 37, 39 and 40]. This requires the development of electromagnetic theory that is suitable for the analysis of RFID systems. The formulations used presently are those used in radar systems [6] and do not allow the optimisation of the coupling links in RFID systems.

One drawback of High Frequency (HF) interrogation systems is very limited read range due to the regulatory limits on power that can be radiated at the HF Industrial Scientific and Medical (ISM) band [9 and 11]. Hence antenna designs that minimise far field radiation but create large volume near fields are required. Recent relaxation of electromagnetic compatibility regulations in the European Union allows a greater degree of radiation [44] and the possibilities for, and consequences of, exploiting this new development to improve the performance of HF RFID systems have not been considered in literature.

Unlike HF interrogator antennas, active development by various companies producing RFID products has developed a broad range of RFID label antennas suitable for tagging a variety of objects. However the understanding and the methodology required to develop label antennas have not been forthcoming and have remained somewhat of a black art [51, 52, and 53]. In addition, the loss of performance from tag antennas placed close to metals and materials of high dielectric losses such as liquids need to be addressed and understood [50 and 231]. There is in the literature an inadequate coverage of antennas suitable for tagging metallic objects. Clearly, an understanding of the issues that need to be considered in tag antenna design and a method for developing tag antennas will serve to remove this black veil and engage a wider community to participate in the tag antenna development process.

The increasing drive towards the identification of smaller and smaller objects has placed much research interest in small tag antennas operating in the Ultra High Frequency (UHF) region [54 and 63]. Such antennas are both physically small and electrically small at UHF frequencies and are very inefficient radiators due to their large losses in relation to their radiation resistance [48, 60, 62 and 63]. Also, it is difficult to achieve adequate bandwidth with small antennas to meet the operational needs required for different UHF frequencies in different jurisdictions around the world. While small antennas have been considered in the past in general [48, 59, 60, 61, and 62], what has not been considered is the design of electrically small antennas and tiny antennas suitable for RFID applications.

Focusing on active RFID labels, it can be concluded that the addition of an onboard battery for such labels creates novel problems dissimilar to those of passive RFID labels [39]. The most unique is that of increasing the life-time of the battery onboard an active label [69 and 70].

The problems associated with performance are not isolated issues, the security and privacy issues resulting from the pervasiveness of RFID technology and the vulnerabilities of low

cost RFID systems are increasingly coming to the fore [85, 86, 87, 92, 93, 95, 96, 102 and 108]. The immediacy of the problem is highlighted by low cost RFID technology that is set to propagate throughout various consumer goods supply chains around the world [94, 97 and 99]. Addressing issues related to security and privacy is important if RFID deployments are to eventuate and become successful.

Despite the considerable progress made over the past decade in the understanding and the development of RFID technology, the aforementioned intriguing problems have remained unresolved, unaddressed, not clearly understood or inadequately considered. The approach taken in this dissertation is to consider each of the areas highlited above in three separate but related parts elaborated in detail in Section 1.4. The following section lists contributions made in this dissertation to address the problems presented in this section.

1.3 Thesis Contributions

A list of publications by the author is given in Section 1.5, while this section summarises contributions made by the author through both publications and the content presented in this dissertation towards addressing the problems enumerated in Section 1.2.

- Publication of material outlining the development of the EPC Network architecture, along with an overview of a web services based implementation and benefits of such a network to supply chain applications whilst also exploring the topic of merging sensor technology with RFID.

- Development of coupling volume theory through experiments and simulations results to extend it to near field electric fields. The application of coupling volume theory as an antenna performance metric for comparing the performance of different antennas and improving antenna performance.

- The use of far field coupling volume theory to consider the feasibility of using miniaturised (tiny) antennas for UHF RFID applications.

- The testing and investigation of large interrogator antenna structures for operation in the HF region. The development of a large wedge above ground plane antenna capable of mid field reading distances in view of the relaxed HF RFID regulations in Europe. The development of empirical results based on Brown and Woodward's results and the author's experimental results to develop an equivalent circuit for large wedge above ground plane antennas. The analysis of a large loop antenna capable of reading tags in the mid field with considerably reduced radiation achieved by uniformity of current distribution on such a large loop antenna.

- Investigating factors affecting the development of electrically small RFID label antennas. Developing a methodology for building antennas for RFID tags. The development and testing of three antennas for RFID applications, where one is a small antenna designed for tagging metallic objects.

- Investigating turn-on circuit concepts and, developing, analysing and testing a zero power turn-on circuit for active RFID labels to increase the lifetime of battery powered RFID labels.

- Investigating and outlining vulnerabilities in low cost RFID systems, the resulting security threats and privacy violations. Providing a framework for researchers to develop solutions for such issues. Developing and outlining a number of practical solutions to address the aforementioned concerns.

- Evaluating the possibility of using metastability and thermal noise for generating random numbers on RFID readers for security mechanisms based on previous work published on Physical Uncloneable Functions. Evaluate the test results of such a generator to assess the quality of randomness and the suitability of the generator for use in RFID interrogators.

- Investigating and evaluating the feasibility of using MEMS in RFID. A MEMS based turn-on circuit used for field sensing with an application to theft detection is analysed and results presented with future directions.

1.4 Thesis Organisation

The first chapter of this thesis has provided an insight into various "Auto-ID" technology and highlited a number of problems that motivated the thesis topics to be covered. Contributions made towards addressing issues mentioned in Section 1.2 and a list of publications by the author are covered in this chapter as well. The final section is a small but important note on notation used used in the discussion that follow.

The following chapters in this dissertation are assembled into three separate but related parts. The parts are related because they focus on improving performance and developing solutions and methodologies to solve the problems hindering the widespread adoption of RFID technology. They are separate because they focus on different aspects of RFID systems outlined in the problem statements described in Section 1.2.

The first part of this thesis begins with an introduction to modern networked RFID systems with the networking aspects covered in Chapter 3. The chapters subsequent to the introductory formalities being dealt with, focus on optimising the coupling link, and coupling elements between a tag and a reader.

The second part is concerned with vulnerabilities of low cost RFID systems that are expected to proliferate in consumer goods supply chains and outlines practicable methods of overcoming such vulnerabilities.

The final part of this work deals with the development of turn-on circuit concepts to improve the operational life of semi-passive and active labels, and the usefulness of turn-on circuits in

the construction of an anti-theft tag for high value items. The various parts are explained in more detail in the sections that follow.

The thesis is concluded with an appendix which describes useful formulae in the analysis of the electromagnetic coupling links in RFID systems.

1.4.1 Part One: Electromagnetic Coupling

In order to introduce concepts developed by the Auto-ID Centre, now the Auto-ID Labs [7], for RFID systems, Chapter 2 provides an overview of the components and concepts underlying modern RFID systems.

Chapter 3 introduces a modern RFID system as interpreted by the author in its process of evolution as an EPC Network along with a number of applications that such a system could bestow to the world at large.

As successful deployment of RFID systems depends on a proper understanding of electromagnetic coupling between interrogators and labels, a formal statement of the laws of electrodynamics is presented in Chapter 4 where they are used to derive boundary conditions that must be observed where fields interact with materials. The retarded potential solutions of Maxwell's equations and their utility in developing near and far radiated fields are developed, especially for infinitesimal electric and magnetic dipoles. Along with transmitting antenna concepts, near and far field description of electromagnetic fields are presented for both electrical and magnetic fields. Existing coupling volume theory for near field magnetic fields is described. The existing theory is then extended to formulate the coupling volume theory for near field electric fields and electrically sensitive antennas, thus providing a comprehensive theory capable of optimising the performance of near field systems that are either electrically or magnetically coupled.

The influence of electromagnetic compatibility constraints on the choice of operating frequency and interrogator antenna style is explored in Chapter 5. A number of near field creation structures are illustrated, along with a detailed design and analysis of a wedge above ground plane antenna and a large loop antenna and their ability to increase the label operation range in the HF region to the near field and far field boundary.

Chapter 6 examines the design of RFID label antennas. The chapter introduces a methodology for developing tag antennas for RFID labels by presenting the design and analysis of two long range, bow tie antenna designs, suitable for tagging cases and pallets.

Chapter 7 of this work identifies the need for electrically small tag antennas. Focus is placed on developing small antennas for RFID labels, along with small antenna design constraints and limitations in relation to RFID labels. Tag antennas such as those presented in Chapter 6 will not perform well when used for tagging metallic objects. Antennas for tagging metallic objects require novel thinking. An antenna design, which is also electrically small, suitable for tagging metallic objects is presented in the same chapter along test results and performance evaluation of the antenna.

The coupling volume theory for near field electromagnetic fields is introduced in Chapter 4 while extending the theory to near field electric fields. Chapter 8 illustrates far field coupling volume theory. Far field coupling volume theory is then applied to coupling link optimisations, and comparison of antenna performance with the aid of antenna design and analysis is presented in Chapter 5 and Chapter 6.

1.4.2 Part Two: Vulnerabilities and Solutions

While previous chapters considered the ingredients of RFID systems, it is important to consider the implications arising from the pervasiveness and the mass utilisation of this technology. Chapter 9 introduces vulnerabilities, leading to both privacy and security issues, of low cost RFID technology and focuses on challenges that are unique to the provision of security and privacy to low cost RFID systems. This chapter also provides a broad overview of cryptographic tools available for consideration in the provision of security services, and an overview of possible attacks on cryptosystems. Later sections of the chapter present a literature review of RFID security schemes and cryptographic systems for limited resource applications, and finally make mention of existing cryptographic system implementations on limited resource hardware such as RFID microchips. The ability to provide solutions and to evaluate their effectiveness requires a simple security model and a framework within which to develop solutions; Chapter 10 considers the latter issues. Chapter 11 proposes a number of solutions for addressing privacy and security concerns outlined in Chapter 9 and evaluates those solutions for their merits for low cost RFID based on the framework developed in Chapter 10.

Chapter 12 presents a hardware random number generator design evaluated for its usefulness in RFID readers for the rapid generation of random numbers in the implementation of security mechanisms outlined in Chapter 11.

1.4.3 Part Three: Turn-on Circuits

Chapter 13 provides two concepts for the development of turn-on circuits for semi passive and active RFID labels based on a turn-on circuit for sensing an electromagnetic pulse. Both employ diode resonance to address the problem of increasing the battery life of these high performance labels.

Chapter 14 is an illustration of the fusion of RFID technology and sensor technology for the creation of more powerful applications in the future.

Despite the shortcomings of low cost RFID technology, RFID has the potential to provide solutions for various problems such as product authentication, anti-counterfeiting and theft detection. Chapter 14 describes the outcomes of a feasibility analysis to evaluate the possibility of developing a theft detection sensor using a turn-on circuit for active RFID devices based on a MEMS device. It is the author's view that the incorporation of sensors on RFID tags will become more commonplace in the future and RFID systems will be able to

benefit from the current enthusiasm, excitement and research outcomes in the Sensor Network community. However the opposite may also be true.

1.5 Publications

P. H. Cole, D. C. Ransinghe and B. Jamali, "Coupling relations in RFID systems", *Auto-ID Center White Paper Series,* June 2002.

P. H. Cole, D. C. Ransinghe and B. Jamali, "Coupling relations in RFID systems II: practical performance measurements", *Auto-ID Center White Paper Series,* June 2002.

D. Ranasinghe, D. W. Engels, P. H. Cole, "Security and privacy solutions for low cost RFID systems", *Proc. of the 2004 Intelligent Sensors, Sensor Networks & Information Processing Conference,* Melbourne, Australia. pp. 337-342, 14-17 December, 2004.

D. Ranasinghe, D. Hall, D. W. Engels, P. H. Cole, "An embedded UHF RFID label antenna for tagging metallic objects", *Proc. of the 2004 Intelligent Sensors, Sensor Networks & Information Processing Conference,* Melbourne, Australia. pp. 343-347, 14-17 December, 2004.

D. C. Ranasinghe, D. Lim, S. Devadas, B. Jamali, Z. Zhu, and P. H. Cole, "An integrable low-cost hardware random number generator", *Smart Structures, Devices, and Systems II*, S. F. Al-Sarawi, Editor, *Proceedings of SPIE International Symposium MSN, & MSS*, Vol. 5649, Part Two, pp.627-639. Sydney, Australia, 12–15 Dec. 2004.

B. Jamali, P. H. Cole, D. Ranasinghe, Z. Zheng, "Design and optimisation of Schottky diodes in CMOS technology with application to passive RFID systems", Smart Structures, Devices, and Systems II, S. F. Al-Sarawi, Editor, *Proceedings of SPIE International Symposium MSN,* & MSS, Vol. 5649, Part One, pp.323-343. Sydney, Australia, 12–15 Dec. 2004.

D. Lim, D. C. Ranasinghe, S. Devadas, B. Jamali, D. Abbott, P. H. Cole, "Exploiting metastability and thermal noise to build a reconfigurable hardware random number generator", *Noise in Devices and Circuits III*; edited by Alexander A. Balandin, Francois Danneville, M. Jamal Deen, Daniel M. Fleetwood; *Proceedings of SPIE* Vol. 5844, p. 294-309, Texas, USA, May 2005.

D. C. Ranasinghe, and P. H. Cole, "Evaluation of a MEMS based theft detection circuit for RFID labels", *VLSI Circuits and Systems II*, edited by J. F. López, F. V. Fernández, J. M. López-Villegas, J. M. de la Rosa, *Proceedings of SPIE* Vol. 5837, Seville, Spain, June 2005.

D. M. Hall, D. C. Ranasinghe, B. Jamali, P. H. Cole, "Turn-on circuits based on standard CMOS technology for active RFID labels" *VLSI Circuits and Systems II*, edited by J. F. López, F. V. Fernández, J. M. López-Villegas, J. M. de la Rosa, *Proceedings of SPIE* Vol. 5837, p. 310-320, Seville, Spain, June 2005.

D. C. Ranasinghe, K. S. Leong, M. L. Ng, D. W. Engels, P. H. Cole, "A distributed architecture for a ubiquitous item identification network", *Seventh International Conference on Ubiquitous computing*, Tokyo, Japan, Sept 2005.

D. C. Ranasinghe, D. Lim, S. Devadas, D. Abbott, P. H. Cole, "Random numbers from metastability and thermal noise", *IEE Electronic Letters*, vol. 41, Issue 16, pp. 11-12, 2005.

D. Ranasinghe, D. W. Engels, P. H. Cole, "Low cost RFID systems: confronting security and privacy", *Auto-ID Labs White Paper Journal Volume 1*, September 2005.

D. C. Ranasinghe, K. S. Leong, M. L. Ng, D. W. Engels, P. H. Cole, "A distributed architecture for a ubiquitous RFID sensing network", *Proc. of the 2005 Intelligent Sensors, Sensor Networks & Information Processing Conference,* Melbourne, Australia. 14-17 December, 2005.

B. Jamali, D. C. Ranasinghe, P. H. Cole, "Analysis of a UHF RFID CMOS rectifier structure and input impedance characteristics", *RF and Wireless Circuits, Proceedings of SPIE* Vol. 6053, Brisbane, Australia, December 2005.

B. Jamali, D. C. Ranasinghe, P. H. Cole, "Design and optimisation of power rectifiers for passive RFID in monolithic CMOS circuit", *RF and Wireless Circuits, Proceedings of SPIE* Vol. 6053, Brisbane, Australia, December 2005.

D. C. Ranasinghe, and P. H. Cole, "Analysis of power transfer at UHF to RFID ICs by miniaturised RFID label antennas", *IEEE International Workshop on Antenna technology: Small Antennas and Novel Metamaterials*, New York, March 2006.

P. H. Cole, and D. C. Ranasinghe "Extending coupling volume theory to analyse small loop antennas for UHF RFID applications", *IEEE International Workshop on Antenna technology: Small Antennas and Novel Metamaterials*, New York, March 2006.

D. C. Ranasinghe, K. S. Leong, M. L. Ng, and P. H. Cole, "Small UHF RFID label antenna design and limitations", *IEEE International Workshop on Antenna technology: Small Antennas and Novel Metamaterials*, New York, March 2006.

D. C. Ranasinghe, and P. H. Cole, "Confronting security and privacy threats in modern RFID systems", *Asilomar Conference on Signals, Systems, and Computers*, California, October 2006.

D. C. Ranasinghe and P. H. Cole, "A perspective on fixing security and privacy holes in low cost RFID", *Auto-ID Labs, White Paper Series on Anti-Counterfeiting and Secure Supply Chain*, 2006.

D. C. Ranasinghe, D. Lim, S. Devadas and P. H. Cole, "A low cost solution to authentication in passive RFID technology", *Auto-ID Labs, White Paper Series on Anti-Counterfeiting and Secure Supply Chain*, 2006.

1.6 Notational Aspects

The notation and nomenclature used in this dissertation for physical quantities will be as defined in ISO 1000, [8]. Sinusoidally varying quantities will be represented by peak (not r.m.s.) value phasors. Lower case variables will be used for instantaneous values of scalars such as, for example, voltage or current as given in (1.1) for a real time-varying voltage. Bold calligraphic characters, for example $\mathcal{E}$, $\mathcal{D}$, $\mathcal{H}$, and $\mathcal{B}$, will be used for instantaneous values of field vectors.

$$v = v(t)\,. \tag{1.1}$$

Upper case upright Roman variables, such as V and I will be used for phasors representing sinusoidally varying scalar quantities, and bold upright Roman characters, such as **E**, **D**, **H**, and **B**, will be used for phasors representing sinusoidally varying field vectors.

PART I: ELECTROMAGNETIC COUPLING

Chapter 2

NETWORKED RFID SYSTEMS

While there is a googol of information on Radio Frequency Identification systems, most of which arose in the last decade, it is important to identify concepts and operating principles that will be discussed in this dissertation and to present the author's views on modern RFID systems. This chapter will provide an overview, however brief, of a modern RFID system as perceived and comprehended by the author.

2.1 RFID Systems Overview

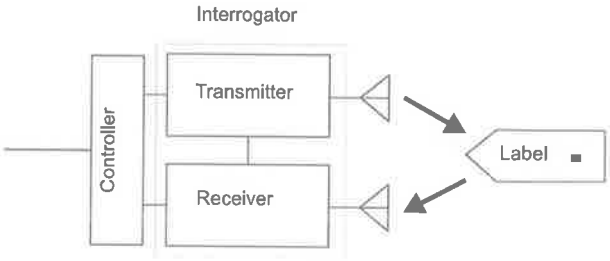

Figure 2.1 An illustration of an RFID system.

A simple illustration of the concept of a Radio Frequency Identification (RFID) system is provided in Figure 2.1. Here, a transmitter of interrogation signals which is contained within an interrogator, communicates via electromagnetic waves with an electronically coded label to elicit from the label a reply signal containing useful data characteristic of the object to which the label is attached. The reply signal is detected by a receiver in the interrogator and made available to a control system.

There is a wide range of operating principles for such a system [6, 9, 10 and 11]. The operating principle and operating frequency are driven principally by the application of the labelling system and the constraints provided by electromagnetic compatibility regulations, environmental noise, and the ability of fields to permeate a scanned region of space or to penetrate intervening materials.

Over the years a number of RFID systems have evolved with advances in material science, microelectronics, and fabrication technology. Irrespective of the underlying technology and the type of labels around which an RFID system is built, (that is, microelectronic labels, surface acoustic wave labels, labels using multiple resonances to encode data and so on [6]) all modern RFID system infrastructures can be categorised into three primary components given below.

1 RFID labels (transponder)

2 RFID label readers or interrogators (transceiver)

3 Backend network (electronic databases)

The material presented in this thesis will consider RFID systems based on using microelectronic devices for RFID labels.

2.2 RFID Labels

Generally, a microelectronic RFID label consists of a small microchip with some data storage and limited logical functionality, and an antenna. The antenna allows the label to couple to an EM field to obtain power, or to communicate with the reader, or to do both. Figure 2.2 provides an illustration of an HF and a UHF RFID label.

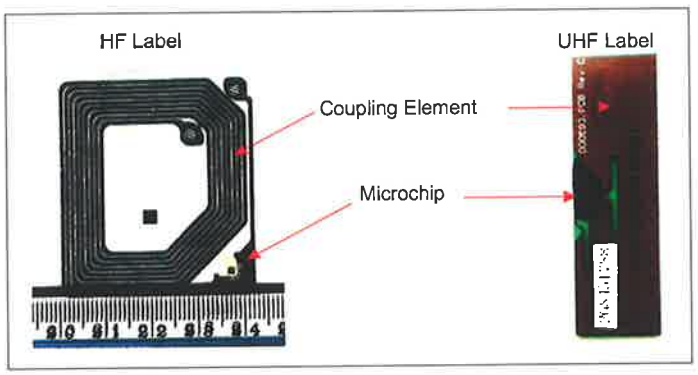

Figure 2.2 Components of an RFID Label.

RFID labels can be distinguished based on their frequency of operation,

- LF,
- HF,
- UHF or
- Microwave

as described in Table 2.1 below.

Table 2.1 Characteristics of tags based on their frequency of operation.

	LF	HF	UHF	Microwave
Frequency Ranges Used	125-134kHz	13.56 MHz	860 MHz – 960 MHz	2.45GHz 5.8 GHz
Data-Transfer Rates	Low	Medium	High	Very high
Advantages	Tags less affected by metals and liquids	High data-transfer rates, cheaper than LF tags	Higher data-transfer rates, cheaper to manufacture. Less sensitive to environmental detuning.	Very high data-transfer rates.

Disadvantages	Low data-transfer rate Tags are expensive because of the copper required for larger antennas.	Affected by metals and liquids	Poor near metals and liquids	Poor near metals and liquids. Line of sight required for long distance communication.
Typical Applications	Access control, animal tracking	Contactless smart-cards, Libraries, and Access cards. Item level tracking.	Item level and pallet level tracking of goods in a supply chain	Electronic toll collection

Labels may also be categorised based on powering techniques as listed below.

- Passive
- Semi-Passive
- Active.

In a primary category of passive systems, the most common operating principle is that of RF backscatter or load modulation [10 and 11] in which a powering signal or communication carrier supplies power or command signals via an HF or UHF link. However, the circuits within the label operate at the carrier frequency or at a lower frequency, and reply via sidebands generated by modulation within the label, or by modulation of a portion of the powering carrier. This approach combines the benefits of relatively good propagation of signals at HF and UHF and the low power operation of microcircuits at RF or lower. Powering at UHF is employed when a longer interrogation range (several metres) is required, and HF powering is employed when electromagnetic fields, which exhibit good material penetration and sharp spatial field confinement are required, or sometimes when a very low cost RFID system implementation is desired.

Table 2.2 Comparison of passive, semi-passive and active labels.

	Passive labels	**Semi-Passive labels**	**Active labels**
Powering	Incident RF signal	Battery	Battery
Transmission	Backscatter	Backscatter	Powered from a battery
Typical read range	3 m – 5 m (UHF)	10 m – 20 m	1000 m

In the category of active labels, the most common objective is to obtain a long range in a battery-assisted backscatter label. Such labels that use backscatter for reply generation and an onboard battery to power the label's logic circuits are called semi-passive labels, while other types of labels called active labels do not use backscatter but instead use a battery for powering and transmitting requirements. In active RFID labels, power conservation is an important issue since the lifetime of the labels will be determined by the lifetime of the onboard battery. A simple solution to conserving power is the use of turn-on circuits.

Chapter 13 of this thesis will examine this issue in more detail. Table 2.2 gives a comparison of the significant characteristics of passive, semi-passive, and active labels.

The data stored on an RFID label can be used to uniquely identify the object, person or animal associated with the label. Unique identification requires a unique identifier defined according to some standard. Among the various identifiers used in different industries, the EPC [12 and 14], which is a unique product identification code format, is being standardised for use in RFID applications. An EPC typically contains information that identifies the manufacturer of the item to which a tag is attached, the type of item and the serial number of the item. This information is also referred to as the label ID.

It should be noted that an EPC is one form of a unique numbering scheme proposed for RFID applications and is a result of an ongoing effort to standardise the data format on labels. The widespread adoption of EPC is a result of its suitability for object identification in the global supply chain. It has been demonstrated that, through supply chain visibility, manufacturers, distributors, retailers and other intermediaries can reduce inventories and ensure product availability, resulting in tremendous cost savings and increases in efficiencies. This is currently the largest market for RFID technology and the potential cost savings and efficiencies are the drivers that are propelling RFID technology into the future. Thus, throughout this thesis, the author will primarily focus on examples related to the global supply chain, while a closer look at the implications of RFID technology to supply chain applications is given in Chapter 3.

2.2.1 Label to Interrogator Communication

Labels perturb the field created by an interrogator to achieve near or far field, label to reader communication. However there are similarities and differences in the way communication is achieved in both the far and the near field by a label antenna.

In near field systems, a reader senses the reactive power flowing in the label. A label is able to control the reactive power flowing through itself by varying the load across its antenna [6, 10 and 11].

However in the far field, labels cause some of the incident RF energy to reflect back or scatter. This is referred to as backscatter. Similar to the manner in which RFID labels in the near field varies the load across the label antenna to vary the reactive power, a label in the far field also varies the load across its antenna to modulate the amplitude and phase of the backscattered wave. In the far field, variation of a label's load impedance causes an intentional mismatch between the label antenna and its load. Backscattering a label response in this manner is akin to the label antenna radiating its own weak signal [6, 10 and 11].

Variations in the load across a label antenna can be achieved by either switching on and off a resistor or a capacitor across the label antenna. The particular mechanism used gives rise to ohmic load modulation and capacitive load modulation, respectively [6, 10 and 11].

In the near field, the real power received by a label is enhanced by using a high quality factor of the label antenna resonance [10], while the reactive power which flows in the label is further amplified above the real power by that same quality factor. Thus the quality factor of the label antenna increases the effect of the changes in reactive power flowing in the label. This is made evident in (4.35) given in Section 4.13.1.

However, in the far field, the power backscattered is not aided by the quality factor of the antenna where the radiation resistance of the antenna is much larger than the antenna losses. Such antennas are generally considered lossless and the best that can be achieved is to reflect all the incident power back to the interrogator while modulating the amplitude or phase of the reflected wave in a time varying manner. This is not fully achievable since some power must be used to power the microcircuits of the label antenna. The latter concepts and coupling relations in the far field are more comprehensively considered in Section 4.13.3 and Chapter 8.

2.2.2 EPC Concept

The concept involves identifying objects through a uniquely formatted number kept on each label as a data field, with associated data, stored in a backend system database. The unique object identifier must have a global scope that is capable of identifying all objects uniquely and act as a pointer to information stored about the object somewhere over the network. The Electronic Product Code (EPC) is a scheme designed for universal object identification with the associated standards developed by EPCglobal Inc [13]. A binary representation of the EPC is shown in Figure 2.3 [14]. The Header identifies the EPC format used by the tag; 96-bit, 64-bit or 256-bit while the General Manager Number identifies an organisational entity. These numbers need to be unique and can be assigned by a standards body such as EPCglobal Inc. The Object Class is used by a General Manager to identify a specific product class. The Serial Number is a unique number within each Object Class.

It is important for serial numbers to be unique for objects labeled by a particular organisation within a product class. However, different objects may reuse the same serial numbers, as the difference in product codes will ensure unique identification of the product. Hence, the triplet of General Manager Number, Object Class, and Serial Number uniquely identifies an object.

Figure 2.3 Bit level representation of an EPC general type identifier format.

2.2.3 Label Hierarchies

A more recent classification of RFID labels proposed by the former Auto-ID Center is based on a functional hierarchy. The functional classes have emerged as a result of discussions between the Auto-ID Centre and sponsor companies [15]. The formulation of a label class hierarchy permits the development of a range of labels to suit different application requirements while setting realistic expectations for functionality. This is an important aspect of the development of this technology, as the environmental, cost, functional, operating range, security and privacy requirements will vary depending on the application employing RFID technology. Figure 2.4 outlines the proposed class hierarchy. The Class 1 labels with simple read-only EPC data form the backbone of the evolving label class hierarchy. The following section provides a brief overview of the functional classes.

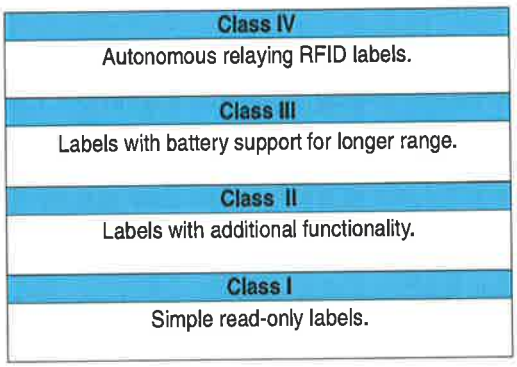

Figure 2.4 An outline of the label class hierarchy.

Class I

Class I labels provide minimal required functionality. In addition to a read only write once EPC code, a label falling into the Class 1 category carries a CRC for transmission verification, and a password that is used in the process of label destruction. As a guarantee of customer privacy, labels can be electronically destroyed before goods are released to the purchaser at their final point of sale (This feature is available as an option on a Class I tag). Class I labels are low cost RFID labels that are typically passive.

The latest standard being ratified by EPCglobal for UHF RFID communication in the 860 MHz – 930 MHz band known as Class 1 Generation 2 (C1G2), is developed to standardise the air interface for Class I labels [16].

Class II

EPC labels with additional functionality, such as read-write ability, provision for data security, privacy, and theft detection ability. Class II labels are likely to be passive with limited read range.

Class III

Class III labels contain battery support for long-range communications. These labels may be active or semi passive and may support broadband communication.

Class IV

These are active labels capable of peer-to-peer communication with other Class IV labels utilising the same frequency. Class IV labels are increasingly being viewed as autonomously networking labels that are able to form ad-hoc wireless networks.

In the scheme outlined above, the higher classes exhibit an inheritance of characteristics of the lower classes; in particular they are all based on the EPC concept. Clearly each class consists of a set of diverse functionalities. For instance, Class II labels are expected to possess a multitude of functionality on a low cost passive label and this requires careful attention in defining its specification.

2.2.3.1 A Classless RFID Label Society

The dilemma of diverse functionalities may be resolved by means other than a rigid hierarchy, as outlined above, by writing standards that allow manufacturers to define and extend functionalities without having to revise or continually rewrite standards.

The RFID label will continue to evolve and be more than just an identifier. As put forward in Chapter 14 the fusion of sensor technology with RFID will create a vibrant and diverse future for RFID with labels capable of a variety of different functionality in addition to those related to security and with other product specific functions. Given such a pool of as yet unimaginable set of infinite capabilities for an RFID label we should not limit ourselves to an inheritance based hierarchy that is inflexible.

Instead it is possible to create a classless society where the functionality of a label can be related to its EPC and defined on a database. It is then possible to cache that database in readers, given that it is relatively small [17]. With this philosophy it is possible for manufacturers to be creative and define methods for accessing those tag capabilities.

If, however, distinction is required, as it is human nature to classify and categorise all that exists, it is possible to have Class I labels, Class II Simplex labels (labels with read write memory) and Class II complex labels (labels processing read write memory and unlimited functionality, in theory) [17].

2.3 Interrogators

The readers communicate with the labels using an RF interface. Either a strong energy storage field in the vicinity of the reader antenna, or radiating EM waves, establishe the RF interface. Communication between a reader and a label process may involve interrogating the label to obtain data, writing data to the label or transmitting commands to the label so as to affect its behaviour. The readers consist of their own source of power, processing capability and an antenna. The general design principle in EPC based RFID systems is to off load silicon complexity of the label to backend systems and to the reader in order that

the cost of the labels may be kept to a minimum. Thus, readers have ample computing power and are capable of acting as proxies to perform computationally intensive tasks for RFID labels. However, increasingly in pursuit of low cost readers, reader architectures are being simplified to manage cost and complex tasks are being pushed further back to the backend systems. A more detailed look at interrogators can be found in [21].

2.4 Back-End Systems

The readers are generally connected to a backend database for storing collected data for processing. The electronic databases are used to collect data aggregated through readers and the electronic database software uses the data for various purposes.

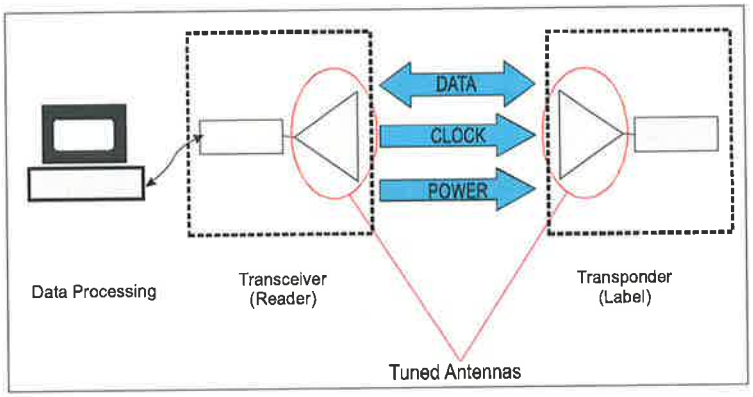

Figure 2.5 A high level illustration of the interactions between RFID components [6].

The interaction of the components outlined above, the backend network architecture and the flow of data within an EPC based RFID system, is discussed at length in Chapter 3.

2.5 Anti-Collision

An important aspect of an RFID system is being able to read multiple labels in a relatively short period of time as in any realistic system there is likely to be more than one label. Implementing such plurality of reading requires a means of overcoming collisions between label replies (refer to Figure 2.6). Here the term 'collision' is used to imply the resulting signal interference from two or more tags replying simultaneously. A collision generally results in a failed communication. There are a number of anti-collision algorithms used to implement such plurality of reading and prevent the occurrence of collisions. These algorithms are embedded in multiple tag reading protocols.

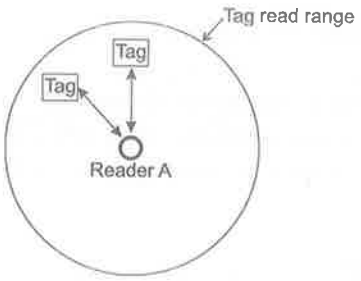

Figure 2.6 Tag reply collision.

The anti-collision methods used in RFID systems are similar to multiple access communication conflict resolution or detection methods used in various computer networks such as Carrier Sense Multiple Access (CSMA) [18]. However RFID anti-collision methods are constrained by limited computational power, memory and power constraints of an RFID label. Anti-collision methods used in RFID must consider the wireless and ad hoc nature of RFID networks along with the necessity to recover from sudden power loss, as is the case with passive RFID systems.

There are a wide variety of anti-collision algorithms. The vast majority of existing anti-collision methods are time-domain methods which are either deterministic or probabilistic schemes. The most widely used deterministic methods are based on the binary tree walking protocol while most probabilistic methods are based on Slotted Aloha [19]. These schemes are intended to reduce the occurrence of multiple simultaneous responses from tags to a reader query.

In addition to the feature which reduces the frequency of collisions, the capacity to detect collisions is a powerful addition to an anti-collision algorithm. Most commonly used techniques rely on line coding schemes. When simultaneously transmitted signals, even of different strengths, coded by line coding schemes, interfere they will always be demodulated by an interrogator attempting to decode a demodulated signal from an RFID label. However an interrogator attempting to decode those signals may fail due to the decoding process producing an invalid symbol as a result of the type of line coding scheme used by the tag to interrogator communication. Level coding schemes such as NRZ and RZ coding schemes are example schemes where collisions will not cause a resulting invalid symbol at the interrogator during the decoding process, while the Manchester coding technique and other transition coding techniques allow collision detection by preventing the proper decoding of the received signal at the interrogator [6] by the resulting invalid symbols that may occur as a result of a collision.

With two messages of unsynchronised bit periods, the level codes will decode to some valid symbol (even though they are not the symbols in either message), but transition codes might lose or develop transitions in the middle of presumed bit intervals and produce invalid symbols (dissimilar to those present in either message). Nevertheless it should be noted here that when a weak tag to interrogator signal collides with a strong tag to interrogator signal, no line coding scheme will allow the detection of a collision. Instead the strong signal will tend to mask the weak signal.

While CRC codes appended to communication between readers and tags are not part of the anti-collision algorithm or a collision detection method, due to the properties of CRC codes, it is possible to detect a possible collision using CRCs irrespective of the properties of the line coding scheme. It is important to stress the word 'possible', since a weak tag reply and thus a poor signal to noise ratio in the received reply from a tag may also cause an invalid CRC. In the event a tag reply to an interrogator is correctly decoded (but the reply is not that intended from either one of the tags attempting to communicate simultaneously), the CRC calculated on the reply may not match that part of the message interpreted at the CRC. This is because the CRC itself may have been incorrectly decoded or in the event the CRC is that which was sent by one of the tags it may not match that calculated by the interrogator for the received message.

In addition to the collision between label responses, applications requiring readers to operate in close proximity to other readers can cause the interrogation signals from one reader to interfere with signals from other readers [20]. There are several instances where readers can be the source of a collision.

A simple scenario is when an interrogation signal from a nearby reader interferes with a weak reply from a tag outside the read range of the interfering reader during an interrogation of that tag by another reader in the read range of the tag. Such a problem can be mitigated by using Frequency Division Multiple Access (FDMA), as has been demonstrated by the C1G2 protocol [16] which calls for the spectral separation of tag replies from the reader transmission frequency, while readers themselves operate at a number of different channels in the allocated band.

Another situation may arise resulting in tag confusion when multiple readers simultaneously attempt to read the contents of a tag within their tag read range. Such a collision can cause the misinterpretation of the reader communication resulting in an incorrect reply or a complete failure to reply because the command can not be decoded. In such a situation, a carrier sensing scheme may be utilised to prevent concurrent access to a tag by nearby readers [18 and 21].

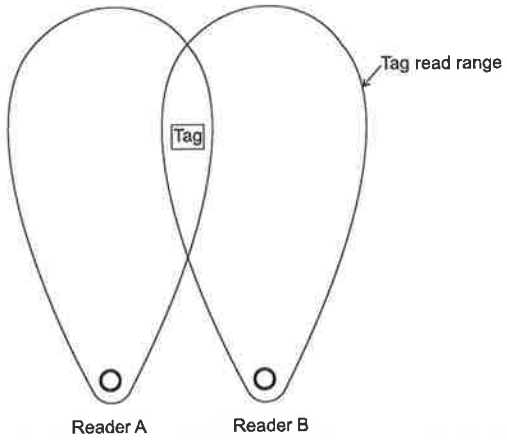

Figure 2.7 A practical interrogator arrangement where carrier sensing can fail.

However, using carrier sensing becomes a difficult issue if the reader antennas are directional, as it is often the case with practical implementations. Hence a reader may not detect the RF field of another active reader because both reader antennas are not in the RF field created, as a result of beam forming to increase antenna gain. While it is possible for a tag to be in read range of both readers as illustrated in Figure 2.7, the RF field of the readers may not be detected by either reader antennas. When such a phenomenon occurs, the commands or queries from different readers collide and the resulting signal will cause the misinterpretation of the intended reader communication. Since tags themselves are made simple, to control costs, they can not be expected to participate and aid in a collision avoidance scheme.

In addition to the phenomenon discussed above, tag collisions, reader collisions, noise from other readers and tags, and echo of tag replies and reader replies has lead to the observation of phantom tags. In this situation, a reader reports an EPC of a tag that does not exist in its tag reading range. However the C1G2 protocol was postulated with the idea of eliminating the phenomenon of ghost reads by the following mechanisms:

- Tags must respond to a reader within a short time frame, so that responses that do not fall into this time frame are eliminated.
- Tags transmit a preamble with every transmission. A preamble is a predefined signal sequence that the reader expects from a valid tag signal. This reduces the likelihood of a noise corrupted tag signal being interpreted as a valid tag by a reader
- The reader performs a validation check of the EPC, to confirm if the reply contains a valid preamble. This is achieved by checking that the data received conforms to a valid EPC format as defined in the tag data specification [14].
- The reader then performs a third validation check to ensure the number of bits received in the response is the same as that reported by the tag. Any mismatch results in the reader discarding the response.
- Finally the reader performs a CRC check on the data received to check for bit errors.

2.6 Conclusion

There are complex hardware devices (labels and readers) and software that bundle together to form an RFID infrastructure. Wide scale adoption of this technology requires an overall reduction in the cost of these devices. Thus, there is a great deal of interest in low cost RFID technology, which essentially involves reducing the cost of RFID labels.

There is a variety of RFID devices designed for various applications. The Auto-ID Centre has taken a modular approach to accommodate this diverse application base and technological possibilities. The EPC concept and the Class structure are a result of this modular thinking.

Class I and Class II labels fall into the low cost end of the RFID label hierarchy. The microcircuits of these classes are extremely resource scarce with only hundreds of bits of storage and only thousands of gates available for logic functions. These circuits are designed with strict limits on power consumption and cost.

This chapter has provided a brief introduction to a very broad subject area but with a number of references where further information can be obtained. While the RF identification aspects are commonly discussed in literature there is generally very little information regarding the backend infrastructure vital to the formulation of a modern RFID system. The following chapter will elaborate on the integration of backend systems to RFID technology, developed under the original vision of a "Networked Physical World".

Chapter 3

EPC NETWORK ARCHITECTURE

The concept of a "Networked Physical World" originated from the Auto-ID Center, now called the Auto-ID Labs [26]. Such a "World" can be realised with the combination of an automatic identification technology and a ubiquitous computer network that will glue the physical world together. Low cost RFID (Radio Frequency Identification) technology can automate identification of physical objects by providing an interface to link a vast number of objects to the digital domain. Thus, RFID as the enabling technology has paved the way forward for the creation of a "Networked Physical World". The ability to form a ubiquitous item identification network has a wide range of applications including automation of manufacturing and supply chain management. The previous chapter provided a brief overview of RFID systems. This chapter describes the backend system components formulating a distributed ubiquitous item identification network enabled by the development of automatic identification provided by RFID technology, and examines the flow of tag data, once obtained by an interrogator. The implementation of such an architecture using a web services based model, as well as the impact of the network on supply chain applications, is also investigated.

3.1 Introduction

The roots of the architecture to build a ubiquitous item identification network originated at the former Auto-ID Center, now called the Auto-ID Labs [26]. Further development of that network and the process of standardization of issues related to that network, are currently managed by a number of working groups at EPCglobal Inc [13].

The Auto-ID Center vision was to create a "Smart World" by building an intelligent infrastructure linking objects, information, and people through a ubiquitous computer network. The creation of the intelligent infrastructure demanded the ability to identify objects automatically and uniquely with the backbone of infrastructure provided by a ubiquitous computing system leveraging the Internet for global connectivity. The components forming the intelligent infrastructure are commonly referred to as an EPC Network where the term EPC (Electronic Product Code) is a result of the unique object identification scheme employed by the system. This new infrastructure enables object-centric computing that will allow universal coordination of physical resources through remote monitoring and control by both humans and machines.

The availability of a Networked Physical World system connecting physical objects to the Internet will have an immediate and profound impact on supply chain management, home automation and manufacturing automation. However, such infrastructure needs to be cost effective to allow its feasibility and large scale adoption.

The EPC Network is assembled from many building blocks representing a number of fundamental technologies and standards. The EPC Network architecture has continued to evolve with technology since its inception. Current developments in the EPC Network are based on an N-tier service oriented architecture and have been published in [27].

3.1.1 N-tier Service Oriented Architecture

Contrary to the component based EPC Network architecture developed initially by the Auto-ID Center, the more modern version is based on an N-tier architecture [28] with an emphasis on defining interfaces. The interfaces define the required standard functionalities and methods by which optional functionalities can be accessed rather than defining components and their associated functionalities. The benefits of such architecture are many; primarily the modularity of the software and the standardised interfaces allows various components to be purchased from different suppliers. Thus, in addition to complexity reduced by modularity, there is also a strong business case for supporting the current architecture of the EPC network. The definition of standard interfaces will ease the process of compliance certification with standards.

The N-tier service oriented architecture approach [28] fits naturally with an object oriented modelling of the architecture because objects encapsulate information and state while

offering functionalities through their interfaces. The modules also have a loose coupling due to the independence of different modules. This reduction in dependency implies that the system is easier to manage and enhance.

The EPC Network architecture is built around a loosely coupled and interoperable system of modules. Many of the concepts for this architecture come from the conceptual architecture called service-oriented architecture (SOA) [28].

Web services are one method of implementing the SOA over standardised protocols and interfaces. There is a strong tendency and a technological trend driving the EPC network architecture towards a web services based SOA.

The following Section presents the system architecture and technologies of the ubiquitous item identification network (EPC Network) and provides a seamless architecture for extending the local area EPC Networks to a wide area infrastructure while the later Sections include an application of such a network. The chapter is concluded with an investigation of the impact of the EPC Network on its most revolutionary application; supply chain management.

3.2 EPC Network

The EPC Network can be described as an intelligent ubiquitous infrastructure that automatically and seamlessly links physical objects to the global Internet. The network of physical objects is achieved by integrating a tag, or an RFID label, into each object. The system networks objects seamlessly by communicating with these labels using interrogators at suitably placed locations, for example: RFID portals; handheld interrogators; and potentially, eventually for some tags, continuously throughout the environment by a network of readers. Interrogators collect data from tagged objects. The RFID labeled objects communicate an EPC code to a reader and thus identify themselves as a unique entity. The data originating from the network of readers is passed to backend systems that control and collect data while providing service layer functionalities.

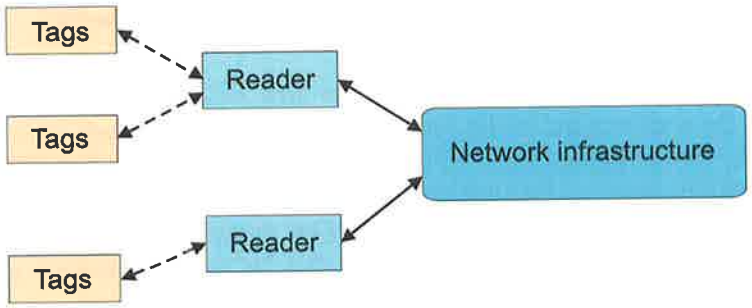

Figure 3.1 An Overview of an EPC Network.

An illustration of the components constituting the EPC Network is shown in Figure 3.1 where the arrows indicate the flow of data from tags to the network support system and the flow of control and data back to the readers and tags.

EPC Networks are significantly different from more traditional computer networks in the sense that the flow of data and information is from many nodes (RFID tags) at the edge of the network towards a number of central servers. In RFID networks, readers detect certain events or readers query RFID labels to obtain event data and forward the resulting information to backend applications or servers. The application systems then respond to these events and application processes orchestrate corresponding actions such as ordering additional products, sending theft alerts, raising alarms regarding harmful chemicals or replacing fragile components before failure.

The EPC Network can be separated into six primary modules, some physical, some logical: (1) RFID tags, (2) RFID tag readers, (3) EPC, (4) the Application Level Event (ALE) Engine systems, (5) Object Name Service (ONS), and (6) EPC Information Service (EPCIS). Figure 3.2 shows the structure of a typical EPC Network where readers communicate with tags and then capture that information to be passed up the information chain to be utilised by the service layers for various applications.

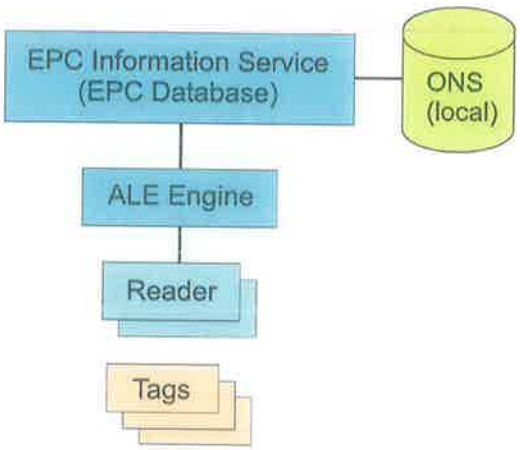

Figure 3.2 The modular structure of a local area EPC Network.

The EPC Network shown in Figure 3.2 is a local area EPC Network akin to a LAN. This model captures the architecture of the system at a local site, company or organization, or a private network. Nonetheless, local EPC Networks can be linked together through the already established backbone of the Internet to achieve a global flow of information and data to extend the reach and the usefulness of the EPC Network. Figure 3.3 illustrates such architecture where a global public ONS system may be used to connect public local area EPC Networks. The subject of ONS is considered in Section 3.5.

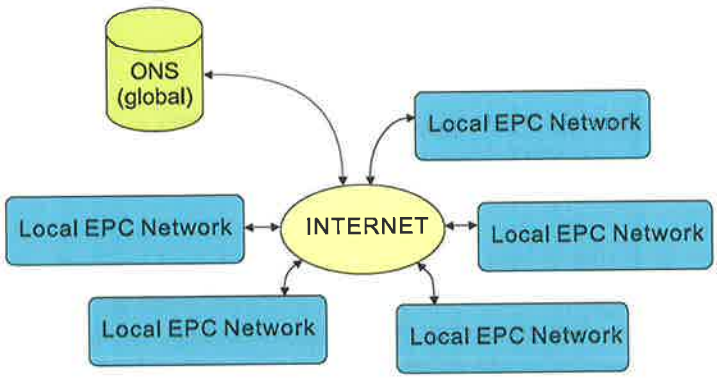

Figure 3.3 Wide area EPC Network overview.

There are ongoing collaborative efforts to standardise the interfaces linking the modules outlined in Figure 3.2 by the various actions groups of EPCglobal, and the Auto-ID labs, to achieve interoperability and to allow hardware and software vendors to be able to compete in a fair and open market in the supply of technology and equipment to establish EPC Networks. Figure 3.4 below identifies the interface layers being standardised by EPCglobal.

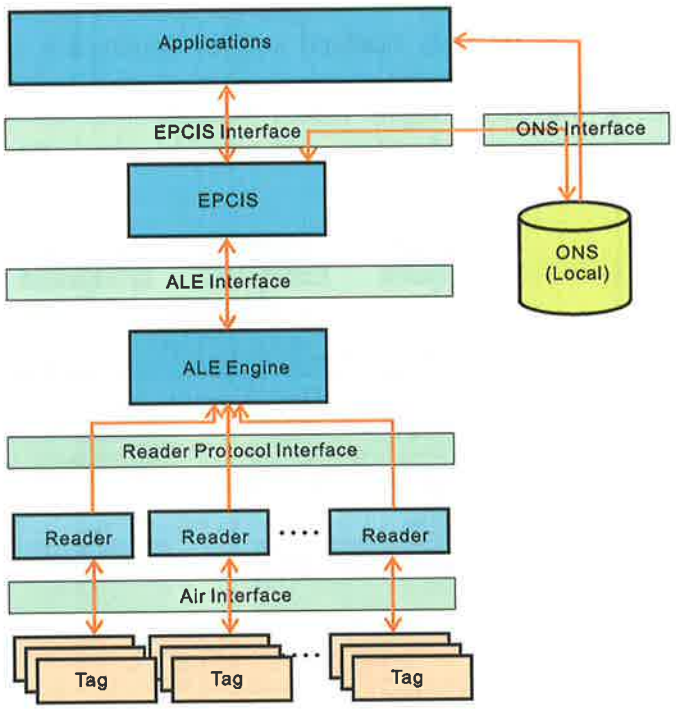

Figure 3.4 EPC Network architecture interfaces.

3.3 RFID Components

The RFID aspects of the EPC Network consist of RFID tags, readers and the unique identifier formatted by the EPC. These modules of the network have been discussed in Chapter 2.

It suffices to note here that a very advanced version of a ubiquitous reader network will allow continuous tracking and identification of physical objects. Reader arrays can, in principle, be fabricated and integrated in floor tiles, carpeting, shelf structures, cabinets and appliances. Similarly to cellular phone grids, the reader network may provide seamless and continuous communication to tagged objects. A data collection and control system must support the reader network to enable efficient use of the continuous, or at least very frequent, object communications.

3.4 Application Level Event (ALE) Engine

An ALE Engine system is a middleware system providing real time processing of RFID tag event data. Conceptually an ALE Engine occupies, as shown in Figure 3.5, the space between a Reader (or multiple Readers) and the application systems. Networked ALE Engine systems form a framework to manage and react to events generated by tag reads by interrogators. The ALE Engine passes requests from the applications to the reader(s) and receives unique tag identifiers and possibly other data from sensors, and passes that information to the applications.

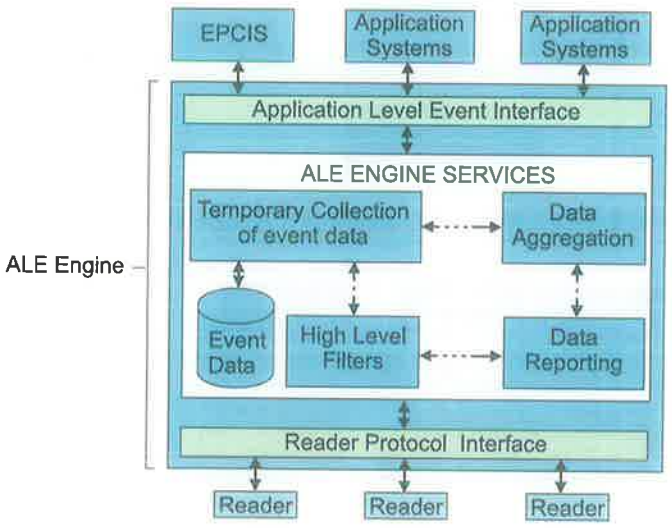

Figure 3.5 Architecture of an ALE Engine System and its interaction with EPC Network components, EPCIS and Readers.

The ALE Engine has several fundamental functions integrated into its design, some of which are data filtering of received tag and sensor data, aggregation and counting of tag data and accumulation of data over time periods. These fundamental functions are required to handle the potentially large quantities of data that RFID systems are capable of generating through continuous interrogation of tags. For instance, ALE Engines enable local applications to state the significance of specific data obtained from RFID tags (for instance a record of temperature variations over a time period) and to report accumulated data using a standard format defined by an XML schema (an existing XML schema definition can be found in [29]). The ALE Engine framework may also implement an application specific XML schema (such as that more suited towards a specific sensor application) or a number of such schemas to allow the capture and reporting of physical world events and measurements.

The ALE Engine possesses two primary interfaces that allow it to communicate with external systems: the Reader Interface and the Application Level Event Interface. The former provides an interface between the ALE Engine and Readers, and the latter between the ALE Engine and external applications [29]. An ALE Engine is composed of multiple ALE Engine Services, each with their own functionality. The ALE Engine Services can be visualised as modules in the ALE Engine. These modules can be combined to perform certain functions for specific applications. Hence one or more applications may make method calls to the ALE Engine resulting in an operation being performed (collection and return of temperature readings from a sensor) and the return of results. Other than ALE Engine Services interacting with each other to perform certain tasks, ALE Engine Services can also interact with services such as the EPC Information Service (EPCIS) to provide services for the framework of global applications (the EPCIS will be considered in detail in Section 3.6). Figure 3.5 shows a conceptual architecture of the ALE Engine system.

Event management is a primary service provided by ALE Engine services. A common event management function is filtering, which is particularly useful in situations where there is heavy data traffic. For example, readers may read data coming in from multiple RFID tags repeatedly. Not all the data from all the tags may be of interest to an application. Filtering of that data can eliminate information that is either redundant (multiple reads of the same data), or that is not required (tags read but not of interest to that application), from reaching an application.

3.4.1 EPC Data Encapsulation and Reporting

The EPC serves as a reference to information. However, the storage, transport and description of that information requires a structured and universal vessel that can be easily understood, stored and transported across the Internet. Previously the Auto-ID Center defined the Physical Mark-up Language (PML) (PML Core specification 1.0, Sept. 2003) to encode captured object information. However, recent developments have retreated from such a rigidly defined schema to the characterization of two instances: `Escape` and `ECReports` instances using a standard XML depiction [29]. Thus requests to the ALE Engine are sent as `Escape` object while data from the ALE Engine is returned as an `ECReports` object. The core XML schemas for these objects are defined with extensions

and rules to accommodate application or manufacture specific XML schema (such as that suited for a specific sensor application) or a number of such schemas to allow the capture and reporting of physical world events and measurements.

The core XML schema is tailored specifically to describing common attributes of physical objects and observables, such as the expiry date, manufacture date, weight, or the time an object was seen at a specific location. XML core does not interpret the data that it handles nor does it promise a universal means for encoding structured information. The XML schema definition is rigid, simple and all the elements can be understood easily because of the use of long descriptive tag names which increase human readability and help to avoid mistakes in the interpretation and the understanding of data, and how that data is to be handled.

3.5 Object Name Service

The functionality provided by the ONS system is similar to the services provided by the Domain Name System (DNS); however instead of translating host names to their underlying IP addresses for user applications, ONS translates an EPC into URL(s). The Object Name Service (ONS) in an EPC Network identifies a list of service endpoints associated to the EPC and does not contain actual data related to an EPC. These service endpoints can then be accessed over a network [30].

The ONS functions like a "reverse phone directory" since the ONS uses a number (EPC) to retrieve the location of EPC data from its databases. The ONS is based on existing DNS systems and thus queries to, and responses from, ONS adhere to those specified in the DNS standards (RFC 1034: Domain names, Concepts and facilities). This can be observed in the ONS resolution process outlined in Table 3.1. However, unlike the DNS, ONS is authoritative, that is the entity that retians control over the information about the EPC placed on the ONS is the same entity that assigned the EPC to the item.

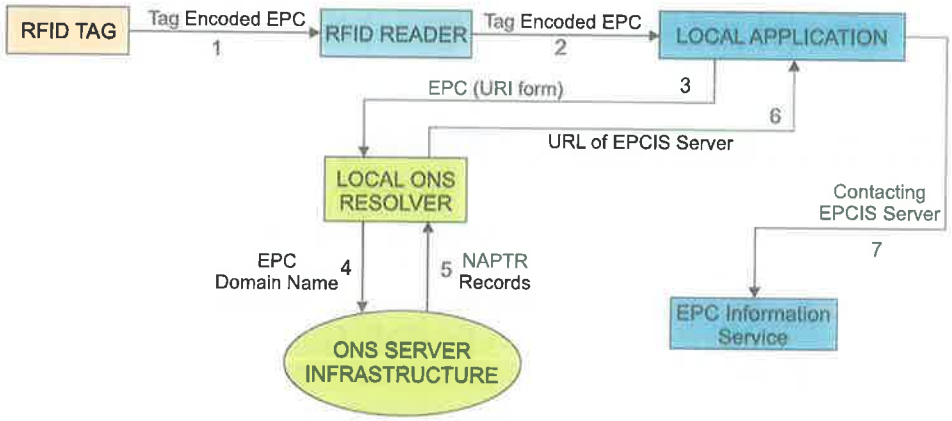

Figure 3.6 An overview of an ONS system functionality [31].

Table 3.1 Outlines a description of the object name resolution process illustrated in Figure 3.6.

	DESCRIPTION
1	A reader interrogates a tag and obtains the EPC in binary form. Example: Assuming that the EPC is of a 96 bit general identity type [14] and the following binary string is returned. This bit string represents an EPC encoded in the form identified in Figure 2.3 where the first 8 bits form the header, the next 28 bits the general manager number, followed by a 24 bit object class and the last 36 bits are the serial number. (00110101 000000000000000000000000010 000000000000000000011000 000000000000000000000000000110010000)
2	The EPC obtained (as a binary number) is passed to the local network application processes. Example: (00110101 000000000000000000000000010 000000000000000000011000 000000000000000000000000000110010000)
3	The EPC is then converted into URI form to provide a means by which application software is able to manipulate the EPC codes independent of any tag level encoding scheme, thus allowing software systems to treat the EPC data in a uniform manner, irrespective of how they are formed or obtained. All URIs are represented as Uniform Reference Names (URNs) using the standard format defined in RFC2141 using the URN Namespace epc. Example: For an EPC general identifier the URI representation is as follows: `urn:epc:id.gid.GeneralManagerNumber.ObjectClass.SerialNumbe r` In the above representation the *GeneralManagerNumber*, *ObjectClass* and the *SerialNumber* refer to the fields identified in Figure 2.3[14]. [urn:epc:id:gid:2.24.400]
4	URI is converted into domain name form so that a query in the form of a DNS query for a NAPTR record for that domain can be issued. In the event that the local ONS can not respond, the request is sent to the global ONS server infrastructure. Remove urn:epc Example: [id:gid:2.24.400] Remove serial number, since resolution down to the serial number level is not undertaken by the ONS and the resolution process stops at the Object Class level. This is a practical implementation as resolution to such granular level adds both complexities, cost and raises questions regarding the scalability of the architecture. Example: [id:gid:2.24] Invert the string (replace ':' with '.') Example: [24.2.gid.id] Append ".onsepc.com" Example: [24.2.gid.id.onsepc.com]

	DESCRIPTION
5	The ONS server infrastructure will generate a set of possible URLs that point to one or many services that are provided for the `ObjectClass` and `GeneralManagerNumber` in the ONS query. Example: In the example the list of URLs returned relate to the `GeneralManagerNumber` 2 and the `ObjectClass` 24. [http://bar.com/epcis.php; http://advark.com/sensor_is.asp; http://foo.com/epc_is.wsdl]
6	The correct URL is picked and extracted from NAPTR record by the application depending on the application type and need. Example: [http://www.foo.com/epc_is.wsdl]
7	The application system contacts the desired service. Example: In the example the application system can examine the wsdl file to obtain the desired service and then using the information provided therein contact that service. [http://www.foo.com/GetAppearancesByEPC] – Returns events for the specified EPC[29].

Figure 3.6 shows the overview of an ONS system where an EPC encoded in an RFID label is read by an RFID reader, where obtaining information associated with the EPC involves the resolution of the EPC through a query of the local ONS server to obtain a location of an appropriate application layer service (such as that provided by an EPCIS). In the event that the local ONS server is unable to satisfy the requests it is forwarded to a global ONS server infrastructure for resolution (the details of an EPC resolution process are outlined in Table 3.1).

The backbone of the ONS server infrastructure is an ONS Root providing a general manager id level (refer to Section 2.2.2) nameserver. The administration of the Root ONS server, according to current proposals, will be carried out by EPCglobal Inc.

The ONS needs to resolve to a greater depth than an IP address. An IP address is only sufficient to discover a location but it is not sufficient to locate a particular service needed by an application. It is possible to serve one service at each IP address and avoid the complications in the resolution process. Alternatively an IP address may host a number of other services. In a scenario where multiple services are provided at a specific IP address, the ONS will need to resolve down to a unique URL with the exact path and name of the service (such as a service provided by an EPCIS).

A challenging aspect of the resolution process is the ability to select the required URL since a list of URLs corresponding to a particular EPC may be returned by the ONS server (as shown in step 5). The format of the choices returned by ONS is defined in the Naming Authority Pointer (NAPTR). The complete definition of NAPTR can be found in [30]. In essence, NAPTR is a collection of information that points to a location on the World Wide Web when only a URI is provided. The NAPTR is formatted as shown below.

Order	Pref	Flags	Service	Regexp	Replacement

In a NAPTR, the URL is located in the [Regexp] field while [Order], [Pref] (Preference), and [Flags] are used to state the preference order of a list of URLs. [Service] is used to specify the type of service that is offered at the URI, such as an EPCIS. The [Replacement] field is not used for EPC Network purposes while the [Flag] field is set to 'u' to indicate that the [Regexp] field contains a URI [30].

It should be noted here that the ONS does not resolve queries down to the level of fully serialised EPCs. The depth of the query stops at the Object Class level (product type) of the EPC. The architectural and economic implications of specifying an ONS request to the serial number level was still an open issue at the time of compiling this chapter [30]. Thus queries directed at the serial number level must be resolved by the service locations obtained from the ONS query. A proposal for such a service is outlined as part of the EPCIS module in Section 3.6.

3.6 EPC Information Service

EPC Information Service (EPCIS) is the gateway between any requester of information and the databases containing that information. It primarily responds to queries from authorised entities that are expressed in a standard format; however the internal storage of that data within the databases may be in any format or standard. The EPCIS is the "interpreter" communicating between database(s) and application(s) and provides a standardised interface to the rest of the EPC Network for accessing EPC related information and transactions.

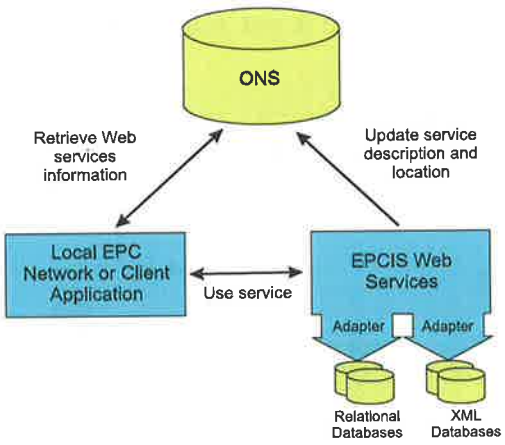

Figure 3.7 Interaction between EPCIS, ONS and external applications.

A possible interface for an EPCIS can be implemented by adopting web services technology. Web services technology based interface allows applications in the wider area network to utilise services provided by local EPC Information Services using a remote method invocation paradigm (refer to Figure 3.7). Such an architecture has the advantage of leveraging standardised XML messaging frameworks, such as that provided by SOAP

(Simple Object Access Protocol), and a description of the available services defined in terms of a WSDL (Web Services Description Language) file. Hence an application requiring information is able to access a WSDL file which has a description of the available service methods, the required input and output parameters to the methods and information to invoke those methods.

EPCIS provides a model for the integration of RFID networks across the globe. However it is important that EPCIS provides a secure communication layer so that local EPC Networks can retain the authority to determine access to information. WS-Security [32] is a candidate proposal for enhancing web services security that describes enhancements to SOAP messaging to provide message integrity and message confidentiality while proposed architectural extensions to the existing WS-Security profile [33] could provide access control as well as a federated security model for EPCIS.

Information about a particular EPC may be spread across a number of local networks (in the event of a supply chain application an object will pass through a number of physical locations, for instance manufacturers, distributors and retailers). The ONS does not resolve to the serial number level of the EPC and the DNS technology upon which the ONS is based also does not allow the fine grain resolution down to serial number levels. Resolution down to serial EPC level (to a specific object) is handled by the EPCIS Discovery Service (EPCIS-DS).

EPCIS-DS is best described as a "search engine" for EPC related data [30]. EPCIS-DS provides a method for custodians of a particular RFID tag data to update a register within the EPCIS-DS to indicate that that they are in possession of data related to an EPC. The register may contain a list of EPCIS URLs where such information may be obtained [34]. However unlike the ONS, the Discovery Service is not authoritative about information that it may contain regarding an EPC.

3.7 An EPC Network Application

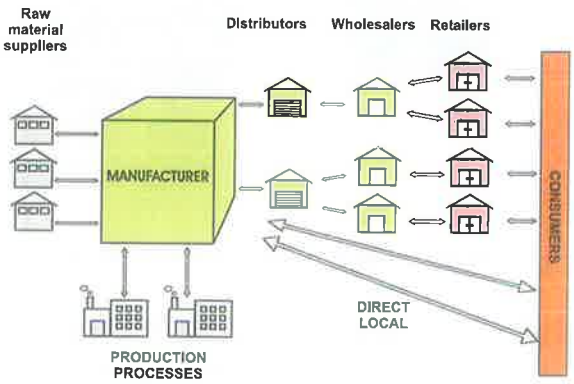

Figure 3.8 A very simple supply chain model.

Figure 3.8 shows an example of a simple supply chain model. The manufacturer is linked to raw material suppliers, production factories, distributors and some direct sales customers. The goods flow is from the manufacturer to distributors and then to wholesalers. The goods then flow from wholesalers to retailers, and then to consumers. In reality, multiple linkages or relationships are present, and the distribution system is much more complex than that shown in Figure 3.8. However, the simple model in Figure 3.8 is adequate to demonstrate a use case scenario.

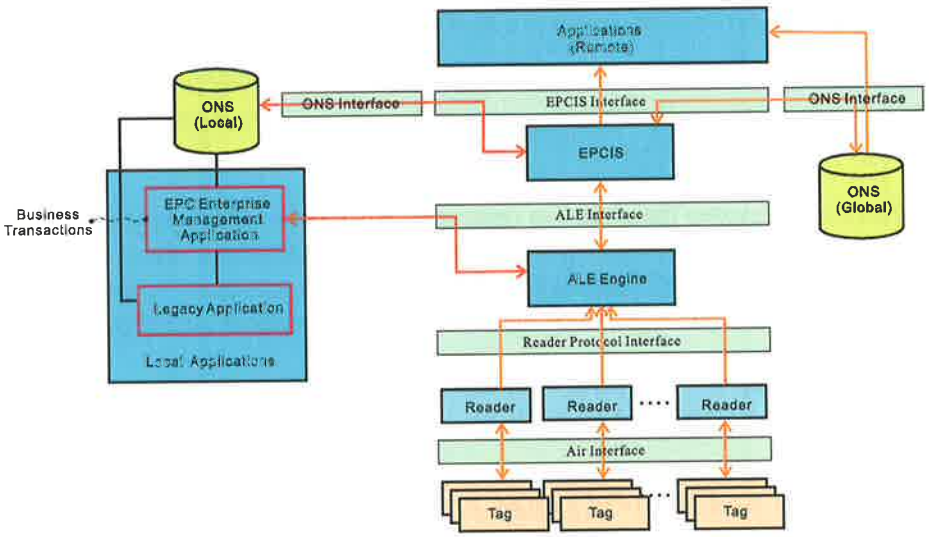

Figure 3.9 EPC Network utilisation.

Each party in this simple model, except for the consumers, is linked to the flow of goods through a wide area EPC network infrastructure. In a supply chain utilizing RFID technology all transactions including individual consumer purchases can be automated. In such an application the architecture of the EPC Network at each party is illustrated in Figure 3.9.

3.8 Supply Chain Management

RFID technology allows transactions to be automatically recorded and inventory levels to be updated in real time. When a customer purchases a product from a retailer, the RFID tag attached to the product will be scanned. The local system of the retailer will be informed of the current status of the product (i.e. product sold). If the stock level of the product has reached a critical level business logic systems at the local system can automatically send a purchase order to the wholesaler. The wholesaler, upon receiving the order request from that retailer, will automatically check through its inventory for stock availability. If the stock level is sufficient, the goods will then be dispatched to the retailer. If the stock level is low, either before or after the goods are dispatched, the wholesaler can place an order to the distributor automatically. A similar process will occur between the manufacturer and the distributor and the manufacturer and its raw material suppliers.

However, it should be noted here that the EPC network services do not allow an external party to influence the collection of data without a prior contractual agreement.

The real time visibility into the supply chain allows the dynamic transfer of supply and demand information along the supply chain and thus prevent what is termed the "bull whip effect" with the consequence of considerable savings to businesses and improved services to consumers [35]. Visibility provided by RFID technology allows the organisations along the supply chain to adjust rapidly to meet market conditions without the burden of large inventories to meet a foreseeable demand that may or may not occur.

Automating inventory control and smart supermarket shelves [36] will also ensure product availability and dramatically reduce customers lost to businesses through product unavailability.

The EPC Network provides the ability to capture an instance of the supply chain in real time. Enterprise applications exploiting knowledge based architectures can utilise the events stream collected from an EPC network to adjust business processes to minimise cost and increase efficiency while perhaps notifying managers when critical events occur.

Below are a few scenarios depicting how the EPC Network can enhance and bring about further improvements to supply chain logistics.

3.9 Solutions to Grey-Market Activity and Counterfeiting

Products falling into grey-markets have been one of the main issues that are of much concern in some industries, such as the pharmaceutical industry. One scenario where grey-market goods appear is in the re-importation into a particular country of products which were previously exported from that country, at prices far below those ruling in the domestic market of the exporting nation. These export prices may be for various reasons, including subsidies, are below the domestic prices. Some operators illegally import products back to the country of origin, where substantial profits can be made by selling the products at the higher domestic price.

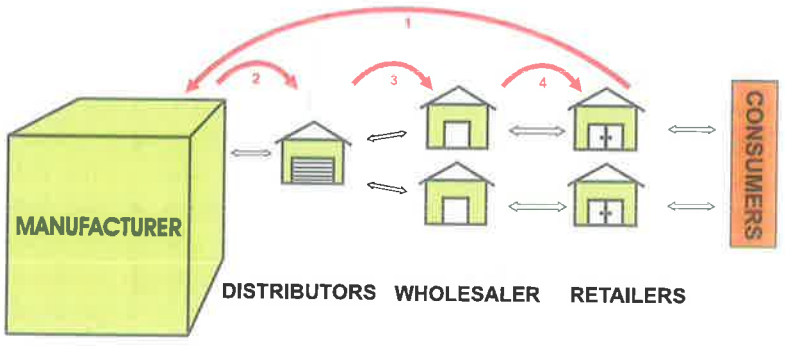

Figure 3.10 Counterfeit goods detection.

Product authenticity is also of concern to consumers. For example, in the pharmaceutical industry, most high value drugs are illegally manufactured by various manufacturers. Many of these products have the same potency as the authentic product, but some provide no benefit, or worse, cause harm. The ability of the EPC network to track and trace a product throughout the supply chain can prevent the entry of illegal counterfeits.

Figure 3.10 shows one possible data flow in an EPC enterprise system to fight grey-market distribution. When a product is sold and the EPC is recorded at the point of sale; 1) the local EPC Network will send a request to the manufacturer to check the validity of that particular item; 2) the manufacturer, upon receiving the request, will make an enquiry to its distributors; 3) the distributor, with the record of that product, will make an enquiry to its wholesalers; 4) the same situation repeats between wholesaler and retailer. If a valid supply path exists for that product, the product is validated. Grey-market goods will not pass this examination and an alert can be sent to the appropriate authority in such a scenario.

For product authentication, a party such as the retailer will only need to send a request to the manufacturer. The manufacturer can check the EPC received with its database. The manufacturer will be able to validate that the product was indeed manufactured and is currently available at a specific location in the supply chain. The retailer will also be able to check for multiple sale records to ensure that the item label was not obtained from a previously sold item. Thus the ability to track and trace provide an electronic history of the item throughout its life cycle through the supply chain and thus help prevent counterfeiting and the re-sale of goods in grey markets. The term "electronic pedigree" has been coined to label the electronic history of an item's life throughout the supply chain.

However such a simple solution may not be sufficient when dealing with counterfeit tags introduced into the supply chain and other appropriate solutions need to be contemplated. Such issues related to security are considered in more detail in Chapter 9, Chapter 10, and Chapter 11.

3.10 Product Recall and Other improvements

Apart from the obvious advantages of supply chain visibility and the reduction in labour needed to check and scan shipments to confirm that the correct goods were received, there are many other advantages arising from the use of EPC Networks. These include, amongst many others, easy and efficient product recall and the convenient identification of returned or rejected goods, at all levels of the supply chain, from retailer back to manufacturer.

The most significant of the above application is product recall. It is possible to flag each EPC of a product destined for recall, and such products reaching each location of the supply chain can be detected and removed from the supply chain.

3.11 Conclusion

This chapter has introduced and elaborated on the technology and on the concepts of the EPC Network. An application of the network to supply chain management was also illustrated with a brief outline of how the services provided can be implemented leveraging the existing technologies provided by web services tools and standards.

The EPC Network is still a concept under development. The functionality of the EPC Network can be summarised as providing the linkages between all physical objects with RFID tags, the management of the vast volume of data generated by readers and the provision of a universal query system for accessing and sharing information that describes objects over the Internet for access by remote services.

With a clear understanding of a modern RFID system, it is possible to focus on addressing various technical issues. The following chapter will illuminate the electromagnetic concepts that need to be understood when dealing with RFID technology and provide and extend the electromagnetic theory of coupling volume useful for analysing and improving the performance of RFID systems.

Chapter 4

ELECTROMAGNETICS AND COUPLING

In optimising the performance of passive RFID labels, the most important issues governing their practical applications are related to the levels of environmental noise and the strength of the electromagnetic coupling links.

Electromagnetic coupling relations explore the concepts behind powering up of labels. There are number of electromagnetic issues to consider and understand before performance improvements can be made to RFID systems. However, before such an understanding can be developed it is important to review and comprehend some of the basic principles of electromagnetic fields and waves.

This chapter will provide an outline of Maxwell's electromagnetic laws, and their effect upon electromagnetic field creation and field propagation. The significance of coupling volume theory in evaluating the performance of RFID systems is introduced using coupling volume theory for near field magnetic fields, while an extension of the theory is presented to cover near field electric fields to complete the previous formulation of the theory first presented in [10].

4.1 Electromagnetic Fields

Though the word "field" has numerous meanings, in the context of electromagnetics it is a vector quantity which has a value (generally time varying) at all points of a region of three-dimensional space.

4.2 Fundamental Laws of Electromagnetics

James Clerk Maxwell was first to correctly assemble the complete laws of electrodynamics which satisfied the continuity equation given in (4.1) for the law of conservation of charge. Maxwell's equations form the four fundamental equations governing the behaviour of electric and magnetic fields. The equations are also known as Faraday's law, Ampere's law as modified by Maxwell, Gauss' law for electric flux, and Gauss' law for magnetic flux [6, 9, 10, 37, 38, 39, 40 and 41].

$$\frac{d\rho}{dt} + \nabla \cdot \mathbf{J} = 0 \tag{4.1}$$

4.2.1 Faraday's Law

Faraday's law defines the relationship between magnetic fluxes and electric fields. The law states that the circulation of the electric field vector **E** around a closed contour is equal to minus the time rate of change of magnetic flux through a surface bounded by that contour, where the positive direction of the surface being related to the positive direction of the contour by the right hand rule.

$$\oint_c \mathbf{E} \,.\, d\boldsymbol{r} = -\frac{d}{dt}\oint_s \mathbf{B} \,.\, d\boldsymbol{s} \tag{4.2}$$

4.2.2 Ampere's Law as Modified by Maxwell

Ampere's law identifies the relationship between current and magnetic field. The law states that the circulation of the magnetic field vector **H** around a closed contour is equal to the sum of the conduction current and the displacement current (time varying electric flux density integrated over a surface) passing through a surface bounded by that contour, with again the right hand rule relating the senses of the contour and the surface.

$$\oint_c \mathbf{H} \,.\, d\boldsymbol{r} = \oint_s \mathbf{J} \,.\, d\boldsymbol{s} + \frac{d}{dt}\oint_s \mathbf{D} \,.\, d\boldsymbol{s} \tag{4.3}$$

4.2.3 Gauss' Law for Electric Flux

The total electric flux (defined in terms of the **D** vector) emerging from a closed surface is equal to the total conduction charge contained within the volume bounded by that surface.

$$\oint_s \mathbf{D} \,.\, d\boldsymbol{s} = \oint_v \rho \,.\, d\boldsymbol{v} \tag{4.4}$$

4.2.4 Gauss' Law for Magnetic Flux

The total magnetic flux (defined in terms of the **B** vector) emerging from any closed surface is zero.

$$\oint_s \mathbf{B} \,.\, d\boldsymbol{s} = 0 \tag{4.5}$$

Using Gauss' and Stokes' laws of mathematics and the definitions

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \text{ and } \mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}), \tag{4.6}$$

these laws may be expressed, when the fields are spatially continuous, in the differential form as given in (4.7).

$$\begin{aligned} \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{4.7}$$

4.2.5 Concept of a Source and a Vortex

Maxwell's equations are often more conveniently interpreted using the simple concepts of a source and a vortex put forward by Helmholtz and using the field pictures of Michael Faraday [37, 38 and 39]. Helmholtz's interpretation of the above equations state that the electric field vector **E** can have vortices caused by changing magnetic flux, the magnetic field **H** can have vortices caused by conduction or displacement currents; the electric flux density **D** can have sources caused by conduction charge density; and the magnetic flux density vector **B** can have no sources.

In linear media, some of the statements about **D** and **B** can be extended to **E** and **H** as well, but when non-uniform fields and boundaries are considered, it can be shown that **E**, **D**, and **H** can have both sources and vortices, but **B** is alone in that it can have no sources.

Figure 4.1 and Figure 4.2 below provide conventional illustrations of the source nature of the electric field and the vortex nature of a magnetic field. The figures also illustrate two of the most important boundary conditions which apply when any electric field **E** or a time varying magnetic field **H** approaches a conducting surface.

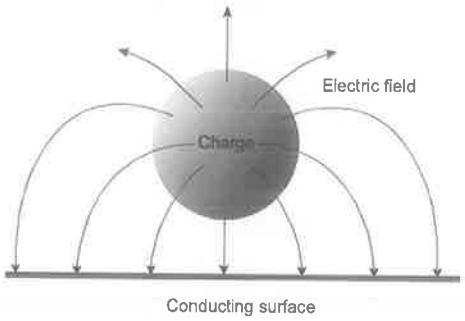

Figure 4.1 Concept of a source illustrated using an electric field near a conducting surface [37].

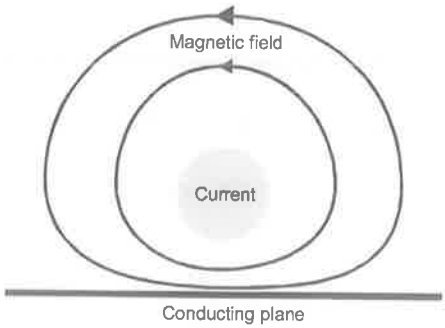

Figure 4.2 Concept of a vortex illustrated by an oscillating magnetic field near a conducting surface [37].

4.3 Boundary Conditions

Some electromagnetic boundary conditions were introduced in the previous section. The following is a complete statement of those boundary conditions derived from the fundamental laws of electromagnetics provided in Section 4.2.5.

Overlooking the properties on any materials involved, the tangential component of **E** is continuous across any boundary; the normal component of **B** is continuous across such a boundary; the normal component of **D** may be discontinuous across a boundary, with a discontinuity being equal to any conduction charge density ρ_s^c per unit area on the surface; and the tangential component of **H** may be discontinuous across a boundary, with the discontinuity being equal to in magnitude and at right angles in direction to a surface current density flowing on the surface.

Accounting for the restrictions imposed by the properties of the materials which may exist on one or other side of the boundary, the following conclusions can be drawn.

- The electric field is continuous across the boundary for all materials and time variations
- There are no electric fields or fluxes, or time-varying magnetic fields or flux densities inside a good conductor
- A surface current density can exist only on the surface of a perfect conductor
- The time-varying charge density cannot exist on the surface of a perfect insulator, although a static surface charge density can

4.4 Electromagnetic Waves

The resulting corrected equations by Maxwell given in (4.7) predicted electromagnetic waves, later experimentally confirmed by Hertz. Maxwell's equations describe the behavior of electromagnetic fields at every point in space and instant in time relative to the position and motion of charged particles. The following equations express (4.7) in time harmonic (sinusoidal) from in the frequency domain using complex phasors representing sinusoidal steady state oscillations at an angular frequency ω.

$$\begin{aligned} \nabla \times \mathbf{E} &= -j\omega\mathbf{B} \\ \nabla \times \mathbf{H} &= \mathbf{J} + j\omega\mathbf{D} \\ \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{4.8}$$

Where:

$\mathbf{E}$ = complex phasor representing the electric field (V/m)

$\mathbf{H}$ = complex phasor representing the electric field (A/m)

$\mathbf{D}$ = complex phasor representing the electric field density (C/m^2)

$\mathbf{B}$ = complex phasor representing the magnetic field density (T)

$\mathbf{J}$ = complex phasor representing the current density (A/m^2)

ρ = volume conduction charge density (C/m^3)

In free space or a homogeneous linear medium (that is in a region with no charge or current density with constant μ and ε and zero conductivity) sinusoidal steady state electric fields and magnetic fields are orthogonal with the direction of propagation being perpendicular to

both fields. Then equation (4.8) has the following frequency domain formulation. Such a wave propagating through a medium is called a Transverse Electromagnetic Wave (TEM).

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{4.9}$$

$$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} \tag{4.10}$$

$$\nabla \cdot \varepsilon\mathbf{E} = 0 \tag{4.11}$$

$$\nabla \cdot \mu\mathbf{H} = 0 \tag{4.12}$$

Combining equations (4.9) and (4.10) yields the Helmholtz equation.

$$\nabla^2 \mathbf{E} = -\varepsilon\mu\omega^2 \mathbf{E} \tag{4.13}$$

Equation (4.13) represents a three-dimensional wave equation which can be used to prove the existence of transverse electromagnetic waves. There are accordingly two independent solutions which can in fact be chosen in many ways. One such choice is linearly polarised waves, and one choice of linearly polarised wave solutions is to make the electric field direction along the x axis. The equations of such a TEM wave can be described by an **E** and **H** equation that describe the propagation of the time and space oscillating electric and magnetic fields. The propagating **E** field vector is described by equation (4.14) where E_0 is a constant defining the wave amplitude.

$$\mathbf{E}(z,t) = E_0 \sin(\omega t - \beta z)\mathbf{a}_x \tag{4.14}$$

The magnetic **H** field is symmetric to the **E** field and 90 space degrees displaced from it. The equation for the **H** field has a similar form as the **E** field:

$$\mathbf{H}(z,t) = H_0 \sin(\omega t - \beta z)\mathbf{a}_y. \tag{4.15}$$

The solution to equation (4.13) is as given in equation (4.14) and (4.15) for free space propagation of electromagnetic waves, where β is the free space wave number given by

$$\beta = \omega\sqrt{\varepsilon_0\mu_0} = \frac{\omega}{c} = \frac{2\pi}{\lambda_0}. \tag{4.16}$$

Where:

c = speed of light (3×10^8 m/s)

ε_0 = free space permittivity (8.85×10^{-12} F/m)

μ_0 = free space permeability ($4\pi\times10^{-7}$ H/m)

The relationship between magnetic flux density and magnetic fields, electric flux density and electric fields is given in equation (4.6) for a general medium; however in free space equation (4.6) will reduce to the following due to the absence of polarization or magnetisation vectors.

$$\mathbf{D} = \varepsilon_0 \mathbf{E} \text{ and } \mathbf{B} = \mu_0 \mathbf{H} \tag{4.17}$$

The ratio between the peak E_0 and H_0 provides a quantity having the units of resistance called the intrinsic impedance of the medium. In free space this is defined as below [37].

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \cong 120\pi\Omega \cong \quad 377\Omega \tag{4.18}$$

4.5 Retarded Potentials

The retarded potentials listed below may be regarded as integral solutions for Maxwell's equations which are available when charge and current distributions are known.

For the calculation of electric and magnetic fields at a point r_2 caused by a distribution of sinusoidally oscillating charge and current at points r_1 over a volume v we may make use of the retarded potentials

$$\Phi(r_2) = \frac{1}{4\pi\varepsilon_0} \int_v \frac{\rho(r_1) e^{-j\beta r_{12}}}{r_{12}} dv \tag{4.19}$$

and

$$\mathbf{A}(r_2) = \frac{\mu_0}{4\pi} \int_v \frac{\mathbf{J}(r_1) e^{-j\beta r_{12}}}{r_{12}} dv . \tag{4.20}$$

The fields in the sinusoidal steady-state can be derived from these potentials by the equations

$$\begin{aligned} \mathbf{E} &= -\text{grad}\,\Phi - j\omega\mathbf{A} \\ \mathbf{B} &= \text{curl}\,\mathbf{A} \end{aligned} \tag{4.21}$$

The formulae given above may be used in the calculation of the electromagnetic fields launched by time varying electric charge.

4.6 Radiation

Electromagnetic fields are launched by sources of time varying electric charge. Radiation is used to describe the phenomenon of electromagnetic waves propagating from a time varying electric charge. Radiation is a process by which electric and magnetic energy can be transmitted. Conducting or dielectric structures called antennas allow the efficient radiation or propagation of electromagnetic waves into free space.

Infinitesimal electric or magnetic dipoles as expressed below [37 and 39] form two ideal and simple conceptual antennas that can be used as building blocks from which fields of more practical antenna can be calculated.

4.7 Electric Dipole

In spherical polar coordinates at a point $P(r, \theta, \phi)$ the non-zero field components of an oscillating small electric dipole of length L carrying a current **I** and of moment **P** where $j\omega\mathbf{P} = \mathbf{I}L$ are given below.

$$\mathrm{E}_r = \frac{\beta^2 j\omega \mathrm{P}\eta}{4\pi}\left(\frac{1}{(\beta r)^2} - \frac{j}{(\beta r)^3}\right)e^{-j\beta r} 2\cos\theta \tag{4.22}$$

$$\mathrm{E}_\theta = \frac{\beta^2 j\omega \mathrm{P}\eta}{4\pi}\left(\frac{j}{(\beta r)} + \frac{1}{(\beta r)^2} - \frac{j}{(\beta r)^3}\right)e^{-j\beta r} \sin\theta \tag{4.23}$$

$$\mathrm{H}_\phi = \frac{\beta^2 j\omega \mathrm{P}}{4\pi}\left(\frac{j}{(\beta r)} + \frac{1}{(\beta r)^2}\right)e^{-j\beta r} \sin\theta \tag{4.24}$$

4.8 Magnetic Dipole

In spherical polar coordinates at a point $P(r, \theta, \phi)$ the non-zero field components of an oscillating small magnetic dipole of moment $M = IA$ are

$$\mathrm{H}_r = \frac{\beta^2 j\omega\mu_o \mathrm{M}}{4\pi\eta}\left(\frac{1}{(\beta r)^2} - \frac{j}{(\beta r)^3}\right)e^{-j\beta r} 2\cos\theta \tag{4.25}$$

$$\mathrm{H}_\theta = \frac{\beta^2 j\omega \mathrm{M}}{4\pi\eta}\left(\frac{j}{(\beta r)} + \frac{1}{(\beta r)^2} - \frac{j}{(\beta r)^3}\right)e^{-j\beta r} \sin\theta \tag{4.26}$$

$$\mathrm{E}_\phi = -\frac{\beta^2 j\omega \mathrm{M}}{4\pi}\left(\frac{j}{(\beta r)} + \frac{1}{(\beta r)^2}\right)e^{-j\beta r} \sin\theta \tag{4.27}$$

4.9 Transmitting Antenna Concepts

Radiation emanating from an antenna can be calculated using the concept of retarded potentials described in Section 4.5 and outlined in more detail in [37], [38] and [39]. These are just the same form as the electrostatic scalar potential and the magnetostatic vector potential, but have the added concept that the potential propagates away from the source (or vortex) at the speed of light.

Antenna characteristics such as radiation pattern, radiation intensity, directive gain and power gain, along with antenna efficiency and lumped element models of antennas are used to distinguish different antennas. These concepts can be studied in more depth with the aid of [38], [41], [42] and [43].

Nevertheless it is appropriate to introduce and define the important antenna parameters, radiation quality factor, and antenna quality factor that will be used throughout this dissertation. The antenna quality factor Q_a of an antenna is defined as

$$Q_a = \frac{\text{Impedence of the self - inductance (or capacitance) at resonance}}{\text{Radiation resistance of the antenna + loss resistance of the antenna}}. \tag{4.28}$$

While the radiation quality factor Q_r of an antenna is defined as

$$Q_r = \frac{\text{Impedence of the self - inductance (or capacitance) at resonance}}{\text{Radiation resistance of the antenna}}. \tag{4.29}$$

4.10 Characteristics of Near and Far Fields

The field equations for a magnetic and an electric dipole show that the distance r, where $r < 1/\beta = \lambda/2\pi$, is of significance in determining the nature of the fields surrounding the dipoles. Within this distance the dominant fields may be recognised as being the same as the energy storage fields (similar to electrostatic or magnetostatic dipoles). This region is known as the near-field region, and the dominant fields in the region are called the near fields, and simply store energy that periodically emerges from, and later disappears back into, the dipole. Near field analysis do not require all of Maxwell's equations but Gauss' law, Ampere's law before being corrected by Maxwell, and Faraday's law are sufficient.

The region beyond the near field region, that is when $r > (2D^2/\lambda)$ where D is the largest linear dimension of the antenna, where the dominant fields are those associated with energy propagation by electromagnetic waves away from the sources, is known as the far-field region. The dominant fields therein are called the far fields, and transport energy continuously away from the dipoles. Contrasting with near field analysis, far field analysis requires the use of all of Maxwell's equations outlined in Section 4.2.

The region between the near field and the far field, is the mid field region. This is the transition region where the antenna radiating pattern is taking shape but is not fully formed

and the wave fronts are becoming approximately planar. This is a region of interest considered in Chapter 5.

EM fields are created (energy storage field) and radiated (propagating field) by reader antennas. This is the function of a reader antenna in its role as a transmitter. Reader antennas designed for labels operating in the near field attempt to create large energy storage fields while reader antennas designed for labels operating in the far field attempt to minimise the energy stored and to maximise the radiated energy.

Labels operating in HF region operate in the near field and the label antennas are all electrically small. Hence label antennas for operation in the HF region need to be designed to couple to either the near field electric or the near field magnetic field and as such the interrogators need to deliberately create large volume electric or magnetic fields in the near field.

4.11 Near and Far Field Measures

A linearly polarised magnetic field expressed by a peak value phasor **H** can be described by the two measures of the exciting field [37] given in (4.30) and (4.31).

$$\text{Radial component of Poynting vector } S_r = \frac{\eta |\mathrm{H}|^2}{2} \text{ in Wm}^{-2} \tag{4.30}$$

$$\text{Volume density of reactive power } W_V = \frac{\omega\mu_0 |\mathrm{H}|^2}{2} \text{ in VAm}^{-3} \tag{4.31}$$

A linearly polarised electric field expressed by a peak value phasor **E** can be described by the two measures of the exciting field given in (4.32) and (4.33) [37, and 39].

$$\text{Radial component of Poynting vector } S_r = \frac{|\mathrm{E}|^2}{2\eta} \text{ in Wm}^{-2} \tag{4.32}$$

$$\text{Volume density of reactive power } W_V = \frac{\omega\varepsilon_0 |\mathrm{E}|^2}{2} \text{ in VAm}^{-3} \tag{4.33}$$

From equation (4.30) to (4.33), it is clear that once in the far field the reactive power in the far field is given by $W_v = \beta S_r$, where β is the propagation constant at the frequency in use, since $\beta = \omega\sqrt{\mu_0\varepsilon_0}$.

4.12 Reciprocity

The integral form of the Lorenz reciprocity relation for two solutions $\mathbf{E}_1$, $\mathbf{H}_1$ and $\mathbf{E}_2$, $\mathbf{H}_2$ of Maxwell's equations in the same region and at the same angular frequency ω is, under appropriate conditions,

$$\int_S (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot \mathbf{ds} = \int_v (\mathbf{J}_1 \cdot \mathbf{E}_2 - \mathbf{J}_2 \cdot \mathbf{E}_1)\, dv \qquad (4.34)$$

where the integral is over a closed surface S bounding a volume v [37 and 39]. In the derivation, which proceeds from Maxwell's equations, it is allowed that the material parameters be characterised by complex (to allow for losses) and possibly symmetric tensor (to allow for anisotropy) dielectric permittivities and magnetic permeabilities, but gyromagnetic behaviour of magnetic materials which is characterised by skew-symmetric tensors is not permitted [37].

In certain cases the right-hand side becomes zero. These cases include those where the surface is a conducting surface, where the surface is at great distance from sources confined to a finite region, where the surface encloses no currents, where currents flow only by the mechanism of drift, (but not by diffusion, as they can in a semiconductor) and where the surface encloses all sources.

The theorem has in electromagnetic theory, and on the simplification of electromagnetic theory known as lumped circuit theory, some profound consequences. These include: the symmetry of impedance, admittance and suitably defined scattering matrices; the interchangeability with matched sources and loads of transmitting and receiving antennas; the gain of a lossless transmitting antenna being related to the effective area of the same antenna used as a receiver; and the propagation loss from an interrogator to label being equal to the propagation loss from a label to a receiver when antennas have single ports. The later is an important result that will be used in the following chapters dealing with electromagnetic coupling between antennas in relation to RFID systems.

4.13 RFID Label Antenna and Reader Antenna Coupling

Analysing and improving the performance of RFID systems requires considering whether the labels are placed in the far (propagating) or near (energy storage) fields of the interrogator antenna. When an antenna is of small gain, the distance which divides the near and far fields is given by the size of the radian sphere of radius $r = \lambda/(2\pi)$, where λ is the free space electromagnetic wavelength at the operating frequency.

Antenna designs are influenced by a range of issues, such as the region of label operation (near or far), the coupling field (electric field or magnetic field), the regulatory constraints, and the environment in which they operate. For example, an environment with many metal structures can affect the time varying EM fields, and thus affect the performance of an RFID system. Illustrations of designing antennas to suit their environment of operations are presented in Chapter 6, Chapter 7 and in [40]. A vital aspect of the design process is to allow maximum coupling between reader antennas and label antennas.

4.13.1 Near Field Coupling - Magnetic Field

In the near field, energy storage fields created by reader antennas excite the labels. Power transfer relations are analysed using the laws of electrodynamics. The interactions between the reader antenna's exciting field and the label receiving antenna can be considered as weakly coupled inductors of self-inductances L_1 and L_2 and mutual inductance M. When both antennas are tuned to resonance with respective quality factors Q_1 and Q_2, it can be shown that the power P_2 dissipated in the losses of the label antenna is related to the power P_1 dissipated in the losses of the reader antenna by (4.35)[37 and 39].

$$\frac{P_2}{P_1} = k^2 Q_1 Q_2 \text{ where } k = \frac{M}{\sqrt{L_1 L_2}} \tag{4.35}$$

The above equation indicates the effect of the quality factor, Q, of the label and the reader antenna resonances in maximising the power transfer from a reader to an RFID label operating in the near field. The latter relationship is useful to show the role of the quality factor, Q, of the resonances in both the label and the interrogator coils in promoting power transfer, but it is not useful in separately optimising the properties of those two widely dissimilar elements.

4.13.2 Near Field Coupling - Electric Field

In RFID systems, the coupling can be via the magnetic field or the electric field. In near field systems it is almost always by way of the magnetic field. Nevertheless it is possible to couple to the electric field whether in the far field or the near field.

The energy transfer is provided by the electric flux terminating on the antenna surface and inducing a charge on the antenna. The induced charge will oscillate as the field around oscillates. The induced charges will thus produce a current.

4.13.3 Far Field Coupling

Far field operation of labels and readers are analysed using electromagnetic propagation theory. A simple approach to calculate the power coupled in the far field to a label with a lossless receiving antenna is using the available source power P_r from the label antenna. For calculation of the power P_r coupled in the far field the usual approach is to derive the available source power from the label antenna from (4.36) where A_e and g_r are the effective area of the label and gain of the label antenna.

$$P_r = S_r A_{er} = \frac{g_r \lambda^2}{4\pi} S_r \tag{4.36}$$

Effective area can be formally defined as

$$A_e = \frac{\text{Available source power of the antenna in a field}}{\text{Power density per unit area of that field}}.$$

The effective area for the far field is a concept unrelated to either a magnetic flux collection area or an electric flux collection area. It is unrelated to any physical area the antenna may possess, but it has the desirable property that it is possible to imagine that the label antenna collects all of the radiated power which flows through that effective area which may be thought of as surrounding the label antenna.

$$S_r = \frac{g_t P_t}{4\pi r^2} = \eta |\mathrm{H}|^2 \ \ \mathrm{Wm}^{-2} \tag{4.37}$$

Equation (4.37), where P_t is the power transmitted and r is the distance from the transmitter antenna to the label position assuming that the label has been placed in the direction of strongest radiation from the RFID interrogator (transmitter) antenna, defines the power flow per unit defined by the radial component of the Poynting vector. Then Lorenz reciprocity theorem of electrodynamics may be used to show that the effective area of a receiving antenna is related to the gain g_r it would have in a transmitting role by the equations

$$A_{er} = \frac{g_r \lambda^2}{4\pi} \tag{4.38}$$

Combining (4.36), (4.37) and (4.38) gives the result in (4.39).

$$\frac{P_r}{P_t} = g_r g_t \left(\frac{\lambda}{4\pi r}\right)^2 = \frac{A_{et} A_{er}}{\lambda^2 r^2} \tag{4.39}$$

Equation (4.39) provides the result usually used to evaluate the power extracted by an RFID label in the far field.

The significance here is to note that improving the power transfer from a reader to a label involves designing efficient radiators, as the gain of an antenna does not vary much with size. That is, unless the antenna is very large, akin to a multi-element array, or a large dish, the gain turns out be close to unity. Commonly used values of gain for RFID label antennas is a gain of 1.5 for a small dipole or a small loop, or a gain of 1.64 for a half wave dipole antenna [38]. However, radiated power must still be within EMC regulations stated for the frequency of operation.

In contrast to near field analysis, antenna quality factor Q_a has completely different effects in far field analysis. While in the near field, a large quality factor of resonance provides increased performance, in far field analysis a large Q_a simply restricts antenna bandwidth (when the bandwidth of the antenna is defined using Q_a or when the ohmic losses of the antenna are negligible), while a smaller Q_a results in broad banding albeit with increased

losses, some or most of which may be ohmic losses. These issues are discussed in more detail in Chapter 7.

4.14 Development of Coupling Volume Theory

The formulation of the near field coupling provided in Section 4.13.1 is useful to show the role of the quality factor, Q, of the resonances in both the label and interrogator coils in increasing power transfer; however it does not allow the individual optimisation of the completely dissimilar elements. The development of coupling volume theory addressed the latter problem for near field magnetic fields [10].

4.14.1 Near Field – Magnetic Field

The disadvantage outlined above was overcome in [10] by the formulation of the coupling volume theory for near field magnetic fields. Coupling volume theory can be described by focusing on the energy storage measure of the exciting field, which is the reactive power per unit volume in the field created by the interrogator at the label position. In terms of that measure, it is possible to define a figure of merit of a label antenna as the ratio [37 and 39]

$$V_c = \frac{[\text{Reactive power flowing in the untuned label coil when it is short circuited}]}{[\text{Volume density of reactive power created by the interrogator at the label position}]}.$$

V_c has the dimensions of volume, and is thus is called the *coupling volume* of the label antenna. The companion concept of a *dispersal volume* V_d can be defined for the performance of the interrogator antenna and it is the ratio

$$V_d = \frac{[\text{Reactive power flowing in the inductor of the interrogator field creation coil}]}{[\text{Volume density of reactive power created by the interrogator at the label position}]}.$$

When both antennas are tuned it is possible to show

$$\frac{P_2}{P_1} = \frac{V_c}{V_d} Q_1 Q_2. \tag{4.40}$$

The benefit of this formulation is that the coupling volume is a property of the label parameters alone, and the dispersal volume is a property of the interrogator antenna parameters alone (and of the label position), and separate optimisation becomes possible, whereas k^2 in (4.35) is a complex function of the entire system geometry.

Coupling volumes for various label antennas are readily determined. The following sections provide the coupling volumes of a magnetic loop and a solenoid.

4.14.1.1 Coupling Volume of a Magnetic Loop

For a planar coil, which in its idealised state has no physical volume, the coupling volume is given by

$$V_c = \frac{\mu_0 A^2}{L} \tag{4.41}$$

where A is the flux-collecting area (incorporating by summation an area for each turn) of the coil, and L is the self inductance [10].

4.14.1.2 Coupling Volume of a Solenoid

For a long air cored cylindrical solenoid of cross-sectional area A and length l, the coupling volume is given by [10]

$$V_c = A^2 l \tag{4.42}$$

Hence for an air cored solenoidal antenna the coupling volume is approximately the volume of that antenna. It can be shown that when a magnetic core is in place that volume becomes multiplied by the relative effective permeability

$$\mu_{er} = \frac{\mu_{ir}}{1 + N(\mu_{ir} - 1)} \tag{4.43}$$

where μ_{ir} is the relative intrinsic permeability of the core material and N is the demagnetising factor in the direction of the interrogator field.

4.14.2 Near Field – Electric Field

The issues of understanding and optimising coupling to the electric field are important. The coupling volume theory for near field magnetic fields in [10] can be extended to address near field electric fields. This will allow comparisons between different antennas sensitive to both or either electric and magnetic fields for their effectiveness and efficiency in terms of their actual physical volumes and their coupling volumes. This section will establish the coupling volume theory for near field electric fields by extending the theory presented in [10] and thus completing the formulation of the theory.

Similar to the description in Section 4.14.1 electric field coupling volume theory can be described by considering the energy storage measure of the exciting field, which is the reactive power per unit volume in the field created by the interrogator at the label position. Thus we can define a figure of merit of a label antenna as the ratio

$$V_c = \frac{[\text{Reactive power flowing in the untuned label capacitor when it is open circuited}]}{[\text{Volume density of reactive power created by the interrogator at the label position}]}$$

which has the dimensions of volume, and thus called the coupling volume of the label antenna. For the performance of the interrogator antenna, we can define the companion concept of dispersal volume V_d given by

$$V_d = \frac{[\text{Reactive power flowing in the capacitance of the interrogator field creation electrodes}]}{[\text{Volume density of reactive power created by the interrogator at the label position}]}.$$

When both antennas are tuned it is possible to show

$$\frac{P_2}{P_1} = \frac{V_c}{V_d} Q_1 Q_2 \tag{4.44}$$

The benefit of this formulation is that the coupling volume is a property of the label parameters alone, and the dispersal volume is a property of the interrogator antenna parameters alone (and of the label position), and separate optimisation becomes possible. The following section illustrates the calculation of the coupling volumes for various electric field sensitive label antennas using the coupling volume theory for electric fields established above.

4.14.2.1 Coupling Volume of a General Shape

In order to derive the coupling volume of antenna structures it is important to define a number of concepts. For a given antenna we can define an electric flux collecting area as the area in space required by the antenna to generate from the displacement current of the exciting field an oscillating current I when the antenna is placed in an oscillating electric field of flux density D. This area may or may not be equivalent to the physical area. This electric flux collecting area is denoted by the symbol A_f.

The formula for the phasor representing the current I, flowing from an antenna placed in an electric field with a flux density represented by the phasor D, and oscillating angular frequency ω, using the flux collecting area A_f of the antenna structure is

$$\mathrm{I} = j\omega \mathrm{D} A_f . \tag{4.45}$$

The reactive power flowing in the antenna with a self-capacitance C when the antenna is open circuit is given by

$$\frac{|\mathrm{I}|^2}{2\omega C}. \tag{4.46}$$

The reactive power W_V per unit volume in the field is obtained by

$$W_V = \frac{\varepsilon_0 |\mathrm{E}|^2 \omega}{2}. \tag{4.47}$$

Thus the coupling volume of an antenna structure with a self-capacitance of C can be obtained as

$$V_c = \frac{\varepsilon_0 A_f^2}{C}. \tag{4.48}$$

This formula provides the coupling volume of any antenna with a self capacitance of C and a flux collecting area A_f. It should be pointed out that A_f can easily be determined experimentally by measuring the current flowing from an antenna placed in an oscillating electric field of known strength.

$$A_f = \frac{\mathrm{I}}{\varepsilon_0 |\mathrm{E}| \omega} \tag{4.49}$$

4.14.2.2 Coupling Volume of a Rectangular Capacitor

The derivation discussed below will prove that the coupling volume of a parallel plate capacitor is equal to the physical volume of the capacitor reduced by a factor given by the relative permittivity of the dielectric. Thus for an air filled capacitor the coupling volume is shown to be that of the physical volume of the capacitor.

Consider a parallel plate capacitor antenna, as shown in Figure 4.3, consisting of two plates, each having an area A and immersed in an external electric field **E**. Then the charge Q distributed on the plate is given by

$$Q = \varepsilon_0 A |\mathrm{E}| \tag{4.50}$$

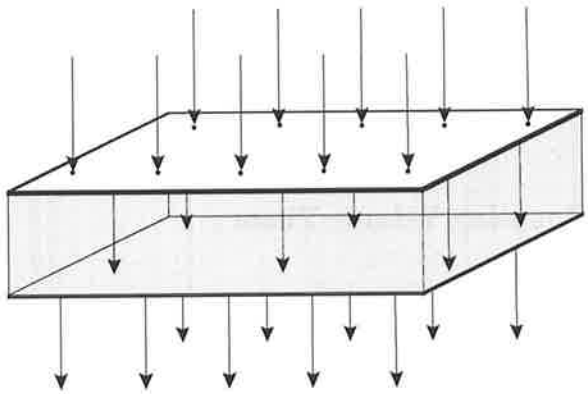

Figure 4.3 Field configuration for a parallel plate electric field antenna.

The capacitance C of a parallel plate capacitor is given by

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d}. \tag{4.51}$$

Now the flux density between the plates is the same as the flux density outside, but the field strength between the plates is $1/\varepsilon_r$ times the field outside. So the voltage V between the plates can be calculated by (4.52).

$$V = \frac{|\mathrm{E}|\, d}{\varepsilon_r} \tag{4.52}$$

The energy stored in the capacitor is obtained as

$$\frac{1}{2}\frac{\varepsilon_0 A |\mathrm{E}|^2\, d}{\varepsilon_r}. \tag{4.53}$$

Equation (4.54) gives the energy stored per unit volume in an electric field.

$$\frac{1}{2}\varepsilon_0 |\mathrm{E}|^2 \tag{4.54}$$

Hence coupling volume V_c of the parallel plate structure can be expressed as that given in (4.55).

$$V_c = \frac{Ad}{\varepsilon_r} \tag{4.55}$$

Examination of (4.55) reveals a number of results. Coupling volume has a dependence on the value of the relative permittivity ε_r, and a dielectric, if present, will reduce the coupling volume. This last result stands in contrast to the case for a rectangular magnetic coil, equation (4.41), where relative permeability μ_r multiplies the coupling volume. However, for an air dielectric capacitor, the coupling volume is equal to the physical volume. It can be observed that the coupling volume of a general structure reduces to the coupling volume of a parallel plate capacitor when

$$A_f = A \text{ (area of a plate) and } C = \frac{\varepsilon_0 \varepsilon_r A}{d}. \tag{4.56}$$

4.14.3 Far Field Coupling Volume Theory

In general it is possible to calculate the power coupled in the far field to a label with a lossless receiving antenna using (4.39), which is the Poynting vector-effective area formulation. However, it is possible to calculate the power coupled in the far field using the ideas expressed in coupling volume and dispersal volume formulations outlined in Section 4.14 albeit useful in different contexts. The formulation of far field coupling volume theory and its applications to RFID is explored in Chapter 8.

4.15 A Relation Between Electrostatic and Electrodynamic Theory

The concepts of electric flux collecting area, effective area and the reciprocity theorem have been introduced in the sections above. The radiation resistance of an antenna, which is a full electrodynamic theory concept also provides for a label antenna a means of calculating, using the Lorenz reciprocity theorem, the effective electric displacement current collecting area of an antenna, which is an electrostatic theory concept. While this is a significant development of electromagnetic theory, the relationship between the two theories and its practical applications is illustrated in Chapter 5 using the electrostatic theory and the full electrodynamic theory and applying them to the wedge above a ground plane antenna.

4.16 Conclusion

This chapter highlighted the fundamental laws that form the foundation for electromagnetic theory and its applications to RFID with emphasis on near field coupling. Previous formulations of coupling relations for near fields proved inadequate for increasing the performance of RFID systems. However coupling volume theory and its completion through to the extension of that theory to near field electric fields provides an opportunity for RFID engineers to optimise RFID systems by considering the label and the reader antennas separately. After considering the subject of label and reader antennas, coupling volume theory is revisited in Chapter 8. The following chapter will consider the design of interrogator antennas for improving the performance of HF RFID systems in view of the relaxed electromagnetic compatibility regulations in the European Union.

Chapter 5

NEAR FIELD INTERROGATOR ANTENNA DESIGN

The previous chapter considered relevant electromagnetic theory for improving RFID system performance by considering the coupling link between tag and reader antennas. This chapter focuses on the actual design of coupling elements, in particular, antennas for interrogators. The design of antennas for interrogators is influenced by regulatory limits, the frequency of operation and the environment in which they operate. A vital aspect of the design process is to allow maximum coupling between the reader antennas and label antennas in arbitrary configurations.

A proposed relaxation of electromagnetic compatibility constraints in the ISM HF region by the European Telecommunication Standards Institute (ETSI) [44] outlined in Annex 9 of the ERC/REC 70-03 provides for interesting consequences for HF RFID system operation whereby substantially increased label reading distances may be obtained. This chapter presents an overview of the changes to ETSI regulations and a number of possible interrogator antenna designs suitable for the HF spectrum while considering in detail the design and analysis of large loop antennas and monopole wedge antennas.

5.1 Electromagnetic Compatibility Constraints

Prior to delving into the effect of electromagnetic compatibility constraints normally applied to RFID systems it is important to remark that all of the regulations, whether at UHF or at HF, are enforced in the far field. However, as shown in Chapter 2, near and far fields scale differently with distance. In particular, the near field energy density per unit volume decreases as the inverse sixth power of distance from the antenna. The result is that close to the antenna, substantial energy densities may be obtained, but these diminish very quickly as distance increases.

Although the HF electromagnetic compatibility regulations, shown in Figure 5.1 below allow only minimal radiation (a radiated power of approximately 5 mW using an antenna with a gain of 1.76 dBi), well inside the radian sphere distance of $r = 2\pi/\lambda$, it is practicable at HF to obtain a sufficient energy storage field for the operation of an RFID label. However, as efforts to increase reading distance are made, the previously mentioned inverse sixth power of the reactive power density is generally sufficient to reduce the label energising signal to a level below that for practical operation before the boundary of the far field, where its less severe inverse square dependence of energy density is reached. Thus under previous regulations operation of HF systems is almost entirely confined to the near field and short distances.

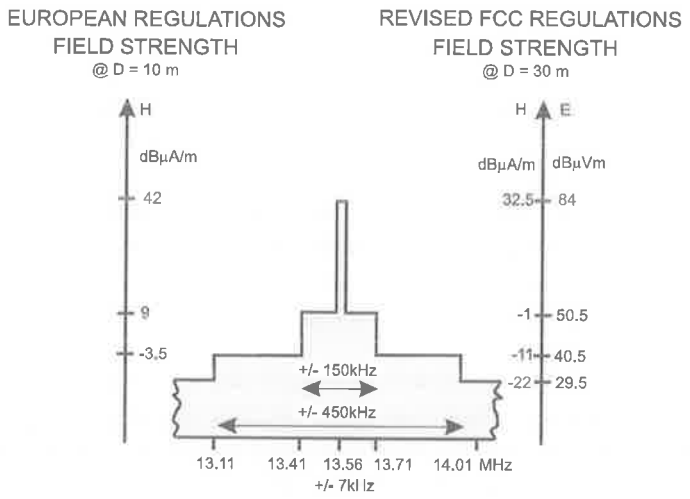

Figure 5.1 Previous HF electromagnetic compatibility regulations.

In an effort to improve the range of HF RFID systems efforts have been made, through the design of clever quadrupole antenna systems, to minimise far field radiation while enhancing close-in near fields. While these efforts do achieve some improvement, maintaining the necessary balance between antenna elements which are intended to produce far field cancellation is difficult to achieve in practice.

Focus on minimizing far-field radiation diverts attention from an alternative productive approach to gaining long interrogation range. This approach is that of using greater label antenna sizes. Laboratory tests with square loop label antennas of sizes in the range of 148 mm x 210 mm along with large interrogator antennas (occupying areas of 1 m x 1 m) radiating at the limits imposed by the revised FCC regulations [45] have shown that reading ranges will increase to approach 0.5 m to 1.2 m range under ideal conditions.

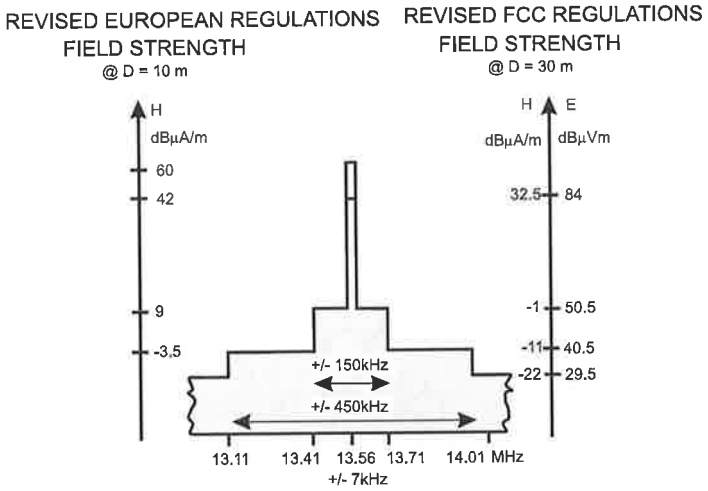

Figure 5.2 Revised HF electromagnetic compatibility regulations.

However laboratory tests conducted in view of recent changes to European regulations depicted in Figure 5.2 (that is radiating about 320 mW of power with an antenna gain of 1.76 dBi) with larger interrogator antennas and larger label antennas of sizes outlined previously have shown that reading ranges will increase to approach the mid field distance. In the mid field region it is reasonable to assume that many antenna fields will, in the ratio of mid field to far field amplitude, replicate the value that can be estimated from the dipolar field expressions given in Section 4.7 and Section 4.8. Then it is evident that it is not possible to have a strong mid field without a related strong value for the far field (that is without radiating some energy). The later conclusion does not significantly change with interrogator antenna size. In consequence, the rational approach to achieving long interrogation range includes canvassing for a greater allowed radiation from HF interrogator antennas, at least in respect of their carrier level, if not in the signalling sideband level.

5.2 Near Field Creation Interrogator Antennas

Figure 5.3 to Figure 5.6 below illustrate structures which may be useful in the HF region. Such structures can generate either electric or magnetic fields, either in the near-field, or the mid-field, that being the field at the boundary between the near field and the far field. As a consequence of the value (22 m) of the electromagnetic wavelength at 13.56 MHz, the structures are always electrically small in practice. Antennas from Figure 5.3 to

Figure 5.5 can be considered as creating mainly electric fields in the near-field, but as a consequence of the expressions given earlier for dipole fields, which apply to some extent to this situation, the structures will also create some lesser value of near magnetic field [37 and 39].

In the far-field these structures will create electric and magnetic fields in equal proportion, in the sense that for radiated fields in the far-field $|\mathbf{E}| = \eta|\mathbf{H}|$ and the stored energies per unit volume $\frac{1}{2}\varepsilon_0|\mathbf{E}|^2 = \frac{1}{2}\mu_0|\mathbf{H}|^2$.

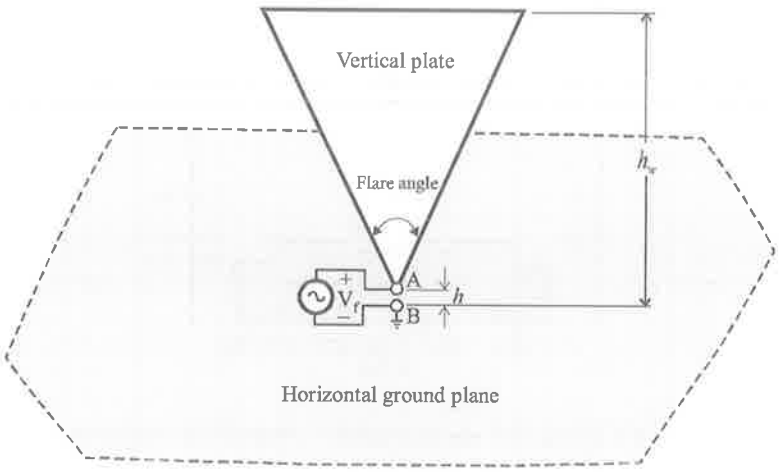

Figure 5.3 A wedge above a ground plane antenna.

Figure 5.4 A meander line antenna.

Figure 5.3 is a wedge above a ground plane antenna and an alternative is the meander line structure illustrated in Figure 5.4. Figure 5.5 below illustrates a top-loaded helical structure which may be useful for creating large volume near-fields. These structures are conceptual antenna designs to have similar field creation properties to that in Figure 5.3, with small variations in the impedance properties to provide simpler driving requirements.

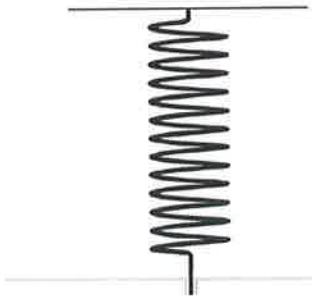

Figure 5.5 A top-loaded helical antenna.

As HF tags often couple to magnetic fields because they are less easily stopped by conducting materials than are electric fields, there is in fact a growing interest in the creation of strong near-field magnetic fields. The usual structure by means of which this is achieved is a small current carrying loop such as is illustrated in Figure 5.6. This loop has tuning and matching elements at the top end. A strip line transmission line (not visible in the image) on the underside of the right hand half conveys the driving signals from the connecting point at the centre of the bottom to the driving point terminals at the centre of the top. The examination of electric and magnetic dipoles in Section 4.7 and Section 4.8, repectively, shows that magnetic dipoles are superior to electric dipoles in the creation of magnetic fields in the near-field.

Figure 5.6 A loop antenna.

5.3 Interrogator Antenna Equivalent Circuits

There are numerous antenna designs, each with its own set of characteristics described by their gain, directivity, radiation pattern, effective length, efficiency and effective area [42, 43, 47 and 48]. In addition to the above characteristics, three primary parameters are often used to describe an equivalent circuit model for the antenna using lumped circuit theory. These parameters are indicated in Figure 5.7. Here R_r represents the radiation resistance; R_l represents the Ohmic losses in the antenna while X_A represents the reactance of the structure.

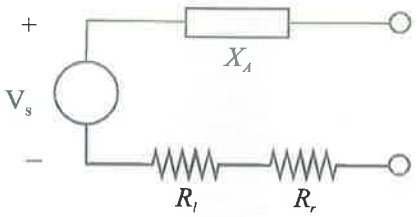

Figure 5.7 Antenna circuit model.

The range of validity of these equivalent circuits is where the reactance properties of the antenna may be described by a single parameter, L (inductance) or C (capacitance) or a combination of the two. When the antenna is large, as is the case for interrogator antennas considered in Section 5.2, reactance properties might well be described by an appropriate mixture of L and C.

5.4 Wedge Above a Ground Plane Antenna

Figure 5.8 A practical construction of the wedge above a ground plane antenna. The antenna was tuned to a 50 Ohm input impedance using a tapped inductor.

A number of near field creation structures were introduced in Section 5.2. The wedge above a ground plane (also referred to as the monopole wedge above ground) structure is investigated in this section as it presents a compact structure that can be incorporated into building infrastructure with the least amount of obstruction. Additionally, analysing and developing an ability to understand the behaviour of such a structure is aided by the work presented in [46]. Figure 5.3 shows a monopole wedge above ground antenna while Figure 5.8 depicts a practical construction of such an antenna in the laboratory. A three-parameter circuit model, which has been obtained by analysing the results for a monopole wedge above ground presented in [46], is shown in Figure 5.9.

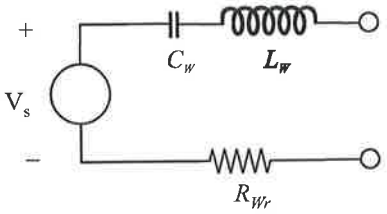

Figure 5.9 An equivalent circuit model for a wedge above a ground plane antenna.

From the circuit model of Figure 5.9 the reactance is given by

$$jX_W(\omega) = \frac{1}{j\omega C_W} + j\omega L_W \,. \tag{5.1}$$

Table 5.1 Expressions for evaluating the antenna parameters of a wedge above a ground plane antenna.

Capacitance (C_W) in Farads	$K_{WC}\varepsilon_0 h_W$
Inductance (L_W) in Henrys	$K_{WL}\mu_0 h_W$
Radiation Resistance (R_{WR}) in Ohms	$K_{WR}(\beta h_W)^2$

The model parameters identified in Figure 5.9 will vary for different flare angles and heights of the wedge. However, within the range of validity of the equivalent circuit, which depends upon the dimensions of the structure in relation to a wavelength, the radiation resistance, the capacitance and inductance can be expected to scale up with increasing height for a specific flare angle. Table 5.1 provides generic expressions for evaluating the radiation resistance, the capacitance and the inductance values for a wedge above a ground plane antenna. In Table 5.1 the constants K_{WC} and K_{WL} are dimensionless quantities while K_{WR} is measured in Ω. The specific values of the constants K_{WC}, K_{WL}, and K_{WR} depend on the flare angle of the monopole wedge above a ground plane antenna.

The parameters C_W and L_W can be conveniently obtained from a reactance $X_W(\omega)$ plot of a wedge above a ground plane antenna. Then the frequency of intersection of the reactance with the horizontal axis (that is where the reactance is zero), and the slope of the reactance curve at that point can be used to generate two simultaneous equations to compute C_w and L_w. Brown and Woodward's [46] experimental results for the radiation resistance of monopole wedge above ground plane antennas can be used to produce a value for the radiation resistance R_{Wr} identified in Figure 5.9. Both of these calculations have been performed for different antenna heights and flare angles to evaluate the constants K_{WC}, K_{WL}, and K_{WR} identified in Table 5.1 for calculating the equivalent circuit model parameters of wedge above a ground plane antennas.

Table 5.2 provides a set of values for K_{WC} and K_{WL}, and a set of values for K_{WR} derived from Brown and Woodward's [46] results. All of these results agree with the results of direct measurements of wedge reactance and capacitance, and numerical analysis performed using the method of moments confirm the measured self capacitance values of the monopole wedge above ground antennas found in the experimental observations.

Table 5.2 Flare angles and radiation resistance constants for a wedge above a ground plane antenna of height h_W.

Flare Angle	K_{WR} in Ω	K_{WL}	K_{WC}
5	15	0.2888	2.11
10	17	0.2823	2.35
30	22	0.2605	3.27
40	23	0.2470	3.95
50	24	0.2349	4.61
60	25	0.2250	5.00
90	30	0.2128	7.60

The significant finding is that, as expected, the low frequency impedance of a monopole wedge above ground antenna is mainly capacitive and thus the value of the capacitance C_W in the model provided in Figure 5.9 can be obtained by calculating the self-capacitance of the monopole wedge above ground.

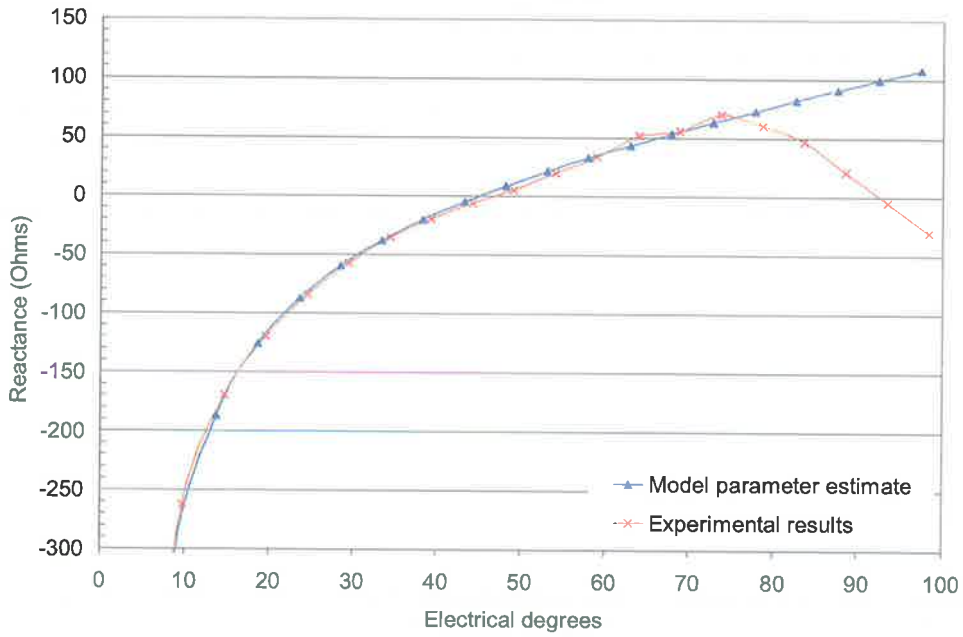

Figure 5.10 Reactance values obtained for a wedge above a ground plane antenna with a flare angle of 90 degrees and height h_W as indicated in Figure 5.3.

However the application of the model in Figure 5.9 and the derived expressions are only suitable for electrically small antennas obeying the strict limit given in (5.2).

$$h_W << \frac{\lambda}{6} \text{ or electrical degrees} << 60 \qquad (5.2)$$

In (5.2), h_W is the height of the antenna as indicated in Figure 5.3, λ is the wavelength, measured in metres, [39] and electrical degrees is evaluated as $(h_w \times 360)/\lambda$.

The graph in Figure 5.10 shows a comparison between the measured reactance values for a 90° flare angle monopole wedge above ground and the reactance determined using the expressions indicated in Table 5.1. It can be observed that the model parameters correctly estimate the measured reactance values within the bounds given in (5.2). Figure 5.11 is an illustration of the radiation resistance values evaluated using the expressions indicated in Table 5.1 and the constants calculated in Table 5.2 within the validity of the model parameters outlined in (5.2).

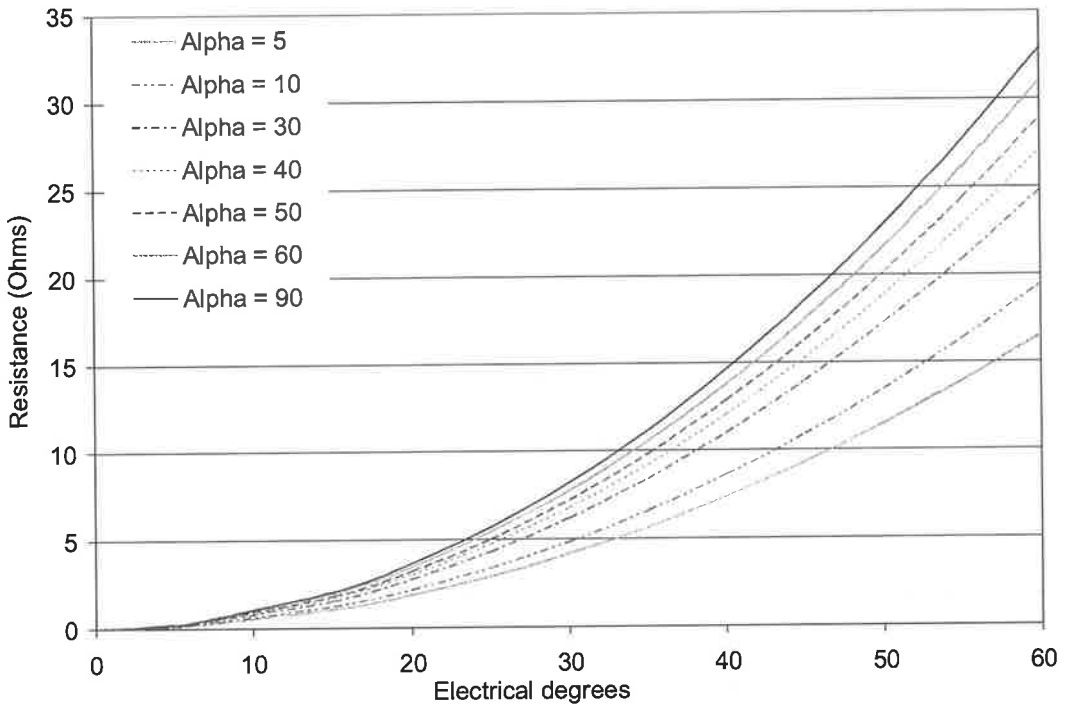

Figure 5.11 Radiation resistance values evaluated from the derived formula for various flare angles, denoted by alpha.

It is important to note that these results can be directly extended to analyse bow tie antennas using the method of images. Bow tie antennas and their application to RFID are considered in Chapter 6.

The radiation resistance parameter obtained has the significance of allowing the amount of radiated power to be calculated for a transmitting antenna using (5.3), where P_t is the radiated power and V_t is a peak value phasor representing the voltage across the radiation resistance R_r.

$$P_t = |V_t|^2 / 2R_r \tag{5.3}$$

5.5 A Relation between Electrostatic and Electrodynamic Theory

The concepts of electric flux collecting area, effective area and the reciprocity theorem were introduced in Chapter 4. The radiation resistance of an antenna, which is a full electrodynamic theory concept, also provides for a label antenna a means of calculating, using the Lorenz reciprocity theorem, the effective electric displacement current collecting area of an antenna. This relation between the two theories is illustrated below using the electrostatic theory and the full electrodynamic theory and applying them to the wedge above a ground plane antenna.

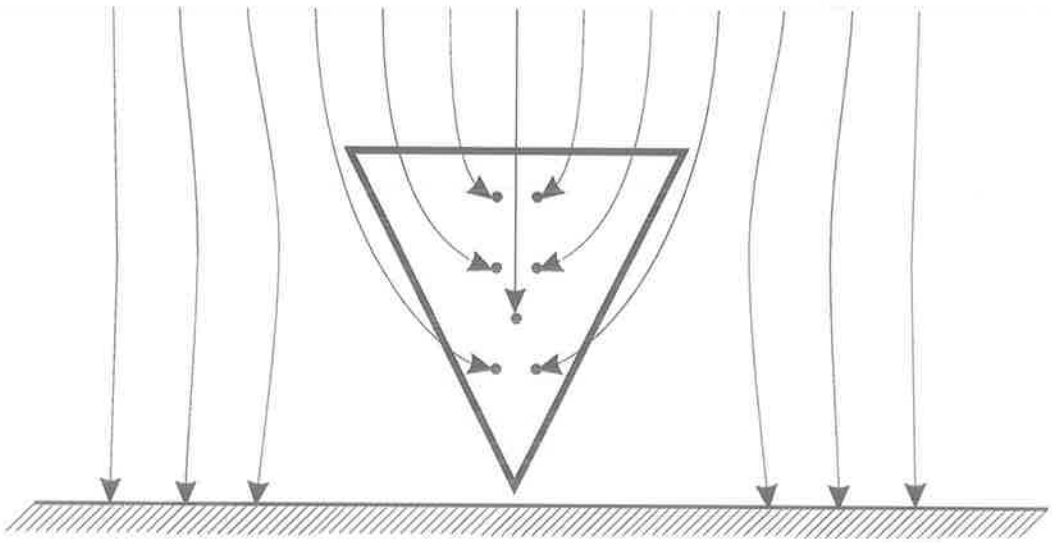

Figure 5.12 An illustration of the electric displacement current collecting area of a wedge above a ground plane antenna.

There are two methods for obtaining the short circuit current of the wedge above ground plane antenna. The simplest of which is an electrostatic experiment, as has been performed, which involved the concept of electric flux collecting area A_f discussed in Section 4.14.2.1 and as shown in Figure 5.12. The second method is based on radiation antenna theory including the Lorenz reciprocity theorem that involves the antenna effective area A_e discussed in Section 4.13.3. Both of the above methods assume an incident electric field, which can be used to calculate the S_r (refer to Section 4.10) in the radiation antenna theory approach, or which appears directly in the electrostatic theory approach. Then it is possible to compare the short circuit current across the antenna terminals from the two approaches as discussed below.

Consider the equivalent circuit of the antenna in Figure 5.9. At low frequencies the impedance due to the capacitor is dominant and $|1/j\omega C_w| >> R_{Wr}$. Then the short circuit

current I, across the antenna terminals when the antenna is modelled by an ideal voltage source V_s in series with a capacitor C_w and is in receiving mode, can be obtained by (5.4).

$$I = j\omega C_w V_s . \tag{5.4}$$

Initially, consider calculating the short circuit current using the radiation antenna theory and the full electrodynamic theory. Then, the available source power from the voltage source can be used to obtain an expression for V_s in (5.4) as expressed in (5.5).

$$|V_s| = \sqrt{8 P_r R_{Wr}} \tag{5.5}$$

Using (4.32), (4.36), (4.38) and (5.4), where g is the antenna gain the following expression for I can be obtained.

$$|I| = 2|j\omega C_W|\lambda\left[\frac{R_{Wr}}{\eta}\left(\frac{g}{4\pi}\right)\right]^{1/2}|E| \tag{5.6}$$

Then consider the electrostatic theory approach. The short circuit current I can now be obtained by a direct measurement of the voltage across the antenna terminals when the antenna is placed in a low frequency electric field, assuming $|1/j\omega C_w| >> R_{Wr}$ and using the concept of electric flux collecting area of the antenna as given in (4.45).

Hence if R_{Wr} is known it is possible to evaluate the flux collecting area A_f defined in (4.49) using (5.6) to produce (5.7).

$$A_f = \frac{|I|}{\varepsilon_0 |E|\omega} = 2\lambda\frac{C_W}{\varepsilon_0}\left[\frac{R_{Wr}}{\eta}\left(\frac{g}{4\pi}\right)\right]^{1/2} \tag{5.7}$$

However if R_{rW} is unknown, then from the electrostatic theory approach it is possible to obtain I from the measured voltage across the antenna terminals (assuming $|1/j\omega C_w| >> R_{Wr}$ and using the concept of electric flux collecting area given in (4.45)) and use (5.6) to obtain a value for the radiation resistance of the antenna.

The concept of flux collecting area appeared at first, in Chapter 4, as an electrostatic concept (refer to Figure 4.3) and has now been shown to be related through the reciprocity theorem to the radiation resistance, which is a concept of the full electrodynamics theory. Thus the effective area of an antenna is related to its flux collecting area through the radiation resistance.

The significant practical application of the above finding is that radiation resistance of an antenna can be calculated using an experiment based on an electrostatic theory approach. It should also be noted that in light of the reciprocity theorem, the radiation resistance is a direct indication of the capacity of the wedge above a ground plane antenna to collect displacement current from a uniform applied vertical electric flux density.

5.6 Large Loop Antennas

Loop antennas are of significant importance at HF frequencies where their primary purpose is to create strong near fields without excessive far field radiation. The analysis of small circular loop antennas, where the term small implies that the dimensions are such that the perimeter of the loop is much smaller than the wavelength λ, can be found in [38], [41], [43], [47] and [48] where it is shown that the radiation resistance R_r of a small loop antenna is given by

$$R_r = 20\beta^4 A^2, \tag{5.8}$$

where β is the propagation constant and A is the area of the loop. Also small loop antennas have an associated gain of 1.76 dBi.

It should be noted here that the radiation fields and the radiation resistance of small loops are independent of the shape of the loop and depend only on the area of the loop [48]. Hence (5.8) can be applied to a loop of any shape, as long as its area is known and its perimeter is only a small fraction of a wave length. Also the radiation from a small loop is a maximum in the plane of the loop and is zero along its axis. The maximum radiated far-field electric field from a magnetic loop in the direction containing the plane of the loop is,

$$E = \frac{\pi\eta I_0 A}{r\lambda^2} \text{ Vm}^{-1}. \tag{5.9}$$

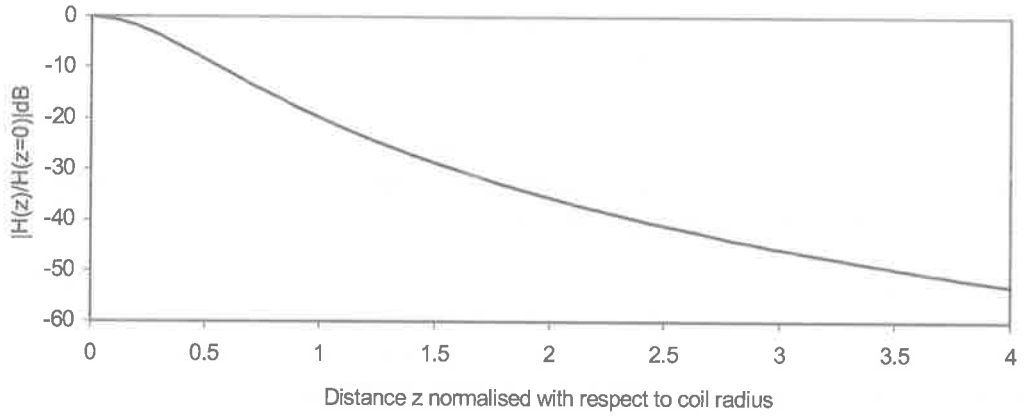

Figure 5.13 Variation of the normalised magnetic field strength with distance (z/a) along the z axis for a loop antenna where a is the coil radius.

Here I_0 is the peak current of sinusoidal excitation, r is the distance from the loop, and A is the area of the loop. Derivation the above result assumes a uniform current distribution over the perimeter of the loop antenna. This assumption is justified if the loop is electrically small, so that the current distribution, which must vary sinusoidally as a function of position around the loop, with the wavelength of that sinusoidal variation being equal to the free space wavelength for electromagnetic fields at the same frequency, is approximately uniform.

As shown in Figure 5.13 loops are known to produce near fields that diminish substantially when the interrogation distance significantly exceeds the loop radius along its axis.

The task of exciting a label in the near-field is substantially one of creation of a stored energy density per unit volume in the volume served by the antenna. As shown in Figure 5.13, the smaller the loop, the more the field is reasonably confined to the volume near the antenna, and the less the range of the antenna, but less total stored energy that needs to be provided.

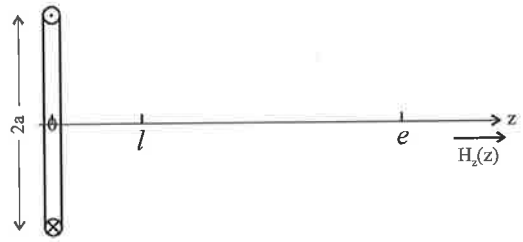

Figure 5.14 Geometry for calculating loop antenna near fields.

But there is another feature of small loops that is of interest. Considering Figure 5.14, and by taking the task of maximizing the magnetic field at the label position *l*, in the near field, for a given electromagnetic compatibility enforcement distance *e*, in the far field, it can be shown that the optimum loop size is one where the radius tends to zero (when no other considerations such as antenna bandwidth or driving power are taken into account).

Clearly this is not a practical solution, as it has neglected important practical consequences and practical constraints. Making the antenna size small increases the concentration of very high and unused (because the tag is further away) energy density per unit volume created closer to the loop centre than at the label. In an attempt to keep the same energy density per unit volume, at the tag position, the total stored energy becomes very large, the quality factor of the antenna becomes unreasonably high, and also the real power density required to drive it becomes too large. An alternative view point is that small loops are badly proportioned to create longitudinal near fields to excite tags.

Hence, practical consequences and constraints, such as the need to constrain the driving power to practical values, indicates that such an optimisation is misconceived, and loops of finite size must be considered. For this reason, attention passes to the construction of larger loops. Such loops may suffer from non-uniformity of the current distribution (due to their electrically large size) as noted by Pocklington's Theorem, with the attendant problems of increased radiation and difficulties of tuning. If the current distribution around the loop is not uniform, the cancellation of the radiation from different current elements which occurs for certain radiation directions, such as the polar direction, does not occur, and the loop begins to radiate in that direction as a result of the current non-uniformity. There is also an increase in the radiation in other directions [48]. The result is that the electric field expressed in (5.9) becomes the limiting far-field electric field, that is, it is the minimum value of electric field that occurs as a result of the loop radiation. The following section considers the construction of large loops that aim to reduce the current non-uniformity and thereby allowing the creation of large volume magnetic fields in the near field.

5.6.1 Practical Construction of a Large Loop Antenna

However it is possible to modify the structure of a large loop to contend with these problems. Figure 5.15 depicts such a structural modification where effort has been made, by breaking the loop resonating capacitance into several series elements which are distributed around the loop, to counteract the tendency of the current distribution on long wire antennas to become sinusoidal. Such structures can be used to obtain substantially increased read range from high frequency labels.

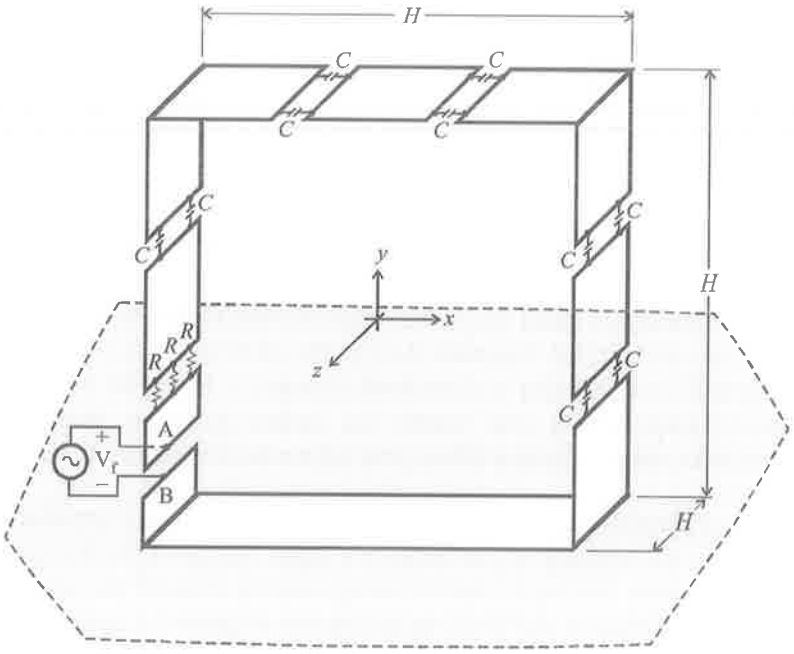

Figure 5.15 A large loop antenna construction.

The large loop structure shown in Figure 5.15 can still be considered as an electrically small antenna at 13.56 MHz, because an effort was made to keep the current distribution uniform despite the loop's considerable size. The construction of large loops according to this design overcomes characteristic limitations of electrically small antennas [48] outlined below and discussed in greater detail in Chapter 7.

- A high radiation quality factor
- A small antenna bandwidth
- A high input impedance sensitivity to changes in frequency
- Practical consequences such as high and possibly unused energy density per unit volume created close to the loop centre, so that the real power required to drive the antenna becomes large.

The analysis of large loops requires careful consideration. The results considered thus far are for electrically small antennas with perimeters much less than a wavelength. However, as the perimeter of a loop antenna becomes a sizable faction of wavelength, it is important to consider the consequences of the results obtained by Pocklington that shows the current distribution on thin wires to be sinusoidal [47], to a very good approximation, with a wavelength of the sinusoidal variation being equal to the electromagnetic wavelength in free space at the operating frequency. Hence, an increase in a perimeter of a loop towards the 22 m value of the wavelength at 13.56 MHz demands that the true nature of the current distribution on a loop be taken into consideration.

As discussed above, a direct result of the non-uniform current distribution is an increase in the far-field radiation. With a balanced feed, the magnitude of the current follows a cosine distribution function with the maximum opposite the feed point. Figure 5.16 illustrates the non-uniformity of current distribution around a large loop where the thickness of the current distribution strip represents the magnitude of the current I along the loop given by

$$I = I_0 \cos(\beta l) \tag{5.10}$$

where l is the linear position on the circumference with $l = 0$ opposite the feed point, and I_0 being the maximum value of the sinusoidal current excitation.

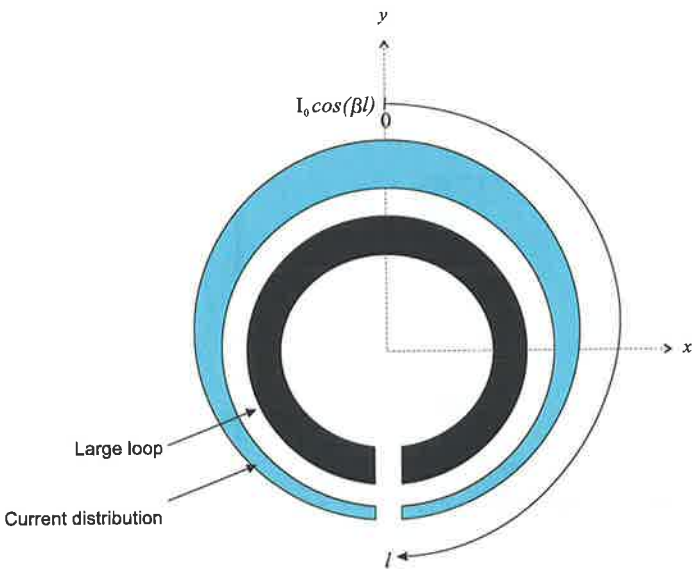

Figure 5.16 Distribution of current magnitude around a large loop.

Figure 5.16 allows the effect of non uniform current distribution on far field radiation to be explained qualitatively. As far as radiation along the polar axis is concerned, the vertical (y axis directed) sections of loop have approximately equal magnitude and oppositely directed currents, and continue to cancel in their effect upon radiation in the polar direction. However, the current at the feed point is substantially less than the current opposite and these currents no longer provide such cancellation, so radiation in the polar direction begins to occur.

Considering now the points to the right of the loop or the left of the loop, that is the radiation in the equatorial plane, do not produce a great contribution in the far field radiation as the current elements on the left and the right sections of the loop have the same magnitude. Hence the net radiation after superposition is due to the fact the two groups of current are at different distances from the far-field point, and suffer a different phase delay in radiating thereto. However, if a far-field position above the loop is considered then the current elements on the bottom of the loop pointing left are further away from the current elements on the top of the loop pointing right, but in this case the elements on the bottom have a lesser magnitude than the current elements on the top, so the net radiation after superposition is due to the residual currents on the top both being closer to the far-field point and also having a larger net magnitude (the reduction from the other side of the loop is less).

Thus, the unbalanced currents on sections of the loop will act in a manner similar to a small electric dipole. Hence the radiation patterns in the far field will be similar to those produced by small electric dipoles and the electric and magnetic field vectors created by the unbalanced current can be evaluated using the far field solutions of an electric dipole [41], with a current element of size IL given by the unbalanced current I and the length of the section of the loop containing the unbalanced current. The radiation produced will be polarised along the x axis.

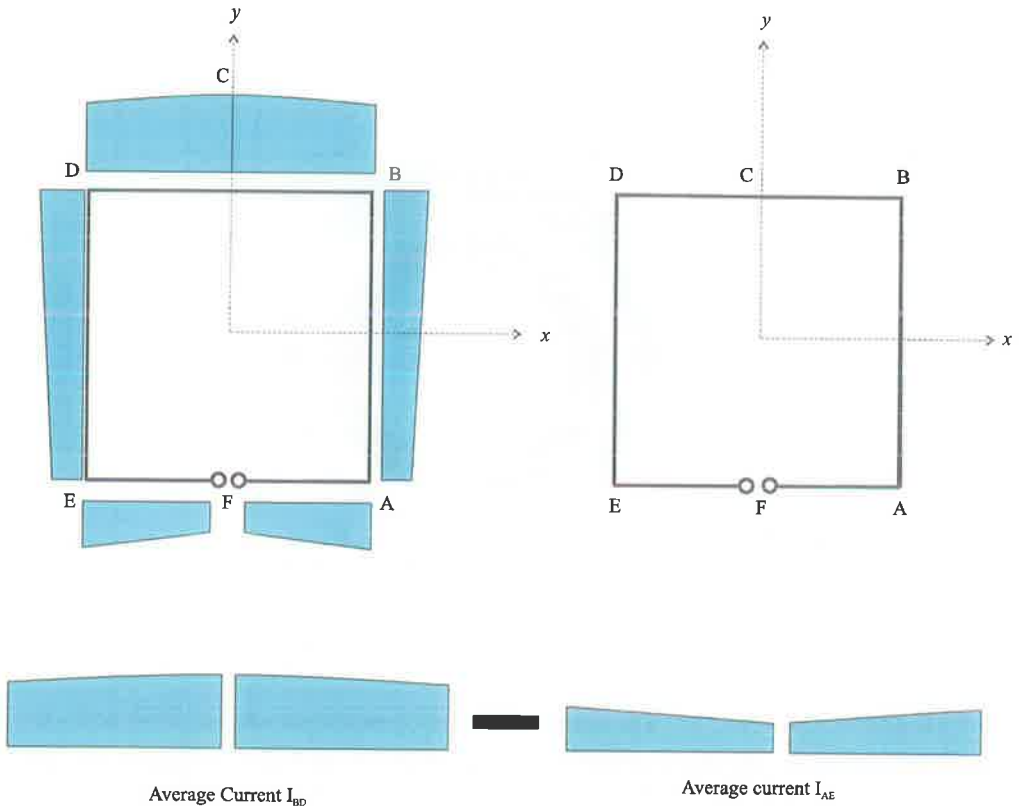

Figure 5.17 The current distribution around a square loop of perimeter λ/4 in free space. The size of the loop considered is similar to the size of the loop discussed in this in Section.

It is possible to estimate the unbalanced current of two parallel sections of a square loop by considering the current distribution around the loop to be sinusoidal. Figure 5.17 illustrates the current distribution around a square loop of perimeter $\lambda/4$ while a complete formulation of the radiated fields and the resulting radiation patterns can be found in [41]. Clearly the currents on the left and the right sections of the loop are balanced due to symmetry. However there is a significant variation in the net current on the top and the bottom section of the loop. The product of IL for the unbalanced current can be found by the difference between the current integral along the top and the bottom of the loop, where the sinusoidal current distribution is given by

$$\mathrm{I}_0 \cos(\beta l). \tag{5.11}$$

The difference in the current length product will radiate in a manner similar to having a short dipole of length $\lambda/16$ at the centre of loop, aligned along the x axis. The unbalanced current IL thus calculated is

$$IL = \frac{0.1\mathrm{I}_0}{\beta} \tag{5.12}$$

Substitution of the current-length product into the short dipole's far field electric field equation, keeping in mind that the current element is now aligned along the x axis, yields the result in (5.13) for radiation in the equatorial plane. As stated previously, the radiation from the short dipole will be polarised in the x direction.

$$\mathrm{E}_\phi = \frac{0.1\beta \mathrm{I}_0 \eta}{4\pi}\left(\frac{j}{\beta r}\right)e^{-j\beta r}\sin\phi \tag{5.13}$$

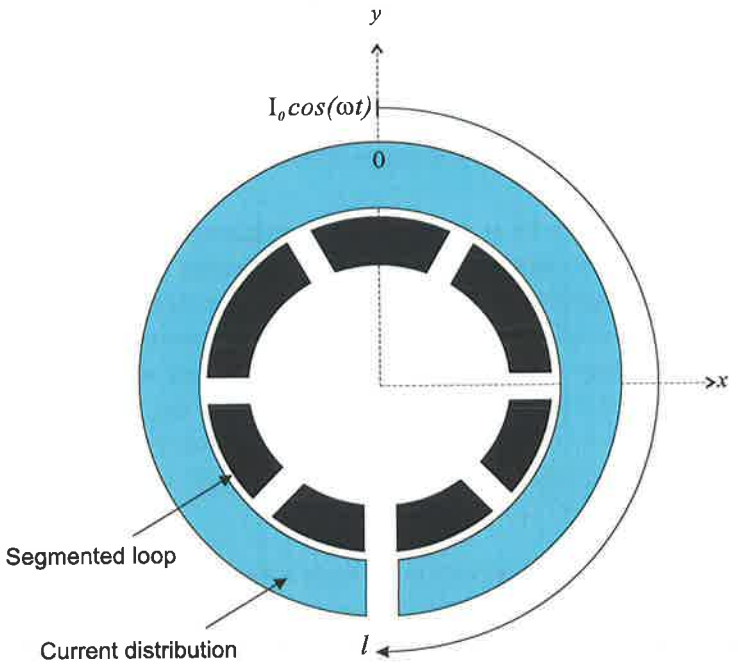

Figure 5.18 The current magnitude distribution around a segmented loop.

The method for reducing the far-field radiation of a practical magnetic loop thus stems from the effort to keep the instantaneous magnitude of the current distribution around the loop uniform. If a loop is segmented as shown in Figure 5.18, and is supplied with equal excitation at each of the gaps, the current distribution of a segment of a loop can be considered as having a maximum at the centre of the segment, taken as being the local origin $l = 0$, and an even distribution function following

$$I = I_0 \cos(\beta l) \tag{5.14}$$

as the linear dimension l is traversed. With this segmentation, the variation in current magnitude in each segment follows that of the cosine function near its maximum, i.e. the amount of change in each segment is small, as the segment lengths are small.

Figure 5.19 A photograph of the large loop construction.

The segmentation of Figure 5.18 is implemented practically by physically cutting the loop into small sections with the electrical connection maintained by the insertion of a capacitor between each section as outlined in Figure 5.15. The value of each capacitor is chosen such that the combined value of all capacitors in series is the capacitance required to resonate the inductance of the uncut loop at the frequency of excitation. By way of approximation, one of the feed points is supplied with a voltage generator that has the job of supplying the voltage drop across the radiation resistance. Since this voltage drop is much less than the voltage drop across the loop inductance, a practical approximation to the symmetrically fed loop described earlier in Figure 5.17 can be obtained.

Figure 5.19 shows a practical construction of such a large loop. As just described, the large magnetic loop with the distributed capacitance should allow the current to be distributed more uniformly so that the assumption of uniform current distribution made in the analysis of physically small loop antennas is still true for physically large and no longer electrically small loop antennas.

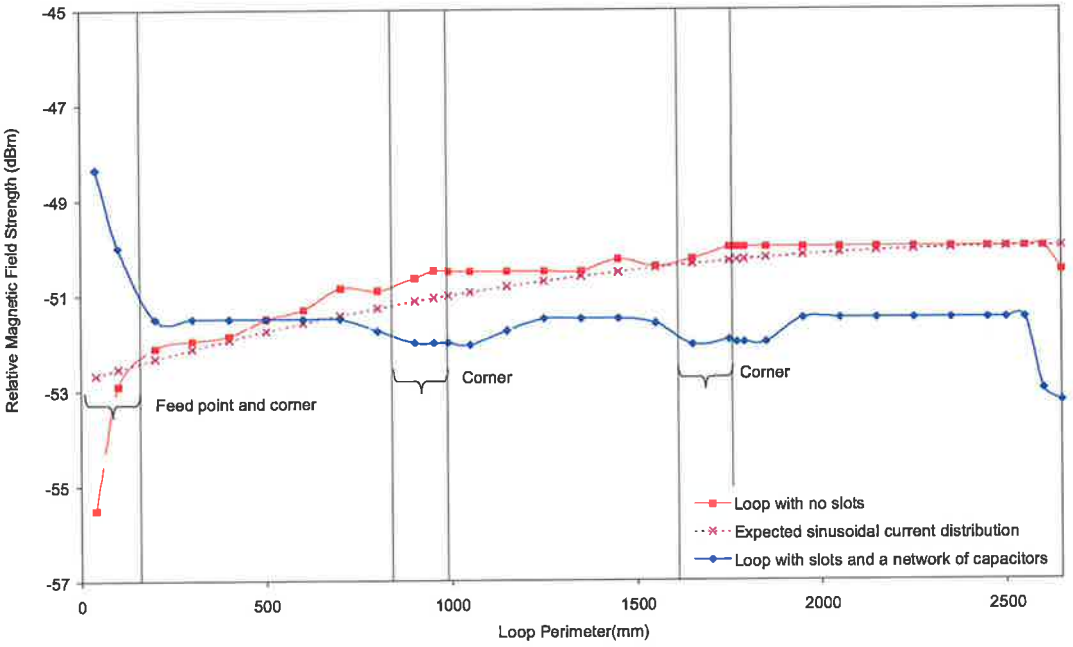

Figure 5.20 Comparison of the current distribution around a large loop and a slotted loop with the expected sinusoidal current distribution.

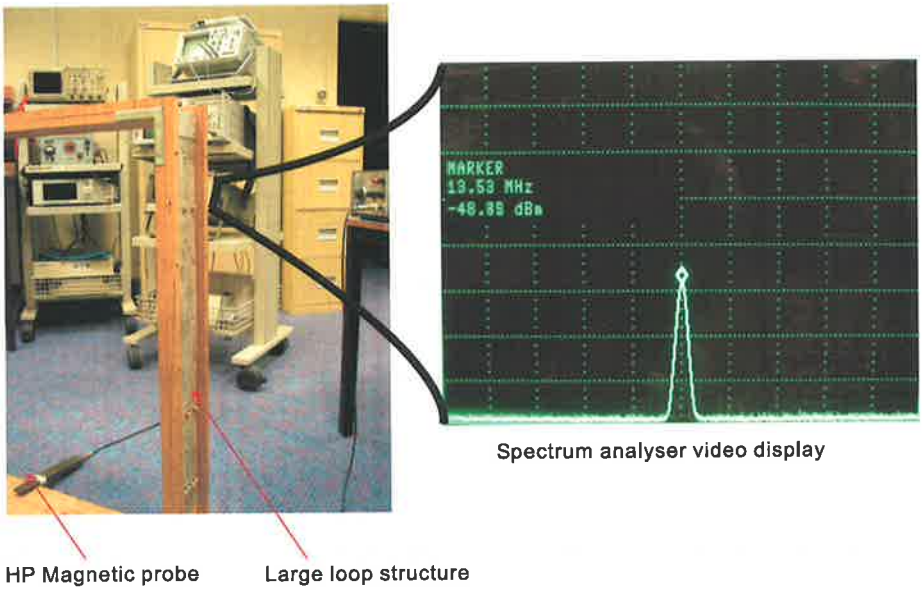

Figure 5.21 Instrument set up used for obtaining the magnetic field around the loop using a close field magnetic probe.

The relative magnetic field strength shown in Figure 5.20 was obtained by using a HP11841A close field magnetic probe connected to a spectrum analyser with a video averaging detector operating over a 10 MHz bandwidth around a centre frequency of 13.7 MHz, to record the variation of magnetic field close to the loop as the probe is moved around the perimeter of the large loop shown in Figure 5.19. Figure 5.21 shows the instrumentation set up used to obtain the graph in Figure 5.20.

The tendency of the large loop with no capacitive slots to follow the sinusoidal current distribution expected around the loop can be easily observed from Figure 5.20. The breaking up of the continuous loop structure into slots has achieved a remarkably uniform current distribution.

The modelling of the loop antenna shown in Figure 5.15 and Figure 5.19 as well as the extraction of useful antenna parameters is discussed below.

5.6.2 Large Loop Antenna Model

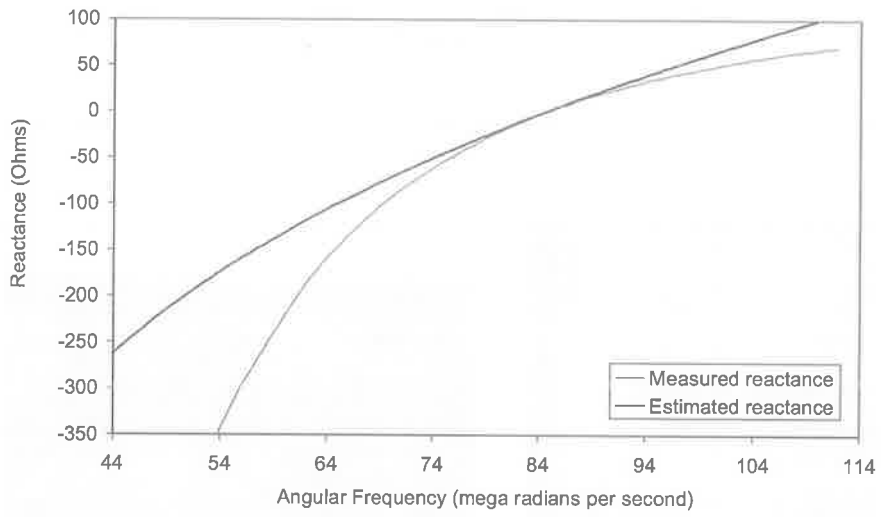

Figure 5.22 Comparison between the measured reactance and the reactance estimated using the equivalent circuit model for the large loop structure.

The loop shown in Figure 5.19 has been tuned to 13.47 MHz and matched to 50 Ohm input impedance at resonance through the insertion of a series resistor at one gap. An equivalent circuit for the loop antenna can be developed by using the measured reactance curve shown in Figure 5.22 to obtain the capacitance, C_L, and the inductance, L_L, of the equivalent circuit given in Figure 5.23. Figure 5.22 also shows the resulting reactance obtained using the model parameters C_L and L_L derived from the measured reactance. Thus the reactance jX_L of this loop can be modelled as

$$jX_L = \frac{1}{j\omega 63.1\text{pF}} + j\omega 2.21\mu\text{H}. \tag{5.15}$$

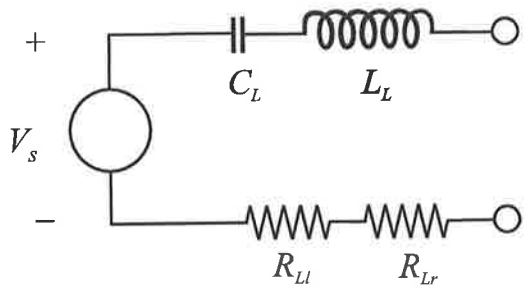

Figure 5.23 An equivalent circuit model for the large loop antenna.

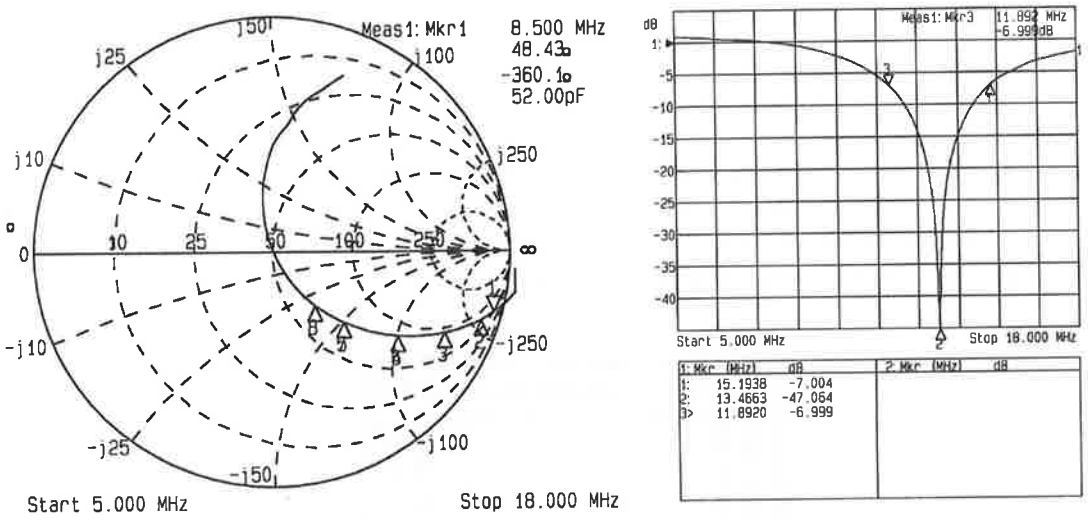

Figure 5.24 The return loss curve obtained for the large loop indicating a resonance at 13.47 MHz and the bandwidth of the loop antenna.

The radiation resistance of the loop can be estimated on the uniform current assumption by using the formula for radiation resistance R_r of an electrically small loop antenna provided in equation (5.8). Since the area of the loop is 0.76 m^2 the radiation resistance of the loop is approximately 0.2 Ohms. The ohmic losses in the loop represented by R_l on the equivalent circuit are due to the added series resistance of 49.8 Ohms. As the input impedance is known, a voltage measurement at the feed point of the antenna can be used to predict the antenna current and hence, on the uniform current assumption, the radiated power from the antenna.

The quality factor for the antenna resonance can be obtained by using a return loss plot from a network analyser as has been shown in Figure 5.24. It may be shown that for deep dips at resonance, the half power points of the resonant circuit correspond to the 7 dB return loss points, from which the bandwidth of the resonant circuit can be computed. Thus the quality factor Q of resonance can be obtained as

$$Q = \text{Bandwidth / resonant frequency.} \tag{5.16}$$

The quality factor, Q of resonance for the loop constructed is about 4. It is important to emphasise that this low quality factor has been achieved through the addition of series damping resistance, and that only a small amount of the power delivered to the loop is actually radiated. As the function of the loop is the creation of a large volume near field, this in fact is a positive outcome.

5.7 Experimental results

Reading range experiments were conducted in a large empty room with the wedge above ground plane antenna and the large loop featured in Figure 5.8 and Figure 5.19, respectively. Measurements were conducted using widely available library tags, and a large tag antenna constructed with such library tag chips (refer to Figure 5.25). The experimental results are summarised in Table 5.3 below. It should be noted here that since both antennas constructed, though physically large, are still electrically small and therefore a gain of 1.76 dBi is assumed throughout the experiments.

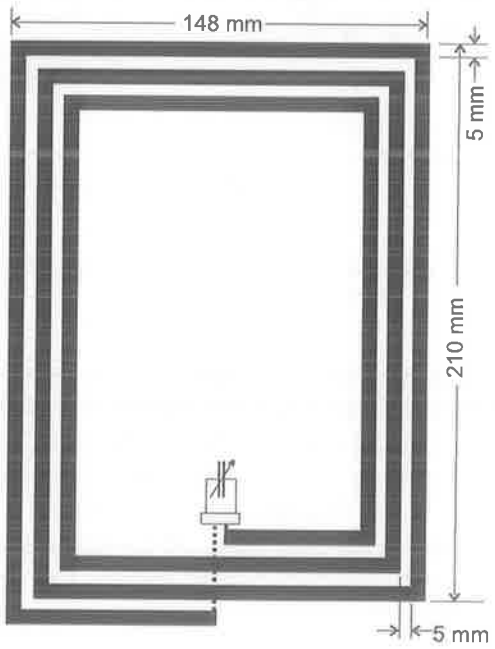

Figure 5.25 The large tag used in laboratory tests.

Table 5.3 Summary of test results.

Antenna	Antenna Gain	Radiated Power	EIRP	Library Tag Read Range	Large Tag Read Range
Large Loop	1.76 dBi	295 mW	442.5 mW	500 mm	2800 mm
Wedge Above a Ground Plane	1.76 dBi	310 mW	465 mW	1000 mm	3200 mm

It is clear from the experimental results that the relaxation of EMC regulations and the use of larger interrogator antennas and label antennas allow a considerable increase in reading range. Extending the reading range pushes the tags into the mid field region. The observed difference in reading range between the large loop and the wedge is as a result of the Wedge above a ground plane antenna's higher Q and its ability to radiate significantly more than that of the large loop antenna, so thus creating a stronger mid field region for tag operation.

5.8 Interrogation at a Large Distance

The Sections above have focused on the structures used to create near fields, and possible structures for increasing HF read range in view of relaxed electromagnetic compatibility regulations. The following section will consider a formulation of the coupling link between tags and readers and consider the issues related to interrogation at large distances in the HF region.

By choice of frequency it is possible to place the large distance in the near field or in the far field, however the HF ISM band centred at 13.56 MHz has a near-far boundary at around 3 m from the field creation structure. Assuming that losses are also minimised, the total reactive power W_t for a power P_t delivered to a field creation structure of quality factor Q_t given by

$$W_t = Q_t P_t . \tag{5.17}$$

The energy density per unit volume, W_v, at the label position is obtained by introducing the dispersal volume V_d defined in Chapter 4, such that

$$W_v = \frac{Q_t P_t}{V_d} . \tag{5.18}$$

Assume that a label exciting field is created at a distance r from an interrogator antenna formed with a circular coil of diameter d. Since the interrogation considered here is at a large distance, analysis of the field properties of such a loop leads, to a good approximation when $r >> d$, to that given in (5.19).

$$V_d = F\left(\frac{4r^2}{d}\right)^3 \tag{5.19}$$

In (5.19), r is the distance from the interrogator to the label, and F is a factor of the order of 2 resulting from evaluating the inductance of the coil using (5.20), where d_w is the diameter of the coil wire.

$$L = \frac{\mu_0 d}{2}\left(\ln(\frac{8d}{d_w}) - 2\right), \text{ such that } F = \frac{1}{2}\left(\ln(\frac{8d}{d_w}) - 2\right) \tag{5.20}$$

Substituting (5.19) into (5.18) gives

$$W_v = \frac{Q_t P_t d^3}{F(2r)^6}. \tag{5.21}$$

This result indicates the advantage of a significantly high Q_t while it also shows how the reactive power at a label position can rapidly diminish with increasing distance from the antenna. In a practical situation the effect of r^6 in the denominator of (5.21) tends to eliminate the benefit of any advantage gained from a high value of Q_t when $r > d$.

Hence it is clear that obtaining large reading distances requires the use of physically large interrogator antennas, as was done in Section 5.4 and Section 5.6. The result obtained in (5.21) also points towards the fact that propagating communication is a far superior choice when $r > d$, which naturally implies that it is also necessary to radiate more if tag reading range is to be increased, as was achieved in the relaxed HF regulation standards outline in Section 5.1.

5.9 Conclusion

A proposal for relaxation of electromagnetic compatibility regulations in a narrow band in the HF region is investigated, and some of the interesting consequences for HF operation at substantially improved distances have been identified.

In view of the new ETSI regulations at the HF ISM band a number of possible antenna designs were proposed with the aim of creating strong mid fields. A detailed analysis of a wedge above a ground plane antenna, and a large loop antenna was presented with methods for matching and overcoming limitations imposed by size. Practical results with these antennas and large label antennas have shown improved performance and the ability to read RFID tags in the mid field region.

These results, while attractive, are unlikely, without substantial increase in label antenna sizes, to challenge the supremacy of far field systems for long-range RFID system operation.

Through the analysis of wedge above ground plane antennas by way of experimentation, simulation and comparison with Woodward's results, the concept of flux collecting area which is an electrostatic concept has been shown to be related through reciprocity theorem to the radiation resistance, which is a concept of the full electrodynamics theory.

The design of interrogator antennas and tag antennas are related, yet, they are completely different disciplines. The design of tag antennas for far field operation in the UHF region has attracted much attention since the adoption of the UHF frequency bands for supply chains applications. Generally tag antenna design has remained somewhat of a "black art" where little is published on their designs and design considerations. The following chapter will illuminate the intricacies of tag antenna design and illustrate a simple methodology for developing antennas for RFID labels that operate in the far field.

Chapter 6

FAR FIELD RFID LABEL ANTENNA DESIGN

Antennas used in the HF region operate at 13.56 MHz which frequency has an electromagnetic wavelength of around 22 m giving a near field boundary of around 3.5 m. Thus, given reading distance requirements of less then 3 m, and using the regulated radiation power at the HF ISM band, reader antennas, as we have seen in Chapter 5, are almost always near field creation structures that aim to create a large energy density fields with the minimum amount of radiation.

However, at UHF frequencies the scenario is different. At UHF frequencies the near field far field boundary is at around 50 mm. Thus the region of operations in the UHF spectrum is almost always in the far field, and therefore reader antenna designs are far field creation structures that aim to operate at the highest possible efficiency.

This chapter considers RFID label antennas for far field operation in the UHF frequency range and the development of antenna equivalent circuits to aid on the development of tag antennas. Designing RFID label antennas for various applications require careful consideration. This chapter presents an RFID label antenna design methodology, illustrated in the far field with the design of two long range, bow tie antennas for tagging cases and pallets.

6.1 RFID Label Antennas

There are numerous label antenna designs, each with their own set of characteristics described by the antenna gain, directivity, radiation pattern, effective length, efficiency and effective area [42, 43, 47 and 48]. Antenna designs are influenced by a range of issues, such as the region of label operation (near or far), the coupling field (electric field or magnetic field), the regulatory constraints, and the environment in which they operate. For example, an environment with many metal structures can affect time varying EM fields, and thus affect the performance of an antenna. Illustrations of designing antennas to suit their environment of operations are presented in [40], [49] and [50]. A vital aspect of the design process is to allow maximum coupling between the reader antennas and label antennas for the coupling field used. Prior to considering the subject matter of label antenna design, the following sections consider a number of different label antenna structures and the merits of their designs.

6.1.1 Magnetic Field Sensitive Antennas

A common example of a magnetic field sensitive HF label is shown in Figure 6.1 below. The label is 42 mm wide by 47mm high. The label is designed to have a sufficient number of turns to provide the resonating inductance for the microcircuit input capacitance, as well as a flux collecting area in the interior which is as large as practicable and consistent with the size requirement for the label.

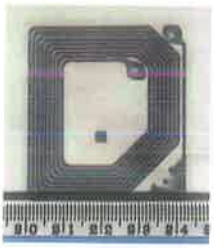

Figure 6.1 A magnetic field sensitive antenna.

Advantages of working in the near field at HF rather than at LF are that the number of turns required to resonate the microcircuit capacitance is small enough for low resolution lithography to be used in antenna construction, and that no additional external resonating capacitance is required.

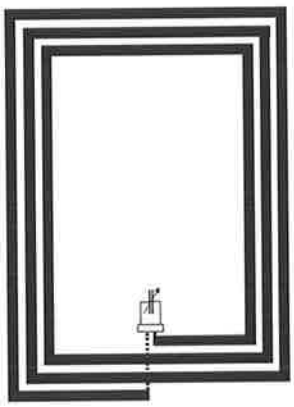

Figure 6.2 A large loop antenna for an HF label.

When a larger space is available for the tag label, a larger coil area should be used. As shown in Figure 6.2, fewer turns are then needed to obtain the required tuning inductance. It can be observed from Section 4.14.1.1 that the figure of merit (the coupling volume) for a planar coil operating in the near field varies as the third power of size since the inductance of a coil is dependent on the equivalent coil diameter. Thus the antenna of Figure 6.2 is about 18 times more sensitive than that of Figure 6.1. Unfortunately, this increased sensitivity does not translate to a corresponding increase in range, as small coil interrogator antennas have an inverse sixth power decrease in energy density per unit volume as distance from the interrogator increases.

Clearly, both of the designs illustrated above are unsuitable for being placed flat against metal, as the boundary conditions shown in Figure 4.1 and Figure 4.2 will not allow a normal component of magnetic flux density at the metal surface. For this situation, the label antenna employing a solenoid with a magnetic core design shown in Figure 6.3 is employed.

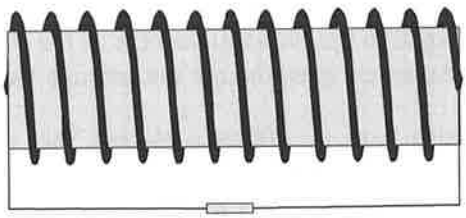

Figure 6.3 An antenna for HF operation against metal.

It has been shown in equation (4.42) that without the magnetic core the coupling volume of a long solenoid is just the physical volume, but when a magnetic core is inserted, the coupling volume increases by a factor equal to the effective permeability defined in equation (4.43).

This behaviour may be contrasted with that of electric field labels, in which in equation (4.55) it has been shown that the inclusion of dielectric material into the interior of the label is not helpful.

6.1.2 Electric Field Sensitive Antennas

Two varieties of electric field sensitive antennas are shown in Figure 6.4 and Figure 6.5. Figure 6.4 shows a small bow tie antenna that is intended to be sensitive to electric fields in the horizontal direction.

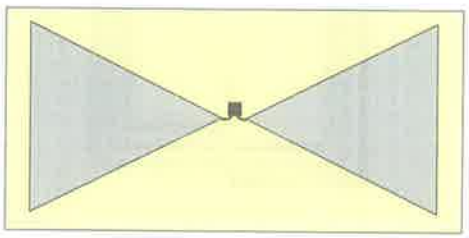

Figure 6.4 An electric field sensitive label.

Figure 6.5 shows an electric field sensitive antenna that is suitable for placement against a horizontal metal plate.

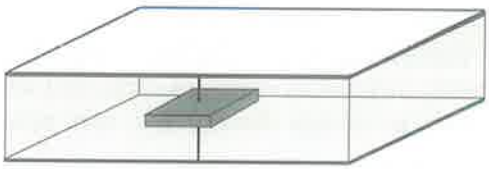

Figure 6.5 A parallel plate electric field sensitive label.

Analysis of the structure in Figure 6.4 is provided in Section 6.4, while the structure in Figure 6.5 is analysed in Section 4.14.2.2. The analysis has a common feature that the figure of merit for these antennas, when placed in the energy storage electric field is a coupling volume, for Figure 6.5 it is equal to the physical volume of the structure, and for Figure 6.4 it is derived from the label dimensions, even though the antenna itself has no physical volume.

Both of the antennas will also have an effective electric flux collecting area but this area should not be confused with the effective area concept of a radiating antenna or of a far field antenna. The effective area for a near field electric field sensitive antenna describes the extent to which the antenna can extract current from the displacement current density of the driving electric field.

6.1.3 Electromagnetic Field Antennas

An antenna can be considered as an electromagnetic field antenna on a couple of different bases. Firstly, if the antenna is capable of responding to both electric and magnetic fields we would consider it to be an electromagnetic field antenna. It is almost

invariably true that unless the antenna is very small, it does have this property. Proper analysis requires that it be analysed using the full set of Maxwell's equations, rather than the subset or simplified versions that pertain to electrostatic or magnetostatic problems. A good example of this phenomenon is provided by the electromagnetic field sensitive antenna shown in Figure 6.6, in which there is no obvious effort to couple to either electric or magnetic field alone.

Figure 6.6 An electromagnetic antenna.

Such electromagnetic antennas are generally useful for operation in the far field, because far field interrogation systems have shorter wavelengths, and antennas of acceptable size can no longer be considered to be electrically very small, but are merely small.

Despite this distinction, there are electromagnetic label reading environments in the UHF region in which, through reflections, either the electric or magnetic field is emphasised at the expense of the other. For such situations it is normally useful to take into account the nature of the driving fields in antenna design, and to shape the design so that it is recognisably attuned to one or other of those fields.

6.2 Label Antenna Design Considerations

Use of RFID in the identification of objects in the various supply chains around the world has created research avenues into consumer product packaging (CPG) to find novel ways of integrating RFID labels into packaging and developing labels to suit packaging and goods. The sections above have described the results of previous label antenna design developments to illustrate the multi-faceted world of tag antenna design.

In addition to the antenna designs presented in Section 6.1, there is an accumulating index of publications on RFID label antennas, for both active and passive tags, such as the slot antenna design in [51], inverted F-antenna design [52] and the folded dipole antenna design [53], to consider a few. Most of these publications only cover aspects of antenna analysis and practical aspects such as the suitability of the antenna for a specific application [54]. However, what is not covered is a methodology for designing a tag antenna and a clear view of tag antenna design criteria. The following sections of the chapter bridge that knowledge gap. Finally, Section 6.4 illustrates two successful antenna designs for a passive RFID label that can be placed on corrugated cardboard boxes containing some form of dry goods. The antenna designs presented are of a 'credit card size', considered generally to be a suitable size for labeling a majority of cases used in supply chain applications.

6.2.1 Nature of Antennas for RFID

This section will consider, in general, types of antennas suitable for RFID applications. The evaluation will be based on both practical aspects and performance aspects. Considering practicable antennas for RFID applications restricts us to mostly planar structures that can be attached to items, cases and pallets. In addition it is important to consider the RFID chip input impedance at the threshold of operation to realise a conjugate antenna impedance to achieve maximum power transfer to the RFID label IC.

Since passive RFID tags operate in a power constrained environment created by electromagnetic compatibility regulations and the power required to operate the tags is obtained from the incident electromagnetic waves, maximum power transfer is of vital importance. Therefore, prior to postulating a theoretical "best" antenna, it is important to consider the load to which an RFID label antenna must provide power. The resulting load impedance presented to a tag antenna is considered below.

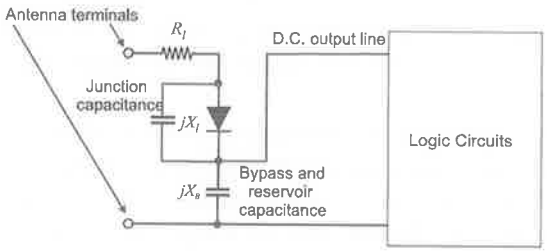

Figure 6.7 A simplified RFID label IC schematic.

A UHF RFID label IC with an antenna terminal and a rectifying circuit that is intended to produce a rectified voltage used for powering the label circuits, can be modelled as indicated in Figure 6.7. Here, X_l represents the reactance of the diode capacitance, X_B is the reactance of the reservoir capacitor that also serves as an RF bypass, R_l represents the loss in bringing reactive power into and out of the diode junction capacitance.

It is clear from Figure 6.7 that the input impedance of an RFID chip is largely dictated by the junction capacitance of the rectification diode. The rectifiers on modern UHF RFID ICs are fabricated using Schottky diodes with a junction capacitance value in the range of a few picofarads or less. Due to the sensitivity of the junction capacitance to the biasing voltage the input impedance of an ASIC RFID chip is a complex function of both the operating frequency and the input power to the chip from the antenna. Thus in general, the chip impedance, Z_c is measured at the threshold of operation so that the antenna impedance is a conjugate match at the lowest power level at which the chip will operate successfully. This ensures that the chip receives the most amount of power possible when the tag is furthest from the powering RF wave.

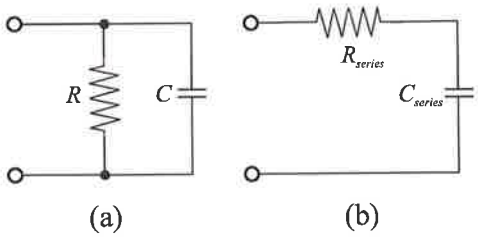

Figure 6.8 (a) A parallel equivalent circuit of an RFID IC input impedance where (b) is a series equivalent circuit of the chip input impedance.

As illustrated in Figure 6.7 the input impedance of an RFID chip at the threshold of operation (minimum input sensitivity) is capacitive. The input impedance of an RFID IC can be modelled as indicated in Figure 6.8 as a series equivalent circuit or a parallel equivalent circuit. Using the series equivalent circuit in Figure 6.8 (b), $Z_c = R_{series} + (1/j\omega C_{series})$. Depending on the fabrication technology and the IC design the typical impedance of RFID ICs will vary. Some of the typical values expected are listed below.

- 6.7 – j197.4 Ω at 915 MHz (EPC Class I Gen I from [55], R = 5800 Ω, C = 0.88 pF)
- 7.4 – j218 Ω at 868 MHz (EPC Class I Gen I from [55], R = 6400 Ω, C = 0.84 pF)
- 36 – j117 Ω at 866.5 MHz (EPC Class I Gen 2 from Impinj [56])
- 33 – j112 Ω at 915 MHz (EPC Class I Gen 2 from Impinj [56])

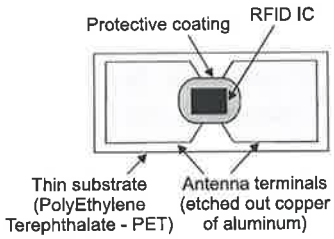

Figure 6.9 A direct chip attachment of an RFID IC.

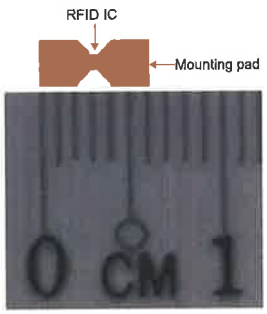

Figure 6.10 An RFID strap.

The final chip impedance seen by the antenna is also affected by the technique used to attach the RFID IC to the tag. Generally there are two different types of attachments possible. When the IC is in a flip-chip package (which is the industry standard technique for low cost packaging), as shown in Figure 6.9, the RFID ICs can be directly attached to the antenna. The RFID ICs may also be obtained as a "strap" where the IC is connected to two mounting pads with a thin superstrate as shown in Figure 6.10.

Typically a resistance R, of around 1300 Ω in parallel with a 1.1pF capacitor C, (which is that quoted for an Alien Class I Gen I RFID IC fabricated with CMOS technology and at the threshold of operation of the IC [57]) resulting in a series equivalent circuit impedance of 18.95 – j155.8 Ω can be expected from an RFID IC strap. Generally it is good practice to measure the input impedance of the chip at various operation frequencies using a network analyser to obtain the impedance characteristics of the chip prior to designing an antenna. Such a measurement method is outlined in detail in [58].

Maximum power transfer requirements dictate that the antenna impedance should be a conjugate match to ensure the greatest possible performance from the RFID label (measuring the performance of an RFID label is discussed in Section 6.2.5). Hence considering the requirements of practicability and maximum power transfer the "best" antenna for an RFID IC is a planar inductive antenna with a reactance that is able to tune out the capacitance of the label IC and also provide an adequate match for the real impedance of the label IC. The following section will consider modelling the input impedance of a typical label antenna by formulating a three parameters circuit model.

6.2.2 Label Antenna Equivalent Circuits

Equivalent circuits for small magnetic field sensitive antennas and electric field sensitive antennas are shown in Figure 6.11 and Figure 6.12 below. The range of validity of these equivalent circuits is where the reactance properties of the antenna may be described by a single parameter, L or C. When the antenna is larger, as was the case for interrogator antennas considered in Chapter 5, reactance properties were described by an appropriate mixture of L and C.

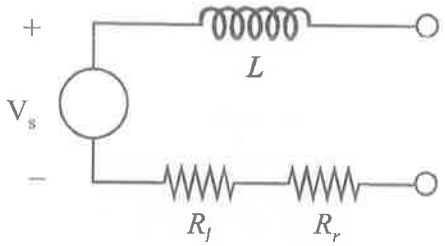

Figure 6.11 An equivalent circuit for a small magnetic field sensitive antenna.

In Figure 6.11, the source voltage is the voltage induced in the flux collecting area of the coil by magnetic fields other than those fields which resulted from current flowing within the coil itself. Those induced voltages are represented by the voltage drop in the inductor L. The parameters R_l and R_r are loss and radiation resistances respectively. Figure 6.1 is an example of a magnetic field sensitive label antenna.

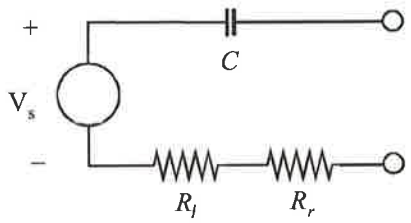

Figure 6.12 An equivalent circuit for a small electric field sensitive antenna.

In Figure 6.12, the source voltage is the voltage developed across the self capacitance of the antenna when it is open circuited, as a result of the current injected into it, when it is short-circuited, by the displacement current density of the electric field in which the antenna is immersed. The parameters R_l and R_r are loss and radiation resistances respectively. Figure 6.4 is an example of an electric field sensitive label antenna.

Calculating the parameters of the magnetic field sensitive antenna is straightforward, the relevant formulae being obvious or contained in Appendix A.1. For the electric field antennas, determination of the relevant parameters is sometimes not quite so simple, as electrostatic field solutions for the relevant geometries are not readily available. Therefore empirical results or numerical modelling are more commonly employed for useful shapes.

Thus far, the nature of RFID ICs and the nature of the impedance of label antennas have been considered. Clearly, matching a label's input impedance to a label antenna's impedance is vital. While an antenna impedance may be adjustable by design variations (as illustrated in Section 6.4.4 and Section 6.4.5), the input impedance of an RFID IC can not be altered without using external circuit components. This is an undesirable result. The following section investigates the practicality of matching to an RFID IC's input impedance at UHF frequencies allocated for RFID around the world.

6.2.3 Matching to an RFID Chip Impedance

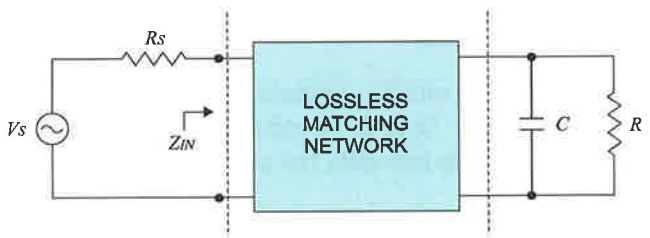

Figure 6.13 A circuit with a lossless matching network and a parallel RC load.

The Bode and Fano theorem can be used to investigate the existence of any theoretical limitations to matching to an RFID IC's chip impedance. Figure 6.13 shows a circuit with a real source impedance, a lossless matching network and an input impedance of an RFID IC by a parallel *RC* load. According to Bode and Fano, the fundamental limitation on impedance matching takes the form [64]

$$\int_0^\infty \ln\frac{1}{|\Gamma|}\,d\omega \le \frac{\pi}{RC}, \tag{6.1}$$

where Γ is the reflection coefficient of the load and its assumed lossless matching network with respect to the source impedance R_S, and R and C is the resistance and capacitance, respectively in the parallel *RC* load (Figure 6.8).

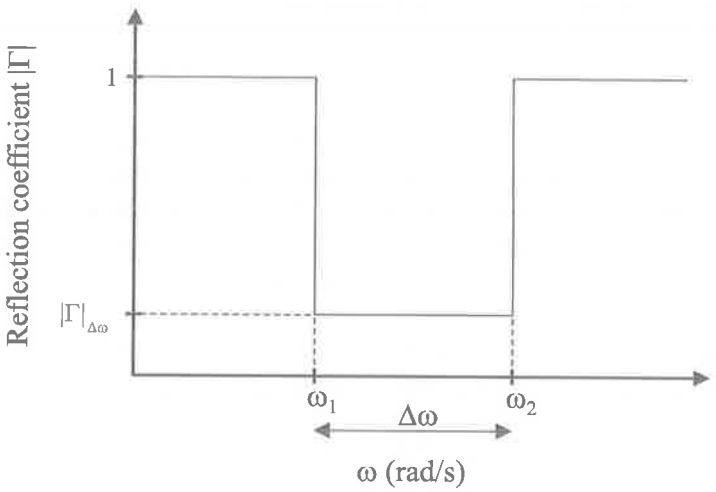

Figure 6.14 Reflection coefficient for the best utilisation of π/RC [63].

Equation (6.1) places a maximum limit on the integral to π/RC. In order to completely utilise the given limit of π/RC for a desired angular frequency bandwidth ($\Delta\omega$), $|\Gamma|$ should be unity along the entire band except for the bandwidth, $\Delta\omega$ under consideration, thus implying a complete mismatch outside $\Delta\omega$. Considering a minimum achievable mismatch over $\Delta\omega$ and thus a minimum bound on the reflection coefficient of $|\Gamma|_{\Delta\omega}$ over the bandwidth $\Delta\omega$ (refer to Figure 6.14) yields (6.2) which reveals that for a given *RC* load, there is a compromise between the maximum matching bandwidth and the maximum power transfer to the load.

$$|\Gamma|_{\Delta\omega} \ge e^{-\frac{1}{2\Delta f RC}} \tag{6.2}$$

If matching is to be performed to satisfy a certain acceptable $|\Gamma|_{\Delta\omega}$ (and hence, amount of power transfer), the bandwidth may have to be reduced. On the other hand, if matching is to be performed over a certain given bandwidth, the amount of power transfer to the load may have to be compromised.

Table 6.1 Regulated UHF frequencies allocated for RFID in major geographic regions around the world.

Region	Frequency range (MHz)	Bandwidth (MHz)
Europe	865 - 868	3
USA	902 - 928	26
Japan	952 - 954	2

Using the bandwidths outlined in Table 6.1, calculations of reflection coefficient limit established in (6.2) are performed for the three regions: USA, Europe, Japan and all of these regions simultaneously. The results are outlined in Table 6.2. In the calculations, a chip resistance R of 1.3 kΩ and a chip capacitance C of 1.1 pF is assumed.

Table 6.2 Minimum achievable reflection coefficients (R = 1300 Ω, C = 1.1 pF).

Region	Theoretical bound on $\|\Gamma\|_{\Delta\omega}$
USA	1.44×10^{-6}
Europe	2.42×10^{-51}
Japan	4.91×10^{-26}
All regions	0.0214

All the values for $|\Gamma|_{\Delta\omega}$ therein are small. This means the allocated bandwidths for RFID usage will not pose any theoretical limitations towards achieving a good impedance match to the input impedance of the RFID IC.

Table 6.3 Minimum achievable reflection coefficients (R = 2500 Ω, C = 500 fF).

Region	Theoretical bound on $\|\Gamma\|_{\Delta\omega}$
USA	2.08×10^{-7}
Europe	1.24×10^{-58}
Japan	1.11×10^{-29}
All regions	0.0123

However more recent advances in the fabrication of Schottky diodes and low power CMOS processes have yielded RFID chips with chip impedances represented by a resistance R of 2500 Ω with a parallel capacitance C of around 500 fF. In this light, the cases presented above are re-evaluated and the results are presented in Table 6.3.

From the values in Table 6.3, it can be observed that in practice, if the tag chip has R = 2500 Ω and C = 500 fF, the theoretically achievable minimum reflection coefficient remains very small and thus presents no theoretical limit to the maximum power transfer to an RFID chip across the UHF RFID bands in the regions considered above.

6.2.4 Environmental Constraints

The propagation of electromagnetic waves is governed by physical laws. RFID is no exception to the laws of physics that define electromagnetic propagation. Hence it is important to understand the effects of various environmental factors on a tag antenna so that suitable antennas can be developed to overcome any difficulties.

Liquids and metals play an important role in the performance in respect to the manner in which they affect electromagnetic waves. High dielectric and lossy materials such as liquids absorb or attenuate UHF RF energy and detune tag antennas reducing radiation efficiency, while metals can either absorb or reflect it, depending on the amount and shape of the metal. The unwanted result of a tag's reduced performance when attached to materials with high dielectric constants and metallic objects need to be taken into consideration during the tag design process.

Even a carton of photocopy paper may prove problematic for RFID labeling because the liquid that affects RFID performance does not have to be an actual liquid. Paper typically has high moisture content, and it does absorb RF energy. Wooden pallets made with green wood and anything but oven dried wood present the same challenge because of the moisture content. Fresh fruits and vegetables and frozen items will also pose liquid related problems for RFID. It is important to evaluate whether the items have the potential to hold or attract moisture when considering the design of RFID tag antennas for tagging the items.

Metal is perhaps more of a challenge because it may either reflect or absorb electromagnetic waves. However the behaviour of electromagnetic waves next to a metal surface is predictable as opposed to the effects mentioned previously. In some situations, the presence of metal can actually improve the performance of an RFID tag. Irregular metal, on the other hand, will either absorb the signal or reflect it in random directions. As with liquids, there is more to metal than may be immediately obvious. Metallised foil bags and even anti-static bags can act as metal. Some materials have metallic contents or coatings that need to be considered. Rice, for example, has been stated as having a high mineral iron content that affects RFID performance [65].

The choice of void fill can affect RFID. Bubble wrap and loose Styrofoam void fill have very little effect on RFID whereas dense foam will absorb some RF energy. Crushed or formed paper, corrugate cardboard will have little effect unless it is very densely packed (and then may pose the potential of the liquid problem).

While it may seem obvious that certain products will have an adverse effect on RFID, it is possible to design antennas that take advantage of the nature of the surrounding material properties. There are RFID tag designs that can be placed directly on flat metal surfaces. These tags employ a relatively thin layer of dielectric insulation between the tag and the metal surface. This effectively turns the metal surface into part of the antenna (a finite ground plane) and can significantly improve performance by using the metal to reflect the RF signal back to the interrogator that would otherwise radiate into the item. Section 7.5 will

look at the development of a label antenna for tagging metallic objects that resulted from the investigation into the tagging of drill strings employed in oil rigs.

6.2.5 Performance Measure

While addressing the topic of RFID label antenna design it is important to consider the practical performances of such antennas when attached to an RFID label IC. The accepted metric for such performance comparisons involves taking a read range measurement. The read range of a tag is the maximum distance between a reader antenna and the tag before the reader fails to decode the tag responses while the tag antenna is favourably oriented to the reader antenna propagation field. These read range measurements may be taken in an anechoic chamber, or may be performed in a more practical environment where the tag is to be deployed.

In the theoretical estimation of read range for systems operating in the UHF spectrum two scenarios can be analysed to estimate the tag read range. These are given below.

- Tag power constrained analysis
- Reader sensitivity constrained analysis

In the tag power constrained analysis, it is assumed that the system is adequately designed so that the sensitivity of the interrogator's RF receiver is not a limiting factor (this might be the case in the event of a reader using a bi-static antenna configuration).

In such a scenario the theoretical read range of a tag with a lossless antenna may be calculated from the Friis equation given in (4.39). Thus the read range of a tag may be evaluated as given below where the radiated power of the reader and the reader antenna gain is replaced by EIRP and the available source power required at the tag antenna is $P_{r(tag)}$.

$$r \leq \frac{\lambda}{4\pi}\sqrt{\frac{EIRP_{reader}\, g_{tag}}{P_{r(tag)}}} \tag{6.3}$$

Equation (6.3) is only useful if an expression can be obtained for $P_{r(tag)}$. This requires careful consideration in case of passive tag technology. Thus if the minimum amount of power required to operate a tag is known to be P_{IC}, and it can be assumed that the tag is receiving that power, and the efficiency of the rectifer structure is η and, k_m is the power transfer factor from the antenna to the tag circuit in the presence of modulation, that is, the ratio the power reaching the tag circuit in the presence of modulation at the greatest mismatch to the available source power from the tag antenna, then (6.4) gives the minimum power required by a tag at its threshold of operation. Hence it is possible to use $P_{r(tag)}$ from (6.4) in (6.3) to obtain a maximum possible read range when having enough power to energize the tag is the constraint.

$$P_{r(tag)} = \frac{P_{IC}}{\eta k_m} \tag{6.4}$$

If however it is the sensitivity of the interrogator's RF receiver that is limiting, it is necessary to find the distance at which the received signal at the interrogator's receiver just meets the SNR of the receiver. Hence the minimum received power $P_{r(reader)}$, at which the SNR of the receiver is satisfied is given in (6.5) where NF is the noise factor of the receiver, B is the bandwidth of the receiver, k is Boltzman's constant, T is the absolute reference temperature used in the definition of the receiver noise factor NF and $(S/N)_{min}$ is the minimum signal to noise ratio needed to decode a tag reply successfully.

$$P_{r(reader)} = (S/N)_{\min} kTB(NF) \tag{6.5}$$

Thus the minimum read range can be calculated by considering the one-way signal strength for a transmission from a tag to an interrogator with the required interrogator received power given by (6.5). The result is the read range givein by (6.6) wherein the power $P_{t(tag)}$ is the power scattered back from the tag and where it is assumed that the tag is just sufficiently energised.

$$r = \sqrt{\frac{P_{t(tag)} g_{reader} g_{tag}}{P_{r(reader)}}} \left(\frac{\lambda}{4\pi} \right). \tag{6.6}$$

To calculate $P_{t(tag)}$ we need the ratio of the effective modulated power radiated by the tag antennas to the available source power from the tag antenna. We call this ratio k_b and the resulting calculation of $P_{t(tag)}$ is given in (6.7). The value of k_b depnds on how good we are in designing the modulator, and exactly how we define the effective modulated power, and that depends in turn on the form of modulation employed. In an inefficient design (in which most of the available source power is going to power the tag, so not much is backscattered), its value could be small. It could also be small if most of the power is backscattered but not in a way that is time varying or is a good expression of the type of modulation desired. Aternatively its value could be up to about 1, in the case where not much of the available source power goes into powering the tag and most is backscattered, and we are successfully using, for example, binary phase shift keying modulation.

$$P_{t(tag)} = k_b P_{r(tag)} \tag{6.7}$$

In practice the read ranges always need to be verified using practical measurements as practical RF propagation losses have not been taken into account in the Friis equation given in (4.39) and it is not easy to model the propagation loss in various environments without extensive experimental data.

One simple method for evaluating tag performance in terms of tag read range in an ideal propagation context is to use an anechoic chamber. The reader antenna can be placed at one end of the anechoic chamber, whereas the tag is placed along the axis of maximum radiation from the reader antenna. The tag should be correctly oriented on a polystyrene stand so that there is maximum coupling between the tag antenna and the reader antenna. The reader output can be monitored while the tag is moved away from the reader antenna to obtain the read range. However, it may be more practical to conduct the read range measurement in an

environment suitable to that in which the tag is to be deployed as it would give a more useful indication of the performance of the tag with respect to the application.

The following section considers identifying necessary tag antenna design requirements for a given application prior to embarking on the antenna design process. Identifying requirements will help the translation of application requirements to design requirements for an RF engineer nominated with the task of designing a tag antenna.

6.3 Label Antenna Design

The sections above considered various aspects of label antenna design that a RF engineer needs to be aware of in the design of antennas for RFID labels. The following sections will consider the label antenna design process in detail.

6.3.1 Design Requirements

Table 6.4 An outline for evaluating antenna design requirements.

Operational frequency band	Operational frequency of the tag will depend on the country or countries in which the tag is deployed. Table 6.1 outlines a list of such frequency bands.
Tag dimensions	Tag size requirements will depend on the application. For instance, the tag may need to be printed on a label for sticking on a cardboard box, embedded in plastic casing or placed within the confined space of a bottle cap.
Label cost constraints	Keeping the cost of a tag to a minimum will limit the choice of RFID ICs that can be used, as well as the type of material that can be used for constructing the antenna. Generally tag antennas are constructed using copper, aluminium or silver ink, while the material used for the substrate may be anything from paper, polyester to FR4 dielectrics.
Read range requirements	Consider the read range required by the particular application. Generally less efficient, smaller antennas may be used for smaller read range requirements. The read range is also affected by the electromagnetic compatibility regulations which control the EIRP or ERP of the reader antenna. Another important consideration is the tag orientation during an interrogation. This requires understanding the radiation pattern of the tag antenna; certain applications may require an almost omnidirectional directivity pattern, while others may only require tags to radiate in a particular direction.
Objects to be tagged	It is important to consider the nature of the item on which the tag antenna is placed as the tag antenna can be designed to be suitable to the surface on which the tag is placed (such as metal or cardboard) or tuned for optimal performance based on the contents of the tagged item (such as liquids).

Operational conditions	The antenna design, material used and the antenna package need to take into account the environmental condition to which a tag may be exposed. For instance tags may be subjected to a range of temperature, pressure or mechanical stresses depending on the application.

Due the nature of RF propagation, the effect of the permittivity of a materials on the performance of antennas, and the environment in which the tags are deployed, RFID tag designs need to take into consideration the application for which the tags are being designed. Identifying application requirements will allow the selection of a suitable tag antenna design. Table 6.4 provides a necessary set of considerations that should be deliberated upon to identify design requirements prior to embarking on the tag antenna design process.

6.3.2 Design Methodology

Generally, RFID label size requirements are restricted by costs and practical aspects that depend on the application and the frequency of operation. Most practical tag antennas for UHF operation are physically small and hence tend to have a gain of around 1.76 dBi. These antennas also have moderately small bandwidths of operation.

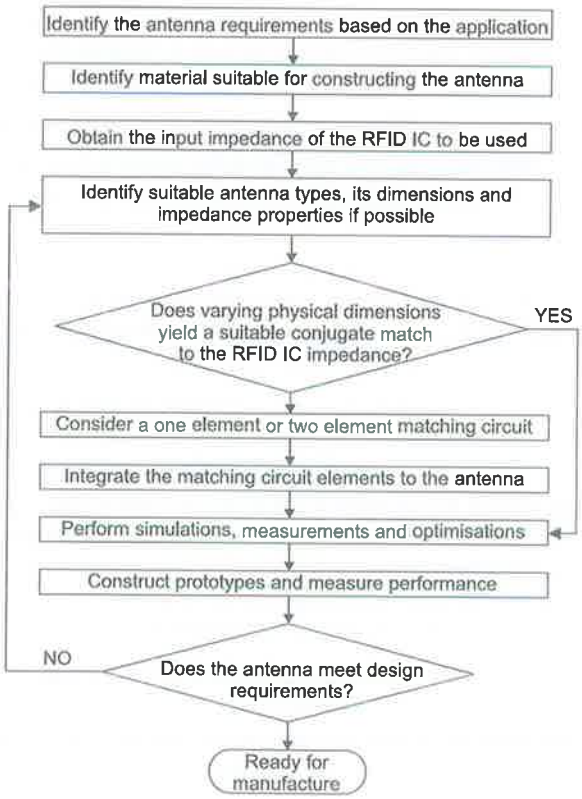

Figure 6.15 A label antenna design methodology.

Generally, designing an antenna that is tuneable in manufacture is highly desirable [235]. This implies that the antenna impedance can be varied easily in a deterministic or a predictable manner by tuning a certain physical dimension or dimensions of an antenna. This will allow the same antenna design to achieve optimal performance when coupled to a range of different RFID ICs, or when the antenna is to be placed on different packaging materials or used in different frequency bands (as may be the case in porting an antenna design suitable for operation in United States to Japan).

A systematic method for designing an RFID label antenna is shown in the form of a flow chart in Figure 6.15. This methodology is an expansion of the approach to tag antenna design found in [235] and addresses the incorporation of using existing antenna designs as well formalising a method of impedance matching by phsycial construction of impedance matching circuits on the antenna.

Once the antenna requirements are established by extracting them from the required application scenario, it is possible to look at the types of material suitable for constructing the antenna based on cost constraints and operational conditions of the application. It is then important to determine the input impedance of the RFID IC in a selected package at the threshold of operation; this might be obtained from the manufacturer or may be obtained by direct measurement using a network analyser.

It is then possible to select a suitable antenna type; there are numerous designs available, ranging from simple loops, dipoles, meanderlines, spirals and patch antennas. However if an antenna parametric study reveals that it is not possible to obtain the required input impedance within the design constraints, a designer is required to think more imaginatively.

A simple approach is to consider a single or a two element matching network that will transform the antenna impedance to form a conjugate match to the chip impedance. Such a matching network may then be physically implemented as part of the antenna, because the use of lumped circuit elements is an expensive, space consuming and a less efficient method for mass production of the antenna. Also, the absence of an efficient method for the manufacture of the antennas will result in an antenna that is too dear and the use of lumped elements is not desirable.

The resulting RFID tags inevitably tend to be too complex for analytical investigation and various numerical EM analysis software based on MoM (method of moments for planar two dimensional structures), FEM (finite element method) or FDTD (finite difference time domain method for more complicated three dimensional structures) may be used. Prior to using design tools, it is important to develop a simulation strategy and evaluate the performance of the tools by comparing simulated results with those from analytical and measurement results. Then, new antenna designs can then be modelled, and simulated to obtain desired antenna gain, input impedance, and to understand the relationship between tag antenna dimensions of the antenna input impedance, which is critical for delivering maximum power to the tag.

Once an optimal antenna design has been developed, prototypes of the antenna can be built and their performance evaluated by taking read range measurements under controlled

conditions (such as an anechoic chamber) or in the practical environment in which the tags are to be deployed. In the event the tag design is unsatisfactory, the whole design process needs to be reiterated to obtain an antenna of adequate performance.

The following sections illustrate, in detail, the design of two tag antennas suitable for tagging cases in the supply chain application depicted in Figure 6.16. Here the RFID portal is constructed by using an array of reader antennas at various orientations to ensure maximum coupling, to tags palced at, possibly different orientations.

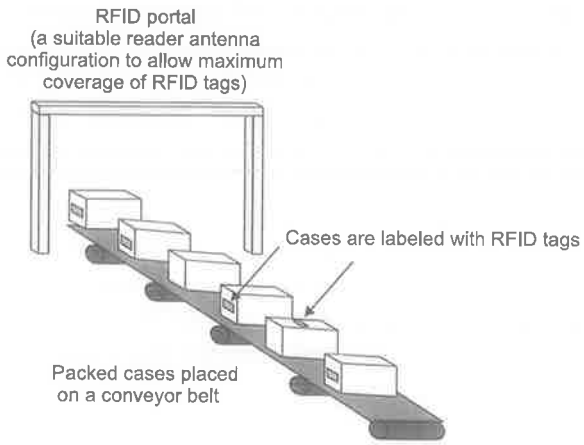

Figure 6.16 An illustration depicting the use of RFID labels in a supply chain application for tracking cases.

6.4 Illustrating a Novel Antenna Design

The antenna design methodology is best illustrated with an example. Considering the design of an RFID label for labelling cases constructed from corrugated cardboard boxes at a distribution centre, a number of antenna requirements can be found.

6.4.1 Antenna Requirements, Material and RFID IC Impedance

Assuming the tags are to operate under the electromagnetic compatibly constraints enforced by the FCC the following requirement can be outlined.

- A convenient size for an antenna for labelling most cases is approximately a credit card size label (86 mm × 54 mm)
- The frequency range of operation required is 902 MHz – 928 MHz
- While using a reader radiating 4W EIRP to meet general application requirements, tags should have read range of not less than 2 metres
- Since tag orientation can be fixed on boxes, there are no constraints on the directivity of the antenna.

Given the above requirements copper was chosen as the material of choice for the antenna, due to its superior conductivity. Considering the skin depth of copper, copper sheets of thickness 32 μm should be used. The substrate considered needs to be flexible, thin, and have a low dielectric loss. Polyesters with low dielectric constants such as PET (Polyethylene terephthalate) of thickness 50 μm are considered for the application.

The RFID straps used will be those from Alien Technologies [57]. The straps have a Class I Generation 1 chip where the input impedance of a strap is typically 18.95 – j155.8 Ω based on the parallel input impedance values of R = 1300 Ω and C = 1.1pF (refer to Figure 6.8 (a)) at 915 MHz.

6.4.2 Antenna Type

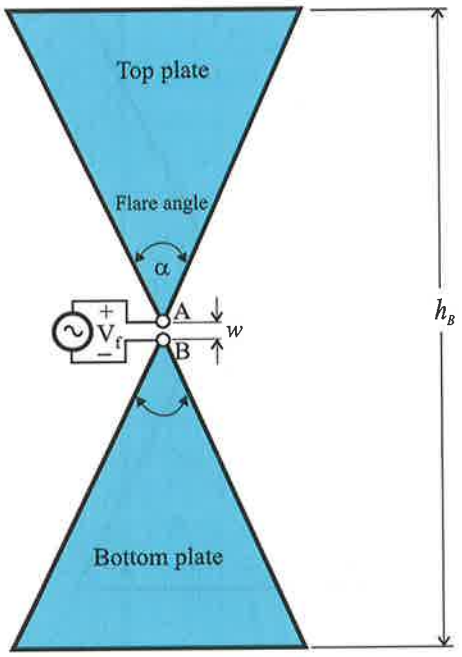

Figure 6.17 A Bow tie antenna with the height h_B and flare angle α.

It is possible to extend the analysis of the wedge above a ground plane antenna considered in Section 5.4 and use Brown and Woodward's [46] result to the analyse bow tie antennas. Then, the Ansfot HFSS simulation tool can be used to fine tune the antenna impedance properties to form a match to the label IC's input impedance.

A bow tie antenna may be thought of as a construction of a monopole wedge above ground (examined in Section 5.4) where the perfect ground plane is removed and the image under the ground plane is replaced by a physical structure as illustrated in Figure 6.17.

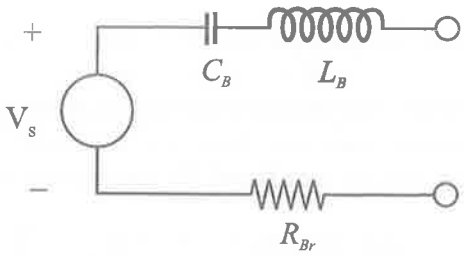

Figure 6.18 A three parameter equivalent circuit model for a bow tie antenna.

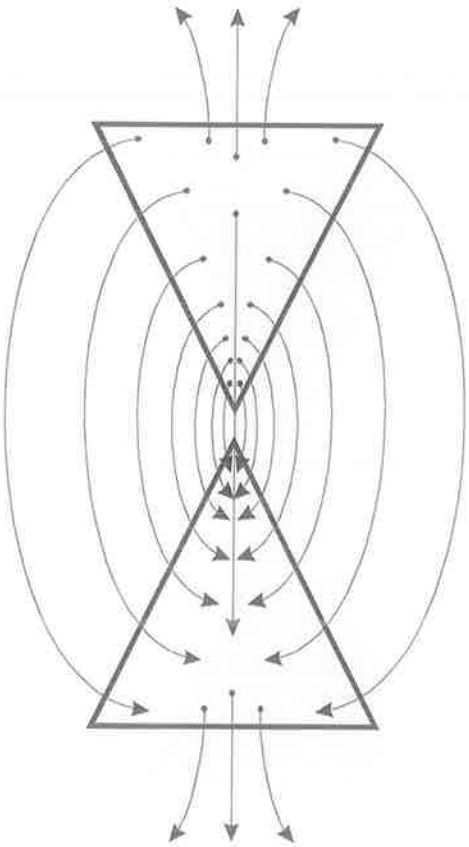

Figure 6.19 Field configuration around a bow tie antenna used for the calculation of its self capacitance.

Similarly to the analysis in Section 5.4 a three parameter model for the bow tie can be developed using a detailed study undertaken of results first published by Woodward [46] for a monopole wedge above ground. The empirical model for an equivalent circuit for a bow tie antenna is shown in Figure 6.18. The model which is derived from experimental results, confirming Woodward's result, has an associated reactance $X(\omega)$, given in (6.7), as a result of a capacitor C_B, whose value is that of the self-capacitance of the bow tie, and an inductor L_B, placed in the series circuit shown. The radiation resistance of the bow tie is represented by R_{Br}.

$$jX = \frac{1}{j\omega C_B} + j\omega L_B \tag{6.7}$$

Calculating the self-capacitance as depicted by the field lines of Figure 6.19 of the bow tie antenna by seeking analytical solutions to Laplace's equation presents a difficult problem. Nevertheless a numerical solution is both tractable and much simpler under the present circumstances and has been performed. The method of moments provides a suitable numerical approximation to the self capacitance of a bow tie antenna.

The significant finding is that, as expected and validated in Section 5.4 for a wedge above a ground plane antenna, the low frequency impedance of a bow tie antenna is mainly capacitive and this value can thus be obtained by calculating the self-capacitance of the bow tie antenna.

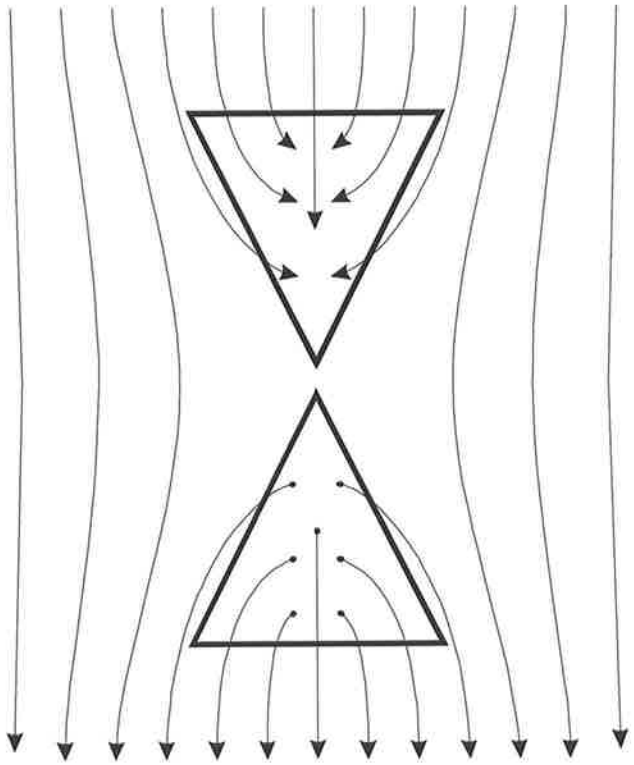

Figure 6.20 Field configuration for calculating the effective area of a bow tie antenna.

Also the radiation resistance outlined in the model parameters has significance in two ways. It allows the amount of radiated power to be calculated for a transmitting antenna, and also provides for a label antenna a means of calculating, using the reciprocity theorem, the effective area as depicted in Figure 6.20, of the antenna. The relationship between the electrostatic theory and the electrodynamic theory that allows the calculation of the effective electric flux collecting area was illustrated in Section 5.5.

The model parameters outlined in Figure 6.18 will vary for different flare angles of the bow tie antenna, as was the case for a monopole wedge above a ground plane antenna (Section 5.4). The radiation resistance, the capacitance and the inductance variation for a bow tie antenna can also be summarised by the general expressions provided in Table 6.5 where the height h_B refers to the height of the bow tie antennas as depicted in Figure 6.17.

Table 6.5 Expressions for evaluating bow tie antenna circuit model parameters.

Capacitance (C_B) in Farads	$K_{BC}\varepsilon_0 h_B$
Inductance (L_B) in Henrys	$K_{BL}\mu_0 h_B$
Radiation Resistance (R_{Br}) in Ohms	$K_{BR}(\beta h_B)^2$

In Table 6.5 the constants K_{BC} and K_{BL} are dimensionless quantities while K_{BR} is measured in Ω. The specific values of these constants depends on the flare angle of the bow tie and they can be derived from the analysis of a monopole wedge above ground antenna (constants described in Section 5.4 and given in Table 5.1) with the same flare angle using image theory. The relationship between these values for the bow tie and wedge antennas are summarised in Table 6.6.

Table 6.6 The relationship between the bow tie antenna constants and the wedge above a ground plane antenna constants.

Bow tie antenna constants	**Numerical value of the bow tie antenna constants in terms of the related monopole wedge above ground plane antenna constants**
K_{BC}	$K_{WC}/4$
K_{BL}	K_{WL}
K_{BR}	$K_{WR}/2$

However, as a consequence of deriving the bow tie antenna constants from the analysis of a monopole wedge above a ground plane antenna, the application of the model in Figure 6.18 and the derived expressions are only suitable for electrically small antennas obeying the strict limit given by (6.8), where h_B is the height of the antenna as indicated in Figure 6.17.

$$h_B << \frac{\lambda}{3} \tag{6.8}$$

6.4.3 Bow Tie Antenna Design

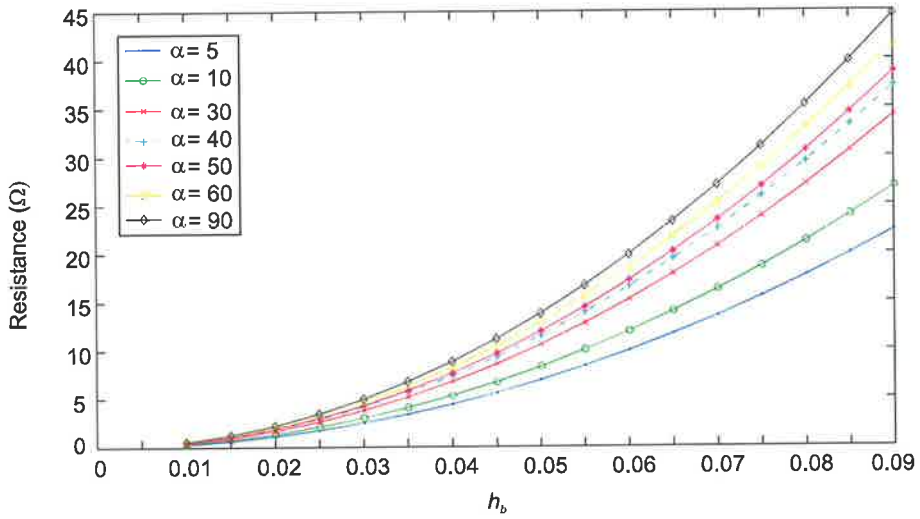

Figure 6.21 R_{Br} of bow tie antennas of various flare angles evaluated using the expressions in Table 6.5.

While it is possible to find a bow tie antenna with an adequate radiation resistance to match to an RFID chip impedance of 18.95 Ω (refer to Figure 6.21) it is not possible to find a bow tie with a flare angle and a height that will provide a conjugate match to the RFID chip's reactance as an examination of Figure 6.22 reveals that all bow tie antennas of less than 90 mm in height appear to be capacitive. Hence any resulting bow tie antenna, while possibly having the correct matching real impedance, will not be inductive to form a conjugate match to the chip impedance.

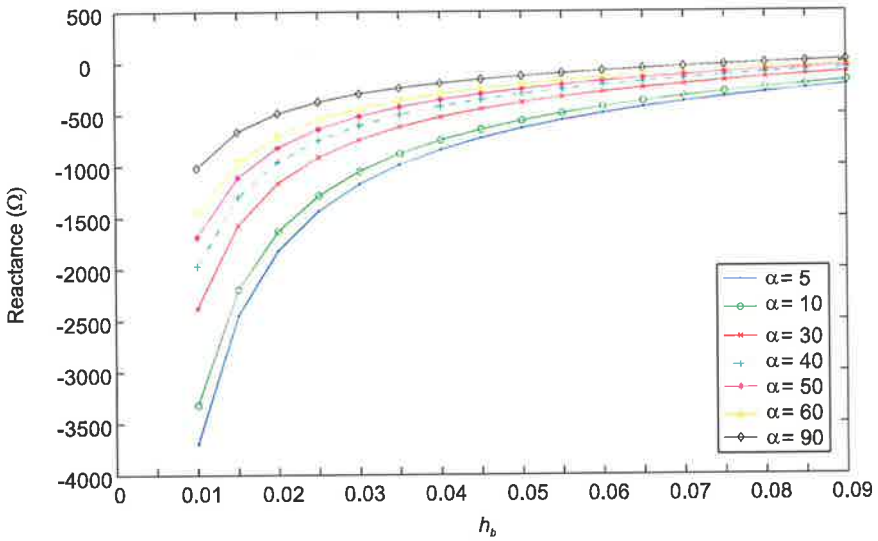

Figure 6.22 Reactance of bow tie antennas of various flare angles (from the expressions in Table 6.5).

It is clear from the discussion in Section 6.4.3 that a bow tie of less than 90 mm in height is capacitive at 915 MHz. However, this structure has the advantage that it is easy to model the resistance and reactance behaviour over a wide range of sizes and flare angles using the expressions in Table 5.2.

Considering a bow tie antenna of 80 mm in height, it is possible to calculate the input impedance of the antenna using the formulas outlined in Table 6.5. While it is possible to obtain a configuration for a bow tie antenna with a real input impedance of the order of 18.95 Ω for a match to an RFID chip's real part of the input impedance, it is not possible to obtain a conjugate match to the reactance of an RFID chip's input impedance.

The following sections will consider two different designs of bow tie antennas. Both designs illustrate the design process for a tag antenna and show how a useful antenna can be designed based on a more easily analysed and understood antenna design, such as the bow tie antenna.

6.4.4 Bow Tie Antenna with a Parallel Tuning Inductor

Considering an approximately credit card size bow tie antenna shows that the input impedance is capacitive. This is true for any bow tie antenna of less than 90 mm in height. It is then possible to engage in the thought process of discovering a suitable matching circuit for the antenna to match to an RFID chip's input impedance. The simplest possible matching circuit design, which is simple to physically incorporate into the antenna, will generally provide the most suitable antenna design.

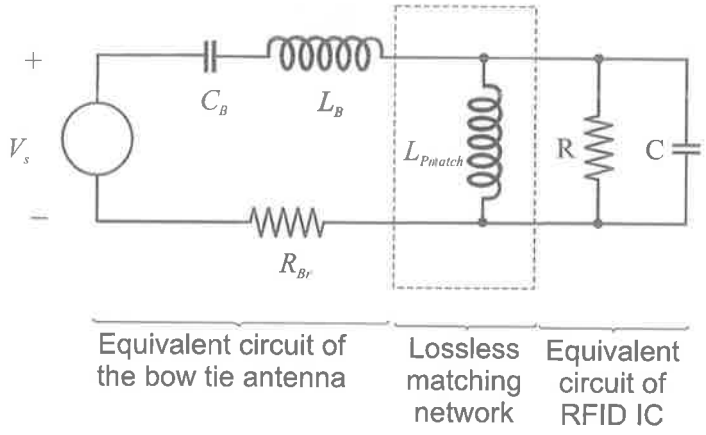

Figure 6.23 An RFID tag with a bow tie antenna and a simple matching circuit.

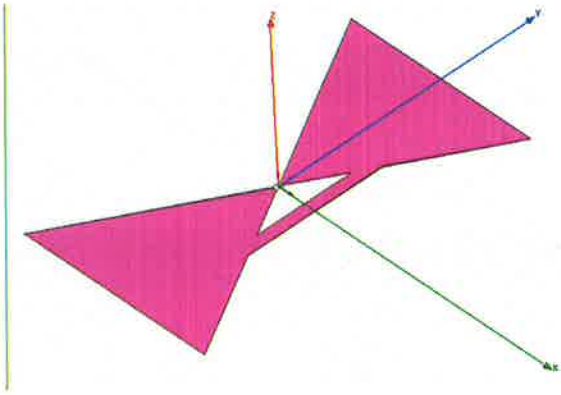

Figure 6.24 A bow tie antenna with a parallel tuning inductor.

Considering the equivalent circuit of the bow tie antenna, it is possible to contemplate a matching element as illustrated in Figure 6.23 to achieve a match to the RFID chip impedance. The tuning element, L_{Pmatch} can be physically incorporated into the antenna design. Such a parallel inductor may be conceptualised as indicated in Figure 6.24. An outline of the antenna design structure is given in Figure 6.25. The copper strip across the wings of the bow tie can then be adjusted to find a suitable inductor value to form a match between the bow tie impedance and the RFID IC impedance.

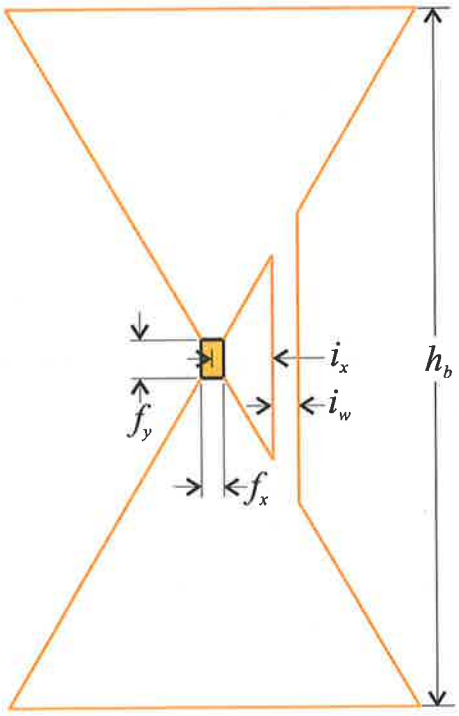

Figure 6.25 Bow tie antenna design structure with a parallel inductor.

Table 6.7 Bow tie antenna input impedance characteristics calculated from the empirical formulas.

R_r	C_B	L_B	h_b	Input impedance
25	0.8 pF	23.1 nH	80 mm	25 – 84j

Table 6.8 Simulation results.

h_b (mm)	i_x (mm)	i_w (mm)	f_x (mm)	f_y (mm)	Input impedance (Ω)
80	5	2.0	2	2	14.50 + 123.51j
80	6	2.0	2	2	29.00 + 172.00j
80	6	2.5	2	2	20.00 + 151.00j
80	6	3.0	2	2	17.00 + 141.00j
80	6	3.5	2	2	15.72 + 144.58j
80	6	4.0	2	2	13.59 + 139.64j
80	6.5	3.0	2	2	25.80 + 177.99j
80	6.5	3.5	2	2	21.38 +167.31j
80	6.5	4.0	2	2	18.38 + 161.25j
80	6.5	4.5	2	2	15.31 + 147.89j

Taking a bow tie antenna 80 mm in height, the properties of the antenna can be calculated as outlined in Table 6.7 where the centre frequency of operation is taken as 915 MHz. However, adding the inductive strip will modify the antenna model developed earlier since the added strip will affect the self capacitance and the inductance of the model. Instead of using an empirical method, simulation results from Ansoft HFSS (finite element method based simulation package) can be used to understand the effect of adding the inductive strip. Figure 6.25 describes the variable parameters used in the simulations while Table 6.8 outlines the results of simulations performed at 915 MHz.

Table 6.9 Tag bow tie antenna configurations.

h_b (mm)	i_x (mm)	i_w (mm)	f_x (mm)	f_y (mm)	Input impedance (Ω)
80	6	2.5	2	2	20.00 + 151.00j
80	6.5	4.0	2	2	18.38 + 161.25j

It can be observed that increasing the length of the inductor increases the inductance of the antenna, while also increasing the radiation resistance of the antenna. Making the inductor smaller reduces the inductive contribution of the strip and also the size of the radiation resistance. These results agree with the manner in which a parallel inductor transforms the antenna impedances as shown in the antenna equivalent circuit in Figure 6.23.

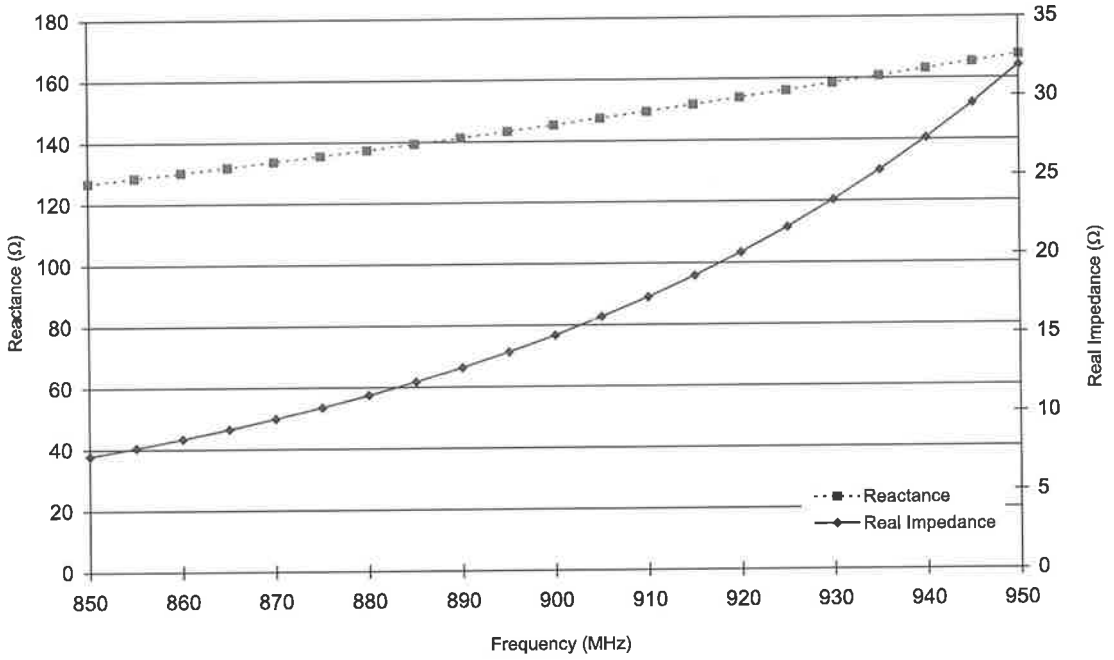

Figure 6.26 The input impedance of the parallel tuned bow tie antenna design obtained from simulated results.

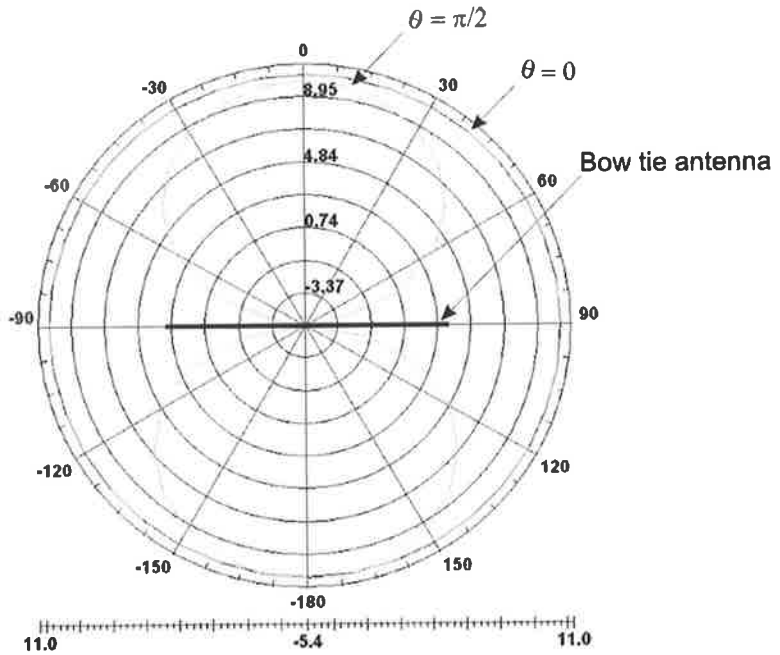

Figure 6.27 The radiation pattern of the parallel tuned bow tie antenna obtained from simulated results.

The simulation results suggest at least two possibilities for constructing the bow tie antenna as given in Table 6.9. The antenna parameters outlined can be further evaluated for their merits in terms of maximum power transfer. Considering the power transfer to an RFID chip input impedance of 18.95 – 155j Ω suggests that the second antenna design will yield better performance in terms of being able to form a conjugate match to the RFID chip's input impedance.

Considering the range of frequencies over which RFID is used around the world (outlined in Table 7.1) the variation in input impedance is plotted as a function of frequency in Figure 6.26. Figure 6.27 shows the radiation pattern of the bow tie antenna. As expected, the radiation pattern confirms that the antenna behaves similarly to an electric dipole.

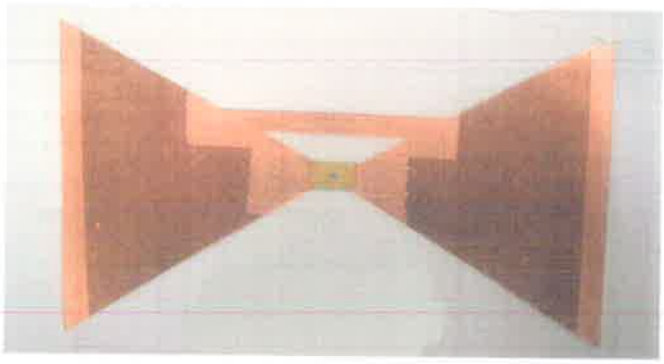

Figure 6.28 A practical construction of a parallel tuned bow tie antenna used in laboratory tests.

A practical construction of the tag can be achieved, using copper sheets of 0.035 mm in thickness or more, to the specifications given in Table 6.9 and as described in Figure 6.25 where an RFID chip can be attached using conductive adhesive. A test tag constructed in the laboratory for read range measurements is shown in Figure 6.28. Read range measurements of such a bow tie constructed and placed against polystyrene foam boxes and corrugated cardboard boxes showed a maximum read range of 6.30 m when the tag is favourably oriented with the reader antenna, and where the transmitted power is 1 W, using a 6 dBi gain reader antenna using an interrogator with a monostatic antenna configuration.

Fine tuning of the antenna dimensions for maximum performance can be easily achieved by changing i_w by stripping away small portions of the inductor until maximum read range is obtained at a required frequency. While simulation results are capable of giving a very accurate result, finer adjustments almost always need to be done manually to obtain the optimal tag dimensions for optimum performance.

6.4.5 Bow Tie Antenna with a Series Tuning Inductor

Alternatively, it is also possible to add a series inductor as a simple matching element as illustrated in Figure 6.30. Then the tuning element L_{Smatch} can be incorporated into the antenna design as shown in Figure 6.30.

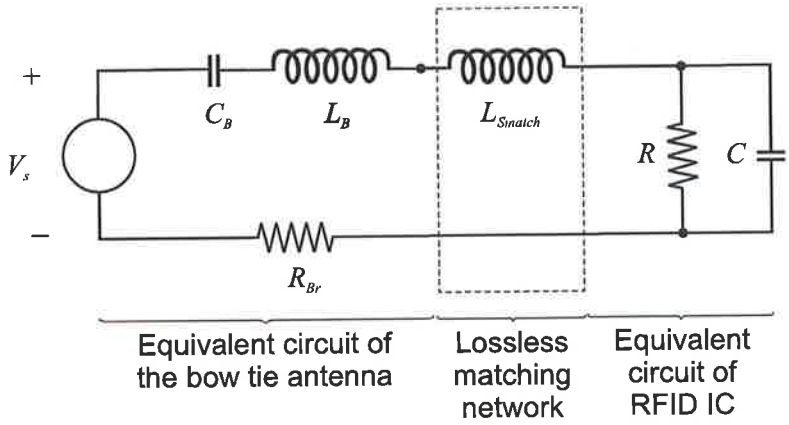

Figure 6.29 An RFID tag with a bow tie antenna and a simple matching circuit.

Similarly to the method used in Section 6.4.4 it is possible to physically incorporate the matching element L_{Smatch} illustrated in Figure 6.29 to achieve a match to the RFID chip impedance as illustrated in Figure 6.30. An outline of the antenna design structure is given in Figure 6.31. The copper strip width, l_w, the size of the gap between the wings, g, and the amount of copper removed, i_c, can then be adjusted to find a suitable inductor value to form a match between the bow tie impedance and the RFID IC impedance.

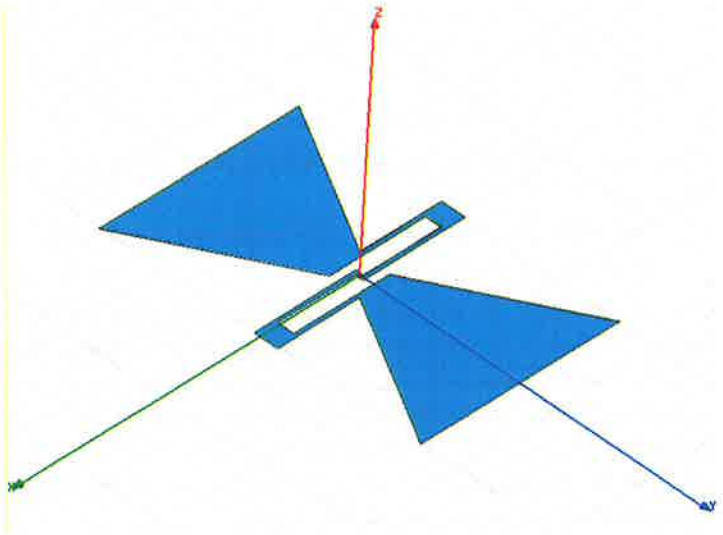

Figure 6.30 Bow tie antenna with a series tuning inductor.

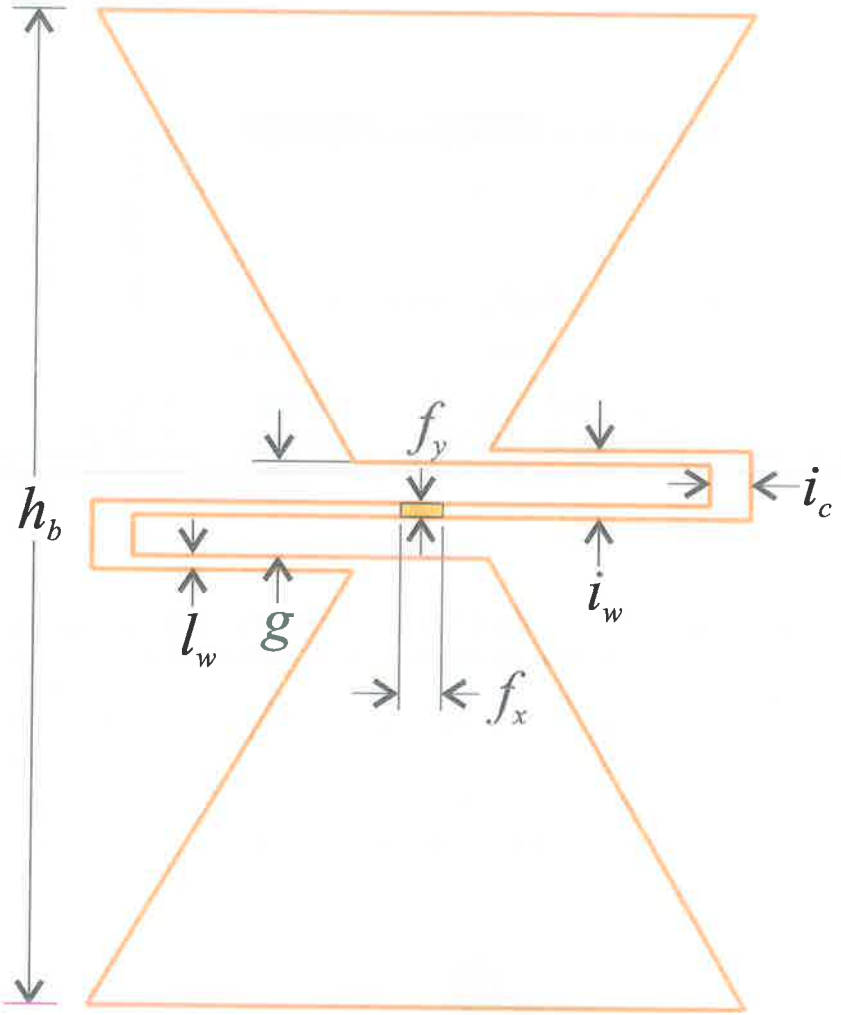

Figure 6.31 Bow tie antenna structure with a series inductor.

Considering a bow tie antenna 80 mm in height, the properties of the antenna can be calculated as outlined in Table 6.7 at a centre frequency of 915 MHz. However adding the inductive strips to form a series inductor will modify the antenna model developed in Section 6.4.2, since the added strip will affect the self capacitance and the inductance of the model. Hence, simulation results are used to understand the effect of adding the series inductive strip. Figure 6.31 describes the parametric variation used in the simulations, while Table 6.10 outlines results from simulations performed at 915 MHz.

Table 6.10 Simulation results

h_b (mm)	l_w (mm)	i_w (mm)	i_c (mm)	g (mm)	f_y (mm)	f_x (mm)	Input impedance (Ω)
71	1	6	2	9	1	1	17.00+122.00j
71	1	7	5	9	1	1	16.47+93.54j
72	1	7	5	8	2	1	16.82+98.50j
72	1	7	2	8	2	1	18.75+142.65j
72	1	7	2	8	2	2	18.97+141.61j
72	2	9	2	8	2	1	16.165+93.14j
72	1	4	3	8	2	1	17.68+80.35j
74	1	6	2	9	1	2	20.35+146.10j
75	1	4	3	5	1	2	19.093+89.86j
75	1	6	2	9	1	2	21.345+154.42j
76	1	3.5	3	4	1	2	19.52+79.33j
76	1	4.5	3	4	1	2	20.5+102.92j
76	1	6	2	9	1	1	22.732+162.33j
76	1	6	3	9	1	1	22+148j
78	1	6	4	7	1	1	17.8+101.83j
78	1	6	3	7	1	1	18.312+114.70j
78	1	6	2	7	1	1	18.95+128.34j
78	1	6	5	7	1	1	17.32+89.35j
79	1	6	6	9	1	1	18.054+90.814j
79	1	6	5	9	1	1	23.929+141.78j
79	1	6	2	6	2	1	19.37+138.00j
81	1	6	5	7	1	1	26.65+152.32j
81	1	6	6	7	1	1	19.433+98.55j

It can be observed that increasing the inductance of the strip by altering the strip size increases the reactance of the bow tie antenna. It can also be observed from looking at electric field vectors on the surface of the antenna in the near field that the strips protruding outwards also affect the self capacitance of the antenna. Hence the addition of the large inductor not only alters the reactance but it also alters the radiation resistance of the antenna.

It can also be observed that the reactance contribution to the impedance transformation from the series inductor can be increased by reducing the width of the inductor, i_w, increasing the size of the inductor by reducing i_c or by reducing the gap, g, between inductance lines. The real impedance of the antenna can easily be adjusted by increasing or reducing the height of

the antenna as outlined in the bow tie antenna model in Section 6.4.2. These results agree with the manner in which a series inductor transforms the antenna impedances as shown in the antenna equivalent circuit in Figure 6.29.

Table 6.11 Tag bow tie antenna configurations.

Antenna name	h_b (mm)	l_w (mm)	i_w (mm)	i_c (mm)	g (mm)	f_y (mm)	f_x (mm)	Input impedance (Ω)
BowAS	72	1	7	2	8	2	2	18.97+141.61j
BowS	74	1	6	2	9	1	2	20.35+146.10j

The simulation results suggest at least two possibilities for constructing the bow tie antenna as given in Table 6.11. The antenna parameters lead to slightly different physical constructions (as shown in Figure 6.32 and Figure 6.33) and both antenna designs were built to evaluate their performance.

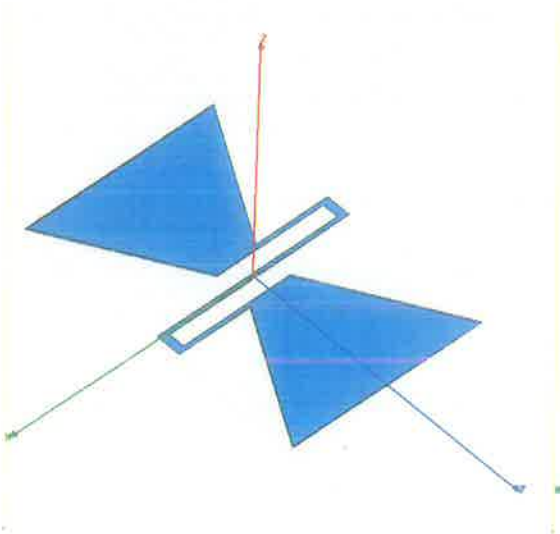

Figure 6.32 Bow tie antenna design, BowS.

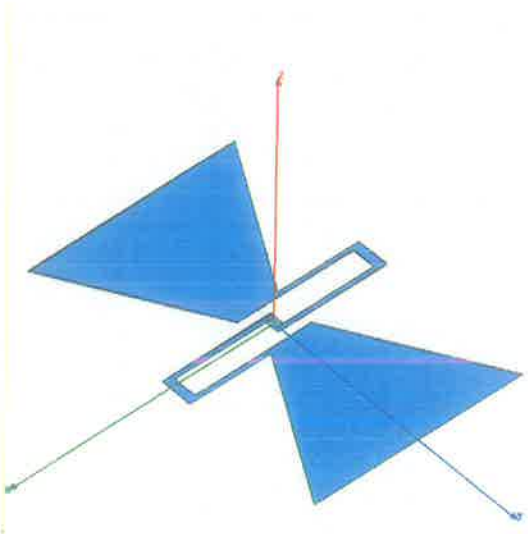

Figure 6.33 Bow tie antenna design, BowAS.

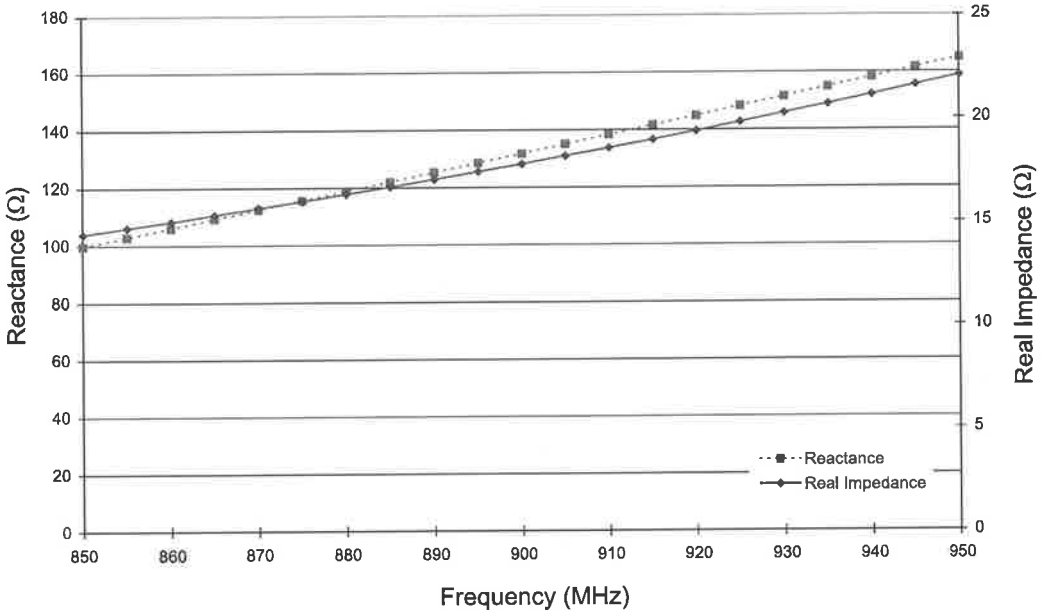

Figure 6.34 BowAS impedance variation over a frequency range of 850 MHz - 950 MHz obtained from simulated results.

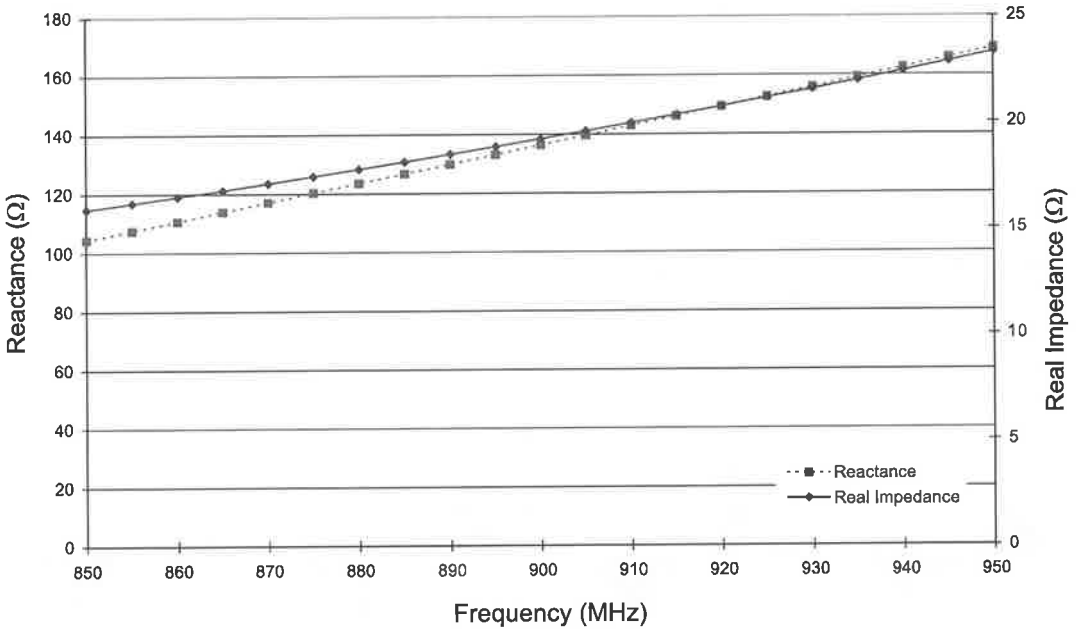

Figure 6.35 BowS impedance variation over a frequency range of 850MHz - 950 MHz obtained from simulated results.

Figure 6.34 and Figure 6.35 show the impedance variation of the two bow tie antenna designs over an operational frequency range of 850 MHz – 950 MHz. While both graphs look similar the change in real impedance over the frequency range is significantly less for the bow tie with a parallel tuned inductor (refer to Figure 6.26). Hence, the parallel tuned antennas can be expected to maintain a reasonably low radiation quality factor over a range of frequencies with the added benefit of being able to better mitigate the detuning effects from environmental factors.

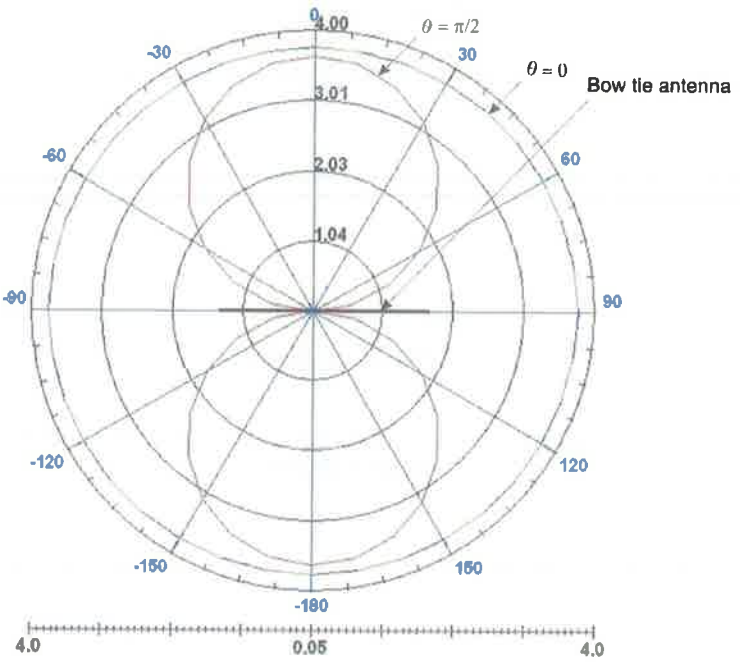

Figure 6.36 Simulated radiation pattern of BowS.

Figure 6.36 shows the radiation pattern of the BowS antenna. As expected the radiation pattern is that of a small dipole.

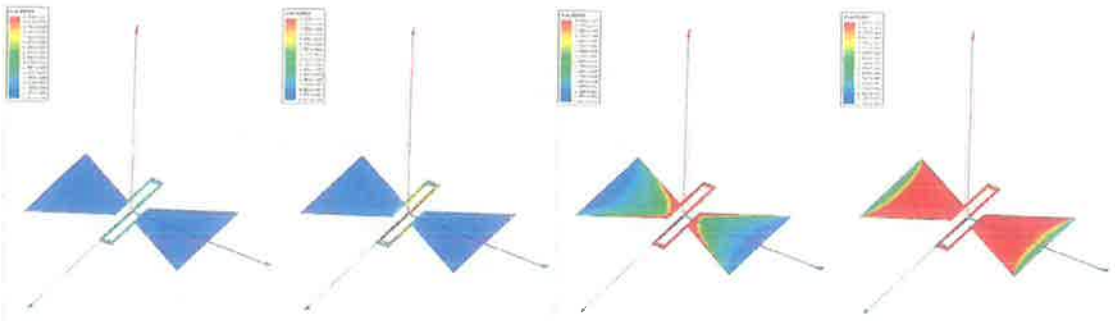

Figure 6.37 Surface current distribution plots of BowS.

The simulation tools can be used to create a better understanding of the functioning of the bow tie antenna and to confirm expectations. Figure 6.37 shows the surface current distribution on the bow tie antenna as it is fed with an oscillating current. The diagrams show the current traversing from the feed point, outwards towards the end of the bow tie wings, over a period of $\pi/2$ of the oscillating current.

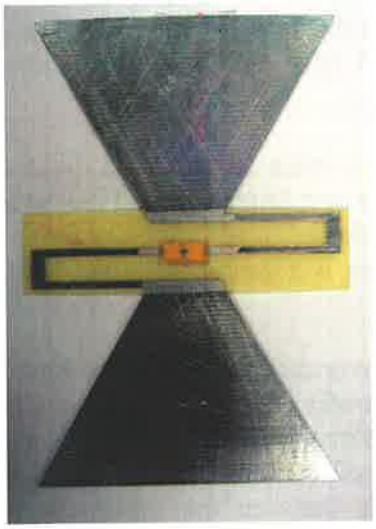

Figure 6.38 Practical construction of BowS.

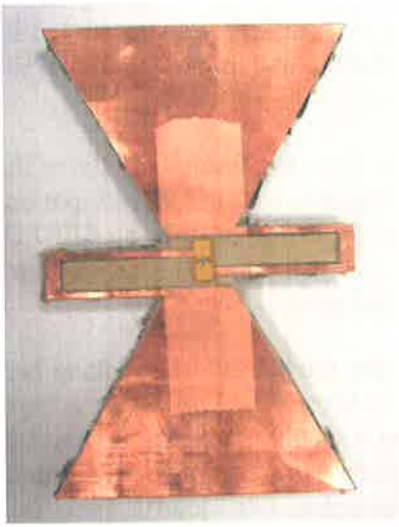

Figure 6.39 Practical construction of BowAS.

A practical construction of the tag can be achieved, using copper sheets 0.035 mm in thickness or more, to the specifications given in Table 6.11 and as described in Figure 6.31 where an RFID chip can be attached using conductive. Figure 6.38 and Figure 6.39 show two of the antennas constructed in the laboratory for testing purposes. Read range measurements (conducted in an indoor laboratory environment) of BowS constructed and placed against polystyrene foam boxes showed a maximum read range of 7.30 m when the tag is favourably oriented with the reader antenna, and where the transmitted power is 1 W, using a 6 dBi gain reader antenna with an interrogator with a monostotic antenna configuration. BowAS showed a maximum read range of 6.70 metres. Similar read ranges were obtained when the tags were placed against corrugated cardboard boxes.

If fine tuning is required for maximum read range, the antenna dimensions can be altered. One such simple adjustment is changing i_c by stripping away small portions of the inductor until maximum read range is obtained at a required frequency. While simulation results are capable of giving a very accurate result, finer adjustments almost always need to be made manually, as described previously, to obtain the optimal tag dimensions for optimum performance. However, as described in Section 7.4, such optimality may not be desirable in practice

6.5 Conclusion

This chapter has explored the subject of RFID label antenna design, and illustrated the requirements that a successful label antenna must posses by exploring the nature of RFID label antennas and by designing two RFID label antennas for tagging cases. The three different structural designs presented have illustrated the antenna design process outlined in the chapter and produced three successful RFID label antenna designs suitable for tagging cases.

It is clear from the discussion above that RFID label antenna design limits the designer to planar structures with inductive input impedance due to cost limitations and the nature of the load impedance presented by an RFID chip. Consideration of an adequate size for an antenna involves designing an antenna with an impedance that is a conjugate of the RFID chip's input impedance.

Considering the subject of matching bandwidth an interpretation of the Bode-Fano theorem provided a theoretical limit to the achievable power transfer to the *RC* load of an RFID label IC. It has been observed that, in practice, if impedance matching is performed over a certain bandwidth, there is a limit to the minimum achievable reflection coefficient. Thus for a given chip impedance (*RC* load), there is a compromise between the maximum matching bandwidth and the maximum power transfer to the load.

The RFID label antennas presented in this section have many advantages. One of the main advantages is the simplicity of the matching network. The bow-tie antenna did not require a complex matching network and both empirical and simulation methods were used in designing the matching network. In addition, an excessively complex equivalent circuit for the antenna was not required due to the smaller size of the antennas considered. Future work may be used to evaluate the performance of the antennas against various packaging material. The antennas are also easily tuneable by trimming the size of the inductors.

However, the label antenna designs considered are relatively large both physically and electrically. New applications of RFID technology and efforts to reduce costs are placing a greater interest in small antennas that are both physically and electrically small. There are various limitations and constraints that need to be addressed before successfully designing small far field label antennas. In addition there is an increasing need to address the problems of tagging objects made of materials that effect RF propagation as mentioned in Section 6.2.4. The following chapter extends the far field antenna design to electrically small antennas and consider constraints that need to be taken into account in the design of electrically small tag antennas.

Chapter 7

SMALL FAR FIELD RFID LABEL ANTENNAS

The previous chapter explored and illustrated the design of RFID label antennas. Demand for small, low cost RFID labels of adequate performance as a result of the growth and proliferation of Radio Frequency Identification Technology has created interest in physically small RFID label antennas. An antenna is called electrically small when its size is a small fraction of the wavelength. This chapter considers physically small RFID label antennas suitable for operation in the far field at UHF frequencies where the antenna is also electrically small.

Limitations inherent to electrically small antennas have been studied in the past, however RFID label antennas are subjected to a unique set of limitations as a result of the load impedance presented to a label antenna and the nature of electromagnetic compatibility regulations which restricts the power a label can obtain at a given position, along with the variety of environments in which the labels have to operate. The subject of electrically small RFID label antennas for RFID applications has not been considered previously in the literature. This chapter explores the topic of small antenna design for low cost RFID, discusses in detail the considerations that antenna designers for RFID applications need to be aware of when RFID antennas become electrically small and presents an electrically small antenna design suitable for tagging metallic objects.

7.1 Introduction

The subject of electrically small antennas has been considered in the past in notable publications such as [59], [60], [61] and [62]. RFID label antennas may be electrically large or small depending upon the frequency band of operation and the physical dimensions of the antenna. Electrically small in this context implies an antenna whose largest dimensions are a fraction of the wavelength under consideration. Generally for an antenna to be considered electrically small, its largest dimension must be approximately $1/10^{th}$ of the wavelength of interest.

The antenna analysis considered in the following sections will be for far field UHF RFID operation, where the frequency of operation is that of the FCC regulated frequency band of 902 MHz – 926 MHz. A centre frequency of 915 MHz in the 902 MHz – 926 MHz band and an associated wave length λ of 327.87 mm is used. Hence the electrically small antennas considered in this chapter will not have a dimension larger than 30 mm. In the view of using dimensionless quantities the term βa, where β is the wave propagation constant, and a is an antenna dimension, is used to describe small antenna sizes. Hence, electrically small antennas considered here has a $\beta a < 0.6$ (using an associated wave length λ of 327.87 mm, and a of $^1/_{10}\lambda$).

A discussion of the nature of antennas suitable for RFID applications has been introduced in Section 6.2.1, and it is sufficient to reiterate here that a small inductive antenna with an antenna impedance which is a conjugate match to the impedance of the RFID IC, formulates an ideal small antenna for an RFID application. This chapter will consider issues that arise as a consequence of designing and using small antennas for RFID applications [63] and illustrate the successful design of a small antenna for overcoming the problems of tagging metallic objects.

7.2 Radiation Quality Factor

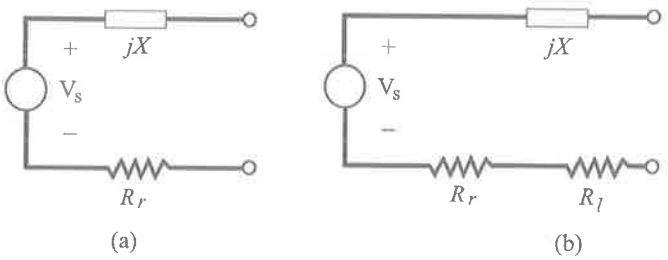

Figure 7.1 Small antenna equivalent circuit, (a) an ideal lossless antenna (b) antenna in which the ohmic losses have been taken into consideration.

Figure 7.1 shows equivalent circuits for an electrically small label antenna with an ideal voltage source V_s, a reactive element X, radiation resistance R_r and ohmic losses R_l. From the elements in Figure 7.1(a) we can construct a radiation quality factor

$$Q_r = \frac{|X|}{R_r}, \tag{7.1}$$

It has been shown that for ideal lossless and electrically small antennas, which would be enclosed completely by a smallest possible sphere of radius, a, (both electric and magnetic dipole antennas), the radiation quality factor scales as follows [61].

$$Q_r = (\beta a)^{-3} + (\beta a)^{-1} \tag{7.2}$$

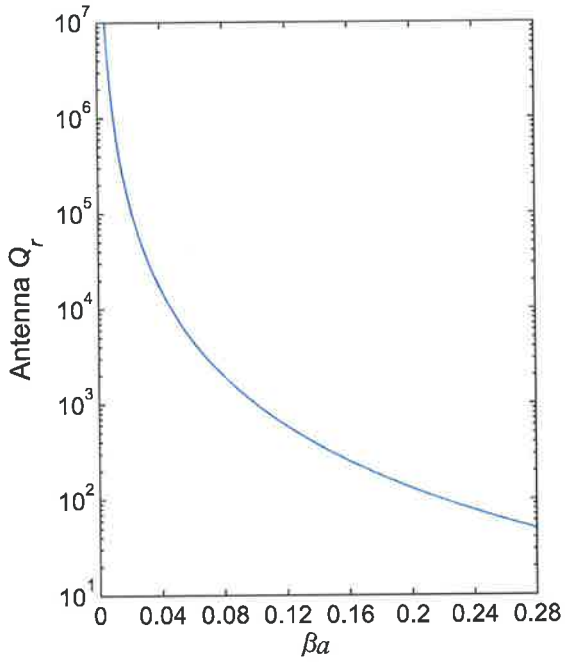

Figure 7.2 The Q_r of an ideal lossless antenna (where β is fixed at the centre frequency of 915 MHz).

Figure 7.2 indicates a characteristically high Q_r for small antennas. This is as a result of a more rapid decline in radiation resistance with respect to the antenna reactance as the size of the antenna is made small.

$$L = \mu_0 a \left[\ln\left(\frac{8a}{b}\right) - 2 \right] \text{H} \tag{7.3}$$

Consider the radiation resistance of a single turn small loop (inductive antenna), given by (5.8) and the inductance, L, of a single turn small loop, given by (7.3) expressed in terms of the antenna loop radius, a, and wire radius, b, where $\mu_0 = 4\pi \times 10^{-7}$ H/m is the free space magnetic permeability [48] and A is the cross-sectional area of the coil. Then Figure 7.3 illustrates the radiation resistance and the inductance of a small single turn coil as a function

of its size. It can be seen from Figure 7.3 that the radiation resistance and the reactance are both sensitive to antenna size but the radiation resistance decreases as a higher power than the inductance. Hence, as the antenna (lossless) becomes very small, a relatively large reactance stands between the radiation resistance and any external load to which we might wish to match. However, it should be noted here that the radiation pattern and the directivity of a small antenna is independent of its size [41].

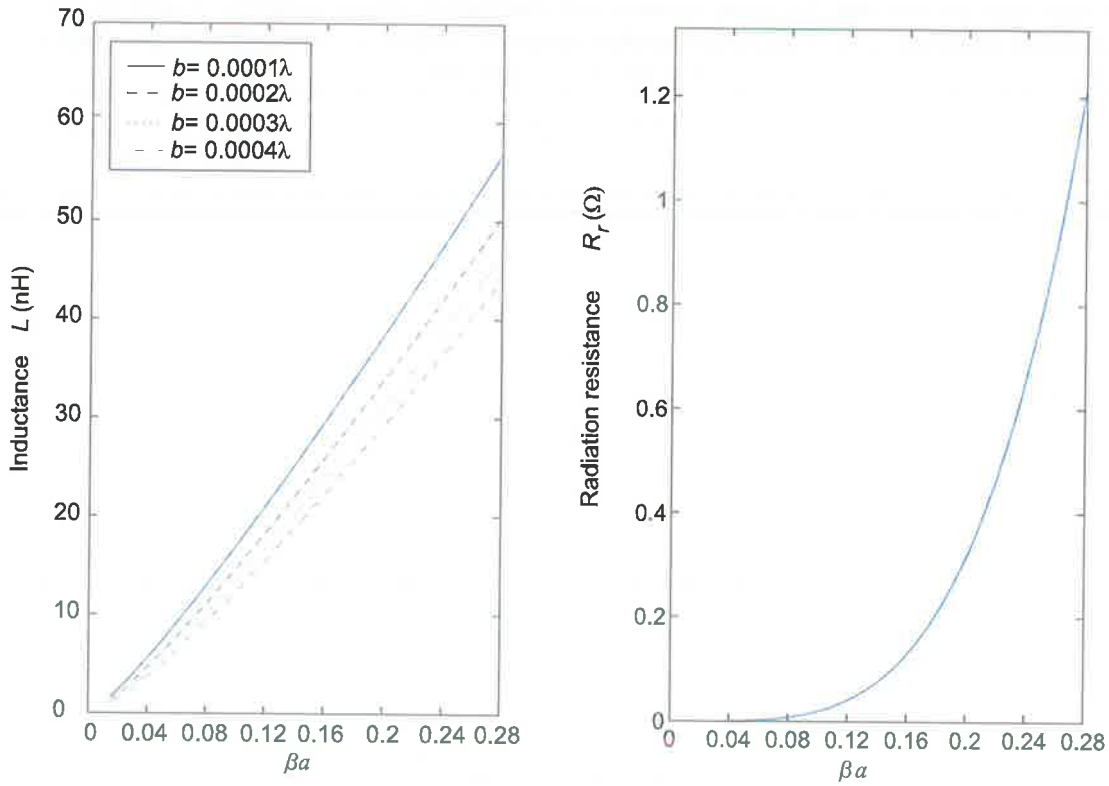

Figure 7.3 Radiation resistance and the inductance of a small loop antenna (where β is fixed at the centre frequency of 915 MHz).

The implications of a high Q_r for a small label antenna are three fold. A large radiation quality factor will affect the antenna bandwidth, create difficulties in matching to the label load and result in inefficient antennas. These implications for RFID applications will be considered in more detail below.

It should also be noted that a large quality factor will significantly increase the effect of detuning by environmental factors. Environmental detuning is a serious concern in RFID applications. These environmental factors may be as simple as the variation in the moisture content of a corrugated cardboard box.

7.2.1 Bandwidth

Assuming the bandwidth of the antenna is regulated by the radiation resistance and not the ohmic losses, antenna bandwidth BW can be defined by (7.4). Hence an increasingly larger Q_r signifies an increasingly smaller antenna bandwidth. Figure 7.4 illustrates the latter fact for a centre frequency f_o of 915 MHz. This signifies a fundamental limitation on the useable bandwidth of small antennas.

$$BW = \frac{f_0}{Q_r} \tag{7.4}$$

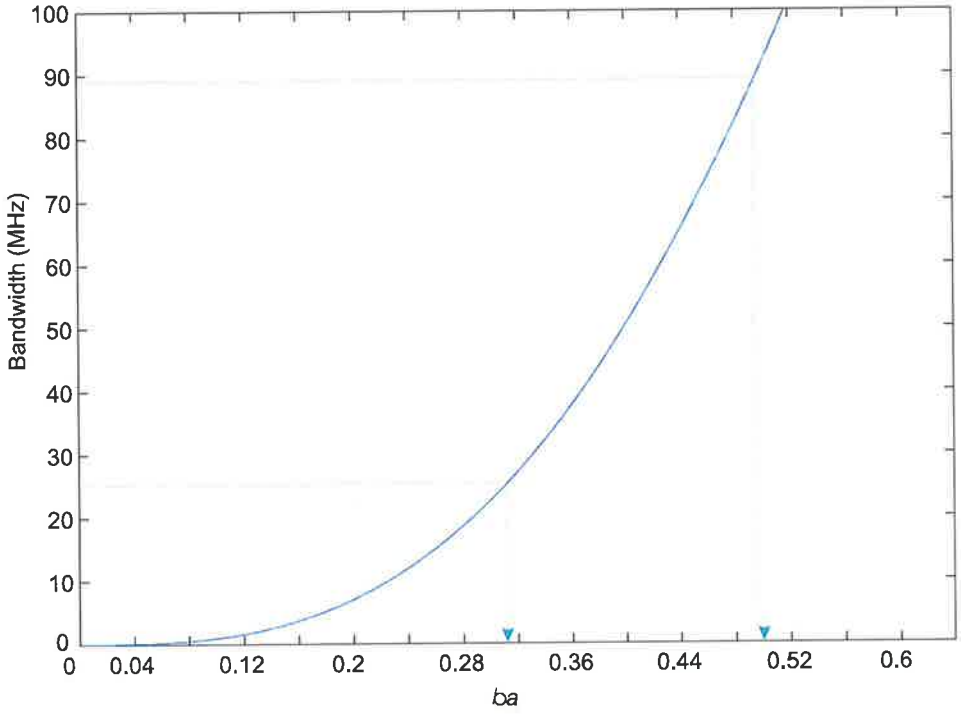

Figure 7.4 Small antenna bandwidths (where β is fixed at the centre frequency of 915 MHz).

Table 7.1 UHF RFID frequency allocations and the implied Q_r.

Region	Frequency range (MHz)	Bandwidth (MHz)	Upper limit of Q_r
Europe	865 - 868	3	288
USA	902 - 928	26	35
Japan	952 - 954	2	476
Europe, USA, and Japan	865 - 954	89	10

The UHF RFID bandwidths for European countries, United States of America (USA) and Japan are as shown in Table 7.1. Despite the limited bandwidth imposed by small antennas, it is clear from Figure 7.4 that to satisfy a bandwidth of 26 MHz (refer to Table 7.1), theoretically, requires an antenna whose largest dimension is greater than 15 mm, while to satisfy the widest bandwidth of 91 MHz (refer to Table 7.1) requires an antenna whose largest dimension is at least 27 mm.

Considering the largest available bandwidth given in Table 7.1, an upper limit of 35 for Q_r can be calculated for an antenna with adequate bandwidth for operation in the United States of America. However if the antenna were to be able to adequately cover the entire spectrum in all the listed countries then an upper limit of 10 for Q_r can be calculated. Figure 7.5 illustrates minimum theoretical sizes of electrically small antennas expressed in terms of the largest antenna dimension a and the wave propagation constant, β, needed to achieve the upper limit bandwidths outlined in Table 7.1. It is clear from the illustration in Figure 7.5 that, theoretically, it is possible to expect an electrically small antenna to have a radiation quality factor that is below the upper limits indicated in Table 7.1. It can also be seen that an antenna with a bandwidth of 89 MHz requires an antenna that is on the borderline of being classified as an electrically small antenna.

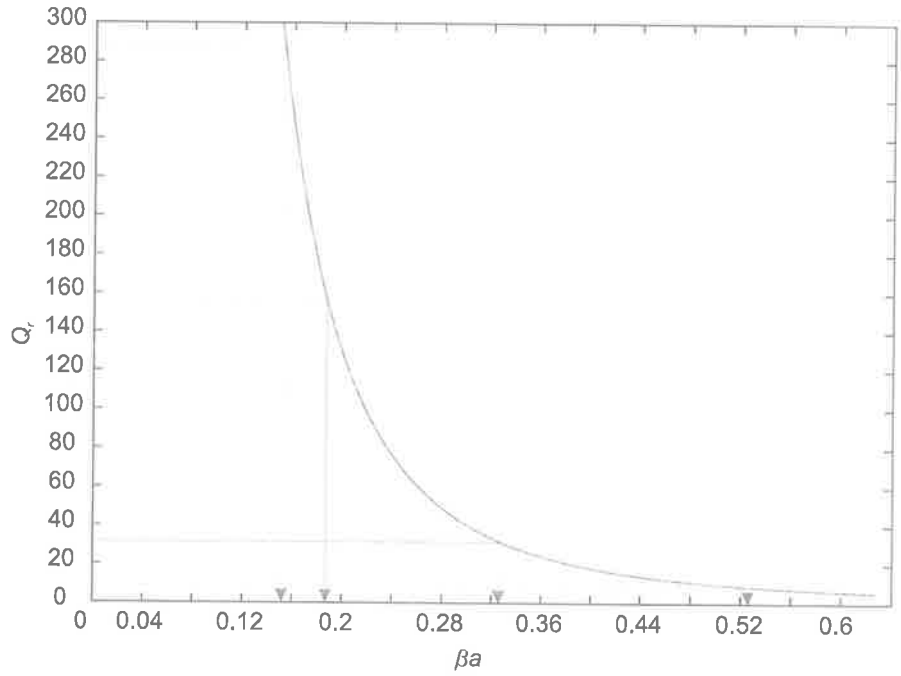

Figure 7.5 Small antenna radiation quality factors (where β is fixed at the centre frequency of 915 MHz).

7.2.2 Matching

Secondly, a large radiation quality factor, Q_r, will result in poor matching to RFID label ICs as they have quality factors in the range of 8 – 10. The large values of Q_r for small antennas are due to their increasingly vanishing radiation resistance in relation to their reactance. Thus small antennas will have an inherent difficulty matching to the real impedance RFID ICs. The subject of impedance matching is considered in more detail in Section 7.4.

7.3 Antenna Quality Factor

The concept of an antenna quality factor was introduced and defined in (4.28). The antenna quality factor takes into account the antenna ohmic losses. Unlike electrically large antennas, practical small antenna structures that are electrically small have ohmic losses that are significantly larger than their radiation resistance. For instance, a single turn loop antenna ohmic losses, R_l, can be characterised by

$$R_l = \frac{a}{b}\sqrt{\frac{\omega\mu_0}{2\sigma}} \tag{7.5}$$

where ω is the angular frequency of operation and σ is the conductivity of the antenna material.

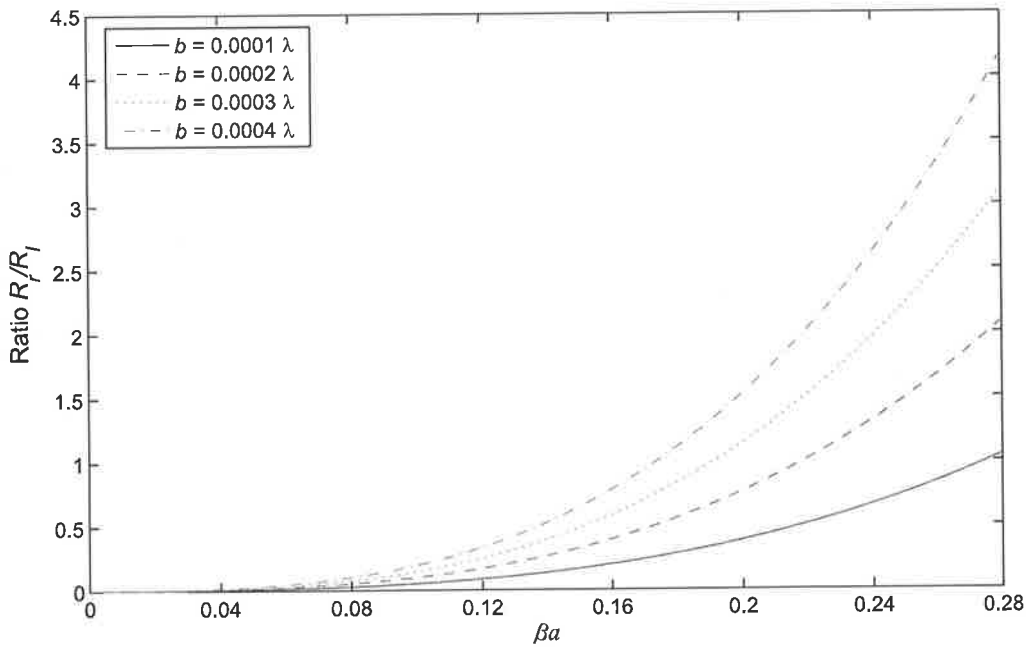

Figure 7.6 Comparison of the radiation resistance and the loss resistance of small loop antennas (where β is fixed at the centre frequency of 915 MHz).

Figure 7.6 shows the relationship between R_r and R_l of an electrically small antenna. It can be observed that the ohmic losses of an electrically small antenna become much larger than the radiation losses of the antenna as the antenna size is made smaller. The dampening effect produced by the increasing losses has two implications for RFID applications and are discussed below.

7.3.1 Bandwidth

Primarily, the loss resistance provides a dampening effect to reduce the antenna quality factor Q_a and thus broadening the bandwidth of the antenna, assuming that the bandwidth of the antenna is now regulated by both the radiation resistance and the loss resistance as given in (7.6). Hence this further reinforces the view presented in Section 7.2.1 that electrically small antennas of adequate size are capable of meeting UHF RFID operational bandwidth requirements, albeit with the detrimental affects of increased losses in relation to the radiation resistance.

$$BW = \frac{f_0}{Q_a} \tag{7.6}$$

7.3.2 Efficiency

Secondly, the increasing loss resistance reduces the antenna efficiency. The efficiency of an antenna e_{ant}, is given by (7.7). It is clear from the discussion in Section 7.3.1 and (7.7) that increasing loss resistance aids in broadening the bandwidth of the antenna albeit at the cost of inefficiency and thus reduced range of operation. Figure 7.7 illustrates the declining antenna efficiency with antenna size for electrically small loop antennas.

$$e_{ant} = \frac{R_r}{R_r + R_l} \tag{7.7}$$

As antenna designers attempt to find the "best" antenna for RFID applications, it is clear from the above discussion that making antennas smaller will not yield impedances of the order 18.95 + j155.8 Ω required to match to a typical RFID label IC with impedance of 18.95 – j155.8 Ω. However, making antennas smaller will lead to the creation of inefficient radiators, as was illustrated with an electrically small loop antenna [63]. Nevertheless, theoretically, electrically small antennas are capable of meeting UHF RFID bandwidth requirements.

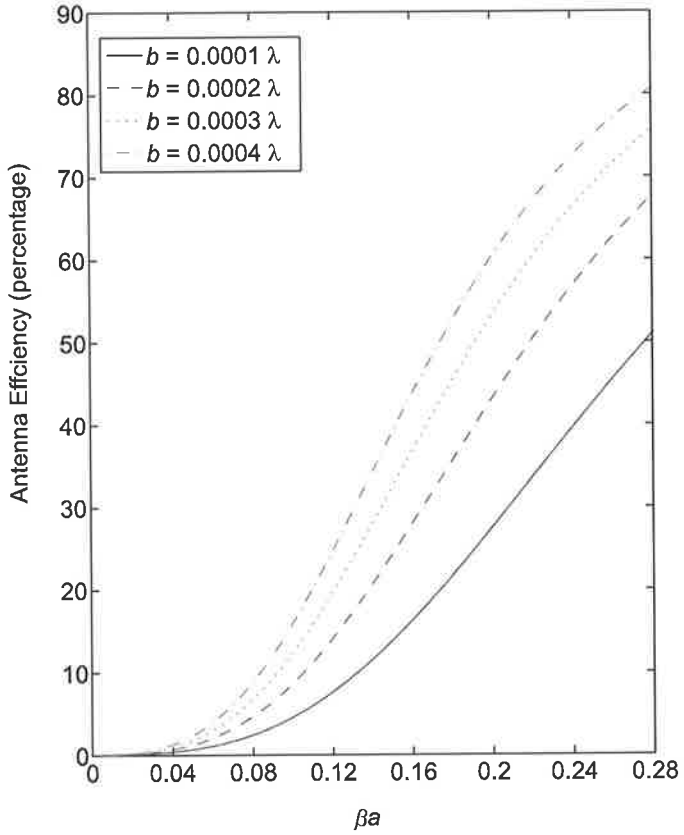

Figure 7.7 Efficiency of a small loop antenna (where β is fixed at the centre frequency of 915 MHz).

7.4 Difficulty: Narrow Bandwidth Antennas and Impedance Matching

Table 7.1 outlines the UHF frequency spectrums used by several countries around the world. It can be seen from the table that an antenna with a bandwidth of 26 MHz is required to cover the spectrum allocation in the US (which is the broadest available UHF spectrum) while a bandwidth of 89 MHz will be required by a single antenna to cover the spectrum allocations of the three regions under consideration. This will require an antenna with a Q_r or Q_a (depending on antenna size) of at least 35 or 10, respectively. However it has been illustrated in Section 7.2 that small antennas are inherently narrow band due to their high quality factors.

The input impedance of UHF chips considered (refer to Section 6.2.3) have quality factors of approximately 8 (R = 1300, C = 1.1 pF) and 7 (R = 2500, C = 500 fF). Thus the chip input impedance is inherently broadband and the available bandwidth adequately covers the cross regional bandwidth of 89 MHz.

This predicament forces an antenna designer who is limited by small radiation resistance values, due to smaller antenna size and cost limitations, to produce less efficient antennas with increased losses to achieve a conjugate match or to produce more efficient narrow band antennas at the expense of limited bandwidth of operation and impedance mismatch. It is never easy to achieve the right balance and it is even more difficult to alter electrically small antenna which also tend to be physically small at UHF frequencies, to achieve impedance matching.

Nevertheless it should be stated here that the imaginary component of the RFID IC input is not negligible in relation to the real impedance, thus it must be matched as well as possible through proper tuning if optimal tag performance is to be expected from an RFID label antenna.

However, due to reasons resulting from the magnitude of the radiation quality factor it is not necessary in practice to achieve a fine grain match to an RFID IC's input impedance as this might be detrimental to the performance of the tag. In a commercial development environment it is generally a good practice to allow the real impedance of a tag antenna to be larger than that of the RFID label IC, and the reactance of the label antenna to be smaller than the reactance of the RFID label IC, so that the antenna will always have a lower Q_r than the RFID IC. This will mitigate the effect of environmental detuning while providing the chip with adequate power during its load modulation cycles.

Chapter 6 has highlighted the problems of antenna performance when placed against metallic surfaces. The sections above have highlighted the difficulties of designing electrically small tag antennas. The following section will present a tag antenna design that performs well when placed on metallic objects and is also electrically small.

7.5 A Novel Electrically Small Antenna for Tagging Metallic Objects

Under the categories of item, case and pallet level labelling, a number of challenges have arisen related to developing RFID label antennas suitable for labelling metallic objects, objects filled with fluids and other such objects posing a barrier to radiation in the UHF region, because fluids tend to absorb radiation while metallic objects tend to reflect radiation. The material on which an antenna is placed also changes the properties of the antenna due to the permittivity and the permeability of the material. It is evident that new antenna designs are needed for identification of items manufactured with materials that affect radiation and the performance of antennas.

RFID tags embedded in metal may be used for metal asset tracking applications. Previous work [50] has been performed at HF frequencies but there is increasing interest in UHF frequencies as the frequency of choice for object tracking. In this section it is shown that suitable antennas can be made from capacitive structures sensitive to electric fields perpendicular to the metal surface.

An HFSS simulation environment for antenna applications was used to relate simulated results to empirical results. Both the simulations and the measurements indicated that the antenna concept is suited to the application of tracking metal assets.

7.5.1 Antenna Requirements, Materials and RFID IC Impedance

The drill strings used in oil drilling have a limited life in that a section may be used 400 times or until the diameter reduces 20 mm from wear against the surrounding rock. These drill string pieces vary in size from 250 mm to 1067 mm in diameter and are made from steel. The RFID tag is to be located in a 20 mm diameter hole at a depth below the surface of at least 10 mm after wear has occurred. RFID labelling the pieces of the drill string allows the recording of usage and location data at a particular depth of the drill head, aiding in the planning of a rotation scheme, so that any one particular section is not constantly at a position of high stress.

Interest in the analysis of electric field coupling to Radio Frequency Identification (RFID) antennas placed on or against metal, and some informal measurements based on some assembled structures, had suggested that the electric field coupling mechanism might be suitable for use in small RFID labels placed on or against metal, and excited in the far field by UHF RFID readers.

Assuming the tags are to operate under the electromagnetic compatibility constraints enforced by the FCC the following requirement can be outlined.

- The frequency range of operation required is 902 MHz – 928 MHz
- Tags should have read range of not less than 1 metre using a reader radiating 4W EIRP to meet general application requirements
- Since tags are to be embedded, the directivity of the antenna should be in a vertical plane above the metal surface.

Given the above requirements copper was chosen as the material of choice for the antenna, due to its superior conductivity. Considering the skin depth of copper, copper sheets in excess of 32 μm in thickness should be used. The substrate considered needs to be hard, be able to maintain the structural integrity of the tags, and have a low dielectric loss. Rexolite with a dielectric constant of 2.53 (up to 500 GHz), with an extremely low dissipation factor and an operating temperature range of -60° C to 100° C was chosen for the substrate.

RFID straps supplied from Alien Technologies [57] will be used in the antenna prototyping phase. The straps have a Class I Gen 1 chip where the input impedance of a strap has a typical value of 18.95 – j155.8 Ω based on an R = 1300 and C = 1.1pF (refer to Figure 6.8 (a)) at 915 MHz, at the threshold of operation.

7.5.2 Antenna Design

This problem has been addressed in the past by embedding 'U' shaped ferrite cores in depressions in the metal, and exciting the connected RFID chips through the creation of tangential magnetic fields at 13.56 MHz, some of which is encouraged to enter the hole by the high permeability of the aforementioned cores [50].

The problem of RFID labelling of large metal objects previously approached by promoting field propagation as described above by an HF technology has been translated to UHF by attempting to exploit the electric field between the antenna and its metal surround rather than attempting to promote the tangential magnetic field to dip down into the "hole" containing the RFID tag.

The antenna of interest consists of a top circular plate of a small diameter on a dielectric substrate, with the sides and base of the substrate copper clad to form a cylindrical tub. This capacitive plate is connected to the edge of the tub by an inductive track. The dielectric substrate used had a thickness of 3 mm, a diameter of 20 mm, and a relative dielectric constant of 2.53, with 1 oz. copper (35 μm). The top plate was 17 mm in diameter, and the spiral track was notionally 0.5 mm in width (refer to Figure 7.8).

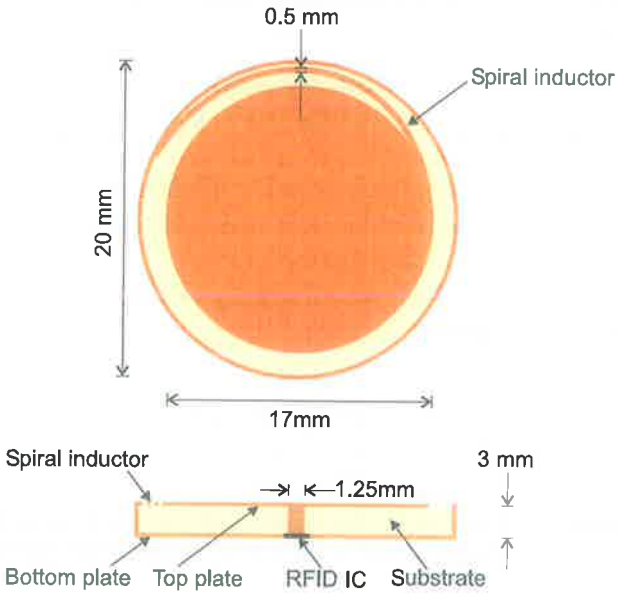

Figure 7.8 Antenna structure.

The inductive track length can be adjusted to vary the antenna reactance by reducing or increasing the inductance. The region where the track joined the top circular plate was filled in with copper tape and solder to raise the resonant frequency (reduce the inductive track length) to within the 902-928 MHz band. The proposed RFID chip connection was chosen to be at the base between the bottom of plate and a pin through a hole at the centre of the substrate to the top plate.

7.5.3 Simulation

The antenna without the external connections was simulated in HFSS using a modified FR4 substrate, to suit the characteristics of Rexolite. Perfectly matched layers (PML) surrounding a vacuum box were used for radiation calculations rather than a 'radiation boundary' in the Ansoft simulation environment.

Figure 7.9 shows the magnitude of the simulated input impedance of the antenna and the real part of the input impedance of the antenna. The large impedance at 905 MHz is suggestive of an intrinsic parallel tuned circuit. Then the simulated structure has a self resonant frequency of 905 MHz.

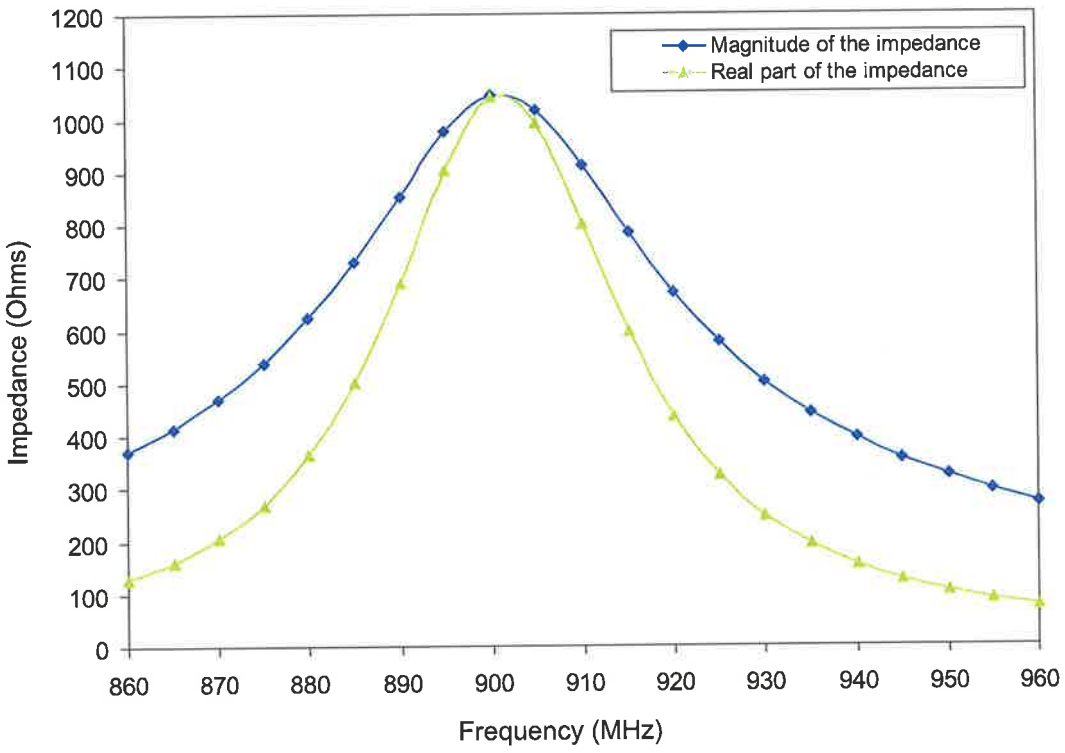

Figure 7.9 Impedance values of the antenna obtained from simulations.

Figure 7.10 shows the radiation pattern of the antenna obtained using the simulation software.

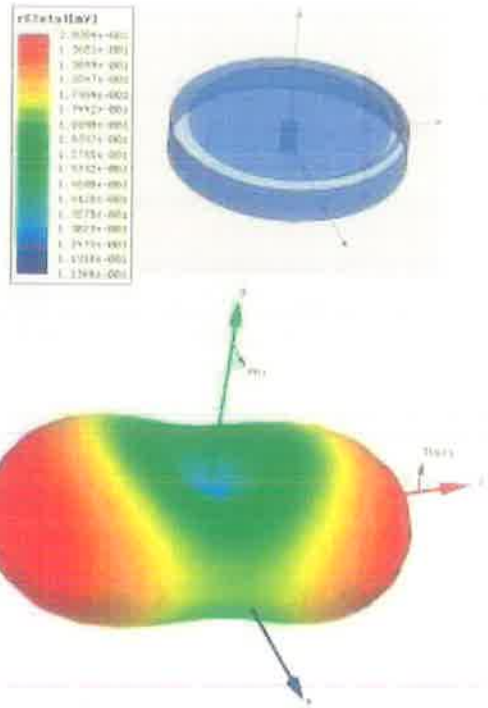

Figure 7.10 3D Polar plot of the radiation pattern obtained from simulations.

The simulated impedance values of the antenna show that the impedance of the antenna at 915 MHz is too large to form a conjugate match to the RFID IC impedance $|Z_c| = 155\ \Omega$. These results can be verified by constructing a prototype antenna and measuring the input impedance using a network analyser.

7.5.4 Measured Results

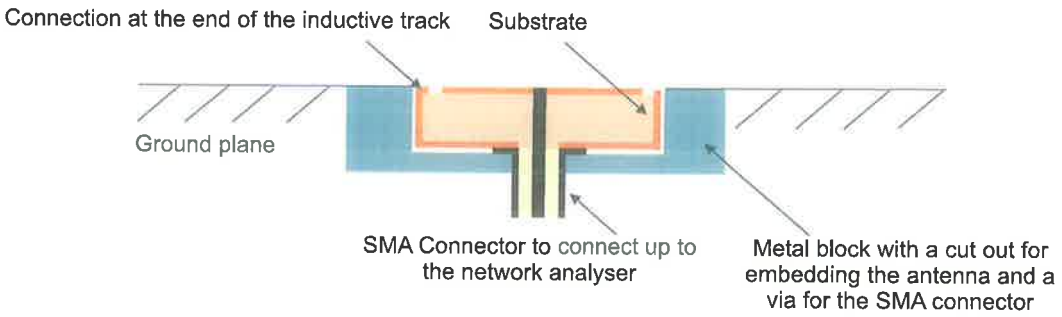

Figure 7.11 Antenna parameter measurement arrangement.

The simulated results were confirmed by constructing an antenna prototype and measuring the antenna reactance over the operational frequency range using a network analyser. The measurement arrangement is illustrated in Figure 7.11. Impedance measurements were performed with the antenna embedded into a ground plane, with a hole in the ground plane, keeping fields away from instrument cables. The network analyser was calibrated at the base of the cable connecting to the SMA connector (at the panel mount plane) to represent where a chip could be protectively located. For this investigation a panel mount SMA connector was used with a 1.25 mm diameter centre pin (via) to provide a connection to those nodes.

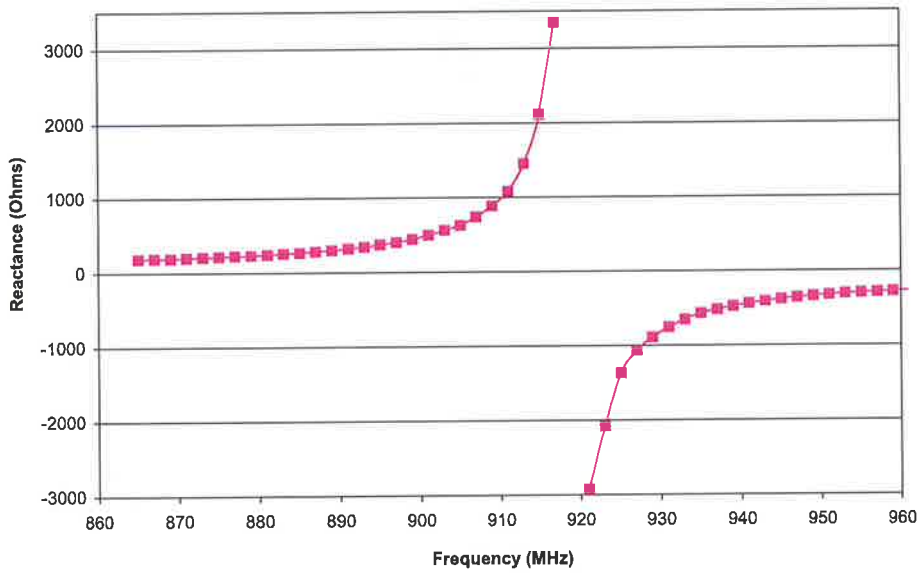

Figure 7.12 Measured reactance of antenna impedance (ohms vs. MHz).

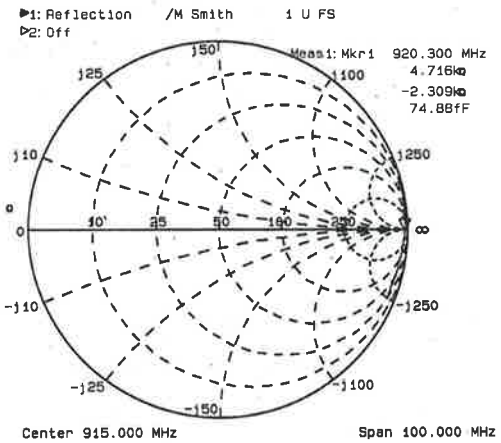

Figure 7.13 Smith chart plot of the measured antenna impedance.

It is clear from the measured antenna reactance values given in Figure 7.12 and the Smith chart trace in Figure 7.13 that the antenna exhibited, as expected, parallel resonance. As the real part of the impedance was much larger than 50 ohms, the trace was on the Smith chart periphery (refer to Figure 7.13) and thus points around the resonance (far right hand side of chart) fluctuate and cannot be trusted in respect of the real part of the impedance. As a result the real part of the impedance is not plotted in Figure 7.12 and only the reactance values which can be accurately extracted from the Smit chart plot with points on the periphery are plotted.

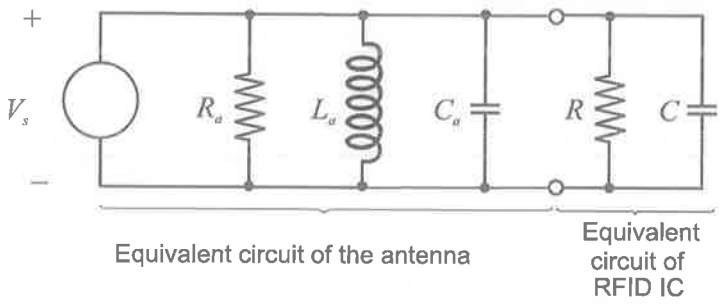

Figure 7.14 Equivalent circuit of the antenna and RFID IC.

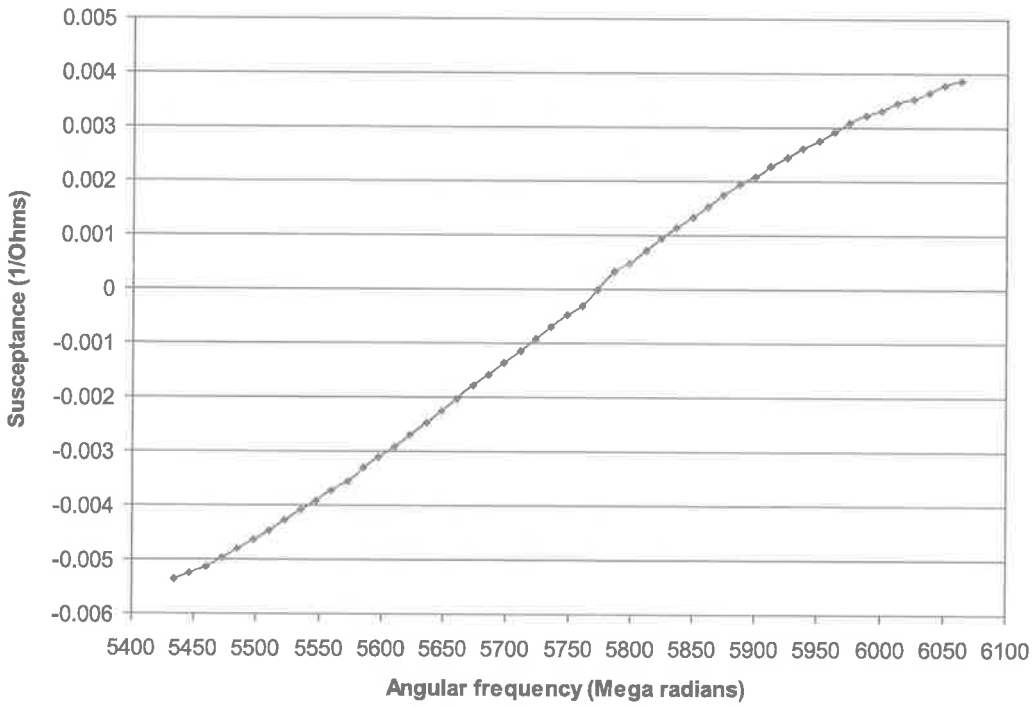

Figure 7.15 Suceptance plot of the antenna obtained from the measured reactance values.

It is possible to model the antenna equivalent circuit as a parallel resonant circuit, as shown in Figure 7.14, to evaluate the parameters L_a and C_a of the equivalent circuit by using the measured reactance values to obtain the susceptance. The susceptance versus angular frequency was plotted over a wide enough span as shown in Figure 7.15 to ensure that the susceptance measurements were stable and away from the "crowded" far right region of the Smith chart. From the susceptance equation of a parallel tuned circuit, $jB = j\omega C_a - j/(\omega L_a)$, the slope of the susceptance versus angular frequency plot around resonance is $2C_a$, which was found by "line of best fit" ignoring values near the resonance, and the resonant frequency taken to be where this line crossed zero susceptance. The value for C_a was found to be 8.8 pF, which for a resonant frequency of 919 MHz yielded an inductance L_a of 3.4 nH.

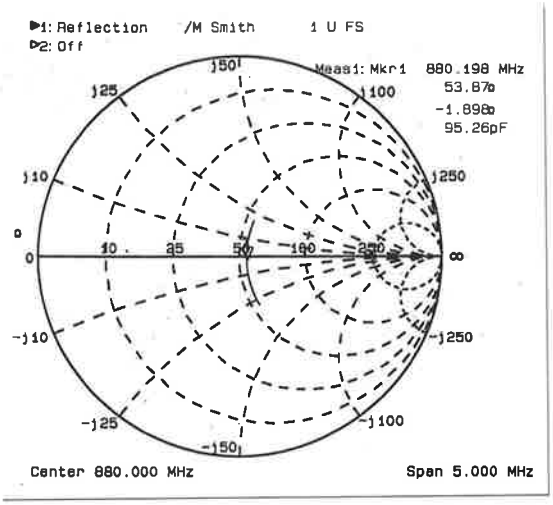

Figure 7.16 Measured antenna impedance.

This parallel circuit was then loosely coupled to a network analyser by removing a ring of copper around 2 mm in diameter from where the SMA centre conductor connected to the top plate of the antenna so that the centre conductor and top plate were capacitively coupled. The result was that the circuit resonated at a lower frequency but was an effective series circuit (detuned open), and the coupling could be arranged for the resistance at resonance to pass through 50 ohms (the Smith chart centre). Figure 7.16 shows the resulting Smith chart observed for the loosely coupled configuration, where the goal was to be close to the centre of the Smith's chart for accurate measurement of resistance. Here, the value of the coupling capacitor is not required to be known, only the resistance at the frequency of resonance.

The following calculation is based on the assumptions listed below.

- The substrate loss at the new lower frequency is equivalent to the loss at the frequency of interest (the substrate is suitable for the frequency of operation and is not being "pushed" so that losses are not highly frequency dependent).

- The copper losses are constant (the difference in frequency is not sufficiently great so as to require manipulation of the losses).

The reason for these assumptions is that the losses were not sought to be separated into different contributions from inductor losses and dielectric losses but rather the intrinsic quality factor, Q, of the antenna needed to be known for the first order calculations to evaluate the suitability of the antenna for an RFID label.

The combined losses were represented by R_a, the resistance of the parallel tuned circuit at resonance (refer to Figure 7.14). In order to extract this loss, the parallel L_a and C_a were combined at the lower loosely coupled frequency to yield an effective inductance L_{eff}, as the circuit was below the intrinsic resonance and is thereby inductive. L_{eff} and R were then related to the parameters of a series equivalent circuit at the lower frequency by the relation

$$R = (Q^2 + 1)\, r, \tag{7.10}$$

where r is the resistance value (close to 50 Ω) at the lower resonance, and Q is given by (7.11).

$$Q = R/(\omega L_{eff}) \tag{7.11}$$

Solving for R results in a quadratic equation with the higher of the two solutions being the one chosen. R was found to be 885.12 ohms, which includes the dielectric and copper losses along with the radiation resistance. This value agrees well with the simulation results shown in Figure 7.9. Thus, the intrinsic Q of the original parallel resonant circuit at its original resonant frequency was calculated to be 45.1.

7.5.5 Performance

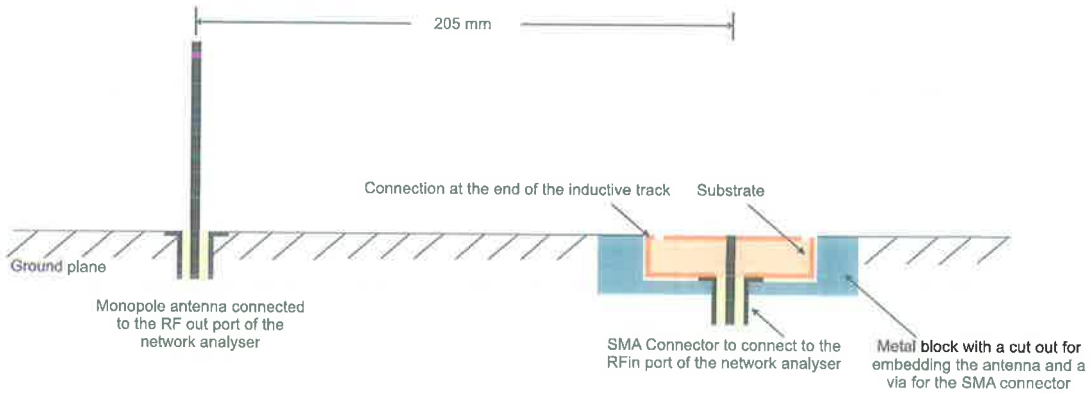

Figure 7.17 Measurement arrangement.

S_{21} measurements between the drill string antenna and a monopole field creation antenna were performed above the ground plane, with a distance of 205 mm between the monopole and plate antenna centre as shown in Figure 7.17. The 1/4λ monopole was made self resonant at 915 MHz. The magnitudes of S_{11} for the field creation monopole

shown in Figure 7.17 (in a 50 ohm system) were -13.15 dB at 884 MHz, and -9.12 dB at 919 MHz. Transmission measurements between the drill string antenna and the 1/4λ monopole (distance 205 mm) above the same ground plane gave a transmission loss S_{21} of –29.6 dB at the loosely coupled resonant frequency of 884 MHz.

The transmission loss between the antennas, when adjusted for a realistic interrogator to label distance of 1 m and radiated power of 4W EIRP suggests that the available source power from the antenna will be 160 microwatts, which is more than sufficient to power an RFID label at this distance. Hence the antenna design is successful and capable of performing adequately for the intended application.

This is a highly tuneable antenna design as the length of the spiral inductor shown in Figure 7.8 can be adjusted to vary the inductance, that is L_a denoted in Figure 7.14 to achieve a conjugate match to an RFID strap's reactance. It is important to note that altering the inductance by adjusting the spiral inductor will not effect R_a significantly as R_a is primarily due to the losses and the portion of the losses attributable to the losses in the inductor are comparatively very small.

Unlike the antenna design presented in Section 6.3 where the radiation resistance was evaluated, there was no attempt to model or calculate the radiation resistance of the drill string antenna, nor is such a calculation required. When antennas become electrically small, far field coupling volume theory suggests that only the coupling volume and the antenna quality factor need to be known. The formulation of far field coupling volume theory and an illustration of the latter mentioned postulate is discussed in Chapter 8.

7.6 Conclusion

This chapter has presented a number of considerations that must be accounted for when designing electrically small UHF antennas for RFID chips.

It is clear from the discussion above that small antenna designs mainly limits the designer to structures with inductive input impedance due to cost limitations and the nature of the load impedance presented by an RFID chip. Consideration of an adequate size for a small antenna involves designing an antenna with an impedance that is a conjugate of the RFID chip. However, the radiation resistance of small antennas are inherently small and thus achieving a conjugate match while ensuring maximum power transfer also effectively increases antenna losses. Nevertheless the added losses then ensure that the antenna has an adequate bandwidth despite being rather inefficient.

It has been shown that a significantly high Q_r or Q_a is not desirable. While a high Q_a is not meritorious to the antenna, it can help to enhance the performance of labels with tiny antennas as will be shown in the discussion on far field coupling volume theory presented in the next chapter.

It has been illustrated that electrically small UHF RFID antennas are inherently inefficient: thus impedance matching becomes vitally important in obtaining adequate performance (read range) from an RFID label. It has been shown that there can be difficulties in impedance matching when the RFID tag antenna size is limited to be electrically small.

Since electrically small antennas are by nature inefficient the best a designer can do to improve performance is to design antennas that deliver as much power as possible to the RFID IC because providing adequate power to the IC is paramount in increasing performance. This involves manipulating antennas to increase losses and to find the best match possible to an RFID chip IC albeit at the cost of reduced efficiency.

Thus understanding these physical and theoretical limitations of electrically small RFID label antennas allows a designer to overcome some of these limitations by manipulating the shape of electrically small antennas to create a matching network as part of the antenna structure. Since electrically small antennas can be viewed as an inductive or a capacitive structure this process is akin to physically creating lumped elements on an antenna structure.

However, due to size limitation only simple matching networks are possible (one or two elements) and it is not possible to implement a complex and low loss structure on an already physically small antenna. A number of simple matching networks based on series or shunt inductors or capacitors, pi-matching networks or L-matching networks can be used in designing the matching networks. The successful design of an electrically small label antenna was given Section 7.5.

The usefulness and a description of the coupling volume theory were presented in Chapter 4 in the context of near fields. It is even more useful in the analysis of coupling links in the far field involving tiny antennas which are electrically small antennas where the ohmic losses are significantly larger than the radiation resistance. The following chapter will consider the application of coupling volume theory in the far field, for tiny RFID label antennas.

Chapter 8

TINY ANTENNAS AND FAR FIELD COUPLING VOLUME THEORY

Coupling volume theory was devised for situations in which labels are placed in the near field, i.e. the energy storage field of a transmitter antenna, and also in the situation in which the radiation resistance of the label antenna is small in relation to the losses in that antenna. For operation in the HF ISM band centred at 13.56 MHz, both of these conditions are normally satisfied. For the situation when labels are placed in the far field of an interrogator antenna, but the labels are so small that their own losses are large in relation to the radiation resistance of the label antenna, it is not appropriate to use radiating antenna theory to evaluate the performance as it does not take into account the predominant ohmic losses of the antenna. It is in this context that the far field coupling volume theory provides an elegant solution.

This chapter outlines the far field coupling volume theory analysis for tiny RFID label antennas with an example of its successful application to analyse the power coupled to a tiny RFID label antenna in the far field as well as depicting its usefulness in the comparison of antennas for performance.

8.1 Far Field Coupling Volume Theory

In general it is possible to calculate the power coupled in the far field to a label with a lossless receiving antenna using (4.39), which is the Poynting vector-effective area formulation. However, it is possible to calculate the power coupled in the far field using the ideas expressed in the coupling volume formulation outlined in Section 4.14 [67 and 66] although it was originally discussed for the different context of near field excitation. The formulation of far field coupling volume theory and its applications to RFID is explored in the following sections.

Far field measures in terms of reactive power density per unit volume, W_v, and the radial component of Poynting vector S_r, were introduced in Section 4.11 where it was shown that if β is the propagation constant at the frequency under consideration then volume density reactive power can be expressed as in (8.1) using S_r.

$$W_v = \beta S_r \quad \text{Wm}^{-3} \tag{8.1}$$

Figure 7.1(a) shows an equivalent circuit for an electrically small lossless label antenna. It shows that there is a radiation resistance in series with an antenna reactance. The same resistance and reactance will be found in both the transmitting role and the receiving role of the antenna. When losses are to be taken into account the antenna will also have a loss resistance R_l so its equivalent circuit becomes modified to that shown in Figure 7.1(b). The optimum load impedance, previously R_r - jX, now becomes R_r + R_l - jX, and the power which can be delivered to that load impedance is reduced.

When electrically small antennas becomes very small (tiny), the radiation resistance of a loop antenna of radius a given by (5.8) reduces dramatically. Moreover it has been shown that $R_r << R_l$. It is in this situation that you have the option of applying coupling volume theory in the far field to determine the circuit behaviour [66 and 67]. In coupling volume theory, the source voltage in Figure 7.1(b) antenna circuit might as well be calculated from Faraday's law, the radiation resistance neglected, the self inductance calculated from the magnetostatic formula, and the loss resistance determined taking into account that conduction will only occur within a skin depth of the metal surface.

An application of the coupling volume theory in the far field is performed in the following section for a magnetic field sensitive antenna consisting of a tiny loop antenna, where the term tiny antenna is an electrically small antenna which satisfies the criterion of having a radiation resistance that is much smaller in relation to the loss resistance ($R_r << R_l$).

8.1.1 Analysis of a Tiny Loop

Consider a tiny loop antenna with a radiation resistance R_r and self inductance L with an ohmic resistance R_l and directivity g_d excited by the magnetic field of an interrogator antenna. This loop has an effective area A_e given by

$$A_e = g_d \frac{\lambda^2}{4\pi}. \tag{8.2}$$

When the loop is in a field of Poynting vector, S_r, the available source power P_a, neglecting losses is given in (8.3). This is the power which a lossless loop antenna would deliver to a load $R_L = R_r$ and is identical to that evaluated in (4.36).

$$P_a = \frac{g_d \lambda^2}{4\pi} S_r \tag{8.3}$$

However with electrically small antennas where the antenna is small in relation to a wave length, it has been shown that $R_r << R_l$. Hence it is possible to focus attention on R_l neglecting R_r completely. In view of the definition of coupling volume in Section 4.14, it is appropriate to calculate the power delivered to the losses R_l without any external load yet having been added.

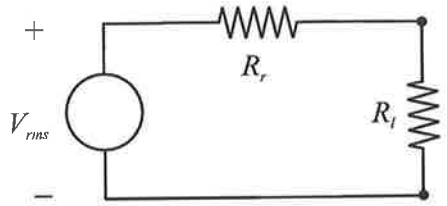

Figure 8.1 An equivalent circuit for calculating the power P_c.

Thus the tag antenna could be replaced as shown in Figure 8.1 by a voltage source V_{rms} with a series resistance R_r with $|V_{rms}|^2 = 4P_aR_r$ where P_a is the available source power. Hence a calculation of the power P_l, delivered to the losses R_l without any external load, from an available source power of P_a, can be formulated as in (8.4) for the case $R_r << R_l$.

$$P_l = \left(\frac{4R_r}{R_l}\right) P_a \tag{8.4}$$

Substituting for R_r, P_a and S_r from (5.8), (8.3) and (4.30) respectively and using the value 1.5 for g_d, provides

$$P_l = \frac{\eta^2 \beta^2 |\mathrm{H}|^2 A^2}{R_l}. \tag{8.5}$$

In order to manipulate this into a more familiar form, replace R_l by $\omega L/Q_L$ where Q_L is the quality factor of the antenna inductor and $\eta\beta$ by $\omega\mu_0$ and obtain

$$P_l = (\omega\mu_0 |\mathrm{H}|^2)\left(\frac{\mu_0 A^2}{L}\right) Q_L. \tag{8.6}$$

Equation (8.6) can be expressed using the definition for coupling volume V_c outlined in Section 4.14 and the formulation of the coupling volume of a planar coil of N turns area A and self inductance L in (4.41) into the more familiar form in (8.7).

$$P_l = Q_L W_v V_c \tag{8.7}$$

In this formula, if the label is in the far field, then far field concepts can be used to evaluate W_v, and near field concepts to evaluate P_l. Equation (8.7) is the standard form of the result from coupling volume theory for coils coupling to the magnetic field. Thus the effective area and coupling volume formulations of loop antenna behaviour are entirely equivalent, as one can be derived from the other consistently. However, the coupling volume theory formulation emphasises the internal antenna losses of the label antenna while the results from the Poynting formulation emphasise the radiation resistance of the label antenna.

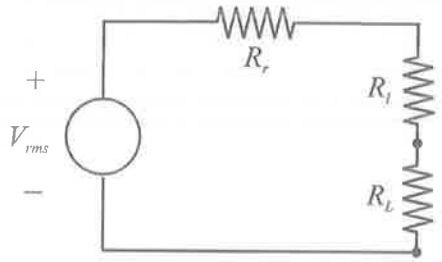

Figure 8.2 Equivalent circuit of the antenna with an external load R_L.

When an external load $R_L = R_l$ is added to the label antenna as shown in Figure 8.2 (such as connecting it with an RFID IC of input impedance R_L), the maximum power that can be delivered to that load is 1/4 the amount evaluated in (8.7) if the definition of Q is evaluated from the loop losses R_l. If however Q was evaluated for the circuit with the externally connected load as being determined by the sum of the external load R_L and the internal losses R_l of the circuit then the power that can be delivered to a matched load R_L is half that given by (8.7).

Clearly, a more useful and a more meaningful formulation of the power available from a small antenna (where R_r is negligible with respect to the losses R_l) may be obtained using the concepts outlined in coupling volume theory.

8.2 Application to Antenna Comparison

Chapter 4 has shown the application of coupling volume theory and its special importance in the design of antennas for RFID systems. While coupling volume theory is useful in the design of antennas, it can also be applied to comparing the performance of different forms of antennas and to illuminate the dependence of antenna performance on different physical parameters of the antenna.

Chapter 5 introduced loop antennas and considered the analysis of those antennas while Chapter 6 introduced and analysed bow tie antennas. The application of coupling volume theory to antenna performance comparison is illustrated in the following sections by revisiting the latter mentioned antenna structures.

An antenna's performance is adequately characterised by its coupling volume, which is derivable from the physical dimensions of an antenna as has been shown in Chapter 4. In addition, the radiation quality factor, which is also a performance parameter, can also be expressed in terms of the physical dimensions of an antenna. The radiation quality factor is important because it establishes the bandwidth over which efficient communication is possible.

Hence the coupling volumes and radiation quality factors of antennas can be used to compare their performance and the influence of physical dimensions on their performance. An application of these concepts to comparing the performance of a well shaped planar electric and magnetic field label antennas can be illustrated by considering a bow tie antenna and a single turn loop antenna structure constructed from a square of size l as shown in Figure 8.3 and Figure 8.4.

8.2.1 Loop Antenna Structure

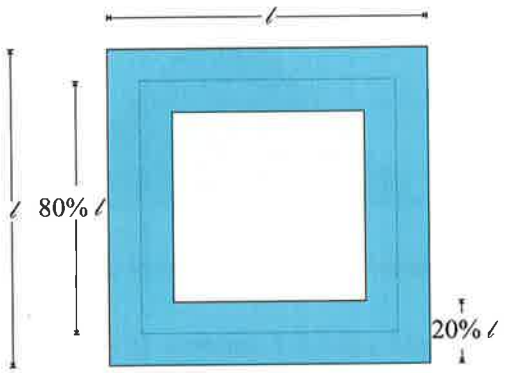

Figure 8.3 A square loop antenna created from a square shaped material of length l.

The self-inductance of a small single-turn circular coil of radius a made from wire of radius b is, given in (7.3).

The formula in (7.3) is assumed to be approximately applicable to a single turn planar square coil provided the square is replaced by a circular coil where: the area of the circular coil is equivalent to the area of the square given by its perimeter marked through the centre of the strips, the diameter D is then the diameter of that circular coil and the diameter of the circular wire d is taken as half the strip width of the square coil. A construction of the square coil is depicted in Figure 8.3.

The radiation quality factor for the magnetic field sensitive loop antenna formed with a square of side l can be calculated, under the above assumptions, to be

$$\text{Magnetic } Q_{Mr} = \frac{40}{(\beta l)^3}. \tag{8.9}$$

The coupling volume of a planar coil, which in its idealised state has no physical volume, is given by (4.41). For the loop presented in Figure 8.3 it can be shown that the coupling volume is that given in (8.10). It should be noted that the factors 40 and 1/2 in (8.9) and (8.10), respectively are approximate numerical constants.

$$\text{Magnetic } V_{Mc} = \frac{1}{2} l^3 \tag{8.10}$$

8.2.2 Bow Tie Antenna Structure

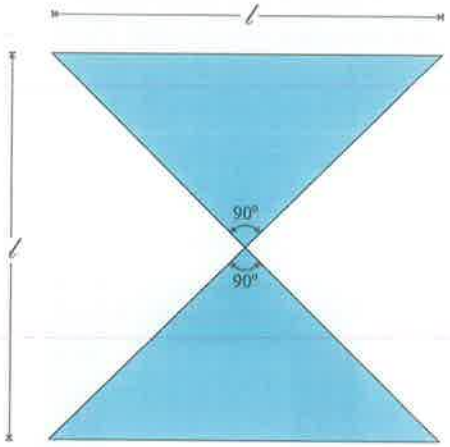

Figure 8.4 A bow tie antenna created from a square shaped material of length l.

Considering an electric field sensitive antenna, it is possible to develop expressions for both the antenna quality factor and the antenna coupling volume. An analysis of bow tie antennas was discussed in Chapter 6, with further results available from [46]. These results can be successfully applied to calculate the radiation resistance and the coupling volume of the bow tie antenna constructed from a square of side l as depicted in Figure 8.4.

The radiation resistance of the bow tie antenna in Figure 8.4 can be shown to be that given in (8.11)

$$\text{Electric } Q_{Er} = \frac{13}{(\beta l)^3}, \tag{8.11}$$

while the coupling volume of the antenna, given in (8.12), can be obtained by using the coupling volume theory for electric fields [39]. It should be noted that the factors 13 and 2/3 in (8.11) and (8.12), respectively, are approximate numerical constants.

$$\text{Electric } V_{Ec} = \frac{2}{3} l^3 \tag{8.12}$$

8.2.3 Comparison

The analysis presented is an illustration of the application of the coupling volume theory to antenna performance measurements and the dependence on physical size for performance. The illustration of the results presented for the loop and the bow tie antenna clearly indicate that the coupling volumes and the radiation quality factor for well shaped planar electric and magnetic field labels are size dependent and similar.

The coupling volume and the radiation quality factor, Q_r, are indicative of the performance of antennas of similar size. These results provide a direct means of comparing the performance and improving antenna performance based purely on antenna dimensions. The formulation above provides clarity to the two most important parameters which determine the utility of the structure by relating them to physical size, and achieves it through equations which consist directly of dimensionless ratios or can be easily put in the form of dimensionless ratios.

It can be concluded that despite the increase in effective area (given in (4.38)) of an antenna with decreasing frequency, efficient communication over an adequate bandwidth becomes impossible with decreasing frequency because Q_r becomes very large due to its cubic dependence on λ as shown by (8.9) and (8.11).

8.3 Application to Power Transfer Analysis

Chapter 7 discussed the repercussions of making electrically small RFID label antennas while the preceding section outlined how a combination of radiation antenna theory and coupling volume theory can be used to analyse tiny RFID antennas. A large portion, approximately several cents [22], of the label cost is allocated to the antenna manufacture, antenna and IC assembly and packaging. Hence there is a keen interest to produce miniaturised, on-chip antennas, to streamline the manufacturing process and to enable tagging of physically small goods such as pharmaceuticals. The following sections of this chapter considers the feasibility of such an antenna for operation in the UHF ISM band 902 MHz – 928 MHz as defined by the FCC, based on analysing the power available to operate an RFID IC, from a tiny antenna, using the far field coupling volume theory.

This analysis makes use of both far field coupling volume theory and radiating antenna theory. Radiating antenna theory is used to calculate the energy density at the label position, and coupling volume theory is used to work out what useful power the label antenna can extract from the field.

8.3.1 Miniature antenna properties

A principal target of the analysis will be the determination of the feasibility of a miniaturised antenna providing adequate reactive power to the junction capacitance of a small rectifier diode. The analysis will be based on an operational frequency of 915 MHz. The antenna material will, for a number of reasons, be assumed to be copper with an electrical conductivity of 5.8 x 10^7 S/m. The antenna is assumed to be a square loop with the dimensions given in Figure 8.3 with length, l, being 2 mm. Where appropriate, the square shape may be replaced with a circle of equal area having a radius, a, as shown in Figure 8.5. Although the antenna coil will undoubtedly be realised as a planar structure with a flat strip conductor, the flat strip may be replaced with an equivalent round wire of diameter half the width of the strip (refer to Figure 8.5).

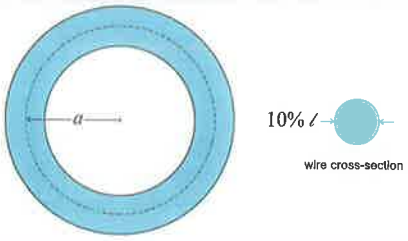

Figure 8.5 Equivalent circular coil replacing a square loop.

The RFID interrogator will be assumed to radiate a power of 1 W through an antenna of gain 4 (6 dBi), as is permitted in the FCC frequency hopping regulations [68]. An interrogation distance of 1 m will be assumed as adequate, which is well into the far field, but is not at such a distance as to represent an unfair challenge to the feasibility of the system under investigation.

The skin depth, δ, for copper at a frequency of 915 MHz is 2.18 μm. It should be noted here that it is feasible to deposit material for an antennas to at least this depth; however depositing more metal will not produce an added benefit as the metal deeper than one skin depth will not contribute to the conduction. It follows that all resistance calculations are based on the assumption that inductors of all sizes are made from strip material of the above thickness.

$$R_s = \frac{1}{\delta\sigma} = \sqrt{\frac{\omega\mu}{2\sigma}} \tag{8.13}$$

Calculation of the coil resistance involves the determination of the surface resistivity, R_s, for the selected material and the frequency of choice. The surface resistivity is evaluated to be 7.909 mΩ per square using (8.13) where σ is the conductivity of copper, μ is the permeability of copper and ω is the angular frequency.

Using the surface resistivity obtained above and a diameter ratio (diameter of the circle to the diameter of the wire) of 9, a coil loss resistance of R_l = 71.18 mΩ can be calculated using (7.5).

In this scenario it is reasonable to use the coupling volume theory, in which radiation resistance is assumed to be negligible. This can be confirmed using (5.8). It can be observed from (7.5) that, provided the number of turns is not changed, an antenna in the form of a single circular coil of round wire in which the diameter ratio is a constant will have a constant resistance, independent of size.

The self-inductance of a single turn coil of radius a, made from a round wire of radius b can be calculated by (7.3). Applying (7.3) to a single turn square coil 2 mm in length (l), with a radii ratio a/b of 9 obtained by approximating it to a circular coil having the same area (refer to Figure 8.5), the coil has a self-inductance L of 2.58 nH. At the operating frequency of 915 MHz, the reactance X of this self-inductance is 14.83 Ω. From the coil reactance and the loss resistance, R_l, an inductor quality factor, Q_L, of 208 can be calculated.

For a coil of N turns, both the self inductance and the coil radiation resistance will scale as the second power of N, and the series induced voltage will scale as the first power of N (provided the same amount of surface area is allocated to the provision of those turns as was allocated to the provision of a single turn). The available source power, the short-circuit reactive power, the coupling volume, and the quality factor will be unaffected by the number of turns, and this is one of the interesting properties of the theory. As a result most of the calculations below will be based on calculations for a single turn coil.

8.3.2 Reactive Power Density per Unit Volume

Since the label is assumed to be in the far field, the reactive power density per unit volume can be calculated using (8.1). The real power flow per unit area is evaluated using the radiation antenna theory formula given in (4.37), where g_t is the interrogator antenna gain, P_t is the transmitted power and r is the distance to a point in the field from the interrogator antenna.

The result, under the assumptions of interrogator power, antenna gain, and interrogator to label distance defined in Section 8.3.1, is a reactive power density per unit volume of 6.1 VAm^{-3}. The power density calculated using (8.1) is an appropriate, power-like, measure of the field strength available for energising the RFID label.

8.3.3 Label Coupling Volume

The coupling volume, V_c, of the label can be calculated using (4.41). For the coil antenna under consideration, with a flux collecting area of 1.6 mm by 1.6 mm, a label coupling volume $V_{c(Label)}$, of 3.192×10^{-9} m^3 can be calculated. As discussed Section 8.3.1, the coupling volume does not change with the number of turns. Hence it is possible to alter the number of turns later to achieve a different inductance, or radiation resistance if that is desired, to produce resonance with the junction capacitance of the rectifier.

8.3.4 Reactive Power in Short Circuit Label

$$W_{R(Short)} = \left(\frac{\beta g_t P_t}{4\pi r^2} \right) V_{c(Label)} \text{ VA.} \tag{8.14}$$

It is clear from the definition of coupling volume in Section 4.14.1 that forming the product of the coupling volume, $V_{c(Label)}$, calculated above and the reactive power density per unit volume calculated earlier gives the reactive power, $W_{R(Short)}$, which would flow within the label coil when it is short-circuited. The result of the calculation is short a circuit reactive power $W_{R(Short)}$ of 19.47 nVA. The resulting formula for this calculation is given in (8.14)

It should be noted here that coupling volume scales as the third power of size, refer to (4.41), hence obtaining a greater reactive power clearly involves increasing the label antenna size.

8.3.5 Power Delivered to a Tuned Label

The power P_c, which will be delivered to the losses that exist within the label when it is tuned to resonance (assuming that the coil antenna is tuned to resonance by the capacitive load provided by the RFID IC), can be obtained using the standard result from coupling volume theory for coils coupling to the magnetic field given by (8.7). The resulting equation is reiterated in (8.15).

If a quality factor Q_L of 208 is assumed from the previous calculations, a power of 4.05 μW can be delivered to the losses of the tuned circuit. If an external load equal to these losses is introduced (that is an external load of R_l) the power delivered to that external load is one quarter of the power P_c just calculated, since the power P_c is now delivered to a load of $2R_l$. The quality factor of the loaded circuit will then be half the figure of 208 calculated previously. Alternatively there may be a more substantial quality factor reduction as a result of the introduction of a real load to which power must be delivered. If the substantial quality factor reduction has occurred through the introduction of a real load (which is much larger than that of the inductor losses), to which power must be delivered, then the calculated power is given by (8.15) with that lower value of Q_L and is going substantially to that real load.

$$P_c = W_{R(Short)} Q = \left(\frac{\beta g_t P_t}{4\pi r^2} \right) V_{c(Label)} Q_L \text{ W.} \tag{8.15}$$

8.3.6 Reactive Power in Tuned Coil

The reactive power $W_{R(Resonant)}$, which flows in the antenna inductor, and also in the capacitance that it feeds, when those elements are resonant, can be obtained by (8.16), where the power calculated in (8.15) is increased by a factor given by the inductor quality factor.

$$W_{R(Resonant)} = \left(\frac{\beta g_t P_t}{4\pi r^2}\right) V_{c(Label)} Q_L^2 = P_c Q_L \text{ VA.} \tag{8.16}$$

The formulation given in (8.16) assumes that the junction capacitance contributes no loss and thus the quality factor of resonance is not affected. For a quality factor of 208, a reactive power of 832 μVA can be calculated.

8.3.7 Reactive Power Needed in the Depletion Layer Capacitance

Assuming that the rectifier requires a 1.5 V r.m.s. voltage across it for a useful output, and that the junction capacitance is approximately 0.1 pF, a reactive power of 1.294 mVA is needed in the depletion layer capacitance.

8.3.8 Analysis Results

Even if a quality factor of 208 can be achieved, the calculations performed above indicate that it is not possible to generate a reactive power in excess of that needed to service the junction capacitance of the rectifier diode. Nevertheless it should be stated here that other sources of loss such as diode losses and losses in the rectification process have not yet been taken into account. The results show that even in the absence of any other losses an adequate power transfer can not take place.

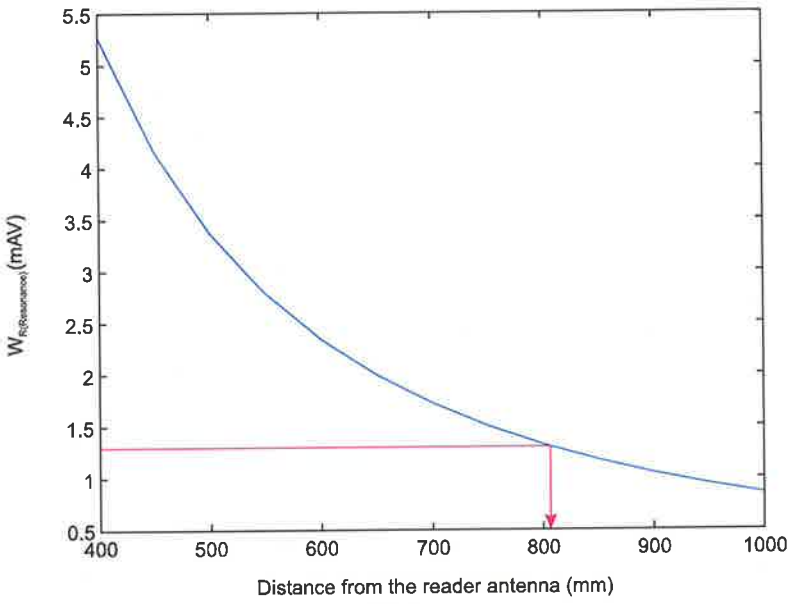

Figure 8.6 Reactive power in the tuned coil at distance r from the reader antenna.

However the scenario becomes quite different if the tag is brought closer to the reader. Considering a minimum distance of 400 mm from the reader antenna to ensure that the tag is in the far field of the reader antenna, it can be observed from Figure 8.6 that it becomes possible to generate reactive power in excess of 1.294 mVA if the antenna is, approximately, less than 800 mm from the reader, albeit over extremely narrow bandwidths.

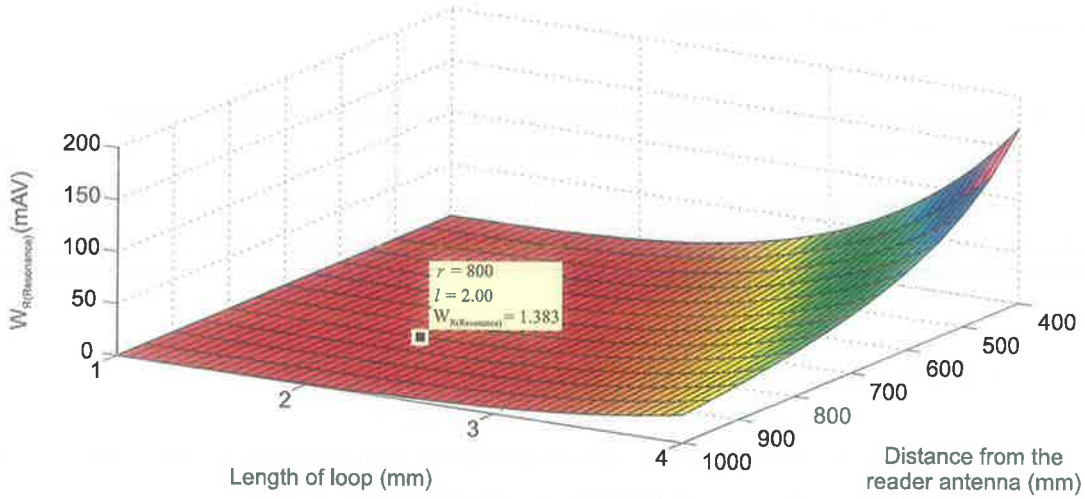

Figure 8.7 Investigation of theoretical reading range and loop size.

Figure 8.7 presents results from investigating the variation in reading distance with respect to the size of the loop with a value of $W_{R(Resonance)} > 1.294$ mVA as a metric for reading range. It can be seen that a loop of length l (refer to Figure 8.3) of at least 2 mm is required to provide the necessary reactive power at a distance of 800 mm from a reader while a loop of length $l =$ 1.54 mm is sufficient to provide the required reactive power at 400 mm from a reader. When real losses in practically available materials are taken into account, or a greater operating bandwidth is desired, the picture will become less rosy as any advantage gained by a large Q_L is rapidly diminished.

An estimate, made by persons familiar with current technology, of the power likely to be required to run a backscatter label circuit is about 5 μW (for read only operations), which is in excess of quarter of the power calculated by (8.15); that being the power available to a real impedance when the chip is a perfect match to the antenna and is in the far field of the reader antenna. Thus with a high-quality factor inductance, it is probably possible to obtain an adequate rectified voltage from an unloaded rectifier, but obtaining adequate output power to run an RFID label in the far field is doubtful.

Figure 8.8 illustrates this difficulty for the coil of length $l = 2$ mm (as identified in Figure 8.3), by evaluating the power available to a load R_l at various distances from the reader antenna. It can be seen that the tiny label antenna needs to be less than 450 mm from a reader and thus the label antenna must be located in the mid field region of the reader antenna.

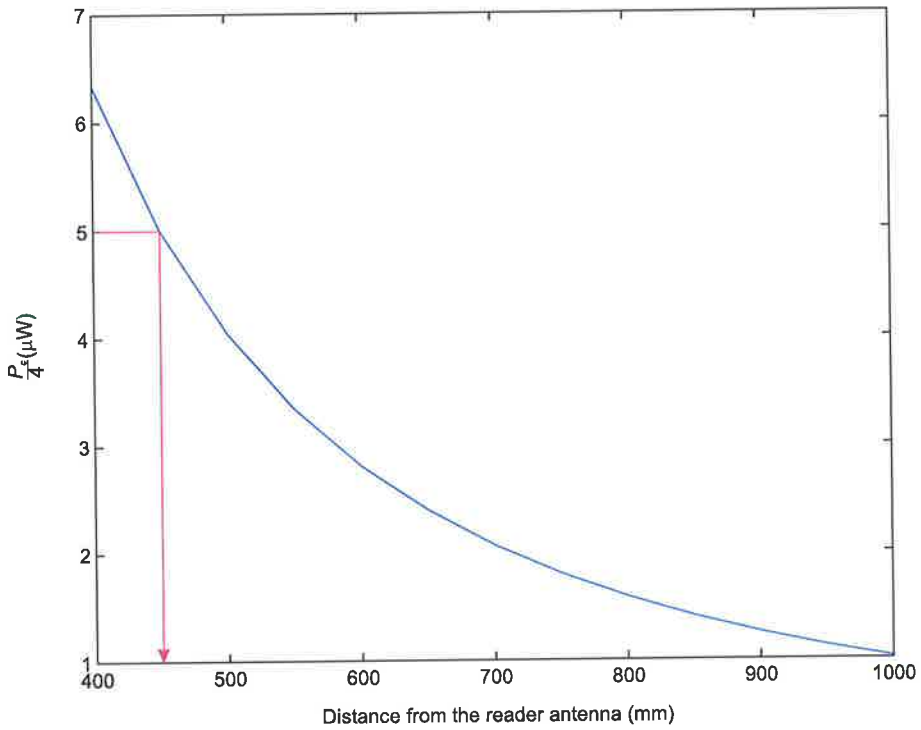

Figure 8.8 Real power delivered to an external load R_l.

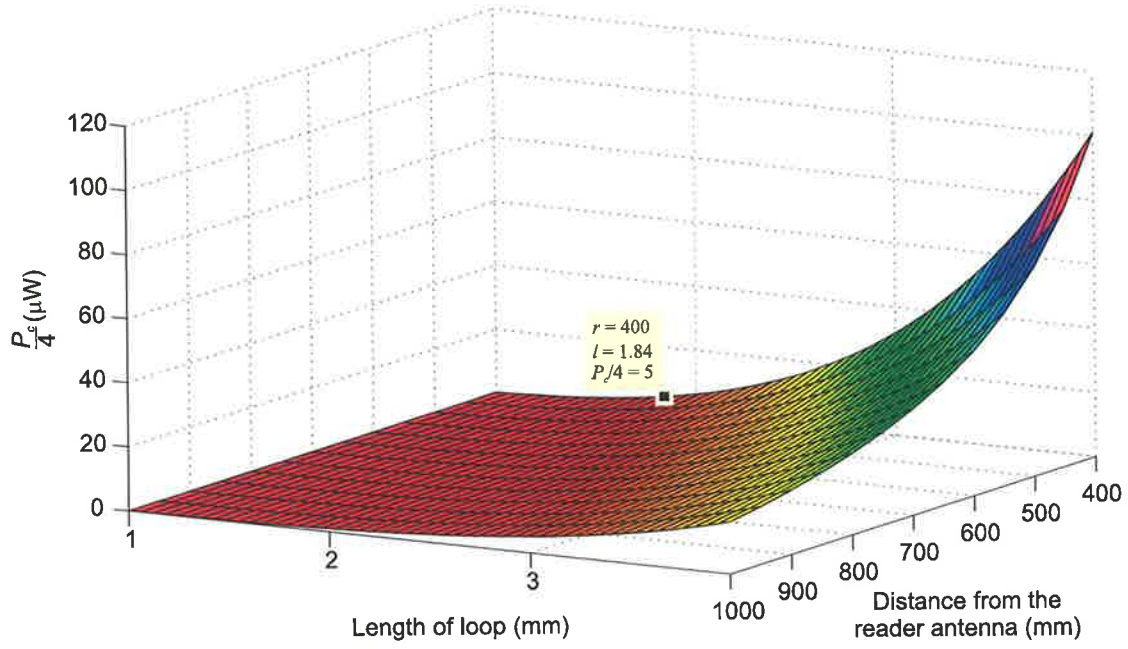

Figure 8.9 Investigation of theoretical reading range and loop size.

Figure 8.9 shows results from investigating the variation in reading distance with respect to the size of the loop with a value of $P_c/4 > 5$ μW as a metric for reading threshold. It can be seen that a minimum loop length, l of at least 1.84 mm is required to provide the necessary power to a label IC to enable the reading of a tag at a minimum distance of 400 mm.

It should be noted here that the power calculated as being able to be extracted from the antenna with a quality factor of 208 is still much less than the theoretically available source power a completely lossless receiver antenna at this distance could provide by radiation antenna theory. Applying (4.39) (formulated from radiation antenna theory) at a distance of 1 m, predicts an output power of 4.10 mW as the power that can be received by a lossless label antenna. However, such an output power is unachievable in practice due to large ohmic losses in the small label antenna.

This analysis explains that it is difficult to contemplate the use of tiny antennas for UHF RFID labels in the far field except perhaps for operation at very close range (in the near field, mid field, and a very limited distance in the far field). Probably the only beneficial effect of such tiny antennas might be that they are easier to prevent from becoming environmentally detuned because their self fields extend only a small distance from the antenna. It should also be noted that the conclusions drawn here depend upon assumptions made about the feasible rectifier capacitance and achievable quality factors in coils.

8.4 Conclusion

This chapter revisited coupling volume theory to outline its application to solve far field problems in the RFID context. It was shown that the Poynting vector-effective area formulation and the coupling volume formulations, apparently dissimilar, are both equivalent and useful in different contexts. The difference between the formulations is whether they emphasise the radiation resistance or the internal losses of the label antenna. Thus the extension of the coupling volume to far fields has provided an alternative and more useful formulation of the far field coupling link for tiny antennas.

While it is possible to conceptualise the use of tiny antennas, the analysis presented on tiny antennas has shown that the limited range of operation results not from being unable to provide the necessary reactive power to service the junction capacitance of a rectifier diode but from being unable to provide an adequate level of real power to power the label.

In addition to optimising antennas, far field coupling volume theory can be used as a tool for comparing antenna performance.

PART II: VULNERABILITIES AND SOLUTIONS

Chapter 9

SECURITY AND PRIVACY

RFID systems, and indeed other forms of wireless technologies, are now a pervasive form of computing. In the context of security and privacy, the most threatening (to privacy) and vulnerable (to insecurity) are the 'low cost RFID systems'. The problems are further aggravated by the fact that it is this form of RFID that is set to proliferate through various consumer goods supply chains throughout the world. This is occurring through the actions of multinational companies like Wal-Mart, Tesco, Metro UPS and of powerful government organizations such as the United States DOD (Department Of Defence) and FDA (Food and Drug Administration). This chapter examines the vulnerabilities of current low cost RFID systems and explores the security and privacy threats posed as a result of those vulnerabilities.

The chapter will also formulate a framework for defining the problem space constructed around low cost RFID systems, and consider the challenges faced in engineering solutions to overcome the defencelessness of low cost RFID systems.

Security issues beyond and including interrogators will not be considered as such concerns may be easily resolved using existing technology and knowledge, and because interrogators are powerful devices where complex encryption and decryption operations may be performed using either the embedded systems, DSPs, or using hardware implementation of encryption engines on a FPGA device onboard a reader.

9.1 Introduction

"RFID" is increasingly used as a common term to encompass a number of different implementations of RFID technology such as VeriChip [74] and SpeedPass payment tokens; however the focus of this chapter will be on low cost RFID systems as identified in Section 9.2, with its primary application being the tagging of cases and pallets in supply chain applications.

One of the inhibitors to wide-scale adoption of RFID technology is the cost of a label. Thus low cost RFID refers to an RFID system based on inexpensive RFID tags. It is imperative that the cost of RFID labels is reduced if RFID technology is to gain any significant market penetration. For example, the current cost of a gate of silicon logic is about one thousandth of a cent [22 and 23]. Thus, a company producing 100 billion units of a product per year would loose $1 million in profits due to the addition of a single logic gate to a label. Therefore, a great deal of attention is naturally focused on low cost RFID.

The proposed Class I and Class II labels by EPCglobal represent the low cost end of RFID labels. These RFID labels are passive transponders and have been discussed in Section 2.2.3. A characterisation of a low cost RFID system with its cost structure is provided in Section 9.2 and such a system will be analysed to highlight its vulnerabilities in the following sections with details of how to such weaknesses can be overcome.

9.2 Characteristics of a Low Cost RFID System

The most dominant form of low cost RFID technology set to spread through out the consumer goods supply chain is that advocated by EPCglobal as Class I and Class II. The low cost RFID labels involving Class I and Class II labels are based on passive RFID technology as has been discussed in Section 2.2. Due to their potential for prolific use in the future, most discussions regarding low cost RFID inevitably consider various aspects of such labels. The following sections provide an overview of low cost label manufacturing costs and IC components in an RFID label and focus on describing low cost RFID systems based around Class I and Class II labels.

9.2.1 A Low Cost Tag

Current fabrications of Class I labels consist of around 7,000 to 10,000 logic gates [22] while Class II labels may have several thousand more gates. An RFID microcircuit can be subdivided into three primary sections: RF front-end, Memory circuitry, and Finite State machine (label logic circuitry). Figure 9.1 is an illustration of a typical low cost RFID transponder (that is, a passive label). The block diagram of an HF chip and a UHF chip varies little, the primary difference being the way in which the local oscillator clock is derived. In a UHF chip there is a dedicated low power oscillator, while in an HF chip, the

clock signal is derived from the received carrier by dividing down the carrier (at 13.56 MHz) in stages.

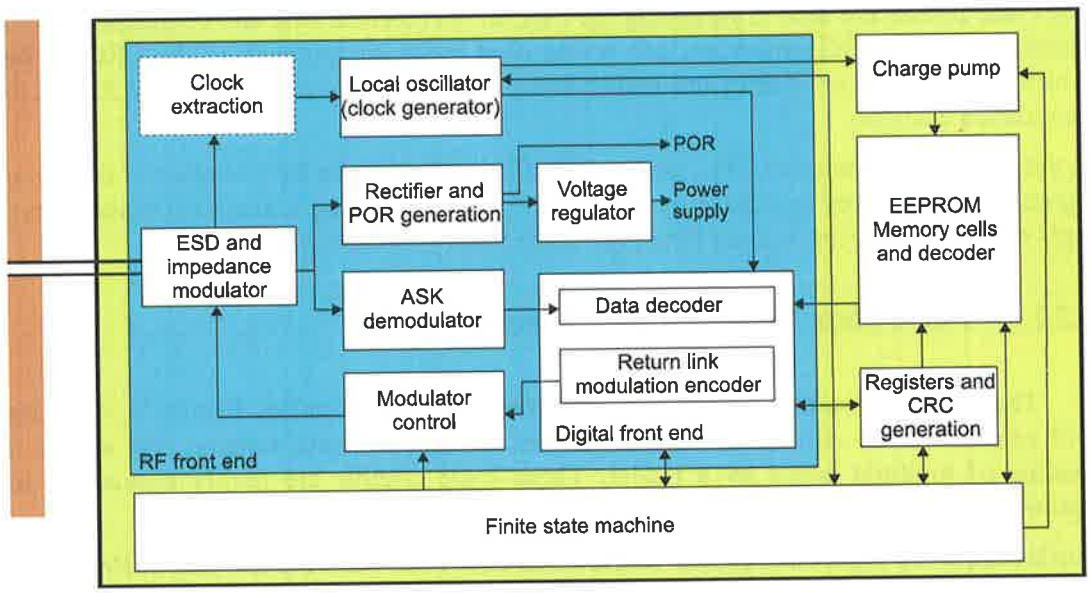

Figure 9.1 A block diagram of a passive UHF/HF RFID label.

9.2.1.1 RF Front-end

RF front-end consists of antenna pads for attaching the terminal of the antenna to the label IC. The antenna input passes through circuits for ESD (electrostatic discharge) protection. The ASK (Amplitude Shift Keying) demodulation circuits extract the modulation dips from the received signal while the Rectifier rectifies the received signal to generate power which must be regulated using a voltage regulator to avoid voltage surges due to variations in RF field intensities.

Passive RFID chips contain a relatively large capacitor following a rectifier for storing charge to power the circuit in the absence of a battery. It is important to note here that the capacitor occupies a relatively large portion of the silicon area and RFID chips consuming larger amounts of power will need higher capacity capacitors and thus will cost more.

9.2.1.2 Memory Circuitry

Low cost tags have limited memory that is either write once or a read-write memory. Class 1 labels have only read only memory while class II labels may have some read-write memory. Read-write memory, at the time of writing, is implemented using EEPROM and thus requires a large voltage before information can be written to memory. Thus a charge

pump, consisting of a series of capacitors, is required to achieve a voltage of about 17 V for writing to the tag's memory.

The CRC circuits are used in validating the CRC in the received data and commands from an interrogator. The CRC generation unit is also used in the computation of the CRC for data sent from the tag to an interrogator before being encoded for modulation by the Return link modulation encoder.

In the implementation of an EPC, tag the E^2PROM will store the EPC number of the tag and the rest of the memory (generally in the order of a few kilobytes) is available to the users. A tag's memory resources account for a significant portion of the tag cost.

9.2.1.3 Finite State Machine (Logic Circuitry)

The logic on board the chip will define the label functionality. Primarily, chip logic will execute reader commands and implement an anti-collision scheme that allows the reading of multiple labels by a reader. These logic circuits are highly specialised and optimised for their tasks.

Furthermore, the logic circuits also control read and write access to the EEPROM memory circuits.

The block diagram of a low cost RFID tag is given in Figure 9.1 along with a description of the various functionalities of the tag components. The following Sections provide quantitative characteristics of low cost RFID systems and reasonable assumptions that need to be taken into consideration when solutions for security and privacy issues are developed.

Computation capability of a low cost tag is limited to a state machine with hard wired logic functionality. The only arithmetic operation performed by current low cost RFID tags is the calculation of a CRC for checking errors in received data and the computation of a CRC prior to transmitting data. Thus, for a low cost tag any additional hardware required to implement security needs to be designed and fabricated incurring additional cost.

9.2.2 Tag Cost

Tag cost is generally based on the evaluation of the area of silicon required for a physical implementation. While this includes the analogue front end of the tag, a reduction in costs has been achieved through the miniaturization of digital functional blocks and not through devices such as capacitors, inductors or resistors. Hence, keeping tag costs low requires focusing on limiting the number of gates on a tag even though the bulk of the tag cost is associated with the analogue components whose costs are difficult to reduce due the nature of passive components.

9.2.2.1 Manufacturing Costs

There are a number of key stages involved in the manufacture of RFID labels after the design of the IC. An outline of the stages is given in Figure 9.2 below. Today, the cheapest RFID labels are passive and cost around 10 US cents in large quantities [24]. Presently RFID read only chips have design sizes ranging from $0.16mm^2$ [24] to $0.25mm^2$ [25] IC foot prints.

Further improvements to IC manufacturing processes will bring the cost of microcircuits even lower. This invariably involves producing more microcircuits per silicon wafer. However, reducing die sizes to very small levels can incur added costs due to the increase in cost of handling smaller die.

A more practical avenue for reducing costs is the use of obsolete IC manufacturing processes and filling up such fabrication pipelines with RFID IC chips. This is a worthwhile consideration as people migrate to smaller and smaller micron processes and larger, highly tuned micron processes such as 0.5 micron become available at a fraction of the cost due to depreciated fabrication equipment and the availability of smaller processes. This will reduce the cost of the IC component of the chip considerably. The older processes have the added advantage of having few or no reliability concerns while being able to provide stable yields.

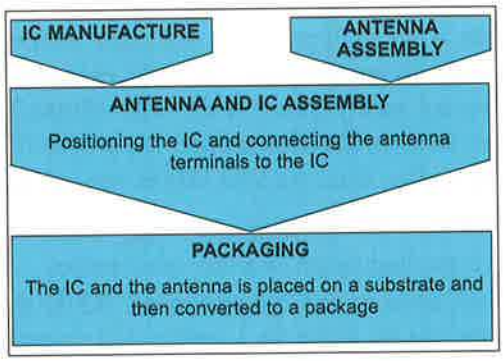

Figure 9.2 RFID Label Manufacturing processes.

9.2.3 Tag Power Consumption

A tag's power consumption will vary depending on whether the tag is just being interrogated or whether the tag is required to perform a write operation. Tag power consumption is also influenced by other factors such as the data transmission rate, the feature size of the fabrication process used, as well as the effort spent in designing low power CMOS circuitry. A tag performing a read operation will require about 5 μW – 10 μW, while a tag attempting to perform a write operation to its E^2PROM will require about 50 μW or more.

9.2.4 Physical Protection (Tamper Proofing)

Low cost tags do not utilize anti-tampering technology due to cost constraints and therefore the contents of a labels memory or the layout of logic circuits are not protected from physical access. Hence the long-term security of label contents cannot be guaranteed.

9.2.5 Standards

There are a variety of standards encompassing all aspects of RFID systems. The ISO 18000 is a multi-part standard that defines the air interface standard of a number of different frequencies from LF, HF to UHF. However for UHF, tags the most prevalent standard is that ratified by EPCglobal, called the Class I Generation II air interface protocol [16]. However the recent amendment to ISO 18000-6 to include Type C has given rise to a protocol specification almost equivalent to EPCglobal's Class I Generation II.

Accordingly, the labels within reading range have a means of revealing their presence, but not their data, when interrogated by a reader. The labels then reply with a non-identifying signal to an interrogation by using a randomly generated number as described in C1G2 air interface protocol [16].

However, for HF tags, there is no such prevalent standard, although EPCglobal is currently developing a HF specification to complement its UHF air interface protocol. Possibly the most prevalent HF tag protocol specification is the ISO 18000-3 Mode 1, most commonly used by tag deployments in various libraries around the world. The existing standards most commonly in use for HF tags, other than the ISO 18000, are listed below.

- ISO 14443 (types A and B). Devices operating under this standard are proximity RFID devices with a reading range of a few centimetres
- ISO 15693 is a recent addition for "vicinity card" RFID devices, where the operating range of the devices can be close to 1 metre (the operating mode I of ISO part 3 specification is based on ISO 15693)

9.2.6 System Operational Requirements

RFID systems are required to meet various minimum performance criteria to justify their benefits to the end user community. Two such important and related performance parameters are the number of label reads per second and data transmission speeds. Performance criteria of an RFID system demand a minimum label reading speed of 100-200 labels per second. In accordance with C1G2 protocol, a maximum tag to reader data transmission rate of 640 kbps and a reader to tag data transmission rate of 126 kbps based on equi-probable binary ones and zeros in the transmission can be calculated.

9.2.7 Communication Range

Considering the current electromagnetic compatibility (EMC) regulations, the operating range of low cost labels is limited to a few metres for those operating in the UHF spectrum and a few centimetres for those operating under the FCC regulations for the HF spectrum [45] (HF systems operating under current European regulations for the HF spectrum can have an operating range well in excess of 1 metre as discussed in Section 9.2.8).

9.2.8 Frequency of Operation and Regulations

Important considerations affecting all EM related issues, especially the powering of RFID labels, are the regulations that govern the operating frequency, power, and bandwidth in different regions of the world. There are a number of regulatory organisations and different EMC regulations around the world. Australian regulators are likely to follow the footsteps of their US counterparts; this is highlighted by the experimental license granted to GS1 (Australia) which stipulates a reader radiated power of 4W EIRP from 920 MHz to 926 MHz by the Australian Communication Authority. Hence, treatment given here for EMC regulations will not focus on Australian regulations. It is important to note here that the EMC regulations are enforced in the far field.

Most RFID systems operate in the Industrial, Scientific and Medical (ISM) bands designated by the ITU [75]. The ISM radio bands were originally reserved internationally for non-commercial use of RF electromagnetic fields for industrial, scientific and medical purposes. The most commonly used HF ISM band in Europe and America is centred at 13.56 MHz and the UHF band in the US is 902-928 MHz [45].

Figure 5.2 shows the revised European regulations at 13.56MHz and the revised FCC regulations [44 and 45]. FCC regulations for the HF spectrum allow only minimal 5 mW radiation when using an antenna of gain 1.76dBi. Hence devices operating under these regulations only have a very small reading range in the order of a few centimetres. However European regulations depicted in Figure 5.2 allow radiating 320 mW of power with an antenna gain of 1.76 dBi. Using larger interrogator antennas and large label antennas have shown that reading ranges under European regulations [44] can be increased to approach the mid field distance (that is, around the 3 metre range).

Near and far fields scale differently with distance and, in particular, the near field energy density per unit volume decreases as the inverse sixth power of distance from the antenna [37 and 40]. The result is that close to the antenna, substantial energy densities may be obtained, but these diminish very quickly as distance increases. The limits on the radiated power generally ensure that the previously mentioned inverse sixth power of the reactive power density sufficiently reduces the label energising signal to a level below an acceptable level for practical operation before the boundary of the far field. Thus, under current regulations, operation of HF systems is almost entirely confined to the near field and short distances. Conversely, at UHF frequencies, the boundary between the near field and the far

field is in the vicinity of the antenna; thus, the operation of UHF systems is almost entirely in the far field region.

Each frequency band provides its own set of advantages and disadvantages. The 13.56 MHz band has a 14 KHz bandwidth. This places a limitation on the bandwidth of the reader to label communication since the central portion of the spectrum shown in Figure 3.2 regulates the operation of RFID equipment in the HF region.

The 902-928 MHz band, under US regulations, allows multiple reader to label communication choices with much higher communication bandwidths and hence data rates. The regulations allowing the longest communication range require the reader to change its communication frequency every 400 milliseconds. The reader may hop between any number of channels, however the maximum bandwidth of a channel cannot exceed 500 kHz [45]. This technique is referred to as 'frequency hopping'. Table 9.1 below highlights the range of frequencies in use in the UHF region around the world.

Table 9.1 UHF RFID frequency allocations.

Region	Frequency range (MHz)	Bandwidth (MHz)
Europe	865 - 868	3
USA	902 - 928	26
Japan	952 - 954	2
Australia	918 – 926 920 – 926 (experimental band till 12th July 2007 with 4W EIRP EMC regulation limit)	8 6

9.2.9 Security Provided by Class I and Class II labels

The most dominant form of low cost RFID technology set to spread throughout the consumer goods supply chain is that advocated by EPCglobal as Class I and Class II. The low cost RFID labels involving Class I and Class II labels are based on passive RFID technology as has been discussed in Section 2.2. Due to their potential for prolific use in the future, most discussions regarding low cost RFID inevitably consider various aspects of these labels.

Led by EPCglobal, the RFID community in its efforts towards standardisation has produced a list of end user requirements for Class I and II labels that has flowed into the current C1G2 protocol standards and is outlined in the following sections. One aim of that list has been to address the privacy and security risks posed by RFID Class I and Class II labels containing an EPC. The security requirements are an appropriate guideline when considering the level of security and privacy that can be expected and required from Class I and Class II RFID labels. The following is an outline of the security features that can be expected from the previously mentioned classes of labels.

9.2.9.1 Security Features of Class I Generation 2 Labels

Class I labels, the characteristics of which have been identified in Section 2.2.3, have only a read only or a write-once memory and are incapable of participating in a complex security mechanism. Hence Class I labels are required to provide "Kill" capability and a password to control access to the kill command, so that consumers have the choice of completely disabling an RFID label at the time an RFID labeled item is purchased.

"Killing" a label involves the destruction of the label thus rendering it inoperable [6] by perhaps setting off a fuse or disconnecting the antenna. Unfortunately, the destruction of the label denies the user the significant benefits that could have been obtained from a "smart object". As a solution, an alternative idea to killing entertained previously involved the removal of the unique serial number of the EPC code in articles that allows the label owners to be tracked, albeit with difficulty in practice. This does not remove all the privacy concerns as tracking is still possible by associating a "constellation" of a label group with an individual. This implies that a particular taste in clothes and shoes may allow an individual's location privacy or anonymity to be violated. However "killing" a label will eliminate privacy concerns and prevent access by unauthorised readers when combined with a password to control access to the kill command.

While throttling is not specified as part of the C1G2 standard, it is reasonable to assume the employment of a delay based throttling mechanism on tags to prevent the guessing of kill or access passwords [76]. The concept behind delay based throttling is that, on the occasions a tag is given an invalid password, the tag enters a sleep state where it will not accept another kill attempt for a specified amount of time. This method can significantly increase the time required by an attacker attempting to kill a tag. In a situation such as a retail environment, the delay factor can be an adequate deterrent to such brute force attacks.

Class I labels should also have the ability to lock EPC data so as to provide one-time, permanent lock of EPC data on the label, so that EPC data cannot be changed by an unauthorised interrogator once it has been written. Interrogators are also prevented from transmitting complete EPC data except when data needs to be written to an RFID label, so that the EPC information may not be eavesdropped upon from a distance without being discovered.

9.2.9.2 Security Features Expected from Class II Labels

Other requirements were identified as necessary for higher class labels since these labels will have greater functionality and thus more hardware. Higher class labels are required to provide a secure forward link for communication with an RFID label while providing access control to label functionalities.

9.2.9.3 Backend System Services: Track and Trace Capability

RFID labels are given a unique identification number: for Class I and Class II labels, the unique identifier is an EPC. Using information technology services offered by backend systems, such as the EPC Network services, it is possible to dynamically generate a profile of the RFID label to create an electronic history of the label as it passes through various stages of the supply chain. The scheme was discussed in detail in Section 3.9. The electronic history, called an electronic pedigree, can serve to thwart cloning attacks.

9.3 Vulnerabilities of Low Cost RFID Systems

Low cost RFID technology described in Section 9.2 has the potential to promote a sound business case because of its potential to save costs and improve processes, while providing certain security benefits. However, as described in Section 9.2, low cost RFID systems generate significant security risks, mainly due to their cost constrained implementations and the insecure communication channels over which tags and readers communicate. The security risks that arise as a result are due to a number of reasons outlined below.

- Communication between a tag and a reader takes place over an insecure channel
- Tags are accessible by any reader implementing the air interface protocol
- Tags are not tamper proof and allow a channel for physical access to tag contents and circuitry (as a result, tags cannot be expected to secure information for long periods).
- IC designs are constrained by cost and are thus minimalist implementations
- Air Interface protocols are designed to reduce tag complexity
- Design flaws in reader implementations due to cost constraints

The reasons above form the basis from which various weaknesses have arisen in low cost RFID systems. The resulting vulnerabilities are examined in detail in the following sections.

9.3.1 Eavesdropping and Scanning

Transmissions from a reader and a tag take place over a clear communication channel which may be observed by a third party. Low cost labels with minimum functionality are only able to identify themselves by transmitting a unique identifier, and these labels can be read by any reader adhering to the air interface protocol used by an RFID tag. Hence a third party may monitor a conversation between a label and a reader to obtain sensitive information. Illicitly obtained information in this manner may be used to create fraudulent labels, unauthorised readers, or used to discover secret information stored on labels (such as a tag password).

Similarly, competitors of an organization (such as a rival supermarket) may, over time, scan another organizations inventory labeled with RFID labels or eavesdrop on the organization's own valid operations to obtain valuable information, such as sales data, to ascertain the performance of its competitors (an act commonly referred to as corporate espionage) [83]. Publications such as [85] have attempted to define various eavesdropping ranges based on the reading ranges of tags. Similar descriptions of the eavesdropping distances possible are stated below so that vulnerabilities of eavesdropping can be better understood. However, it should be stated here that, while it is useful to define terms to explain ideas, the fact that a third party can eavesdrop on a conversation between a tag and reader from a distance still remains a fundamental vulnerability.

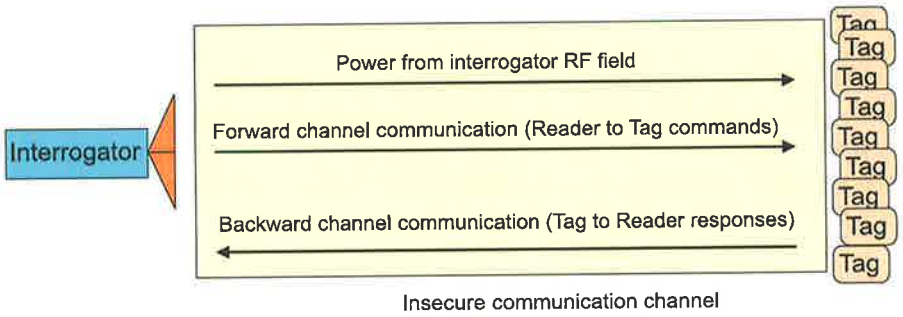

Figure 9.3 A passive RFID communication channel model.

Figure 9.3 illustrates a simple model for a passive RFID communication channel. It is possible to consider the distances at which a third party can listen to a conversation between a tag and an interrogator to formulate the following general classification of eavesdropping distances. Figure 9.4 gives an illustration of the latter distinctions discussed and explained below.

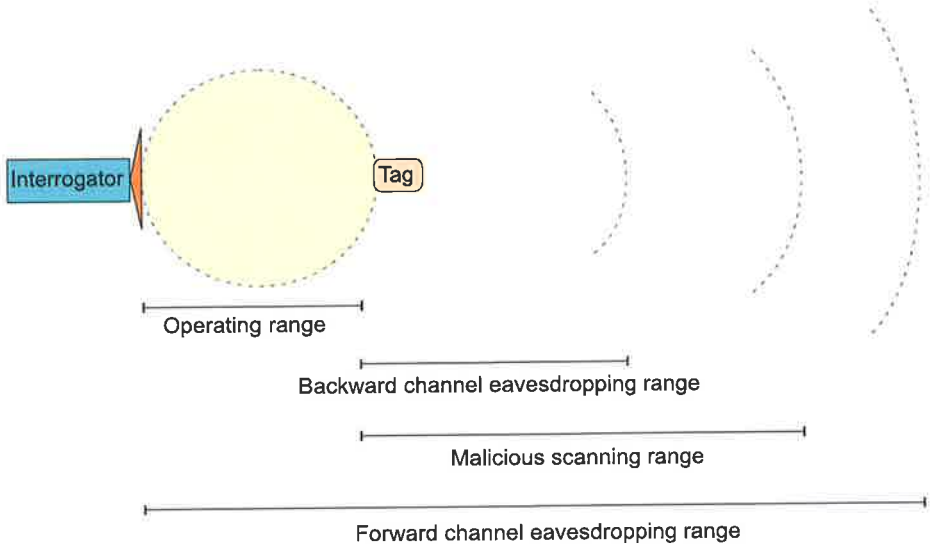

Figure 9.4 Eavesdropping range classification.

Operating range: Tag operating range is either defined by product specification based on user requirements or it may be based on a certain standard. The operation range of a tag will also be application dependent as tag reading distances are affected by various environmental factors. The operating range of tags is then the maximum distance at which any given tag will read to a given reliability, when illuminated by a reader operating under the electromagnetic compatibility regulations of that region for a particular frequency band of operation.

Backward channel eavesdropping range: The backward channel refers to the communications sent from a tag to a reader. In a low cost system where the tags are passive, this reply is weaker in signal strength than a reader transmission as it is achieved by reflecting some of the incident RF energy at the label antenna.

The backward channel range is generally much greater than the operating range of the tag since a third party is capable of using a narrow beam antenna with an RF receiver of higher sensitivity and because the third party does not have to use the same antenna for powering and receiving (unlike most low cost systems which use a monostatic antenna configuration; that is, a single antenna for powering, transmitting and receiving).

Forward channel eavesdropping range: The forward channel refers to the transmissions from a reader to a tag, and the forward channel eavesdropping range is the maximum distance at which a third party with a high gain antenna and a highly sensitive RF receiver can correctly record a tag transmission.

Malicious scanning range: This read range is derived by considering an adversary with no regard for electromagnetic compliance or standards and whose only intention is to both power and read the tag at any cost or to eavesdrop on a conversation between a tag and a reader at any cost. Combined with the prospect of a highly sensitive RF receiver, a narrow beam antenna and the willingness to break electromagnetic regulations, malicious scanning range will have reading distances in excess of that possible for backward channel eavesdropping.

There are generally two forms of eavesdropping possible with low cost RFID systems; passive eavesdropping and scanning (active eavesdropping). The following sections will discuss the previously mentioned forms in an RFID context.

9.3.1.1 Passive Eavesdropping

As the names suggests, passive eavesdropping relates to the observation and, or, recording of communication between a reader and a tag by an unintended recipient. Passive eavesdropping may be performed by a third party in the operating range, the backward channel eavesdropping range or the forward channel eavesdropping range.

9.3.1.2 Scanning (Active eavesdropping)

In this situation, a third party or an adversary is actively attempting to read the contents of a tag without the authority of the tag owner. In a scanning scenario with respect to a low cost RFID system, an adversary is using a rogue reader to power the tag and communicate with the tag without raising the suspicions of the tag owner. An active eavesdropper will have a working range within the malicious scanning range outlined in Figure 9.4.

9.3.2 Cloning

Devices designed to impersonate tags or readers (imitating the behaviour of a genuine label or a reader) present a serious threat to an RFID system. Impersonation will add a new dimension to thieving as attackers are able to write EPC data onto devices that function like RFID tags. A direct consequence of cloning is the possibility for counterfeiting, where a genuine article tagged with an RFID label may be reproduced as a cheap counterfeit and tagged with a clone of the authentic RFID label. The 'track and trace' concept outlined in Section 3.9 is one possible solution to detecting cloning in a supply chain application.

At the time of writing, there is no mechanism for a reader to verify that it is communicating with a genuine RFID label and not a fraudulent label. Thus a thief may replace a tag of a valid item with a fake tag or replace the tag of an expensive item with that of a fake tag with data obtained from a cheaper item. Hence the lack of a means for authentication allows an adversary to fool a security system into perceiving that the item is still present or fool automated checkout counters into charging for a cheaper item. Such fake labels may also be used to create imitation items. There is presently no mechanism for a reader to authenticate itself to a label or a label to authenticate itself to a reader. Thus labels and readers are constantly in a vulnerable environment where the integrity of messages is doubtful and there are no means for establishing the legitimacy of a reader by a label or the legitimacy of a label by a reader.

Clearly, more expensive RFID system implementations are also not immune from cloning as shown by a recent cloning attack published in [84] where a cloned tag was used in the purchase of fuel at a service station and to start an automobile locked with an RFID based car immobiliser. A similar example of cloning of proximity cards is given in [85] while the possibility of cloning the VeriChip [74] in a discussion of its possible use to tag employees was outlined in [86].

The EPC Class I tags have no mechanism for preventing cloning as the tags are simple bit storage devices that transmit a string of bits on request from any valid reader. All that is required by an adversary is to scan a tag and copy its EPC number onto another tag or another device that is capable of impersonating a tag. The EPC Network architecture aims to remedy the problem by creating an electronic history of the product's life through the supply chain by way of an electronic pedigree. While the access to an electronic pedigree of a product by way of a secure database only solves the problem of confirming the existence of

an illegal clone, it may not be possible to distinguish the original tag from its illegal duplicate(s).

9.3.3 Man-in-the-Middle

An RFID system is constantly under threat from man-in-the-middle attacks resulting from eavesdropping on reader and tag transmissions. A third party may monitor a conversation between a label and a reader, record and use or alter parts of the conversation and retransmit messages to illicitly obtain information from RFID devices or to command RFID devices to the detriment of the system.

Retransmission of such recorded information may be used to query RFID labels or fool RFID readers. Such an attack has the ability to fool personal access control systems and contactless payment systems based on RFID technology [87].

For instance the EPC C1G2 protocol uses a "kill" command [16], protected by a password, to disable the label so that it cannot be read. It is possible for a third party to record the conversation between a reader and a tag that performs a kill operation and use that information to kill other tags provided that they are protected with an identical kill password.

9.3.4 Denial of Service

An adversary may initiate a denial of service (DoS) attack to bypass or avoid security systems. A DoS attack is easily carried out by placing a large number of fake labels for identification by a reader. Persons also have the ability to disrupt an RFID system implementation by destroying or corrupting a large batch of labels. Labels are also vulnerable to protocol attacks. DoS attacks may also be performed by exploiting weaknesses in the air interface protocol or weaknesses in the implementation of a tag's finite state machine. A simple scenario of such an exploit may involve labels being repeatedly asked to perform an operation, thus making them unavailable to an authorised reader.

In addition, tags may be prevented from being read by using the simple concept of a Faraday cage or by jamming the RFID interrogator signals, for instance by intentionally creating noise in the frequency band in use. For critical applications, a DoS attack may have devastating effects.

9.3.4.1 Code Injection

RFID interrogation signals can be disrupted or blocked, or RFID readers can be attacked using RFID tags designed to manipulate weaknesses in the air interface protocol or the implementation of the reader to create a denial of service attack during an RFID interrogation process by creating situations of system unavailability. The possibility of

RFID viruses has been highlighted in [86] where a more sophisticated tag may exploit interrogator or protocol vulnerabilities to affect a number of systems by using a reader to cause a system failure by way of a code insertion attack caused by creating a buffer overflow in a reader's memory stack using carefully constructed SQL instructions disguised on the tag as the data that gets transmitted to an RFID reader. This vulnerability is only present if the middleware is intentionally made vulnerable by accepting any data transmitted by the tag without checking for validity of the format of the data sent. Also, modern SQL servers are guarded against malformed SQL instructions.

9.3.5 Communication Layer Weaknesses

The recently ratified EPCglobal C1G2 air interface protocol [16] has a number of security features based on the use of tag specific passwords. Probably the most important feature that is protected is the *KILL* command by using a kill password. There is also a means for access control on the tag using an access password.

A recent publication in [87] has shown how the kill password of a tag can be deduced by the careful analysis of the tag power consumption to a series of well constructed test passwords. This highlights a particular vulnerability of low cost tags to power analysis attacks and the vulnerabilities of storing long term secret information on a tag. However, it is possible to prevent such an attack as power analysis attacks have been well studied in the context of smart card devices. The RFID ICs in the future will need to be designed to avoid such an attack but this will take place at added cost to an RFID tag.

While power analysis attacks may be prevented in the future, the fact that each RFID tag has at least two unique passwords will create both potential security and logistical nightmares if the problem of careful key management is not considered. This problem will be aggravated in the future as item-level tagging begins to proliferate through the global supply chains and PoS (point of sale) devices may need real time access to passwords as consumers purchasing goods may want their tags deactivated at the point of sale. Hence the problem of careful key management needs to be considered in the context of low cost RFID systems where the potential for key discovery is highlighted by the global aspects of supply chains. It is not difficult to imagine a scenario in the future where a list of kill passwords anonymously appears on a public web site.

The recently ratified C1G2 protocol also relies on the tag generating a random number to be used as an input to an exclusive-or operation. The risks associated with inefficient or inadequate random number generation in RFID tags (that is, a high correlation between the random numbers, in a pseudo-random number sequence) is emphasised in [88]. The consequences are two-fold for tags using the C1G2 protocol. Primarily, the lack of randomness may cause particular tags to respond with an identical time slot during the execution of the slot selection process. Thus an attacker may be able to track a tag depending on the time slot it selects in a seemingly random manner. Since the security of the information sent to a tag relies on the randomness of the number that the tag generates, a lack of randomness may allow an adversary to easily decrypt information transmitted, once the

attacker successfully decrypts an encrypted message, or discover the seed used in the pseudorandom generator.

Though with great technical difficulty, [88] also points out the possibility of identifying a tag using a "radio fingerprint". For instance, manufacturing variations that may cause physical glitches in the signals may be used to distinctly identify tags. In such a situation, even strong cryptography protocols and primitives would be ineffective.

9.3.6 Physical Attacks

In addition, the labels themselves are exposed to physical attacks due to the absence of tamper proofing as dictated by the cost limitations of low cost tags. Physical attacks are possible irrespective of whether measures are in place to protect labels. However, the ability to gain useful information from a protected label is a much more difficult problem. A physical attack on an RFID label or a reader may yield an adversary secret information, such as passwords (in C1G2 labels), providing security to an RFID system. The importance of physical attacks is more prominent in cases where RFID tags are used as a means of authentication. The problem is compounded when a physical attack leads to the construction of a clone.

An insight into physical attacks can be gleaned from an increasing body of work in the area of smart cards. A complete overview of possible physical attacks and countermeasures is outlined in [89] while specific lower cost physical attacks are presented in [90].

The majority of physical attacks possible on devices in general can be bundled into two broad categories based on the means used for accessing the device. These attacks are relevant to RFID devices, especially since they have no tamper protection to safeguard label contents.

9.3.6.1 Non-Invasive Attacks

These attacks are as a result of timing analysis, power analysis, analysis of certain glitches (radio fingerprinting), and exploitation of data remnance. Non-invasive attacks are low cost and require little expertise to execute. While non-invasive attacks are generally thwarted by increasing chip complexity in most devices, it is not the case with RFID chips with minimalist implementations that may have design flaws as a result of human errors or insufficient error checking. Non invasive attacks are particularly dangerous as there is no physical evidence and the owner of the tag may not be aware that such an attack has taken place.

9.3.6.2 Invasive Attacks

In addition, an adversary may simply reverse engineer labels to create fraudulent labels for cloning or DoS attacks or use probing techniques to obtain information stored in memory (microprobing and Focus Ion Beam editing) or alter information stored in memory (using a laser cutter microscope [90]). A recent exploitation by reverse engineering of a more costly implementation of an RFID device with added security to carry out a fraudulent payment was published in [84]. Use of microprobe needles to read out the memory contents of a smart card is published in [91].

Attacks such as optical probing and fault injection attacks where the chip is removed from its packaging with the passivation layer still unbroken, are also invasive attacks but these attacks are may be further qualified as semi-invasive attacks.

9.3.7 Privacy Violations

The mass utilization of RFID labeled items creates an imminent and potentially widespread threat to consumer privacy. The privacy issues raised by RFID labels have been receiving a wider audience as a result of the popular press. The mass movement by civil libertarians has seen RFID trials cancelled [92] (despite misunderstandings of the company's intentions [93]) and negative press coverage for other manufacturers causing delays in RFID test trials [94]. Press coverage on privacy issues has also managed to tarnish the image of RFID with a satanic persona and nicknames such as "spy-chips" [95 and 96].

It is possible to imagine various scenarios of privacy violations and most of those are already existing concerns from technologies such as credit cards, browser cookies, mobile phones and Bluetooth devices. However, RFID, due to artifacts resulting from its cost constraints, presence of a unique identifier readable by anyone, and the encoding of product information on the unique numbering scheme such as the EPC, creates two possible scenarios; profiling and, tracking and surveillance, where the privacy of people as well as corporations may be infringed. These scenarios are discussed in the following sections.

9.3.7.1 Profiling

There are clear possibilities for unauthorised interrogators to read label contents from unprotected RFID labels due the lack of a mechanism for authentication and the fact that low cost RFID labels as well as interrogators broadcast unique item identifiers such as the EPC. Even if labels are protected, a traffic analysis attack (or predictable label responses) may be used. Hence an individual with a number of labeled items may be scanned by a third party to identify individual possessions or "taste", and specific EPC numbers on products may then be associated with an individual.

The data obtained can be misused to violate an individual's wishes to remain anonymous. For instance, persons carrying religious material or material related to a certain political affiliation, may no longer be able to privately pursue their beliefs or interests and in addition to their reading material potentially becoming public knowledge, their beliefs and opinions may be used in acts of persecution, jealousy or hatred. At the same time, data collected and associated to individuals can be valuable to market researchers or even thieves in search of wealthy victims. The personal information collected regarding individual preferences will act as a powerful tool for marketing products as more targeted marketing to individual tastes and affordability becomes possible by scanning the RFID tagged possessions of an individual.

It is possible to imagine a variety of plausible ways of using such information. For instance, if Bob purchases a brand named jacket using a credit card, the shop can immediately associate "Bob" with the tag id of the apparel. When Bob enters the store again, the shop has the ability to automatically establish his identity along with a history of his spending habits and tastes. While this information may prove positive for Bob, Alice, who might enter the same store, may be wearing cheap shoes and the shop assistants then have the ability to provide preferential treatment to Bob while perhaps neglecting Alice. Similarly, a thief hiding in the corner of the store may read the tag id of Bob's jacket and conclude from the tag id (by way of careful observation and without having access to any backend databases) that the tag id is indicative of an expensive apparel, then Bob might become the unfortunate victim of a theft.

9.3.7.2 Tracking and Surveillance

A further privacy concern resulting from the association of unique identifiers to individuals and the unobtrusive scanning of RFID labeled items carried by an individual is posed by the possibility of tracking, albeit with technical difficulty. Correlating data from readers obtained from multiple locations can reveal the movement, social interactions or financial transactions of an individual once an association is made between a unique tag identifier and a person. In response to such concerns, there have been suggestions to remove the unique identifier in an EPC to prevent a specific EPC from being associated with an individual. Even if such a scheme is implemented, individuals may be tracked through a "constellation" of predictable label responses. Hence, a person's unique taste in items may betray their location, movements, or identity.

9.4 Addressing Vulnerabilities

Issues resulting from vulnerabilities discussed in Section 9.3 can be divided into two broad categories of security related issues (exemplified by eavesdropping, cloning, man-in-the-middle, DoS, communication layer weaknesses and physical attacks) and privacy related issues (profiling and, tracking and surveillance). Overcoming these seemingly divergent issues can be achieved by the provision of services to enforce measures to address both the privacy and security related issues. These services can be implemented on

low cost RFID systems by identifying existing mechanisms, inventing new mechanisms or by re-engineering existing mechanisms to meet the required security and privacy objectives.

However, there is a notion among the advocates of RFID technology that the general nature that is partly hindering the mass scale deployment of RFID technology - that is, the unreliability of low cost systems mainly due to the reasons given in Table 9.2 - makes the exploitation of vulnerabilities such as profiling and tracking discussed in the previous section impractical.

Clearly, low cost RFID tags are unreliable. For instance an RFID tag placed on your jacket may work while it is on the shop shelf, but it may stop working once the jacket is worn as your body will affect the properties of the RFID tag antenna.

Table 9.2 Sources of unreliability.

	Description
1	Effects of metal and liquids on the propagation of electromagnetic waves
2	Effects of permeability of materials on tag antennas
3	Interference and noise from other users of the RF band
4	Tag orientation with respect to the reader propagation field
5	Distance of the tag from a reader
6	Electromagnetic compatibility regulations
7	Cost and power constrained implementation of RFID chips

It is due to the reasons given in Table 9.2 that some of the vulnerabilities discussed in Section 9.3, in practice, are far from being feasible. Ironically though, the unreliability of RFID tags has prevented much of the security and privacy violations from being realised, with the possible exception of laboratory experiments or in that realm of possibility. While this is the present reality of low cost RFID technology, it is expected that the cost benefits of RFID technology will eventually propel the research community to solve the technical issues outlined above. Hence the idea of using unreliability to brush aside the possible threats is not a long term solution.

Generally, it is clear that the technology of tomorrow is what is being developed currently. Even though deployments of current RFID technology do not adequately satisfy expectations, despite various mandates for RFID compliance, it is gradually beginning to proliferate [97, 98 and 99]. Hence, it is important to address vulnerabilities discussed in Section 9.3, despite some being implausible, so that the systems deployed today do not become problems of tomorrow.

The following sections consider the measures required for addressing security related issues and privacy related issues identified in Section 9.3.

9.5 Addressing Security Issues

Eliminating security related concerns regarding RFID systems illuminated by way of the examples in Section 9.3 require the enforcement of suitable security measures. Before deciding on a set of security measures, the security objectives that need to be satisfied must be identified. Table 9.3 lists a necessary set of security objectives that will be required to address the potential security threats.

RFID systems must employ mechanisms to achieve one or more of the above security objectives to alleviate various concerns cited in Section 9.3. As security cannot be solely accomplished by security mechanisms, it should be mentioned that proper legislation, procedural techniques and enforcement of laws is also required. The following sections describe the security objectives outlined in Table 9.3 and demonstrate that meeting these security objectives eliminates the security threats posed by inherent weaknesses in low cost RFID systems.

Table 9.3 List of security objectives.

	Security Objectives
1	Confidentiality
2	Message content security
3	Authentication
4	Access control
5	Availability [100]
6	Integrity [100]

9.5.1 Confidentiality

The term 'confidentiality' can be used to describe a mechanism to keep information from all but those that are authorised to see it [78].

In an RFID system, the communicated information between a reader and a tag needs to be confidential when sensitive data such as secret keys or other such information, which must not be collected by an eavesdropper, is communicated. The confidentiality of any secret information stored on a tag is also at risk and needs to be secured.

Confidentiality may be achieved by having the communication link between tags and readers encrypted, thus establishing a secure communication link. Confidentiality of tag contents may be achieved by tamper proofing the tag to prevent physical access to tag contents. Currently however, there is no secure means of establishing a secure communication link between a tag, and tamper proofing a tag has cost implications that will hinder the economics of low cost RFID technology.

9.5.2 Message Content Security

Providing message content security or data integrity involves making certain that the data contained in a communication is not altered by unauthorised or unknown means [78]. Alteration in an RFID context may involve the capture, substitution, deletion or insertion of information and the retransmission of that altered information to a reader or to a tag. Ensuring message content security will prevent man-in-the middle attacks involving the retransmission of altered messages. Present low cost RFID systems have no means of providing message content security.

9.5.3 Authentication

The simple objective of meeting authentication can be expressed as authenticating the devices involved (the tags and the reader) or, in a supply chain application where the tags are used to label products, as product authentication. In some applications where perhaps the tag is an integral part of the tagged object, authentication of the tag may be adequate to guarantee the authenticity of the object to which it is associated. In other applications where tags are placed as an external label to a high value item, authentication soley of the tag may not be adequate. The objectives of tag and interrogator authentication and, product authentication are discussed below.

9.5.3.1 Tag and Interrogator Authentication

In an RFID context, authentication simplifies to the corroboration of the identity of a tag or a reader. Authentication is an important RFID security measure for preventing counterfeit manufacture or substitution. It is also important for controlling access to label contents. Use of authentication may also be required in other applications of RFID technology such as baggage reconciliation or secure entry systems. Authentication of a tag is useful in addressing vulnerabilities posed as a result of cloning.

9.5.3.2 Product Authentication

While authentication described above has the objective of establishing that a tag is legitimate and a reader is authorised, in certain application use case scenarios, authentication of the tag is not sufficient to guarantee the authenticity of the product to which the tag is attached as brand or goods substitution may have taken place. Hence in the case of using a low cost RFID tag to label a product, product authentication refers to the establishment of the authenticity of a product by the corroboration of the identity of a tag and/or the legitimacy of the product by creating an irrefutable link between the product and the tag that can be verified by a third party.

9.5.4 Access Control

In the context of interaction between RFID interrogators and tags, access control implies a mechanism by which a tag or an interrogator grants access or revokes the right to access some data or perform some operation. Generally, tags will require access control mechanisms to prevent unauthorised access to tag contents.

9.5.5 Availability

Ensuring availability in RFID systems is an important issue since readers need to be ready to detect tags that may enter their reading range at ad-hoc intervals of time (depending on the application). In an RFID context, availability applies to ensuring that the services offered by a reader to an RFID tag or the services offered by a tag to an RFID reader are available when expected [100]. RFID systems meeting the availability criteria will ensure that there are services in place to thwart or prevent a DoS attack.

9.5.6 Integrity

Integrity of an RFID system applies to the integrity of the devices, such as the reader and the tags where it implies that a reader or a tag has not been malevolently changed. A reader receiving data from a tag needs to be able to trust that the information received is correct, while a tag needs to be able to trust that the information it receives from a seemingly authentic reader is trustworthy [100]. Ensuring the integrity of a system is an important consideration in addressing physical attacks.

9.6 Addressing Violations of Privacy

Table 9.4 An elaboration of privacy.

Privacy Interests	Description
Privacy of personal behaviour	As the name suggests, privacy of behaviour encompasses all aspects of a person's manner. In reality, this is narrowed down to areas that are sensitive to individual people such as political activities, sexual orientation or religious conduct.
Privacy of personal data	Personal data privacy refers to the more commonly used term, data privacy. In essence, data associated with a person should not be accessible by a third party without the consent of that individual. This applies to cases where the data is collected, or processed by a third party.

While it is difficult to define privacy, and a number of different interpretations can be found, it can be most simply stated as the *interests* that a person or persons have in "sustaining a 'personal space' free from interference by other people and organisations" [101]. The ideas captured by *interests* that a person has in an RFID context can be further elaborated as given below in Table 9.4 [101].

It is not possible to describe the number of privacy violations RFID technology can potentially cause, since they are numerous as described in [102]. However, it is sufficient to realise that the root cause of such violations stems from the potential to automatically associate human identification information with object identification information and thus addressing privacy requires certain goals to ensure that the latter association is not possible. Privacy goals outlined in Table 9.5 are an adequate set of goals for addressing the issue of associating object identification data with human identification data and the related concerns outlined in Section 9.3.

Table 9.5 List of privacy objectives.

	Privacy Objective
1	Anonymity
2	Untraceability (Location privacy)

It is important to note that privacy is a multi dimensional issue involving many areas. The successful implementation of the privacy objectives outlined in Table 9.5 will not only require security mechanisms but will also require the formulation of public policies, legislation and the enforcement of the law by the relevant law enforcement agencies. The latter statement is especially important in order to ensure privacy of personal data [103 and 104].

Public policy is a vital aspect because the security mechanisms used to ensure privacy are most effective when implemented in conjunction with a well-formed policy. There are existing privacy policies that can be applied directly in the context of RFID [105]. However, these may need to be clarified, refined or amended to cover aspects specific to RFID systems. Significant issues that must be dealt with by policy formulation or amendment in relations to RFID are those generated by the following items.

- Unique Identification of all label items
- Collection of information (who collects data generated from RFID systems, how do you exploit that data, ownership of information obtained from the data)
- Dissemination of that information
- Mass utilization of RFID technology

It is important to note that existing barcode systems have many of the same risks; they can be read by a simple bar code reader, can be destroyed easily and can be cloned. However, there is not the potential for these operations to be performed wirelessly and unobtrusively on an immense scale.

While public policy and legislation is an ongoing topic of discussion, it is beyond the scope of this thesis to address policy tools and legal tools for addressing security and privacy issues. Nevertheless, technical solutions for addressing previously mentioned issues are considered in Chapter 11. The following sections discuss in detail the privacy objectives introduced in Table 9.5.

9.6.1 Anonymity

While anonymity can be described in a number of ways, the most appropriate is probably the concealment of the identity of a particular person involved in some process, such as the purchasing of an item, visit to a doctor or a cash transaction [101].

Mitigating the problem of anonymity in an RFID context will involve the prevention of associating an EPC of an item with a particular individual as the EPC can be used to obtain information regarding a particular process, or an object, and that information may be associated with a particular person's identity.

For instance, a person walks into a book store, purchases a book of their choosing and pays for the purchase using a credit card. Immediately, this transaction allows a relationship to be created between the identity of the individual and the EPC of the book. The person may then walk on the street, now it may not possible to conceal their identity with regards to the purchase from a third party scanning the book's RFID tag, provided that the third party has access to the relationship between the object identification information and the human identification information. The same person may carry an expensive medication which could be scanned by thieves or by potential employers to his or her detriment.

9.6.2 Untraceability (Location Privacy)

Untraceability in an RFID context is aimed at addressing location privacy issues. Location privacy is an issue that has surfaced more recently with the availability of reliable and timely information about the location of people as a result of pervasive computing. It is also an issue associated with mobile users and other users of wireless devices. While this is not an issue specific to RFID [106], it does apply to modern RFID systems that are being developed because of their pervasive nature and their ability to leverage the Internet to form a global network that can receive and transfer data in real time.

There are number of ways of defining untraceability and, in an RFID systems environment, it can be stated as a means by which the ability of other parties to learn or track the location of people or transactions from a current or present location, based on information obtained from one or many RFID tags in the possession of that person(s) or party to that transaction, is prevented.

Hence, providing untraceability in an RFID system requires the provision of a mechanism to prevent other parties from obtaining RFID tag data without the tag owner's consent and/or to

prevent the association of an EPC of an item with a particular individual and/or to prevent tags from emitting any kind of a unique identification signal when performing a tag query by an authorised reader. Hence a mechanism is required by which a person can hide his or her true identity from devices that scan our personal RFID tags while still being able to take advantage of the benefits of RFID for the consumer.

9.7 Cryptography

Achieving the security and privacy objectives outlined in Table 9.3 and Table 9.5 respectively, require an enormous anthology of technical and legal tools. While legal tools are not considered in this dissertation, the required technical tools may be provided through cryptography. The following sections of the chapter consider cryptography, the science from which a plethora of technical tools for providing services to achieve the privacy and security objectives identified previously can be obtained.

Cryptography is defined as the study of mathematical techniques related to aspects of information security in [78]. However, cryptography is not the only mechanism by which information security may be provided.

Security and privacy issues concerning RFID may be solvable using a set of security mechanisms derivable from various cryptographic primitives. A security mechanism is a collective term used to refer to a combination of cryptographic primitives and protocols used to provide security. Hence, it is appropriate to briefly consider the subject of cryptography in the following sections to examine the range of cryptographic tools available for various applications, the level of security provided by such primitives and a simple classification of the vulnerabilities of various security mechanisms.

9.7.1 Cryptographic primitives

Cryptography is an ancient art that has been used throughout human evolution to provide security and to protect the privacy of individuals or organisations. Providing security and privacy for RFID systems will inevitably involve using some cryptographic primitive already in existence, or newly defined, along with suitable protocols that take into account the unique nature of RFID systems. Figure 9.5 gives a classification of a broad range of cryptographic primitives. A more complete description of these primitives can be found in [78, 79, 80 and 81].

Most modern cryptosystems, such as the RSA cryptosystem (with a few exceptions such as one-time pads), are based on some mathematically hard problem and the level of security provided by the system will depend on the difficulty of the mathematical problem.

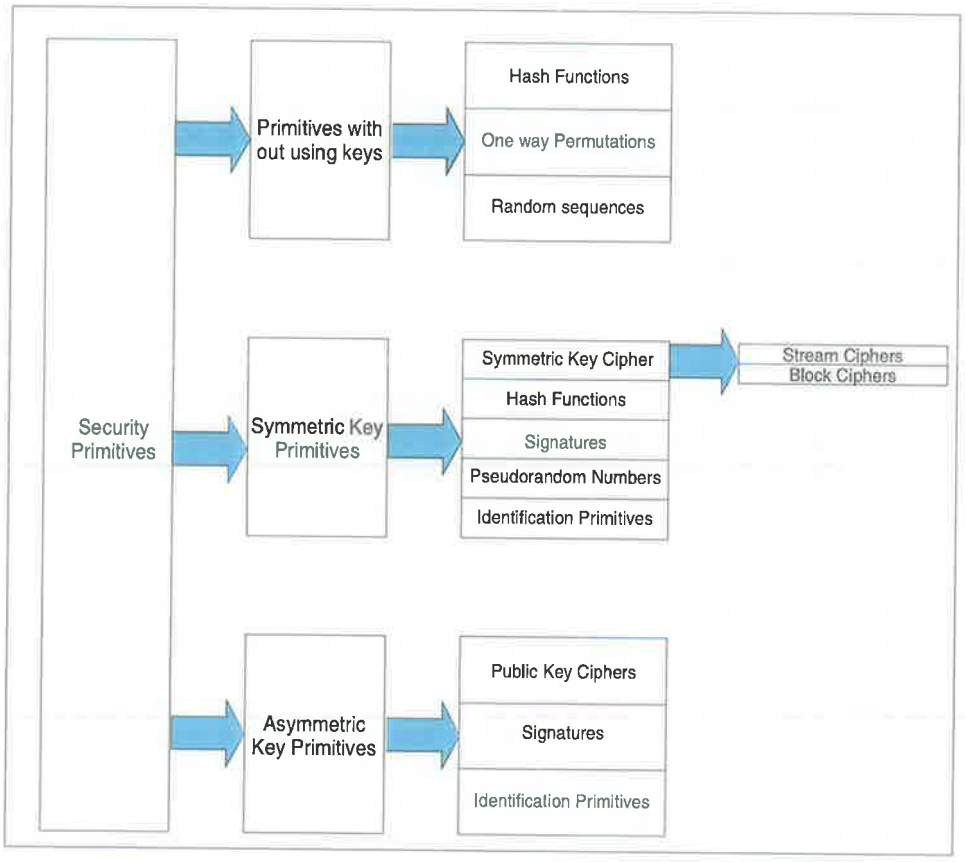

Figure 9.5 Classification of cryptographic tools [78].

It is important to define the difficulty of a problem before the level of security provided by a cryptographic system can be discussed. A mathematical problem is said to be difficult if the time it takes to solve the problem is immense compared to the size of the inputs to the problem. Modern cryptographic systems are based on mathematical problems where the fastest known algorithm takes exponential time to find a solution. This implies that the time taken to solve the problem increases exponentially as the size of the inputs to the problem increases linearly. Thus the level of security provided by a cryptosystem is often expressed as the number of operations required to break the cryptosystem or the time taken. Generally, the level of security provided by a cipher complements the commercial value of the information protected by the cryptographic system. A discussion on quantifying the security provided by a security mechanism is considered in Section 9.7.3.

While there are numerous cryptographic systems in use based on various primitives as outlined in Figure 9.5, all such systems are not without their own set of weaknesses. There are specific attacks on any cryptosystem or protocol employed by a security mechanism to provide security. These weaknesses are a result of certain vulnerabilities in the cryptographic scheme or due to certain flaws that may have entered into the protocol employed in the security mechanism. A classification of attacks on cryptographic systems in general is discussed in Section 9.7.2.

9.7.2 Classification of Attacks

Cryptanalysis is the art of recovering the plaintext of a message without the key by using an algorithm to infer the plaintext from a given ciphertext or by deducing the key so that ciphertext can be decrypted to obtain the plaintext. An attempt made by an adversary at cryptanalysis is termed as an attack [79]. The following sections consider the possible attacks on cryptographic primitives and protocols to defeat a security mechanism.

9.7.2.1 Attacks on Cryptographic Primitives

There are various forms of attacks possible on cryptographic primitives, with various names. However, the most common forms are outlined in Table 9.6 (refer to [78, 79, and 80] for more details).

Table 9.6 Attacks on cryptosystems.

Cryptanalysis Method	Description
Ciphertext-only	This type of attack is carried out by an adversary who is in ownership of a string of ciphertext [80]. Here the adversary tries to deduce the decryption key or the corresponding plaintext from the string of ciphertext [78]. An encryption process that can be broken by a ciphertext-only attack is considered to be completely insecure [78].
Known plaintext	In this type of attack, a cryptosystem is attacked by an adversary in ownership of both a string of plaintext and the analogous ciphertext [80].
Chosen plaintext	In this case, the adversary is in possession of a ciphertext string corresponding to a plaintext of his choosing (perhaps as a result of gaining temporary access to the encryption engine) [80]. Then the adversary uses any information deduced from the analysis of the plaintext and ciphertext pair to obtain the plaintext of another ciphertext message.
Adaptive chosen-plain text	It is a chosen-plaintext attack where the choice of the plaintext used by the adversary may depend of the ciphertext observed from previous requests [78].
Adaptive chosen ciphertext	It is a chosen-ciphertext attack where the choice of the ciphertext used by the adversary may depend of the plaintext observed from previous requests [78].

9.7.2.2 Attacks on Protocols

Similar to attacks on cryptographic primitives, there is a vast array of attacks on the protocols used and the number of attacks has grown with the emergence of new protocols. Table 9.7 is a summary of a prevalent list of possible attacks.

Table 9.7 Attacks on cryptographic protocols.

Protocol Attack Method	Description
Replay	As discussed in Section 9.3.3, in a replay attack, the adversary records a conversation between trusted parties and then replays a section or all of the recorded information at a different time to break the security.
Known key	An adversary attempts to determine the secret keys to be used in the future based on certain secret keys obtained in the past [78].
Impersonation	This term is taken to refer to an attack in which the attacker creates a misleading context to trick a legitimate party into making an inappropriate security-relevant decision, or fooling the legitimate party into believing that the attacker is legitimate.
Dictionary	The adversary uses a dictionary consisting of a large number of probable keys or passwords and applies it to defeat the security by guessing the accepted key or password. This type of attack is usually applied to defeat the security of password protected systems [78].

9.7.3 Level of Security

Many cryptographic systems have been broken because of increased computational resources, development of faster and better algorithms or problems which are proven to be easier than when they were first conceived. This is the reality of any cryptographic system. However, the concerning issue for modern cryptosystems is not that the system will eventually be broken, but that the range of possible attacks on a security mechanism to breach security and the time taken to break the security system using the best possible attack

It should be noted here that, in general, the security of a system is difficult to quantify. The usage of the term, 'level of security,' is generally used to refer to the number of operations required or the amount of time taken to break the security of a given system using state of the art technology and the best available algorithms.

However, it is possible to evaluate the security provided by certain mechanisms and describe them using a number of classifications, some of which are published in [78], [79] and [80]. Terms used to describe the level of security of a system are outlined in Table 9.8.

Table 9.8 Defining levels of security.

Level of Security	Description
Unconditional Security	A cryptographic system is described as being unconditionally secure if the security of the system cannot be broken if an adversary is given unconditional resources. Encryption systems with perfect secrecy (where the observation of the ciphertext does not provide any information regarding the plain text) are unconditionally secure [78]. An example of an unconditionally secure encryption system is the one-time pad.
Computational Security	Computational security is given as a measure of the amount of computational work required, using the best available methods, to defeat the security of a system. A system is considered computationally secure if the amount of computer resources or time required to break the system is far more than that available to an adversary considered in the analysis of the system. Computation security is also termed Practical security [78]. As described in [78] computational security can be evaluated in terms of the number of computational operations (as measured by clock cycle times or the number of fundamental operations) required to defeat an intended security objective. As such, the level of security can be defined as the minimum amount of work required break the security of the system. The amount of work thus required is termed as the "work factor" [78]. Clearly, as technology and algorithms improve, the work factor will vary. Thus a more practical definition is the "historical work factor" [78] which estimates the amount of work (in terms of time) taken to defeat a security objective using the best available methods at a given point in time. Complexity theory provides a means of analysing the computational complexity of different algorithms and problems by expressing the space time requirements of algorithms and then classifying problems based on the algorithms required to solve them. Thus complexity theory generally provides avenues for demonstrating the security of a cryptographic system by estimating the time required to break the system or by proving that cryptographic mechanism is based on a problem related to a class of problems in or beyond the domain of NP in the complexity hierarchy of problems [79].
Ad-hoc Security	Systems are classified as having had-hoc security when postulations are made using any number of apparently convincing arguments that all possible attacks on the system require a level of resources (computational and time) that are beyond the level of resources available to a hypothetical adversary. Security systems of this type are generally designed to counter some well known attacks and, where they survive such attacks, they are also said to have "heuristic security" [78].

Provable Security	A system is described as having provable security if it can be shown that breaching the security of the system involves evaluating the solution to a problem belonging to a class of problems (such as NP-hard problems) that can not be calculated in polynomial time, such as the integer factorisation problem or the discrete logarithm problem. As pointed out in [78] it should be noted here that the proof here is that the given problem is at least as difficult as an existing problem, and not an absolute proof of security.

9.8 Low Cost RFID and Cryptography

The plethora of available security primitives are too excessive in terms of cost to be implemented on a cost constrained RFID chip with around 4000 gates for logical functions (refer to Section 9.2.1). Low cost labels are also not self-powered and only consist of limited logic functionality, unlike smart card processors. However, they may be more suitable for higher class labels with a greater opening price point. For instance, private key cryptosystems such as AES are not suitable since a commercial implementation of AES typically requires 20,000 - 30,000 gates [107]. This is far more than the number of gates on an entire low cost label. However the SHA-1 specified by the US Department of Commerce is a possible candidate for an encryption rule but hardware implementations of SHA-1 are currently too costly to meet the cost budget of low cost RFID labels [108]. Cryptographic systems and protocols need to fit into a label footprint without dramatically increasing the cost of a label.

Considering cryptographic solutions for RFID requires a careful understanding of low cost RFID, underlying assumptions of the system, limitations and expectations from the end user community. There are particular challenges that need to be considered as a result of the nature of low cost RFID systems. These challenges are discussed in the following section.

9.8.1 Challenges

Challenging aspects to providing security and privacy for low cost RFID systems using traditional cryptographic mechanisms and existing hardware are outlined in Table 9.9. Each of the listed constraints needs to be considered before designing a practicable security or privacy measure.

It is evident from the description of low cost RFID systems provided in Section 9.2, and their associated implementation in supply chain applications as Class I and Class II tags, that the main constraint hindering the adoption of more traditional cryptographic solutions is the scarcity of hardware resources as a result of cost limitations. Nevertheless, cost is not the only limitation. There are many other such restrictions and difficulties that result as a consequence of the nature of electromagnetic waves and the constraints placed by end users and electromagnetic compatibility regulations.

As outlined in Table 9.9, EM regulations pose restrictions on the isotropic radiated power at stated distances. This implies that there is a maximum limit on the power available at a given label distance from a transmitter. Thus, passive labels with size limited by a particular label class or an application receive power from a stated power flow per unit area. The power available to the label is one factor contributing to the determination of the type of security scheme and the cryptographic hardware used in a label. Cryptographic hardware consuming considerable power (in the range of tens of microwatts) will significantly diminish the label reading distance and degrade the performance of the whole RFID system implementation. Furthermore, a security mechanism employing a memory write will have to account for the additional power required to operate a labels E^2PROM.

Table 9.9 Challenges facing the implementation of strong cryptosystems on low cost RFID.

Challenge	Description
Cost	Minimizing cost implies limited memory and silicon area constraints. Tag costs are expected to be less than 5 US cents. The cost of tags has reduced over the years and the trend is expected to continue thanks to Moore's Law. This has two implications; more hardware intensive cryptographic functions will slowly enter low cost RFID chips and the cost of an RFID will continue to decrease. Unfortunately, analogue devices fabricated on IC's do not scale in the same manner as the digital devices so RF front end on chips will still remain a cost factor.
Regulations	Transmit power restrictions, spectral masks, frequency of operations, available bandwidth, and time available for computations.
Power consumption	Important to minimize the power consumption of the label IC circuit to gain maximum performance. A cryptographic device which is the highest consumer of a passive chip's power will adversely affect its performance as it will reduce a label's read range.
Performance	Label performance and system performance goals (data transmission rates, number of label reads per second, percentage of correct reads). Performance goals also place a limit on the time available for any computations by cryptographic hardware. Cryptographic systems requiring access to backend systems will need to take into consideration network delays associated with a security mechanism as such delays will affect system performance.
Power disruptions	Sudden loss of power is a practical reality and any security mechanism should not leave the chip in a vulnerable state during such an event.

The power utilization of any security related hardware should not exceed the typical tag power consumption of 10-15 microwatts required for writing to a passive RFID label, as explained in Section 9.2.3. Ideally, the power consumed should be a fraction of this value for any security related hardware to be viable as considerable power requirements will constrain the label performance by limiting the operating range of the label. However reducing power consumption of any encryption hardware is a challenging prospect.

Power dissipation in integrated circuits is a function of many factors; the fabrication technology, the layout of the design and the scale of the fabrication process. Static CMOS

technology is very attractive in low power devices due to the almost negligible power consumption in steady-state operation. Power dissipation in CMOS circuits is mostly due to the charging and discharging of capacitances during dynamic operation. Power consumption can be reduced by the proper choice of circuit, logical or architectural structure. This might come at the expense of silicon area, which is critical to controlling tag costs.

The power consumption in static CMOS circuits is due to the static (or steady state) power consumption and the dynamic power consumption (power consumption during the switching of logic levels). The dominant power dissipation in CMOS circuits is caused by the switching of logic levels, while the static power consumption due to the leakage of current flow through the reversed-biased diode junctions in the transistors is almost negligible.

Equation (9.1) illustrates the total power consumption of a device *i* while (9.2) expresses the static power dissipation as the leakage current I_{static} and the supply voltage to the device V_{dd}. Equation (9.3) formulates the power consumption during $f_{0\rightarrow1}$ switching operations (logic $0\rightarrow1$ an $1\rightarrow0$ transition) per second, where C_L represents the sum of the intrinsic capacitance (junction capacitance and other parasitic capacitances) and the extrinsic load capacitance (due to the wires and connecting gate) of the device [110].

It is clear from (9.3) that higher throughput from a device leads to more frequent signal transitions and results in increased power dissipation. There is always a trade off between the power dissipation and area of silicon used (and hence costs) as use of parallel architectures can reduce the power consumption by reducing the rate of switching components and the supply voltage required. Reducing the supply voltage alone is not sufficient as this reduces the latency of the circuit and any security related hardware may not be able to meet timing constraints or performance constraints. Use of parallelism allows the trade off of silicon area for power. Design methodologies for reducing power consumption in CMOS logic are an active area of research and the reader is referred to [111] for more details.

$$P_i = P_{i\_static} + P_{i\_switching} \quad \text{W} \tag{9.1}$$

$$P_{i\_static} = I_{static} V_{dd} \quad \text{W} \tag{9.2}$$

$$P_{i\_switching} = C_L V_{dd}^2 f_{0\rightarrow1} \quad \text{W} \tag{9.3}$$

However, read range might not be a concern in certain applications and thus it is difficult to set a bound on the required power level, except to state that it should not exceed the power required by the tag during the writing of data to the memory as this is the most power consuming task a low cost tag is likely to perform. In addition, requiring more power would imply that a tag in its current position of being just able to operate (as it can be read by an interrogator) may not be able to complete a security related function, causing that operation to fail. This failure may expose or lead to vulnerabilities in the security mechanism. Hence power consumption is an issue that needs to be carefully considered.

Security mechanisms and communication protocols also need to be carefully designed to avoid leaving the label in a vulnerable state during sudden loss of power or interruptions to

communications. It is also important for security mechanisms to take into account the more powerful signal strength of the forward channel (reader to label transmissions), which can be detected hundreds of metres away, compared to the tag to reader communication channel which can be received from no greater than 20m using highly sensitive receivers.

The sections above have considered the nature of low cost RFID systems, its various vulnerabilities and the unique set of challenges to providing cryptographic solutions to alleviate these vulnerabilities. There are various solutions published in literature to address the weaknesses outlined in Section 9.3. The following section provides a survey of published solutions for addressing various privacy and security concerns related to low cost RFID systems.

9.9 A Survey of Solutions

Section 9.7 considered the subject of cryptography in a general perspective and introduced concepts that will be useful in the discussion of security mechanisms for RFID. The following sections consider cryptographic primitives, protocols and security schemes proposed for low cost RFID systems.

It is important to note here that, in addition to the possible vulnerabilities discussed in Section 9.3, there will be specific attacks on any cryptosystem or protocol employed by a security mechanism used to provide security. These attacks are a result of certain weaknesses in the cryptographic scheme or are due to certain flaws that may have entered into the protocol employed in the security mechanism. A description and examples of such attacks can be found in [78 and 80].

The following sections detail more recent developments addressing the issue of security and privacy for resource intensive environments.

9.9.1 Cryptographic Hash Functions

There have been a number of security schemes outlined in [108 and 109]. A proposed scheme for controlling access to a label uses the difficulty of inverting a one-way hash function [108]. This mechanism, called the 'hash-lock scheme,' is based on a hash value generated from a random message sent to a tag for locking a tag. Until the tag is unlocked the tag only responds with the hash value stored on the tag called the MetaID. The tag can only be unlocked by an authorised reader by sending the original random message to the label, where it is hashed and compared with that stored on the label's memory.

The primary flaw in this approach lies in the fact that a successful discovery of a MetaID and a label ID pair will allow an adversary to engage in a cloning attack. The hash locking method requires the implementation of a suitable hash function and the appropriate logic to implement the details of a communication protocol. The greatest challenge lies in the successful implementation of a hardware efficient hash algorithm on the label IC. Since any

reader can obtain the MetaIDs from labels, this scheme does not solve the problem of location privacy violations. The scheme is also susceptible to man-in-the-middle attacks since an adversary can query a label, obtain its MetaID, retransmit the value to a reader, and later unlock the label with the reader response.

However, the hash based access control can be extended to provide access control to multiple users, or control access to label functionalities such as write access. It is also possible to allow a third party to process labeled items using the MetaIDs and a database lookup scheme without having to unlock the labels. However, it should be noted here that any system that will function using the MetaIDs alone will suffer from the same security flaws as an unprotected label since the MetaID will act as a unique identifier (similar to an EPC on an unprotected label).

Randomised access control is another variation of the above scheme described in [108] and [109]; with the difference to the previous scheme being that a tag always replies with a random MetaID. However, the randomised hash lock scheme has similar flaws and difficulties. The emphasis is placed on removing the predictable nature of the label responses to reader interrogations. A detailed description of this scheme can be found in [108].

Readers are still susceptible to replay attacks. An adversary only needs to obtain a label response and the corresponding reader response to create a fake label. An important consideration in this scheme is the number of labels that can be successfully supported, since a large number of labels will cause increasing processing delays at the back end database systems when performing brute force searches of databases to obtain the matching tag ID for a given MetaID. It is also not known whether keyed pseudorandom number functions required to implement the scheme are a more efficient hardware implementation than a symmetric key encryption such as a hash function. The hardware complexity of keyed pseudorandom number functions is still an active area of research.

Analysis of the random hash lock scheme in [108] and [109] provided in [112] has also concluded that location privacy is only ensured in the scheme if an adversary cannot tamper with the tag to obtain the static tag identifier (such as the EPC).

The 'K-steps ID matching scheme' in [113] has outlined a method that utilizes representation of N tags on the N leaves of a tree of depth K where traversing the nodes of the tree down to the leaves yields the unique label identifier. This method allows to reduce the algorithmic complexity of searching a backend database from $O(n)$ to $O(\log n)$ where n is the number of tag identifiers stored in the database.

The hash locking method requires the implementation of a suitable hash function and the appropriate logic to implement the details of a communication protocol. The greatest challenge lies in the successful creation of a hardware efficient hash algorithm. However, it has been shown in [114] that implementations of hash lock schemes are generally not cheaper than symmetric key encryption schemes using an AES and a SHA-1 implementation as an example of a symmetric key encryption scheme and a hash scheme, respectively.

9.9.2 Cellular Automata

The theory of Cellular Automata (CA) [115] developed by Wolfram has been used to develop a number of different cryptographic systems. Cellular Hash (CH) [116] is one such outcome and there is a rich variety of inexpensive encryption mechanisms developed based on the chaotic nature of CA systems [117 and 118]. CA may be built out of a feedback shift register and a single pair of gates providing a compact solution for low cost RFID. In addition, CA based hashes scale well as the size of the hash digest increases but CA hashes require many parallel calculations and thus they may impose considerable demands on a tag's available power. However, it is possible to perform CA operations in series but that will be at the expense of RFID system performance.

The CA cryptosystem encountered in CAC [118], while being promising, has been shown to be vulnerable to differential cryptanalysis or has been shown to form an affine group [119]. However, the estimated size of the un-optimised pre-layout area is about 4.25 sq. mm, which is far bigger than a typical RFID silicon design, which is about 0.25 sq. mm. Even if optimisation halves the design, the silicon cost is too high for a low cost RFID chip. Nevertheless, improvements and a scaling down of the design may be possible since the analysed design was for a 128 bit key.

The use of CA generators in the formulation of stream ciphers has also been proven to be insecure. It was shown in [120] that the output of a CA generator is identical to the output of a LFSR and hence CA systems are as insecure as LFSR based systems.

9.9.3 Linear and Non Linear Feedback Shift Registers

While LFSRs are capable of generating pseudorandom sequences they are insecure due to certain non random properties. Part of the problem is the linearity of the bit sequence which makes them of little use for encryption. Even in the event that the internal structure of the LSFR scheme is kept secret, an attacker only requires $2n$ output bits of the generator to determine the entire output sequence of a LSFR of length n [121 and 122].

Use of non linear feed back shift (NLFSR) registers to design a hash by using a complicated feed back function is a possibility, since a shift register implementation does not require complex hardware. However an important consideration should be whether the additional cost of a NLFSR provides an adequate level of security, considering the vulnerabilities of various non linear feedback shift register based schemes in literature [78, 79 and 80].

9.9.4 Message Authentication Codes

The use of Message Authentication Codes (MACs) has been discussed in previous literature. Takaragi [25] and his team of researchers have been the first to make an RFID chip (μ-chip) equipped with a MAC commercially available. The chip manufactured using

a 0.18 micron CMOS technology occupies less than 0.25 mm^2 of silicon wafer, placing the IC in the low cost end of the RFID labels.

The MAC implementation adopts a very simple approach. The security of the μ-chip relies on a 128 bit ID stored permanently on the chip at manufacturing time. This ID is a concatenation of a previously encrypted MAC and chip data, the MAC being derived by encrypting a portion of the data using a hash function and a secret key, where the secret key is known to the manufacturer and the client. This mechanism does raise the difficulty level for forgers as the process of eavesdropping and creation of fake labels is made more complex. However it does not provide privacy as the ID code embedded in the chip will breach anonymity and location privacy. There is also the risk of the key, which is common to many labels, becoming known.

9.9.5 NTRU

The NTRU cipher appeared in 1995 [123]. NTRU is based on the Closest Vector Problem which involves finding the closest vector given a lattice L, and a target vector y [80]. It is similar to the knapsack problem. A lattice is defined as "the set of intersection points of a regular (but not necessarily orthogonal) n-dimensional grid". This is an NP-Hard problem where there is no known algorithm for solving it in polynomial time [124].

There are several cryptosystems based on this problem [125 and 126], but they have not gained in popularity due the excessive size of the keys needed to provide security comparable with other public key cryptosystems. The main advantages of NTRU are that it requires moderate resources and it is a faster operating algorithm. However, it is difficult to make accurate comparisons with other algorithms, as NTRU depends on many parameters that govern its behavior. Early research indicates that NTRU is generally faster, relatively easier to implement both in hardware and software than other public key cryptosystems such as RSA, and needs only a modest size memory [123]. Its simple implementation and limited demand on memory have already proven its relevance in RFID applications [127].

Nonetheless, NTRU is susceptible to brute force attacks and multiple message transmissions [123]. A more detailed treatment of attacks on NTRU can be found in [124]. In addition, NTRU has a relatively large message expansion. Encrypted messages are almost twice the length of the plain text messages. This may not be a pressing concern as RFID messages are not of very long length.

9.9.6 Tiny Encryption Algorithm

TEA is an encryption algorithm designed for simplicity and ease of implementation. The encryption algorithm is based on the Feistel cipher [78] and a large number of iterations to gain security without compromising simplicity. A description of the algorithm is provided in [128].

TEA can be effortlessly translated into any language as long as 'exclusive or' is an available operation. A hardware implementation of the algorithm is stated to have the same complexity as DES [128]. Despite its simplicity and the ease of implementation, TEA is a relatively recent invention and the level of security or its vulnerability to attacks is still not very clear.

9.9.7 Scalable Encryption Algorithm

The authors in [129] have noted that resource constrained encryption using symmetric cryptography does not have a long history. They cite TEA above and indicate the vulnerabilities of TEA to linear and differential cryptanalysis attacks.

SEA (Scalable Encryption Algorithm) is a scalable encryption algorithm for small embedded applications [129]. Typical performances of the SEA algorithm on encryption and decryption using a 128 bit key and 1 MHz 8-bit RISC processors can be undertaken in a few milliseconds, using a few hundred bytes of ROM [129].

9.9.8 Re-encryption

In [130] an unorthodox re-encryption mechanism is proposed for providing privacy and security protection to banknotes embedded with RFID labels. In a traditional setting, the entity conducting the re-encryption will not be aware of the plaintext. However, in the re-encryption scheme discussed in [130], the plaintext is known to the entity performing the re-encryption. The scheme is elaborate and the details are complicated, thus, only an overview of the scheme is given below.

The security of the mechanism is based on the ciphertext created by encrypting the digital signature stored on the RFID chip by a central bank authority, the serial number of the bank note and a random number. The authenticity of the banknote can be verified by comparing the ciphertext stored on the banknote to the ciphertext obtained by encrypting the digital signature, the serial number, and the random number using a public key stored on an RFID reader. A match indicates an authentic banknote. Figure 9.6 provides an overview of the data placed on each banknote. In addition, an access control mechanism prevents the data on an RFID label from being read without making optical contact first. This prevents remote alteration and interrogation of an RFID label's memory contents.

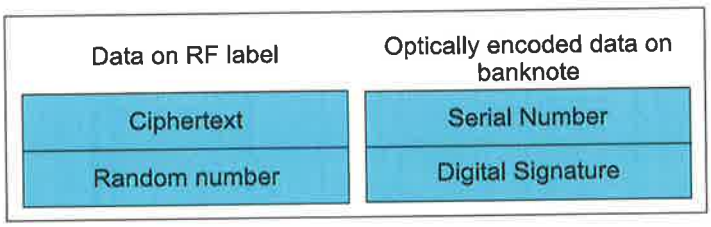

Figure 9.6 Data on a banknote.

The significant privacy and security achievements outlined include consumer privacy (tracing of individuals or banknotes is only possible with the use of a key), forgery resistance, fraud detection and tamper resistance. In contrast with the aforementioned mechanisms outlined in Section 9.9.1, the primary significance of the re-encryption mechanism is that a banknote is not in possession of any secret keys and the RFID label is not required to perform any resource intensive operations. The encryption engines and secret keys have been shifted away from the RFID label to more secure locations, such as the readers and the central bank authority.

Despite the inventiveness of the re-encryption scheme, the significant drawback is the adequacy of the information obtainable from a banknote to create fraudulent banknotes. The digital signature is not verified during a transaction; hence, the fake banknotes can be created with ciphertext obtained from a collection of believable serial numbers. Other shortcomings that might be exploited by a resourceful adversary are provided in [130].

9.9.9 Lightweight Cryptography

Lightweight cryptography is a branch of cryptography that aims to develop fast and efficient cryptographic mechanisms for resource constrained environments. Hence this branch of cryptography has been the most promising avenue to generate secure cryptographic solutions to low cost RFID systems. Several lightweight cryptographic models relevant to RFID are summarised below.

Building cost effective cryptographic hardware for RFID is still not a reality. Although certain advances have been made towards the development of hardware optimised encryption engines in [131], [132], and [133], they still present a performance hindrance and an expensive solution to current RFID systems.

9.9.9.1 Lightweight Hardware

The process of developing or optimising the existing hardware of security mechanisms has been an active area of research with respect to smart card processors. However, there has recently been some focus on more resource-constrained environments such as low cost RFID ICs. There have been a number of advances towards developing such low cost hardware [131, 132, 134, and 135]. Elliptic Curve Cryptography (ECC) has presented itself as a public key cryptosystems for RFID [133] due to the smaller key sizes required to provide an adequate level of computational security. A relatively low cost implementation of an ECC processor suitable for RFID can be found in [136].

9.9.9.2 Lightweight Protocols

Building on prior work, Hopper and Blum have suggested two shared-key authentication protocols, HB and HB+ protocols [137]. HB protocol is proven secure against a passive (eavesdropping) adversary. The HB+ protocol is proven secure against active attacks. Security of these protocols are based on the conjectured hardness of the "learning parity with noise" (LPN) problem. Their extremely low computational overhead makes them very suitable for low power, bandwidth and low cost devices such as RFID. In [138] it has been proven that the security of these protocols only holds for sequential executions and the question of whether the security also holds in the case of parallel or concurrent executions is explicitly left open.

Katz and Shin [139] suggest that, in addition to guaranteeing security against a stronger class of adversaries, a confirmation of the security in parallel and concurrent operations would allow the HB+ protocol to be parallelised. This would also reduce substantially its round complexity. Katz and Shin prove the security above. They also suggest simpler security proofs for these protocols which are more complete. In effect they also explicitly address the dependence of the soundness error and the number of iterations.

An improved version of the HB+ developed by [140] is analysed in [141] where a number of improvements have been made against various vulnerabilities outlined in [140]. HB++ [229] is a modified version of the HB+ protocol designed to thwart a wider adversarial attack in [230].

9.9.10 Minimalist Cryptography

The formulation of mechanisms to achieve security and/or privacy objectives under the constraints presented by low cost RFID systems to real-world tags using a weak, but perhaps a realistic, security model form the basis for minimalist cryptography.

9.9.10.1 Pseudonyms

An early version of minimalist cryptography was proposed in [142] where a list of randomly generated tag identifiers was used on a tag. On querying a tag, a reader is able to hash the response and access tag related data on a secure hash table. The idea of using completely random EPCs and an outline of such a scheme was given in [142]. A similar version was also published in [143] with a minimalist security model and accompanying protocols for low-cost tags. The proposed method has every tag containing a collection of pseudonyms; it releases these pseudonyms on each interrogator query. Both schemes have left open the possibility for a valid reader to renew the list of pseudonyms on a tag.

The use of pseudonyms in [143] is based on the assumption that the intruder only comes into the scanning range of a tag on a periodic basis, as a complete analysis of the limited number

of pseudonyms will allow the identification of the tag. The security model is also based on the underlying assumption that the tags release their data at a limited rate [143]. The minimalist model sets an upper limit on the number of times an intruder or an adversary can scan a given tag or try to spoof a valid reader.

9.9.10.2 One Time Pads and Random Numbers

In addition to the schemes presented above, there are many security schemes in the patent literature. Information security is a secretive realm, with many holding firmly onto their intellectual property with a whole array of patents. It is the nature of the beast.

Most methods outlined in patent literature are too complicated for low cost RFID. However, [144] demonstrates a very simplistic approach. The patented scheme relies on a simple one-time pad concept, where the intended application is that of bank notes. The scheme involves the recording of a random number, a time, and a date stamp on an RFID label of a bank note on release of the note for circulation. The bank note keeps a track of the number of times it has been scanned and this number is used as part of its authentication process. When a bank note is read by a bank teller, the random number, date, time stamp and the number of scans are sent to a central bank computer to verify the authenticity of the note based on comparing the same information securely stored on the computer.

This scheme is subject to imitation, simply because the label can not be trusted as a secure place of storage for valuable information because bank notes provide adequate incentives for a physical attack on the RFID label.

A different application of one-time pads can be found in another patent [145]. In the novel scheme, labels are equipped with a small rewritable memory. Prior to the release of a label, a set of random numbers (authentication keys) generated by a completely random physical process is stored into the label along with a label ID. A back end database stores a copy of the random codes and the associated label IDs.

The label ID may be read from the label during an interrogation. The ID provides knowledge of the label being authenticated by the reader by consulting the relevant records in the secure back end database. In consultation with the database, the interrogator may transmit one or more of the random numbers stored in the database. One of the numbers should match a series of random numbers stored on the label. If a match occurs, the label responds with a return authentication code known exclusively to the database and increments a counter to select a set of new random numbers for the next authentication procedure. An identical counter that determines which of several authentication numbers is next in force is incremented at the database to synchronise the database entries with that of the label.

Hence, the above mechanism prevents an eavesdropper from obtaining any information regarding the next correct authentication key, or the next label authentication response. The only available information to an eavesdropper is an apparent burst of random numbers.

In the event of an unauthorised reader, the label will not respond unless the reader knows the next random number expected by the label. In case a counterfeit label is interrogated, the label may respond with a random number but the interrogator will fail to find a match, and thus detect the counterfeit.

Nevertheless, this scheme still leaves the possibility of a physical attack where the contents of the label may be discovered. However, in the worst case, this information cannot be used to counterfeit labels in massive quantities as the set of authentication keys and authentication responses are all different and completely random on each individual label.

9.9.11 Exploiting Noise

RFID systems are based on limited computing power and are not suitable for public key cryptography. Hence a protocol with low computation burden is outlined in [146].The protocol in [146] takes advantage of signal noise on a communication channel to secretly exchange a key in the presence of an eavesdropper.

An alternative proposal was presented in [147]. Similar to the blocker tags [148], the special tag in this proposal is named as the noisy tag. Noisy tags are owned by a reader's manager and set out within a reader's field. They are regular RFID tags that generate noise on the communication channel between the reader and the queried tag. This is done in such a manner that the intruder or eavesdropper cannot differentiate the messages sent by the queried tag and those sent by the noisy tags. Hence the intruder is unable to identify the secret bits that are sent to the reader. Afterwards, the secret shared by the reader and the tag can be used to launch a secure channel in order to protect communications against eavesdroppers. In addition, the tag's identifier can be refreshed by exclusively-or'ing the new identifier with the exchanged secret [147].

9.9.12 Radio Fingerprinting

If tags have distinct "radio fingerprints" that are difficult to reproduce, then these fingerprints, on their own, could help strengthen device authentication [149]. This technique, while sound in theory, is not a practicable avenue because obtaining such a radio fingerprint is expensive and difficult.

9.9.13 Distance Implied Distrust

This scheme is based on the assumption that an unauthorised reader attempting to read a tag will generally be more physically distant from the tags than a legitimate reader. The latter assumption is based on the realisation that a closer and more visible reader will draw greater investigation by tag owners or tag bearers. Thus, the measurement of distance of a reader to a tag is proposed as a measure of trust [150].

9.9.14 Authentication Protocols

The YA-TRAP protocol proposed in [151], based on [152], provides location privacy and allows the authentication of the tag by using monotonically increasing timestamps stored on the tag which are in synchronicity with timestamps on a secure backend database. This protocol requires the implementation of an iterated keyed hash function on the tag. However, the proposal is vulnerable to a DoS attack initiated by desynchronising the timestamp between the tag and the backend databases. Nevertheless by using hash tables that are pre-computed, the search time required for correlating a tag response is reduced to an $O(1)$ operation which requires less workload than the randomised hash lock scheme outlined in Section 9.9.1.

Another version of the YA-TRAP protocol in [153] also addresses some of the weaknesses in [151] albeit at the cost of increasing the workload of backend systems.

9.10 Conclusion

Despite the vast array of RFID systems, those that are within the low cost spectrum pose the greatest threat due to the possibility of wide scale deployment and inherent constraints that place limitations on the number of possible solutions. This chapter has introduced in detail systems characterised as low cost RFID systems and identified numerous vulnerabilities of low cost RFID systems operating under the UHF EPCglobal air interface protocol. These vulnerabilities lead to both security and privacy related issues. Addressing these issues requires implementing various security and privacy objectives. Providing services to achieve those objectives traditionally requires the implementation of various cryptographic primitives.

However, low cost RFID, due to its resource intensive environment, presents a number of difficult challenges to more robust cryptographic mechanisms with a high level of security. Most of these robust cryptographic mechanisms are too area or power hungry to fit well within the limitations of RFID systems, and much of the encryption hardware available is for smart card technology. Even though the solutions can be applied directly to RFID, the main obstacle is that smart card processors are much more powerful than a typical RFID label consisting of only 200 - 4000 gates. Thus, the solutions are not portable to an RFID platform if we expect the cost of the secure labels to remain below the 5 cents mark.

It is also clear from the discussion on the level of security that all security system designers aim to develop a system where the only possible means of attack is by way of an exhaustive search. Such a system will then have a computational security determined by the size of its key space. Thus any security system considered must surely have a large enough key space to ensure security, especially considering the fact that CPUs used in personal computers in this century have benchmark performances rated in terms of teraflops (floating point operations per second).

Considering Moore's Law, cryptographic solutions that seem too expensive for low cost RFID may become the solutions in the future. However, it is not possible to wait for a future time frame while the deployment of RFID systems is taking place around the world. Since advances in cryptography are slow to arise, due to the time taken to scrutinise new mechanisms and find faults, the best option might be to fall back on simple and proven techniques, such as those presented in minimalist encryption and lightweight cryptography.

The survey of various research efforts to address the security and privacy issues provides an insight into current developments in the area of security for resource intensive, low computation capable devices.

It is important to recognize that the resource limitation of low cost labels suggests that the simplicity of small one-time pads, which involve one or more small shared secrets between a label and an interrogator, and relatively simple chip implementations, should also be considered and must not be discounted. Some of the concerns arising from privacy and security may also be removed by occasional use of shielded electromagnetic communications between the label and the reader system.

There are unique opportunities within the label class hierarchy to develop various schemes for meeting the security and privacy levels expected by labels belonging to their respective classes. This opens the gate to a vast number of research avenues that could be pursued with regard to providing both security and privacy to low cost RFID systems.

It must be realised that security will come in many flavours and strengths, but 'low cost' implies that we find mechanisms that are 'good enough' and are deterrents, rather than mechanisms that are impossible to crack.

Perfect secrecy is a fine mathematical concept; in reality, there will always be a human element that is difficult to quantify into any mathematical formulation. Thus, it is practically impossible to have a perfectly secure system. Once this is understood, it is possible to move onto addressing realistic security and privacy issues overshadowing RFID.

While this chapter has illuminated the weaknesses of low cost RFID systems, resulting problems and services required to address such problems, it has not provided any technical solutions to the implementation of the required services. Prior to discussing solutions, it is important to define a framework within which such solutions are sought so that realistic and reasonable assumptions can be made about the requirements of those services. It is also important to develop a security model of the system that is being secured and a method of analysing the appropriateness of the solutions to meeting the needs of low cost RFID systems. The following chapter will develop an evaluation framework and an appropriate security model, before considering the subject of developing technical solutions.

Chapter 10

EVALUATION FRAMEWORK

The previous chapter examined low cost RIFID systems and various vulnerabilities of such systems that need to be addressed along with various solutions proposed for overcoming resulting security and privacy issues. This chapter will formulate a framework for defining the problem space constructed around low cost RFID systems, so as to enable the engineering of solutions to overcome the defencelessness of low cost RFID systems and be able to evaluate those solutions for their effectiveness.

10.1 Evaluation Framework

Thus far, the focus has been on considering various aspects of low cost RFID technology and considering its various vulnerabilities along with developing security and privacy objectives to address such vulnerabilities. As has been discussed in Section 9.7, achieving security and privacy objectives using cryptographic solutions in low cost RFID is a challenging but a necessary proposition. Nevertheless, the design, analysis and the evaluation of security mechanism is another challenging aspect due to the lack of established evaluation criteria or a clear security model. The following sections develop simple evaluation criteria for security mechanisms and a simple, yet sufficient model of a low cost RFID system for analysing security mechanisms.

Table 10.1 An outline of low cost RFID system characteristics.

Class of labels	Class I and Class II type of labels, (as they are low cost RFID labels)
Unique Identifier	EPC of length 96 – 256 bits. As defined in by EPCglobal in their tag data specification standard.
Read range	3 m – 5 m for UHF and 200 – 500 mm for HF operation under FCC regulations.
Label reads per second	200 – 1500 as demanded by end user performance requirements.
Hardware Cost	250 – 4000 gates in order to keep tag costs low and close to the 5 US cents target value.
Power consumption	10s of microwatts, and should not exceed that required for E^2PROM operation, so the tag read range requirements can be maintained.

It is important to define certain boundaries and assumptions taking into account the challenging aspects of implementing cryptographic primitives on low cost RFID because implementing mechanisms to address security or privacy otherwise is inconceivable. Defining and modelling the problem space will also aid future research in low cost RFID Security. Table 10.1 summarises the important aspects of low cost RFID, introduced in Section 9.2, that needs to be understood and reasonable assumptions that need to be made prior to implementing any cryptosystems to address the vulnerabilities outlined in Section 9.3.

10.2 Evaluating Security Measures

While designing cryptographic solutions is challenging it is important to be able to evaluate security measures devised to ensure that various goals outlined in Table 9.3 and Table 9.5 suggested for providing security and privacy are satisfied while meeting expected performance levels and costs. Table 10.2 outlines a security evaluation matrix to

appraise the suitability of various mechanisms for providing security and privacy to low cost RFID and various applications constructed around low cost RFID.

Table 10.2 Criteria for evaluating security mechanisms.

Achieved Security Objectives	Confidentiality
	Message content security
	Tag Authentication
	Reader Authentication
	Product Authentication
	Access control
	Availability
	Integrity
Achieved Privacy Objectives	Anonymity
	Untraceability
Cost and Performance Estimates	Tag implementation cost estimate (gate count estimation)
	Back end resource requirements (online or offline)
	Overhead costs (initialiaation costs, requirements or time)
	Time estimate (time to complete a process or hardware throughput or clock cycles)
	Estimation of power consumption (maximum bound)

10.3 Evaluating Cost and Performance Objectives

The security and privacy objectives were discussed in detail in Section 9.5 and Section 9.6. However estimating the cost of security mechanisms, and their power consumption and performance was not discussed. The following sections consider the cost and performance objectives in Table 10.2.

10.3.1 Tag Implementation Cost

Evaluating the tag cost of a security measure generally refers to the cost implication for its implementation on a tag IC. It is generally not easy to carry out such an implementation, and an estimate may be found by implementing the hardware required on a FPGA. However, it is generally possible to evaluate the cost of an IC in terms of the number of gates required for its implementation.

One NAND gate is considered to have a unit area in CMOS standard cell based hardware and it is common to express the area evaluation in terms of the number of gates (NAND)

required. Implementing a NAND gate in hardware requires at least four FETs. Typical cost estimations in terms of the gate count are given in Table 10.3.

Table 10.3 Cost estimation guide for cryptographic hardware based on static CMOS designs.

Functional Block	Cost (gate count equivalent)
2 input NAND gate	1
2 input XOR gate	2.5
INV (Inverter)	0.5
2 input AND gate	2.5
FF (Flip Flop)	12
D latch	2
2-1 MUX	5
n-bit LFSR	$n \times 12$
n-byte RAM	$n \times 12$
HA (Half Adder)	1 XOR + 1 AND
FA (Full Adder)	3 AND + 2 OR + 2 XOR

When referring to the cost of a tag, for security purposes the associated cost implied is the cost of the digital components required to implement the security measure on chip. In respect to the cost constraints placed on these labels and taking the current cost of fabricating a transistor to be $1/1000^{th}$ of a cent, the low cost labels can be expected to have 250 - 4000 gates available for security purposes, although the number of gates available is expected to increase over the years as manufacturing techniques and processes improve or as RFID IC manufacturing begins to fill the excess capacity of obsolete processes that are still in operation.

10.3.2 Backend Resources and Overhead Costs

It is generally difficult to implement a security mechanism without the aid of proxy systems or secure backend system for storing secret information such as keys. Security mechanisms of this kind require online and real time access to secure resources. The monetary and time cost of implementing such a mechanism must be accounted for in the evaluation process. Trimming constraints placed on RFID security mechanisms may require expensive database system implementations and expensive networking infrastructure. However, it may be possible to design a security mechanism that performs its operations off-line (that without requiring online access to secure resources). Hence backend resource costs can be generally expressed as those requiring online access or those that can be performed off-line.

Overhead costs may result from the need for initializing tags with secure information, or the need for performing some operations prior to their use or periodically during their use. For instance a security mechanism may require the replenishment of secret keys on a tag.

10.3.3 Power Consumption

Any security mechanism design will eventually involve an IC implementation. Currently, static CMOS is the choice of most digital circuit designs built for low power consumption and robustness. Hence, it is appropriate to consider power consumption analysis based on an implementation using static CMOS technology. However a drawback of this technology is the extra silicon area required for implementing logic compared to dynamic CMOS.

An important aspect of the design process and the establishment of its suitability is to ensure that the power dissipation of the integrated circuits do not exceed that outlined in Table 10.1. There are several techniques for measuring power consumption [154 and 155] and most methods rely on simulation techniques based on various models using simulation tools such as HSPICE [156].

While the use of direct methods to measure power dissipation may be possible a simple method for estimating the dynamic power dissipation is based on formulating the power loss during the charging and discharging of capacitances. Equation (10.1) models the power dissipation of a node that may consist of a number of logic gates and it is probably most practicable when the logic circuit complexity is a minimum.

$$P = p_{0\to 1} C_L V_{dd}^2 f_{clk} \quad \text{W} \tag{10.1}$$

In (10.1), C_L is the output load capacitance along the critical path and $p_{0\to 1}$ is the fraction of time the node makes a power consumption transition (that is a logic 0→1 and 1→0) in a single clock cycle. The combination of $p_{0\to 1}$ C_L can also be stated as the average capacitance switched during each clock cycle. In (10.1), f represents the clock frequency and V_{dd} is the maximum signal swing, more generally taken as the operating supply voltage. It is difficult to apply this formula to ICs of large sizes. However, it is adequate for estimating the power consumption in small hardware, especially if circuit design tools can be used to evaluate the capacitance estimates.

10.3.4 Performance

Furthermore, it can be assumed that the time available for a label operation (and thus, any implementation of a security mechanism that needs to be performed in real-time) is in the range of 5 - 10 milliseconds considering the performance criteria of an RFID system that demands a minimum label reading speed of 100-200 labels per second. In accordance with C1G2 protocol, a maximum tag to reader data transmission rate bound of 640 kbps and a reader to tag data transmission rate bound of 126 kbps, based on equiprobable binary

ones and zeros in the transmission, can be calculated. Hence a typical passive RFID label data transmission rate of the order of 100 kbps is a reasonable assumption.

10.4 Security Model

Given the list of vulnerabilities posed by low cost RFID systems and the unique nature of the technology, it is natural to consider the nature of adversaries in an RFID systems context with its various limitations, and to provide a clear security model. It is important to develop the nature of various adversaries in an RFID systems context in order to be able to analyse the level of protection provided by various security mechanisms and the various circumstances under which the security mechanism may be deployed to achieve the security and privacy objectives outlined in Table 10.2. Various ideas modelling possible adversaries and system models have been published in [109], [143], [112] and [184]. However, analysing the security of an RFID security mechanism is still challenging, partly because of a lack of a universal model or the narrow scope of the models developed.

The following sections develop a simple model of a low cost RFID system and a number of adversaries with various capabilities and goals.

10.4.1 Authorised and Legitimate

A discussion regarding security and privacy requires a clear understanding of the trusted parties. The term "authorised" will be used in relation to readers or interrogators who are registered in a given RFID system's database and are equipped with the necessary security mechanisms to access secure resources of that system. The term "legitimate" will be used in relation to tags that are registered in a given RFID system's database as verified by an authorised reader. Given the above concepts of authority and legitimacy, a reader that is not authorised will be referred to as an "unauthorised" reader while a tag that is not legitimate will be referred to as a "fraudulent" tag. In the event a legitimate tag is copied to produce a copy that may for all purposes has the ability to be verified as a legitimate tag will be called a "cloned" tag.

10.4.2 Tamper Proofing

The long-term security of label contents cannot be guaranteed since these contents are vulnerable to physical attacks. Tamper proofing to prevent physical attacks is an expensive option. Hence, it is assumed that labels cannot be trusted to store long-term secrets such as secret keys that apply to a range of RFID labels, but secrets pertinent to an individual label that is unrelated to another label are considered acceptable as long as the information obtained is unhelpful in defeating the security mechanism of another tag.

10.4.3 System Model

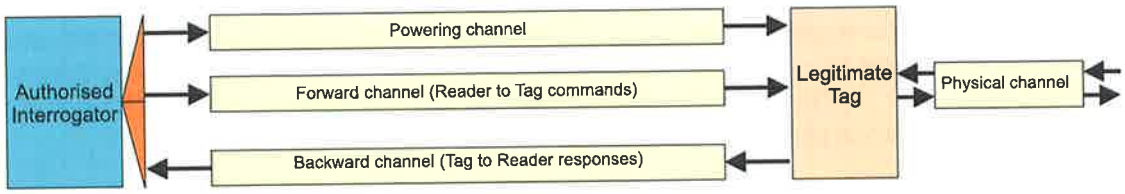

Figure 10.1 System model describing the information channels of an RFID system.

While RFID systems consists of various components a system model need only be concerned with interrogators, tags, and the communication channels between them. The problems involved with securing reader communications with a secure backend database will not be considered; instead, it is assumed that legitimate readers have secure connections to backend systems. The general notions of eavesdropping as described in Section 9.3.1 will be assumed. In such a context it is clear that that the forward channel (interrogator to reader) is exposed to undetectable eavesdropping from several hundred meters away, while the backward channel (label to interrogator) can only be monitored from close range (several meters away in case of UHF systems and a few centimetre in case of HF systems).

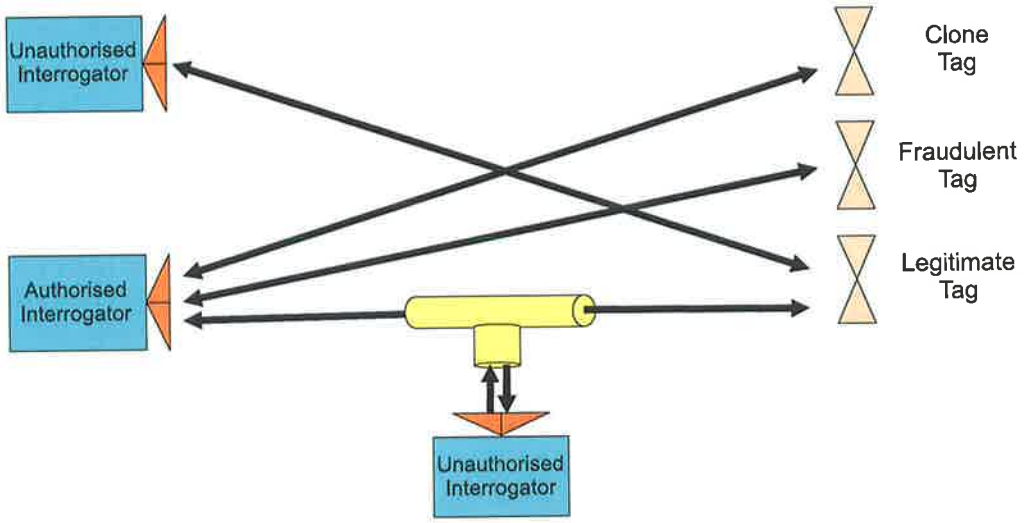

Figure 10.2 Possible interactions between communication participants.

Some implementations of readers may store security related information on board (providing an offline security mechanism), while other readers may simply obtain security related information from a backend system (as required in an online security mechanism). In either scenario a reader has adequate resources to secure the stored information or use a secure communication channel for obtaining the information.

Since other components of the system are not confronted with any constraints with regards to implementing necessary cryptosystems to secure information it is possible to consider an interrogator and the rest of the back end systems as a single entity, which can be referred to as an authorised interrogator. Now it is possible to outline the communication participants; authorised interrogators, legitimate tags, fraudulent tags, clone tags and unauthorised interrogators (refer Section 10.4.1). Figure 10.2 shows the possible ways in which the various communication participants can interact.

A communication model describing the various information channels is shown in Figure 10.1.The sources of information available to an adversary in a low cost RFID system are restricted to those which can be obtained over the insecure communication channels, the contents of the tag memory by way of the memory channel from tags which are constantly in an untrusting environment and from the interrogation of legitimate tags as shown in Figure 10.2. The eavesdropping distance of the forward channel, and the backward channel was discussed in Section 9.3.1. While the forward channel is generally longer than the backward channel, in a UHF RFID system, the backward channel is in the range of several metres and can be easily observed by a hand held reader, or an RF receiver with a higher sensitivity from a greater distance. Hence in a UHF RFID context it is not appropriate to consider the two channels separately while it is worthwhile considering the two channels separately in an HF RFID systems context, where a tag antenna must be inductively coupled at very close range for a reader to be able to decode the backward channel. Hence protocols based on considering the forward and backward channels separately in their security model cannot be expected to be secure for systems operating in the UHF region while the model may hold true for HF systems.

Although it is theoretically possible for an adversary to read or write to the forward channel, backward channel or the memory channel, in practice such duality is not possible. It is possible for an adversary to provide a powering channel and observe the powering channel as in the case of a power analysis attack. The adversary may read or write to the forward and or the backward channel, depending on the frequency of operation and the nature of the adversary being modelled, as in the case of a man-in-the-middle attack. The physical channel is considered to be read only once, as it may be used to read the memory contents of the tag, and such an operation is destructive to the legitimate tag. However, it may then be possible for the adversary to create a clone of the tag using the information obtained, but creating many clones of the same tag is considered to be not possible in supply chain applications due to the availability of a track and trace facility. However for a general application such as an access control system, such an assumption is not valid.

The complete set of data stored on a legitimate tag will be unique, and not dependent on other tags. This assumes that a tag's complete set of memory contents is unique (such as secret keys, unique tag identifiers and passwords) and it is not identical to any other tag legitimate tag.

10.4.4 Adversary Model

Threats exist to RFID systems because of the value of the object to which an RFID tag is attached or the assets that the RFID tag is attempting to protect and the value derived from correlating object identification information with human identification information. Who are the adversaries and what goals do these adversaries have? An RFID system deployed in a consumer goods supply chain application will surely have different adversaries from those in a supply chain application for the Department of Defence. An adversary looking at the system will only see the weaknesses outlined in Section 9.3 and will aim to exploit them to achieve various objectives: theft, defeating an access control system or mass production of counterfeit goods to illustrate a few. Such objectives are only limited by imagination and they depend on the application. However it is possible to have a simple classification of adversaries according to their objectives, level of interference, presence, and available resources.

10.4.5 Objectives of an Adversary

It is clear that an adversary model should be determined with respect to applications. However, achieving an adversary's objective will ultimately involve using various methods, such as cloning, or eavesdropping to defeat one of the security or privacy objectives outlined in Table 9.3 and Table 9.5.

For instance an adversary may violate confidentiality to obtain a tag's access password by eavesdropping on a conversation between a tag and a reader, or an adversary may collect tag data to associate object data with human data or use a physical attack to obtain secure information stored on a tag to create counterfeit products. If there is a gain to be made by disabling a system, the adversary may initiate a DoS attack (defeating the security objective of availability).

10.4.6 Level of Interference

It is possible to simply take the more traditional notion of a passive and an active adversary in an RFID context. A passive adversary is capable of eavesdropping on both the forward and the backward channel or just the forward channel (in case of HF systems) without being detected, then record and analyse the data in real time or at a later time. An active adversary may play a number of different roles.

A malicious adversary has no respect for electromagnetic compatibility regulations and has the ability to communicate with tags without being discovered by avoiding raising any suspicions but with a bound on the number of possible tag queries. The limitation on the number of tag queries comes from the assumption that tags will implement a throttling mechanism to prevent repeated tag queries from being executed in a brute force style attack.

A disruptive adversary can attempt to hinder the process of an interrogation by an authorised reader or the responses from a legitimate tag. Such an attempt will amount to a DoS attack. The adversary may also interrupt protocols, alter authorised reader commands or legitimate tag responses.

10.4.7 Presence

The presence of an adversary is an important defining aspect of the adversarial character. A stationary adversary may eavesdrop on RFID communications in a confined space such as a store, a warehouse or a room.

A mobile adversary seeks to clandestinely track the items. The adversary can maintain logs of all observed tag identifiers and locations while being able to interact with the tag using its standard air interface protocol. Such an attacker has the ability to use the recorded information to interact with legitimate readers to obtain tag related information. However a mobile or a local attacker may be a proximity adversary or a distant adversary.

A proximity adversary is able to record both the forward channel, the backward channel and utilize the powering channel while a distant adversary is only able to record the forward channel and utilize the powering channel.

10.4.8 Available Resources

It is necessary to assume various levels of capability of an adversary. Generally, these capabilities can be expressed as the level of resources available in the following categories.

- Financial
- Equipment
- Knowledge
- Time

The aim of research presented in this dissertation is to achieve a practical and adequate level of security by designing security measures to address various security and privacy goals to defend against realistic attackers, and thus attackers with unlimited resources such as highly funded government or non-governmental organisations are not considered. Equipment availability is another important consideration, especially when dealing with physical security of devices. Generally computing power is considered in the same category. Deciding on adequate key lengths based on current computing power is not difficult but it is difficult to estimate future computing power to forecast the length of time a key size or a cryptographic algorithm may be used. However, the general rule of thumb is that the

efficiency of computing equipment divided by price, increases by a factor of ten every five years [234]. This is only true of PCs (Personal Computers) [79].

A difference may be made with respect to the type of knowledge available to an adversary: outsider, insider, or a permanent insider. An outsider is generally not in possession of any special knowledge beyond that available, for instance by eavesdropping on messages between tags and readers. An insider however, is an adversary with access to additional information such as private keys or other securely stored information. An insider may be a permanent insider if the adversary has continual access to secure information.

10.5 Conclusion

Various solutions proposed for RFID systems have been discussed in Chapter 9. Prior to discussing security services, this chapter has formulated a simple, yet adequate, security model for evaluating security and privacy services proposed in Chapter 11. This chapter also developed an adversary model suitable for analysing the level of security provided by various services to be presented in Chapter 11. Using various characteristics of the adversaries outlined above it is possible to consider likely adversaries to a given RFID application.

The following chapter considers mechanisms for addressing the vulnerabilities discussed in Chapter 9 and presents designs of security services. These services are then analysed using the evaluation framework developed in this chapter to establish their merits for low cost RFID and the level of security provided.

Chapter 11

SECURITY AND PRIVACY BASED ON LIGHTWEIGHT CRYPTOGRAPHY

In the context of security and privacy, the most threatening (to privacy) and vulnerable (to insecurity) are the 'low cost RFID systems' as illustrated in Chapter 9. Various proposals made to address issues regarding information security and end-user privacy have been discussed in Chapter 9. However, some of these ideas are not applicable because they were based on a particular protocol that will soon be obsolete with the new generation of low cost RFID systems that are being developed and others are not practicable on account of their demand for circuit size and operational power while others fail to meet various security and privacy objectives adequately.

This chapter aims to propose a number of practicable solutions based on lightweight cryptography that address the security objectives outlined in Section 9.5 and privacy goals addressed in Section 9.6 and on the low cost RFID framework outlined in Section 9.2. The proposed solutions are then evaluated for their merits using the evaluation framework developed in Chapter 10.

11.1 Introduction

It is clear from the discussions in Chapter 9 that it is not possible to incorporate any encryption engine of significance on a Class I or a Class II label. It is in this view that the majority of proposals below will aim at removing complexity from the label to other proxy systems and limit any security related computation on the chip to simple operations.

The proposals below will also be considered on their merits based on meeting the performance metrics and the framework outlined in Table 10.1 while implementations of the mechanisms will be considered in view of the C1G2 air interface protocol.

Prior to proceeding with the proposals a note on notational aspects and an introduction to the hardware and operations used in the mechanisms outlined in this chapter is given in the following sections. The rest of the chapter will outline and evaluate various mechanisms for achieving the security and privacy objectives discussed in Chapter 9.

11.1.1 Notation

It is appropriate before proceeding any further to discuss a number of notational aspects to improve the clarity of the discussions below.

An encryption function performed using a key K will be indicated by the expression $e_K(<plaintext>)$ while a decryption function using the same key will be given by $d_K(<ciphertext>)$. If the pair of keys used is public and private they will be distinguished as $K_{private}$ and K_{public}. However a hash function operation on a string of *plaintext* using key K will in particular will be expressed as $hash_K(<plaintext>)$ and where a keystream is used, as is the case with a stream cipher, a unique sequence of the keystreams will be indicated using italic roman character such as *Ks* and different segments of the same keystream will be noted by appending a number to *Ks* such as *Ks1*, *Ks2* and *Ksn* for the nth segment. Two different keystreams will be distinguished by using a subscript, such as Ks_1 and Ks_2.

A signature scheme used to sign a message using the signing algorithm *sig* and key K will be given as $sig_K(<plaintext>)$ while the corresponding verification algorithm *ver* using key K will be given as $ver_K(sig_K(<plaintext>),<plaintext>)$.

The exclusive-or operator will be notated in diagrams and equations using the '$\oplus$' symbol throughout this chapter, while its usage in a sentence will be termed as XOR.

A random number or a nonce will be denoted by *RN* where there is a series of random numbers used it will be denoted as *RN1*, *RN2*, *RN3*, ... , *RNn*, while the notation $RN(i)$ will note the ith random number chosen or used.

The CRC (cycling redundancy check) of a number will be noted by preceding the number with the string *CRC_*. For instance the CRC of a random number *RN* will be denoted as *CRC_RN*.

In order to distinguish commands specified in the C1G2 protocol specification, or commands proposed for usage in a protocol, the usage of these commands will be in bold, italic, small capital, roman characters. For instance the command issued by a reader to write data to a tags memory will be expressed as ***WRITE***.

11.2 Related Work

The following sections provides a familiarisation with various lightweight hardware and concepts discussed in the security proposals, and also highlight the significance of these in the context of low cost RFID.

11.2.1 XOR Operation

Table 11.1 XOR properties.

$A \oplus A = 0$
$A \oplus \overline{A} = 1$
$A \oplus 0 = A$
$A \oplus 1 = \overline{A}$

The 'exclusive-or' operation (XOR) is a function that requires minimal hardware to implement. The XOR operator is both commutative and associative, and it satisfies the properties outlined in Table 11.1 for a Boolean variable identified by the symbol 'A'. The XOR operation will be used extensively in the following security proposals as a lightweight hardware for encrypting data.

11.2.2 CRC Generation

RFID tags falling into the Class I and II category include CRC generation hardware (refer to Figure 9.1) utilised for error detection. A number of proposals outlined below also use the CRC generator on a tag as part of a security engine. A CRC is a type of a hash function used on fixed length bit strings to generate a message digest. It is probably important here to stress the properties of CRCs.

Mathematically, an n-bit CRC is the remainder of a division operation generated by performing a division (modulo 2) of a message bit string of length m bits by a predefined bit stream of length n ($m > n$), where the predefined bit stream is defined by a generator polynomial of degree n. In order to guarantee the n consecutive bit error detection property of a CRC, the polynomial chosen must be primitive. While any primitive polynomial of the same degree will guarantee the error detecting ability, the polynomial chosen may not provide a good hash function (a discussion of choosing a primitive polynomial is discussed in detail in [223], from pages 130-132).

CRCs by themselves can not be used for data integrity checks because of the linearity of the division process which permits changes to the data in a message string, with relative ease, without altering its CRC. This can be easily illustrated by a property of a CRC given as its Hamming Distance (HD), which is the minimum possible number of bit errors that is undetected by a message's CRC. Hence if a CRC has a HD of 4 then there are no possible combinations of 1, 2, or 3 bit errors that will pass undetected by the CRC, however there is at least one possible combination of CRC with 4 bits, such that the bit errors will not be detected by the CRC check.

CRCs are also not collision free. The collision rate of an n bit CRC is the probability that two messages will have the same CRC. The theoretical collision rate of an n bit CRC can be given as $1/2^n$. While collisions may be a problem in error detection, they can be seen as an advantage in a security scheme as an attacker may be faced with the reconstruction of the message given only its n bit CRC.

An n bit CRC generator consists of n shift register with XOR gates that form a linear feed back shift register with an input for the message to be checked at the first stage of the register. While it is possible to use additional gates to reconfigure the CRC to produce a LFSR for generating pseudorandom numbers as part of a security mechanism, this provides no added security due to the weaknesses of LFSRs discussed in Section 9.9.3. However it may be possible to use LFSRs or NLFSRs in a context where the initialisation key and the connection polynomial are kept secret. The security of a LFSR scheme is also increased if the encryption scheme can not be subjected to a known plain text attack. However generally it is almost always assumed that the encryption scheme can be subjected to a known plaintext attack. Nevertheless, low cost RFID provides a unique environment in which such an assumption is not always valid.

11.2.3 Stream Ciphers

As suggested in [160] there are four essential approaches to stream cipher design listed below.

1. Information-theoretic approach
2. System-theoretic approach
3. Complexity-theoretic approach

4. Randomised approach

Keystream generators have proliferated as a result of the information-theoretic security possible with one time pads. However, this is impractical in most implementations due to the problems associated with large purely random key sizes, key generation and distribution. Keystream generators provide a scalable means of generating a keystream that is of the same size as the plaintext. However contrary to one time pads the key is not purely random as the sequences are pseudorandom, and hence key stream generators do not provide the same perfect secrecy as one-time pads. A detailed analysis of LFSRs and desirable properties of stream ciphers based on LFSRs can be found in [78] and [81]. Prior to a discussion on LFSRs it is important to define the term linear complexity which is an important concept in the study of stream ciphers [163].

Definition 11.1. Assume all sequences are binary sequences where s denotes an infinite sequence with terms $s_0, s_1,...$; s^n refers to a finite sequence of length n. Then the linear complexity of a finite binary sequence s^n denoted $L(s^n)$ is the length of the shortest possible LFSR that can generate the sequence s^n as the first n terms.

Since no mathematical proof of security can be found for feedback shift register based keystream generators, system theoretical designs based on established guidelines and testable security properties have been developed. Developments over time have led to a number of desirable criteria for keystream generators outlined in [78], [157], [158], [159] and [160] listed below.

- Large period before the keystream repeats itself.
- Large linear complexity.
- Good statistical properties so that the pseudorandom sequence satisfies statistical tests for randomness.
- Confusion (so that the output keystream bits are some complex transform of all or most of the bits of the key stream).
- Diffusion.
- Meet nonlinearity criteria to provide correlation immunity.

While the above design criteria are not proven to provide a secure keystream generator, the criteria above have been proven to be sufficient or necessary for security. There are various stream ciphers designed around these criteria based on non linear feedback shift registers, linear combination generators (the use of several LFSRs to build a single stream cipher), nonlinear filter generators and clock controlled generators which eliminate the linearity properties of the LFSRs. The shrinking generator is such a stream cipher where the generator can be implemented using simple shift registers and yet is secure provided that it is implemented prudently (discussed in more detail in Section 11.2.3.2).

More secure stream ciphers are built using a complex-theoretic approach (or number theoretic approach) where breaking the generator is based around a NP-hard problem. Generally these generators are more complex, slow and result in greater silicon cost for implementation [79]. However, the knapsack generator is an exception where the generator can be implemented using simple shift registers. The latter generator is discussed in more detail in Section 11.2.3.3.

The final approach to designing stream ciphers is based on the randomised approach where the aim is to assure security by ensuring that breaking the cipher requires an adversary to perform an impractical amount of work. The security of such a cipher can be evaluated as the average number of bits an adversary has to examine before improving his probability of ascertaining the key by way of random guesses. An example of such a cipher can be found in [161]. Generally these schemes require very large random bit sequences of the order of 10^{20} and thus require massive communication overheads, as well as computations [79].

An examination of the stream ciphers available in literature summarised in [78] and [79] reveals that there is a large body of stream ciphers based on feed back shift registers despite the availability of a variety of different theoretical techniques and methods for constructing stream ciphers.

11.2.3.1 Linear Feed Back Shift Registers

Feedback shift registers form the basic building block of a number of keystream generators. Figure 11.1 gives a schematic of a LFSR with L registers and a connection polynomial defined by the coefficients in $\boldsymbol{c} = \{c_1, c_2, \ldots, c_L\}$. LFSRs consist of a set of shift registers whose input is computed based on a linear recursion of the current state of the shift registers as shown in (11.1) where s is the current state of the shift register i at time t and $c_i, s_i \in$ GF(2).

$$s_{L-1}(t+1) = \sum_{i=0}^{L-1} c_{L-i} s_i(t) \tag{11.1}$$

LFSRs form the backbone of many stream ciphers or keystream generators. They provide a number of important advantages outlined below.

- Maximum length LFSRs can produce pseudo-random sequences (sequences with statistical properties satisfying various randomness postulates such a Golomb's randomness postulate) [78].
- LFSR can be easily implemented in hardware.
- Maximal length LFSRs are capable of generating sequences of length 2^{L-1} where L is the length of the LFSR [78].
- Since the generators are linear they can be easily analysed.

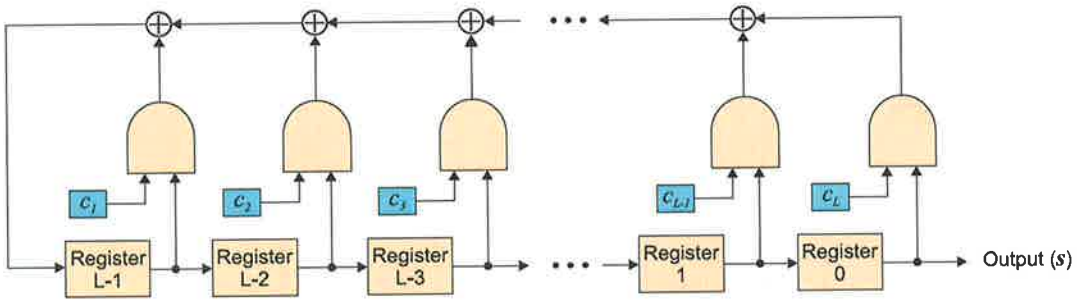

Figure 11.1 Schematic of a LFSR of length L.

However the output bit string of LFSRs are not secure even if the feedback scheme is not known [78]. It only takes a substring t of length $2L$ consecutive bits of an output sequence s from a LFSR to calculate the feedback scheme of the generator (commonly defined as a connection polynomial) using the Berlekamp-Massy algorithm of time order complexity $O(n^2)$ [78]. Once the connection polynomial is determined the LFSR obtained can be initialised with any substring of t with length L and used to generate the remaining bits of the sequence s.

Then the remaining question is regarding procurement of the substring t. This requires an adversary attacking the generator using a known plaintext or a chosen plaintext attack. Thus if the adversary knows the plaintext sequence m and the corresponding ciphertext sequence c, the corresponding keystream k can be obtained as $m \oplus c$. Hence using LFSR in RFID will demand in addition to the LFSR being maximum length, that an adversary is not given the opportunity to perform a chosen plaintext attack or a known plaintext. The latter may be a difficult task even in a well designed system; however it is possible to prevent a known plaintext attack by a passive adversary or an active adversary who is not capable of a physical attack, in an RFID context. The mechanisms presented in this chapter illustrate the latter remark.

Based on a system-theoretic approach the most common practice of making LFSRs secure is to use a nonlinear Boolean function to generate nonlinearity in the output or the irregular clocking of LFSRs. Two generators based on the previous ideas and suitable for RFID applications are considered in Section 11.2.3.3 and Section 11.2.3.4 below. Unfortunately nothing can be proven regarding the level of security they provide. However these generator have survived much public scrutiny and they can be concluded to be computationally secure (refer to Table 9.8). Nevertheless implementation of LFSRs must be done with care. There are a number of practical guidelines that should be followed to avoid stream ciphers based on LFSRs falling to the prey of adversaries. These guidelines are discussed below.

11.2.3.2 Implementation Considerations

In the choice of a connection polynomial it is important to choose a primitive polynomial (irreducible polynomial) that will lead to a large number of feedbacks

connections or in other words a primitive connection polynomial of the same degree as the length of the LFSR where only a number of the coefficients are zero. This will make attacking the stream cipher more difficult [78] while ensuring that the resulting LFSR is maximum length.

Definition 11.2. An nth degree polynomial $f(x)$ is irreducible if no polynomial of degree k, where $0 < k < n$, divides f(x). There are a total of $(2^n-1)/n$ primitive polynomials for a LFSR of length n.

The secret key is used to initialise the LFSR provide the initial state of the registers. However it is possible to have secret connections that can be initialised using a secret key that consists of the connection polynomial coefficients and the initial state of the LFSR registers. Keeping the connection polynomial secret provides greater security [78] but with the added cost of extra hardware for implementation.

An extensive coverage of LFSR based generators can be found in [161]. The nonlinear filter generator and the clock controlled generator are two such generators based on LFSRs and are discussed below.

11.2.3.3 Nonlinear Filter Generators

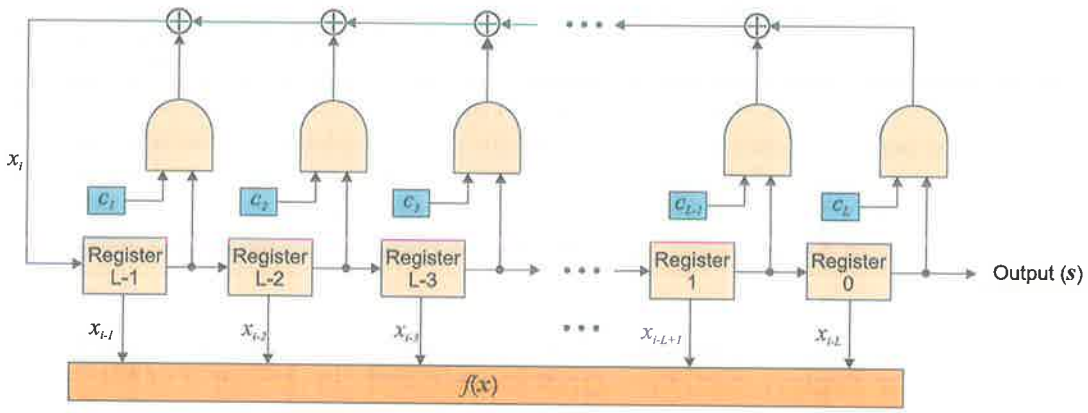

Figure 11.2 Nonlinear filter generator based stream cipher.

Figure 11.2 shows a general structure of a nonlinear filter generator as given in [78] where the function f is a nonlinear Boolean function called the filtering function that operate on the vector $\boldsymbol{x}$ consisting of the outputs from the registers at various stages. The stream cipher generates a keystream based on some nonlinear Boolean combination of the output of various stages of the LFSR.

A stream cipher suitable for RFID applications can be found in a nonlinear filter generator called the knapsack generator. While its classification may be questioned as the security of the generator is based on a NP-hard problem, it is nevertheless based around a LFSR and a function f which acts on $\boldsymbol{x} = \{x_1, x_2, \ldots, x_L\}$ which are the outputs of each stage of the shift

register and the key stream consist of selected bits from f. The knapsack generator is discussed in more detail below.

Knapsack Generator

A key stream generator based on the summation of a set of weights selected based on the register values of a LFSR to generate an integer sum S is called the knapsack generator in respect of the subset sum problem which involves the determination of a subset of weights which when added together equals a given integer S. Provided that such a sub set exists the problem is proved to be NP-hard [78].

Integer addition over GF(2) is non linear as shown by Rueppel in [164]. The knapsack generator based on mapping the state of a LFSR at time t to a knapsack sum S_t is also non-linear. The key stream is generated by calculating the knapsack sum S_j as given in (11.2) at time j by stepping the LFSR forward by one step, where $[x_L,\ldots, x_2, x_1]$ is the state of the LFSR registers at time j and $[a_1,\ldots, a_L]$ are the integer weights each of bit length L. Here $Q = 2^L$ where L is the length of the LFSR. The generator is illustrated in Figure 11.3.

$$S_j = \sum_{i=1}^{L} x_i a_i \bmod Q \tag{11.2}$$

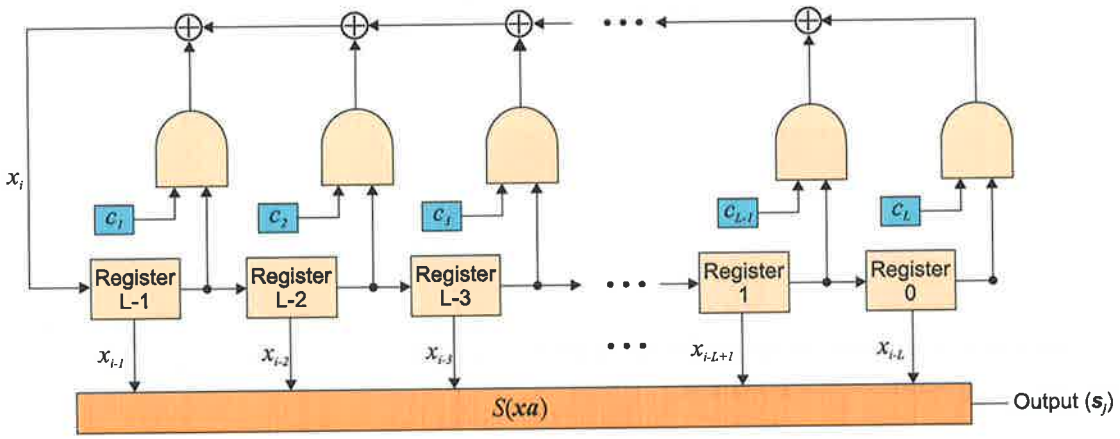

Figure 11.3 Schematic of a knapsack generator.

While the knapsack generator has been extensively analysed by Rueppel [159] there are no published weaknesses of the knapsack generator available in literature. Unfortunately there are also no concrete suggestions on selecting the length of the LFSR or the knapsack weights.

11.2.3.4 Clock Controlled Generator

In general the registers in a LFSR are all clocked using the same clock signal, thus all the registers are clocked at regular times, and the contents of the registers are updated at each rise or fall of a clock signal. However, in clock controlled generators the idea is to use

a combination of LFSRs so that the output of one LFSR controls the clocking of a second LFSR. This introduces nonlinearity as the second LFSR is clocked irregularly and hence the stream cipher designer hopes to defeat attacks based on the regular clocking of LFSRs. The shrinking generator proposed in [162], described below is an example of a clock controlled generator that is suitable for implementation on an RFID label.

Shrinking Generator

This generator is a more recent proposal outlined in [162] that utilizes two LFSRs, *R1* of length *L1* and *R2* of length *L2* clocked in parallel. At any given transition of the clock the output of the generator is that of *R1* given that the output of *R2* is a logical one. If *R2* output is a logical zero then nothing is output from the generator. Hence the output from *R1* is shrunk to produce an irregularly decimated subsequence [78] as shown in Figure 11.4.

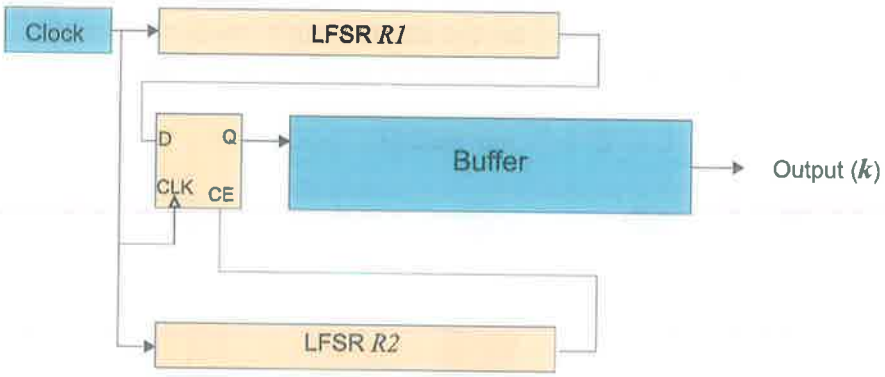

Figure 11.4 Configuration of a shrinking generator.

An outline of the properties of the shrinking generator can be found in [78]. The security of the generator has survived many known attacks on LFSR based systems, especially due to the very long period of the generator.

Attempts at breaking the shrinking generator have been based on an exhaustive search using a divide and conquer attack [166], a correlation attack [167 and 168], a distinguishing attack appropriate for when *R2* has a sparse polynomial [169] and more recently using linear hybrid cellular automata to characterise the shrinking generator provided the connection polynomials, the lengths *L1* and *L2*, and as many keystream bits as the linear complexity of the shrinking generator (that is $2^{L2}.L1$ bits)[170] are known. The order of complexity of known attacks has exponential time complexity as they are a function of the length of both LFSRs. The complexity of known attacks are summarised in [78 and 162].

Despite all of the above attacks, shrinking generators are still considered resistant against efficient cryptanalysis attacks due to the difficulty of the attack scenarios and the time order complexity of the algorithms. It should however be stated here that for maximum security the following implementation consideration should be satisfied.

- Use secret connection polynomials that are not sparse
- Use maximum length LFSRs for *R1* and *R2*
- The lengths of LFSR should be such that $\gcd(L_1,L_2) = 1$

One draw back of this generator is its irregular output but this may be solved by buffering the key stream prior to its use. In [171] an estimate of buffer size calculated using a Markov chain to model the buffer has shown that the probability of not having a byte of keystream data available is very small even when a buffer size of only several bits is used. Generally running *R1* and *R2* at twice the throughput required, using a small buffer size of 16 bits yields a probability of 5×10^{-3} of not having a byte of keystream data available [162]. The probability of not having a byte of keystream data decreases exponentially with buffer size and the rate at which *R1* and *R2* are running with respect to the required throughput [171].

11.2.3.5 Power Consumption

Power consumption of CMOS digital circuits was discussed in Section 9.8.1 and estimating power dissipation in modern CMOS digitial circuits was presented in Section 10.3.3. Equation (10.1) provides a method of estimating the power dissipated by considering the average capacitance switched per clock cycle. Hence the power consumption of a LFSR based stream cipher is directly dependent on the frequency of operation, the supply voltage and the physical parameters of the gates (equivalent output capacitance). The L shift operations and XOR operations performed during each clock cycle results in a significant high switching activity and hence a considerable power dissipation. The estimation of switching activity is investigated in [188] and [187] while [189] gives power estimation of a number of stream ciphers including the shrinking generator. A power consumption of 65 μW was estimated for a shrinking generator of length 65 bits with programmable connection polynomials [189]. This is comparable in magnitude to the power requirements for writing to the EEPROM memory of a passive tag.

Considering (10.1) it is clear that reducing the equivalent output gate capacitance will reduce the power dissipation but this is a fabrication process based parameter that is difficult for a digital circuit designer to control. However there are low power CMOS digital circuit design techniques that can be implemented to reduce the power consumption of LFSRs. There are two major principles used in low-power digital circuit design: supply voltage scaling and minimizing average switched capacitance per clock cycle [190 and 191]. These techniques are briefly discussed below while a detailed coverage of the topic can be found in [110] and [190]. Also it is clear from (10.1) that the most effective method of reducing power dissipation is by reducing the supply voltage as the resultant reduction is proportional to the square of the supply voltage while the relationship between average switched capacitance is linear.

Supply Voltage Scaling

It is clear from (10.1) that the power dissipated scales with the square of the supply voltage. Therefore a simple method of reducing power dissipation is to reduce the supply voltage. An unfortunate consequence of reducing the supply voltage is the resultant increase in the propagation delay, T_d of a gate given in (11.3) as a simple first order derivation which ignores the nonlinear characteristics of a CMOS gate [190].

$$T_d = \frac{kV_{dd}}{(V_{dd} - V_t)^2} \tag{11.3}$$

In (11.3), k is a parameter dependent on the size of the CMOS transistor and the fabrication process, V_t is the threshold voltage and V_{dd} is the supply voltage. Hence the supply voltage can only be reduced until the delay along the critical path (the slowest path) of the circuit is the same as the clock period required to achieve a given throughput from a stream cipher. Further power reductions may be obtained by minimising the delay of the critical path by using pipelining techniques albeit at the cost of extra hardware. The critical path of LFSRs are generally limited by the XOR summation operations [189], however for stream ciphers based on LFSRs the critical path may also be the clock control circuit as in the threshold generator or the output combining function as in the case of a non linear filter generator.

Reducing the Switched Capacitance

Minimizing the average switched capacitance per clock cycle achieved by lowering the switching activity by reducing the number of logic transitions per clock cycle also reduces the power dissipated. Therefore circuits should be clocked at the slowest rate possible and as seldom as possible to achieve the required functionality.

The average switched capacitance of LFSR circuits may be reduced by the choice of LFSR architecture, logic design style, layout style and the used of gated clocks [190 and 191]. The LFSR architecture must be designed to minimise the possibility of unnecessary transitions. For instance, a XOR tree may be used in a LFSR instead of a chain of XOR gates to reduce the number of state transitions. In Figure 11.5 the output of the final XOR in the chained OXR structure can change three times before a valid output is available where as in the tree architecture of the XOR gates, the final output will change only once while the XOR gate outputs propagate up the tree. Similarly portions of the circuit that are not in use may be powered down to attain the same goal of minimising unnecessary transitions.

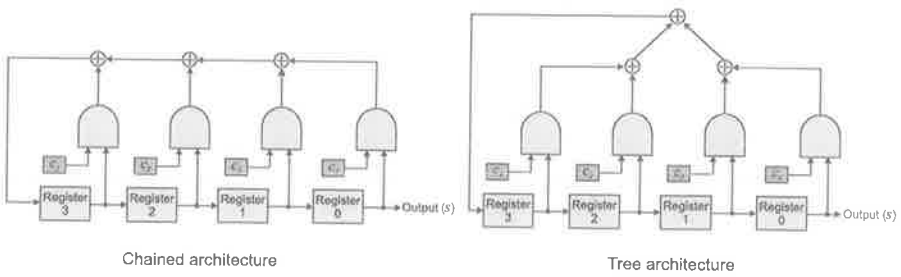

Figure 11.5 Comparing a chained XOR architecture and a tree based XOR architecture.

As discussed previously reducing the clock rate also reduces the average switched capacitance. This can be achieved by using parallelised LFSR hardware [190 and 191] but this is accomplished at an additional cost and therefore is not a suitable method for reducing the power consumption of stream ciphers for low cost RFID.

Finally gated clocks to prevent the clocking of portions of logic that are not immediately being used can reduce the average switched capacitance of the stream cipher. This can be done with the knapsack generator where the LFSR can be halted from being clocked while the adder circuit computes the integer sum of the knapsack weights.

11.2.4 Physically Uncloneable Functions

Attacks such as micro-probing, laser cutting, glitch attacks and power analysis attacks along with reverse engineering techniques [89] used to reconstruct the layout of circuits have enabled adversaries to extract digital keys stored in the memory of integrated circuits. Security systems based on keeping a key secret have thus been broken as a result.

While various tamper-proofing methods have been developed over the years to counter such physical attacks they might be considered to be a costly solution for RFID applications. Such an example is the tamper sensing technology [165]. Using a sensor based on additional metallization layers allows interruptions and short circuits to be detected in the event of an attempt to tamper with the IC. However, such sensors only work while the IC is powered and such a sensor technology can only cause a degree of difficulty to an adversary attempting to obtain the key while the IC is powered as the key can still be extracted when the IC is powered off.

Alternatives to storing keys on insecure hardware devices have been developed. Such an alternative is the introduction of physical one-way functions (POWFs) in [172 and 173]. The solution presented used a laser beam as an input to a transparent optical medium with 3D microstructures and the output was a quantification of the resulting interference pattern. The resultant output is dependent on the frequency and the angle of the laser beam entering the optical medium and the optical characteristics of the medium.

The concept of using physical uncloneable functions (PUFs) is published in [174] and is a result of the early work on POWFs. Below is a general definition of a PUF.

Definition 11.3. A Physical Uncloneable Function (PUF) maps a set of challenge inputs to a set of responses utilizing some physical characteristic incorporated in an object. A PUF should also satisfy properties listed below.

1. *Easy to compute*: The time taken for generating the response set should be acceptable or be computable in polynomial time.
2. *Difficult to model*: The amount of information that can be obtained about the response to a randomly chosen challenge by an attacker without access to the physical device based on a polynomial number of measurements conducted

previously using only a polynomial amount of resources, such as time, is negligible [175].

The ability to construct a PUF on silicon was outlined in [174], [175] and [176]. A PUF structure that can be easily fabricated into an IC using standard CMOS fabrication processes has far reaching consequences. Below is a definition of an integrable physically uncloneable function.

Definition 11.4. An Integrable Physical Uncloneable Function (IPUF) is a PUF as identified in Definition 11.2 and also satisfying the following properties.

1. *Inseparability*: An IPUF is integrated and fabricated as part of an ASIC design such that any physical attack would lead to the destruction of the IPUF.

2. *Secure communication*: It is not possible to tamper with the measurement data from an IPUF.

The idea is based on using process variations, which are beyond a manufacturer's control, in wires and transistors on an IC to obtain a characteristic response from each IC when given a certain input. The IPUF circuit is able to uniquely characterise each IC due to manufacturing variations [174]. These individual characteristics then become similar to the secret keys used in a symmetrical encryption scheme. Thus, it is possible to identify and authenticate each IC reliably by observing the IPUF response. The observation of IPUF results reveal that a string of challenge bit sequences can be used to generate a response string unique to each IC.

The particular advantage in this technique lies in the fact that an adversary cannot construct a model or a device to clone an IPUF as there can be a number of possible challenge-response pairs, exponentially dependant on the number of bits in a challenge. Hence the system has computational security because a model based on an exhaustive search is impractical. However, the IPUF based structure in [174] is sensitive to noise, especially thermal noise, as wire latencies and gate delays depend on the operating temperature of the device. This leads to reliability issues when trying to obtain consistent responses for a given challenge.

Unreliability due to such environmental variations have been addressed in an IPUF configuration given in [175] and [176], wherein a challenge response pair is created using an IPUF circuit based on a differential topology, using only 100s of gates. The design of such an IPUF is considered in the following section where the word PUF will always refer to an IPUF.

11.2.4.1 Circuit Implementation

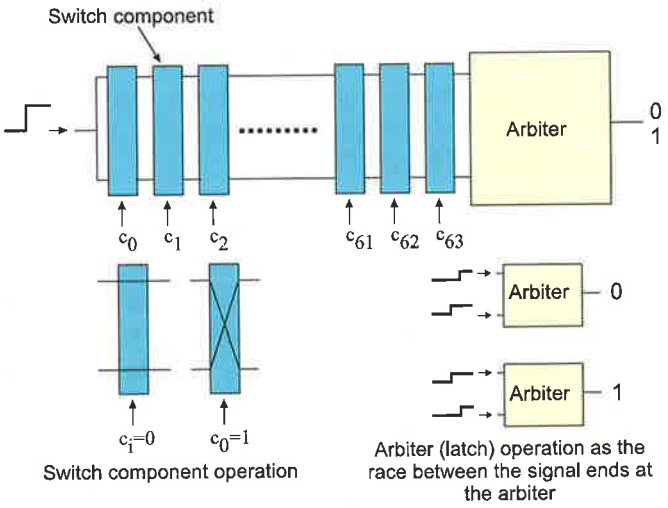

Figure 11.6 Arbiter-based PUF circuit implementation [178].

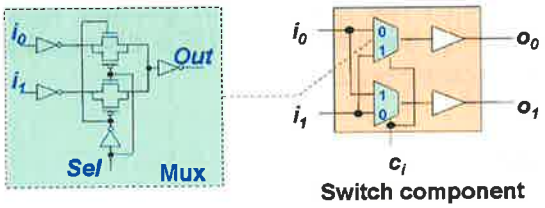

Figure 11.7 Switch component implemented using two-to-one multiplexers to swap two delay paths [178].

The block diagram in Figure 11.6 depicts the structure of a PUF circuit which is based on the arbiter-based PUF in [175] and [176]. The circuit accepts a n bit challenge b_0, b_2, b_3, ..., b_n to form two delay paths in 2^n different configurations. In order to generate a response bit, two delay paths are excited simultaneously to allow the transitions to race against each other. The arbiter block at the end of the delay paths determines which rising edge arrives first and sets its output to 0 or 1. The actual implementation of arbiter-based PUFs in [175] and [176] uses 64 bit challenges. The details of the switch component are given in Figure 11.7.

The switch component indicated in Figure 11.6 is implemented using a pair of two-to-one multiplexers (refer to Figure 11.7). Depending on the select bit C_i, the switch either allows the signal to travel straight through or it swaps the delay paths. The arbiter is constructed using a simple transparent latch with an active-low enable input. The arbiter favours the path to output zero since it is preset to zero and requires a setup time constraint to switch to a logic one. Fixing a small number of most significant challenge bits can compensate for this

skew by effectively lengthening one delay path. The layout was carefully done to ensure that both paths are symmetrical and arbiter responses are not biased to a binary 0 or a 1.

The chip used in testing was built in TSMC's 0.18 μm, single poly, 6-level metal process with standard cells [175]. The chip contains eight sets of the arbiter-based PUF circuits capable of generating an 8 bit response for a given challenge and a JTAG-like serial interface for communication. The total area of the eight PUF circuits is 1212 μm x 1212 μm and the chip can be operated at 100 MHz [175 and 176].

Manufacturers always attempt to control process variations to a great degree, however, these variations are largely beyond their control and hence it is not possible for an adversary to fabricate identical PUF circuits. It is estimated in [175] that there is a strong enough variation between two chips fabricated from the same silicon wafer for a sufficient number of random challenges to identify billions of chips. The probability that the first measured response bits to a given challenge (set of bits) on a chip is different from the measured response for the same set of bits (challenge) on a different chip is estimate to be 23% to 40% depending on the PUF circuit architecture [175]. A symmetric layout will increase this probability to 50%, and subsequent FPGA implementations have increased this probability further. It has been estimated that about 800 challenge response pairs are sufficient to distinguish 10^9 chips with the probability $p \sim 1 - 5\times10^{10}$ [175].

A PUF structure on an RFID IC can provide a suitable solution to create a low cost security engine that is both cost effective and is capable of guarding against tampering to extract secret keys stored on the tag. The usefulness of a PUF is discussed in Section 11.3.1.2 in the context of a challenge-and-response protocol.

11.3 Authentication

Authentication was discussed in Chapter 9. The following sections provide a number of security services for providing authentication in a low cost RFID system, limited by constraints identified in Section 9.7.

The goal of an authentication scheme in RFID is to prevent an adversary from creating a clone of a tag to misrepresent the legitimate tag (and hence the authenticity of the object associated with the tag) by a carefully planned attack on the RFID system (refer to Section 9.3.2 for details on cloning). In supply chain applications authenticating the tag will allow the prevention and detection of counterfeit goods.

There are a number of possible ways in which a low cost RFID system may be attacked to obtain the necessary information to clone a tag. Present systems based on Class I and Class II tags only require passive eavesdropping or a scanning of an RFID tag to carry out a cloning attack (refer to and Section 9.3.1).

11.3.1.1 Challenge-and-Response Protocols

Practically all identification schemes or authentication schemes use a challenge and response protocol to provide strong authentication. Other identification schemes such as the Schnorr Identification Scheme [177] and the Okamoto Identification Scheme [178] are examples of more complex challenge and response mechanisms. The mechanism for authentication using challenge-response protocol in an RFID context is described in Figure 11.8.

1. Reader choses a challenge, x, which is a random number and transmits it to the reader.
2. The label computes $y = e_K(x)$ and transmits the value y to the reader (here e is the encryption rule that is publicly known and k is a secret key know only to the reader and the particular label).
3. The reader then computes $y' = e_K(x)$.
4. Then the reader verifies that $y' = y$.

Figure 11.8 Challenge-response protocol.

In the context of an RFID system, where there is no secure channel for communication, the security of the mechanism relies on the secure storage of the key K and the inability of an adversary to compute the key K given both the ciphertext and the plaintext.

11.3.1.2 Constructing a Challenge-and-Response Protocol

It is possible to construct a challenge-and-response protocol using a variety of cryptographic tools. Most symmetric key encryption algorithms, such as AES, are suitable candidates. However, in terms of silicon they present expensive solutions, while at the same time the security provided by such schemes remains vulnerable to various invasive and non-invasive physical attacks (refer to Section 9.3.6). Security systems based on keeping a key a secret on devices left in untrusted environments have thus been broken as a result. However, the concept of using physical uncloneable functions (PUFs) discussed in Section 11.2.4 can be used to securely store a key on an RFID IC. The following sections will look at leveraging a PUF to construct a challenge and response protocol.

11.3.1.3 Tag Authentication

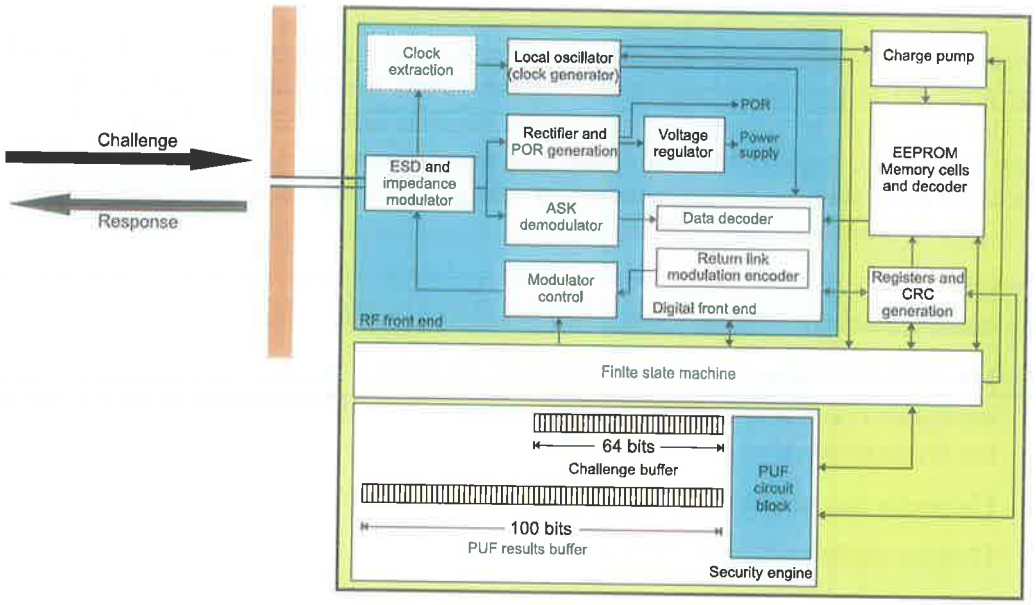

Figure 11.9 A model of a passive RFID chip with an integrated PUF for authentication.

Figure 11.9 depicts a model of an RFID chip with an integrated PUF circuit while Figure 11.10 illustrates the use of a PUF based RFID system. The discussion below using PUF security engines will assume using 800 challenge-response pairs (CRPs) as a sufficient number of challenges in a single set to uniquely identify around a billion chips.

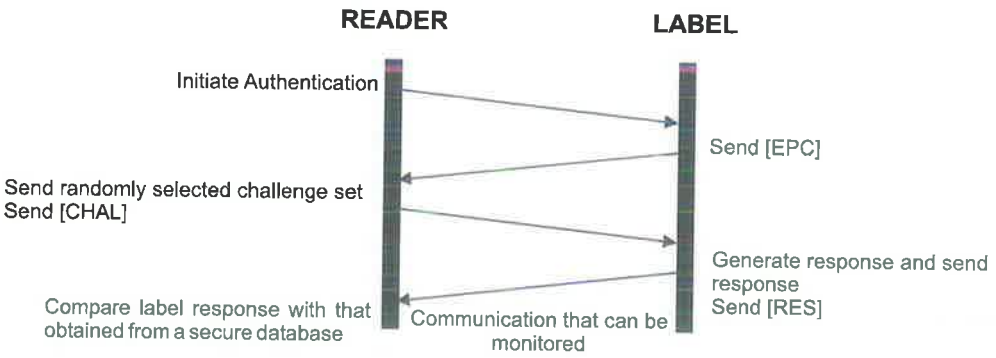

Figure 11.10 Message exchange between a reader and an RFID label during an authentication process

Building a symmetric key engine is still not a cost effective solution and, although certain advances have been made towards the development of hardware optimised encryption engines in [131], [134] and [135], they still present a performance hindrance to current RFID systems. Hence instead of using a PUF to obtain a secret key, a PUF can be directly utilised as illustrated in Figure 11.10.

It is clear that once a challenge has been used it cannot be used again since an adversary may have observed it. However it is possible to have a list of CRPs (challenge-responses pair) or use an encrypted communication link to deliver challenge and obtain the responses. Then there remains the question of delivering a secure communication channel between a reader and the tag. A possible mechanism for obtaining a communication layer encryption scheme is given in Section 11.4.1. However, not all the challenges need to be discarded the following method allows the limited reuse of CRP pairs.

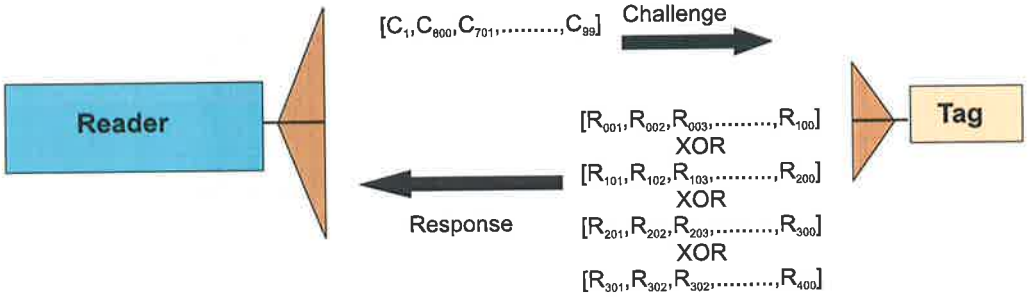

Figure 11.11 Using randomised challenges and XORed responses to allow the limited re-use of challenges.

A simple alternative to the mechanism discussed above requires the reader to randomly alter the order in which challenge set $[C_1, C_2,\ldots,C_{800}]$ is sent to the tag. The tag is then required to store the response, *RES* = $[R_{001},\ldots, R_{800}]$ in a buffer (involves buffering the 800 bit long response). The *RES* to the challenge set is then subdivided into four different blocks, *RES1*, *RES2*, *RES3* and *RES4* of length 100 bits each. The resulting blocks are then XORed together as *RES1* XOR *RES2* XOR *RES3* XOR *RES4* as illustrated in Figure 11.11. Thus a third party observing the communication between a tag and reader is unable to formulate the correct challenge response pairs. This random organisation of the challenge string will allow the challenges to be reused even over an insecure communication channel as shown in Figure 11.11.

It is not possible to use the CRPs indefinitely without recharging PUF responses periodically in a secure location as an adversary collecting the challenge and XORed response pairs can construct a set of equations and solve the Boolean satisfiability problem to discover the *RES* vector with minimal effort.

The above scheme will allow a label to authenticate itself to a reader before any sensitive information passes between the devices, but the fact remains that a reader still needs to identify a tag by requesting its unique identifier (such as the EPC in case of Class I tag implemented using the C1G2 protocol). The scheme also implies that the RFID tags be characterised with a number of challenge response sets. Thus in a supply chain environment a manufacturer might have to perform individual tag characterisations using randomly selected challenges in a secure environment such as a Faraday's cage.

11.3.1.4 Tag and Reader Authentication (Mutual Authentication)

It is possible to extend the above scheme to enable a tag to authenticate a reader and for a reader to authenticate a tag. This involves sending a randomly selected challenge set for which a tag generates a response string. The reply string will be used to authenticate the tag while the reader is authenticated using a one time pad that gets updated at the end of the session.

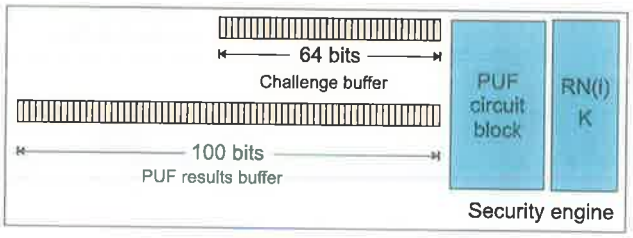

Figure 11.12 RFID label with a PUF and additional memory for storing a secret key and *RN(i)*.

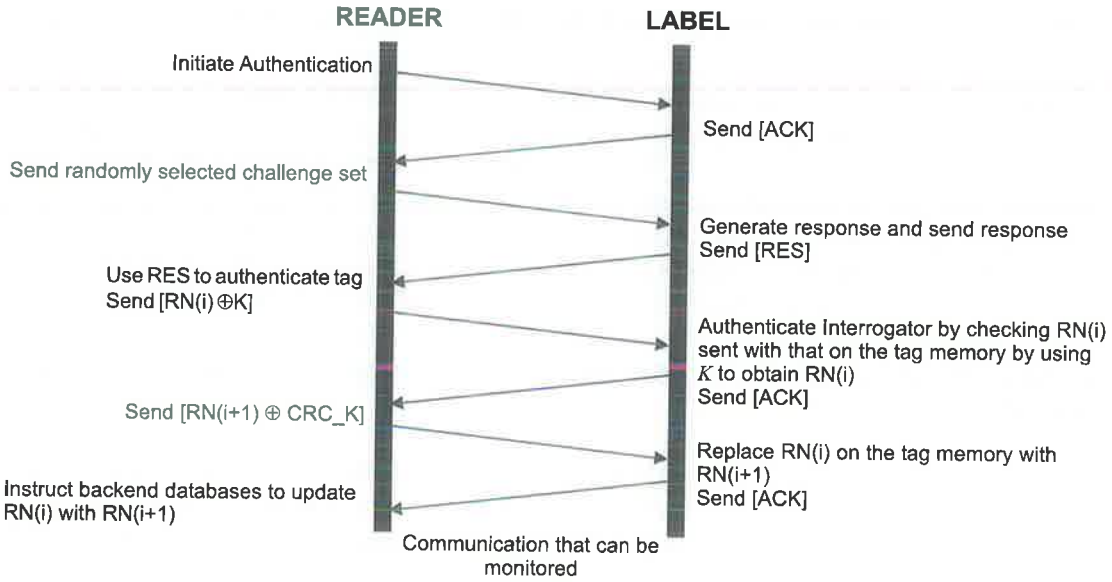

Figure 11.13 Protocol for mutual authentication using a PUF.

This scheme requires that the tag stores a one time pad *RN(i)* and a secret key *K* (refer to Figure 11.12) for encryption and that the reader has access to the tag related information stored on a secure database. The message exchange protocol is outlined in Figure 11.13. The key *K* is unique to each tag and is stored in a tag's memory at the time of manufacture. It is possible to construct the key *K* so as to provide product authentication by replacing the key *K* with an EPAC (Electronic Product Authentication Code), either encrypted or not, based on measurable or observational product specific features or feature, so that a third party can verify the EPAC independently to establish the authenticity of the product. Such a

mechanism is a possible extension that can be adapted to the current protocol. A detailed description of the use of EPACs are described in Section 11.6.1.

Once the tag is identified using its EPC, a reader can request the tag related data set $[RN(i) \oplus K,\ RN(i+1) \oplus CRC\_K]$ from a secure database to complete the authentication process. The reader does not require any information regarding the key K, the number $RN(i)$ or the number $RN(i+1)$ in the entire process. The use of a CRC of the key K and the number RN is achieved at no additional cost, since CRC operation can be performed by the CRC block of a tag (refer to Figure 9.1). The use of the CRC generation hardware extends the usefulness of K and RN, especially K from what will otherwise be a 'use once only' number requiring an extended protocol with greater overheads to refresh two tag related keys. Thus, the CRC generator effectively aids the creation of a lightweight protocol.

The above scheme will require that at least a sufficiently large (about 800 challenges) *CHAL* set be used to be able to effectively identify billions of chips. However measurement noise may cause errors in the response vector and readers will need to access a sequence of redundant information along with a CRP to be able to correct the *RES* vector prior to authenticating a tag. Alternatively a reader can transmit the *RES* vector to the secure database where it can be corrected for errors.

11.3.1.5 Hash Based Tag Authentication

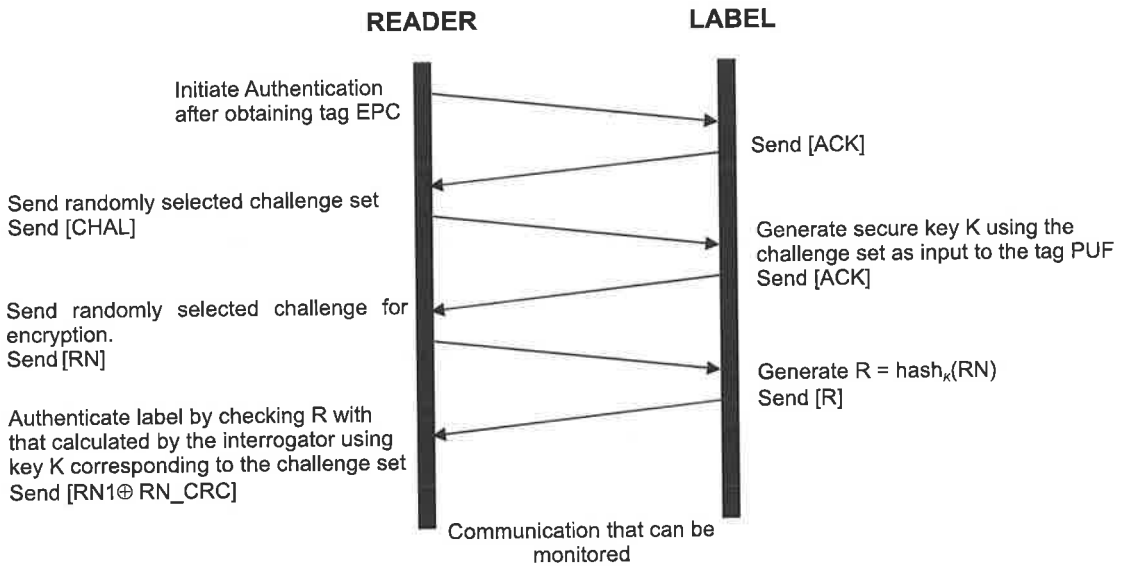

Figure 11.14 Challenge-response protocol using a hash function.

While the nature of a PUF is ideal for the secure generation of a private key, without having to store the key in the memory of a device in an untrusting environment, its direct application in low cost RFID is limited by the fact that implementing a hash function or another asymmetric encryption engine requires valuable silicon area. Until the optimised

cost of hardware implementations of symmetric key cryptosystems becomes a reality such an application context is not possible. However, Figure 11.14 illustrates a tag authentication scenario when the physical implementation of a hash function achieves the critical cost effectiveness required for low cost RFID. Currently there are a number of candidates for such a hash function as outlined in Section 9.9.9.1.

11.3.1.6 Evaluation

The schemes outlined above based around a PUF is evaluated using the matrix outlined in Table 10.2 and the results are summarised in Table 11.2. Outcomes summed up in Table 11.2 do not consider the network delays associated with accessing backend databases as such problems may be solved by using adequate bandwidths, and caching tag related data onsite. Clearly the memory space required to store at least 128 challenges incurs a far higher gate cost than that suitable for low cost RFID. Thus, if the number of challenges required is in excess of 128, the challenges almost always need to be transmitted. The transmission delay bottlenecks will then reduce the performance of the system and limit the number of tags that can be authenticated on average to around 2.5 tags per second.

Table 11.2 Evaluation of authentication mechanisms.

Achieved Security Objectives	Tag authentication
	Reader authentication
Cost and Performance	PUF block: 856 gates Buffers (for storing 164 bits): 246 gates *If the CHAL set is to be stored in memory* Memory cost of storing 128, 64 bit challenges: 12,288 gates
	Scheme requires online resources such as access to secure backend databases containing tag profiles.
	Tags need to under go a verification phase prior to deployment to generate an adequate number of CRPs for each tag.
	Transmitting 800, 64 bit challenges remain the bottleneck. The transmission from an interrogator to a tag will take 400 milliseconds at the highest possible speed. Thus using 800 challenges will only allow on average the authentication of around 2.5 tags per second. However, it is possible to use perhaps a 128 challenges instead of 800 as the tag EPC will be enough to uniquely identify the tag. The reduced overhead will allow the authentication, on average, of about 16 tags per second.
	No power consumption estimates are available but it is not expected to be greater than that required for writing to EEPROM memory.

The need to transmit a large number of challenges from the reader to the tag remains the primary obstacle, especially given that that the maximum possible transmissions speed from a reader to a tag possible in the C1G2 specification is about 126 kbps given equi-probable ones and zeros. However, higher interchip variations of around 50% [228] discussed in Section 11.2.4 and the fact that the EPC can be used for unique identification suggest that for the scheme outlined in Section 11.3.1.4 and Section 11.3.1.5 a smaller challenge set (of around 128 challenges to generate a 128 bit response) may be used. Using a smaller challenge set will significantly increase the number of tags that can be authenticated to over 15 tags per second.

11.3.1.7 Removing Barriers to Performance

The primary performance obstacle in the tag authentication and mutual authentication protocol presented previously is the excessive overhead of transmitting a large number of challenges, where each challenge consisted of 64 bits. There are several methods by which performance of the above protocols can be improved.

Instead of choosing to transmit the challenges, it is possible to use a linear feed back shift register (LFSR) to generate the challenges once initialized with a seed. Then only the seed to the LFSR needs to be sent to a tag as a challenge *C*, from a reader. This arrangement is outlined in Figure 11.15. The additional hardware of a LFSR will allow the protocols in Figure 11.10, Figure 11.13 and Figure 11.14 to be executed with greater efficiency.

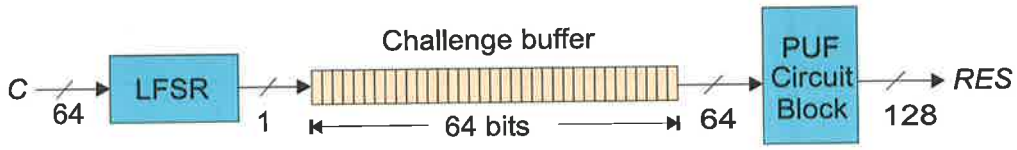

Figure 11.15 Improved security engine design of the lightweight primitive to reduce overhead.

11.3.1.8 Evaluating the Improved Performance

The schemes outlined above can be evaluated to consider the overall performance of PUF based authentication mechanism with the alteration to the security engine discussed in Figure 11.15. The evaluation of the proposed method is detailed in Table 11.3. Unlike the authentication schemes outlined previously, where a large number of challenges need to sent to the tag, the current transmission requirement is that of a single challenge C, from which other challenges are derived.

Table 11.3 Evaluation of improved authentication mechanisms.

Achieved Security Objectives	Tag authentication
	Reader authentication
Cost and Performance	PUF block: 856 gates 64 bit LSFR: 768 gates Buffer (for storing 64 bits): 768 gates Total cost of hardware: 2392 gates
	Scheme requires online resources such as access to secure backend databases containing tag profiles.
	Tags need to under go a verification phase prior to deployment to generate an adequate number of CRPs for each tag.
	Transmitting *C*, a 64 bit challenge, from an interrogator to a tag will take about 0.5 milliseconds at the highest possible speed. The reduced overhead will allow the authentication, on average, of about 2000 tags per second, ignoring the delay on tag for evaluating the PUF responses for the 128 challenges generated on the tag. The mutual authentication protocol execution time will be slightly greater due the requirement for transmitting random numbers. The transmissions from the reader will then take about 1.5 milliseconds, provided RN(i) used is 64 bits in length. This will still allow over 600 tags to under go the mutual authentication process.
	No power consumption estimates are available but it is not expected to be greater than that required for writing to EEPROM memory.

Using a smaller challenge set as well as the incorporation of a LFSR has allowed the lightweight primitive to be implemented in a low cost RFID tag while meeting requirements of both performance and cost.

11.3.1.9 Addressing Reliability Issues

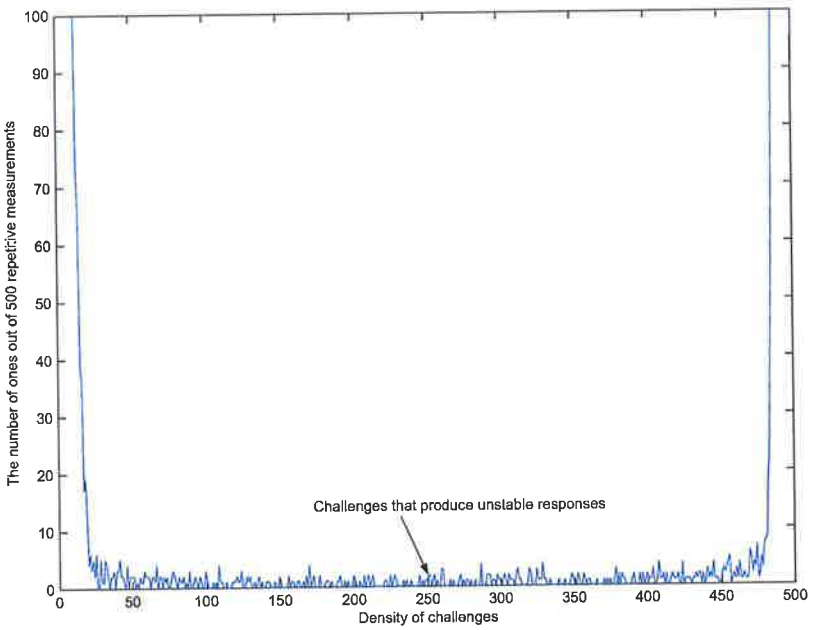

Figure 11.16 The density function of the random variable k, where k is the number of 1's out of 500 repetitive measurements.

As it has been mentioned previously in Section 11.2.4, the responses from a PUF are sensitive to environmental conditions such as temperature and power supply voltage [175]. Variation in environmental conditions exhibits itself as measurement noise in the *RES* output leading to inconsistent responses. In an alternative statement, a challenge set that generates a particular response vector *RES* may not generate the same *RES* vector if environmental conditions change beyond a tolerance level. This issue is highlighted in Figure 11.16 where results of 500 repetitive measurements of 1000 challenges are shown.

The design in [175] and illustrated in Figure 11.6 is used to mitigate the effects of environmental noise (especially temperature changes) based on the fact that noise would affect both signal propagation paths in an identical manner and thus the final results of the circuit are unchanged. Since the circuit measures the relative delay difference, the PUF is robust against environmental variations. For realistic changes in temperature from 20°C to 70°C and regulated voltage changes of ±2%, the output noise is 4.8% and 3.7%, respectively. Even when increasing the temperature by 100°C and varying the voltage by 33%, the PUF output noise still remains below 9%. This variation is significantly less than the inter-chip variation of 23% (or 40%), allowing for the reliable identification and authentication of individual chips [228].

The problems caused by operational voltage changes can be minimised by the fabrication of a voltage regulator on the PUF and authentication process can be refused if the PUF is

inadequately powered. Another simple solution to overcome measurement noise is to calculate and transmit redundant information extracted from a *RES* output along with the *CHAL* set to ensure that the response obtained can be evaluated and corrected for errors. Improving the reliability of PUF systems is an active area of research.

The use of redundant information to improve the performance can be easily demonstrated by using an error correcting code to encode the measured PUF responses in the verification phase. This will allow errors in the PUF responses received from tags as a result of measurement noise (discussed in above) to be corrected as far as that possible with the chosen error correcting code. Noise is a randomly distributed variable, thus it makes sense to use an error correcting code that is best suited for correcting random errors. BCH code named after its discoverers (Bose, Ray-Chaudhuri, Hocquenghem) is a possible choice for the purpose. BCH codes are a generalization of the Hamming codes.

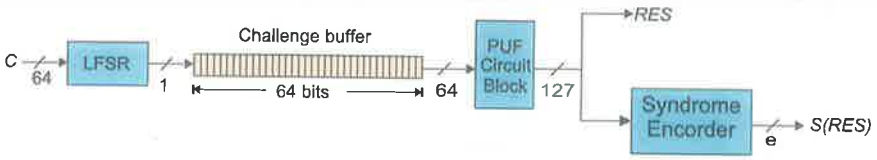

Figure 11.17 Encoding of PUF responses using BCH codes at the verification phase.

For example, Figure 11.17 illustrates the use of a BCH(n, k, d) code where n bits of the PUF output is used to calculate a syndrome *S(RES)* of e bits. The syndrome is evaluated during the verification phase and stored securely for use in one of the authentication protocols discussed previously.

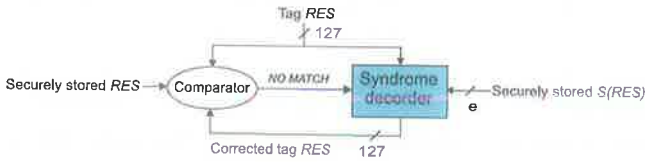

Figure 11.18 Comparison of the received response from tag. If the RES received from the tag is not a match an attempt is made to correct to a challenge at the interrogator in the authentication protocols.

The use of the *S(RES)* at the interrogator is illustrated in Figure 11.18. Upon receipt of the *RES* vector from a tag, a reader compares that to *RES* vector evaluated during the verification phase. A failure in the comparison stage results in the attempt to correct any measurement noise using the stored syndromes. The corrected *RES* vector is again evaluated to ensure the comparison failure most likely did not arise from measurement noise but from a fraudulent tag or a malicious user attempting to clone a tag.

Measurements of noise in PUF circuits have shown a bit error rate of around 5%. Thus we can expect about 7 random errors in a 127 bit PUF response vector. However the BCH code parameters used has an ability to correct up to 15 errors. This is more than double the bit error rate expected from a PUF response and is adequate to ensure the reliability of the authentication process.

11.3.1.10 Practical Issues

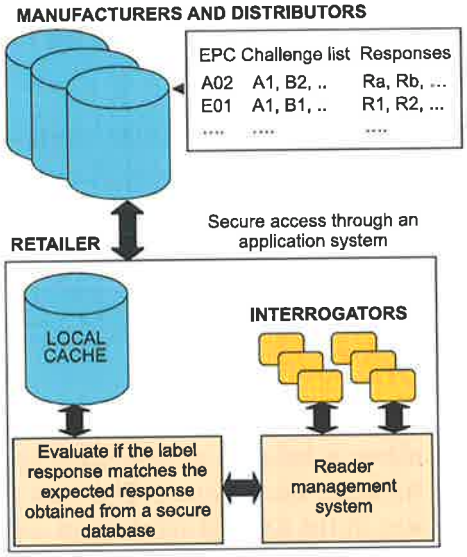

Figure 11.19 An overview of an implementation of a PUF based RFID system.

Figure 11.19 gives an overview of the system implementation required. Clearly a practical implementation will require that tag IC manufacturers or product manufacturers in the supply chain initially obtain and securely store tag specific sets of challenge response pairs. This will be an expensive task but it can be automated similarly to the test and verification of ICs. While the mechanism for authentication appears simple, issues regarding the secure storage of challenge sets, and the corresponding responses along with a method for distributing the challenge sets and responses to parties in a the supply chain need to be addressed. There are various existing solutions to such issues and they are not considered in this dissertation.

What has not been addressed thus far is the compatibility of the mechanism with Class I and Class II labels. The schemes described above require the transmission of at least 800 challenges. This comprises of 51,200 bits of information that needs to be sent to a tag. There is no command in the current specification of C1G2 to transfer such a quantity of data, and at a maximum data rate of 128 kbps it would take more than 400 ms to transfer a complete challenge set to the tag. While it is possible to read tags at the rates of 100 to 200 tags per second it will not be possible to read and authenticate tags within that time frame. However since the authentication procedure does not interfere with the tag query process the above schemes will meet the performance metrics outlined in Table 10.1.

The current version of C1G2 protocol does not have commands that could be used to implement the mechanism outlined above however the specification does allow proprietary commands to be defined. Implementing the tag authentication mechanism and the mutual authentication scheme would require the definition of at least two proprietary commands after initiating an authentication: ***WRITECHL*** (instructs the tag of the number of ***WRITECHL*** commands to expect and transmits a challenge to the tag), ***READRES*** (instructs the tag to

scroll out the response to the challenges). In order to prevent the challenge lists from being discarded in the event of sudden power loss, it is important for the tag to transfer the results from the PUF output buffer to the tag memory and only to transmit back the response once the entire challenge sequence is transmitted and the reader has requested the response using a ***READRES*** command.

In addition to these commands, the mutual authentication mechanism in Section 11.3.1.4 will require a further command; ***SENDENC(<flag>,<data>)***. The reader command ***SENDENC*** will be used to send encrypted *RN(i)* to the tag, where the *flag* value will be used to denote the authentication step. Implementing these additional commands will increase IC cost, albeit being only a small cost increase.

11.3.1.11 Possible Attacks

The above scheme will allow a label to authenticate itself to a reader before any sensitive information passes between the devices. The mechanism relies on a PUF to securely store a unique secret key in the form of fabrication variations. The security of the systems based on PUFs will depend on the difficulty of replicating a PUF circuit and on the difficulty of modelling the PUF circuit successfully. This is not a simple process and is therefore an adequate deterrent depending on the value of the article being authenticated by the reader. Details of probable attacks on a PUF based system are outlined in [175].

11.3.1.12 Conclusion

The PUF provides a cost effective solution to low cost RFID Systems. This security engine can be easily constructed using standard digital gates and layout tools and fabricated using standard CMOS technology. A 64-stage PUF circuit costs less than 1000 gates. Additionally, various kinds of low power techniques such as sub-threshold logic design and multi-threshold CMOS design can be utilised to reduce the power consumption to make it suitable for use in devices sensitive to low power consumption.

The effects of environmental conditions on the measurements obtained from a PUF have been studied but future work should involve the investigations into the effects of voltage on the performance of the PUF. Future work is also required to investigate the effects of the generator throughput to study the time taken for the execution of a challenge response protocol.

Future work should also focus on the performance issues related to the large amount of data that needs to be transmitted to the tag and from the reader, and the time taken in memory storage and retrieval will also need to be investigated to further refine the analysis of performance issues related to using a PUF security engine on current Class I or Class II tags.

11.4 Confidentiality and Authentication

While authentication was discussed in the previous section the following sections will consider mechanism for providing a service to deliver confidentiality to low cost RFID systems. This is a difficult task as the tags and readers communicate over an insecure communication channel, where tags can not be trusted to store secret information and strong cryptographic mechanisms can not be used. Achieving confidentiality involves creating a secure communication channel in an untrusting environment over an insecure channel. This is not a novel problem (refer to [78, 79, 80 and 163]) but the solution space has little offerings for a minimal resource intensive environment.

11.4.1 Secure Forward Link

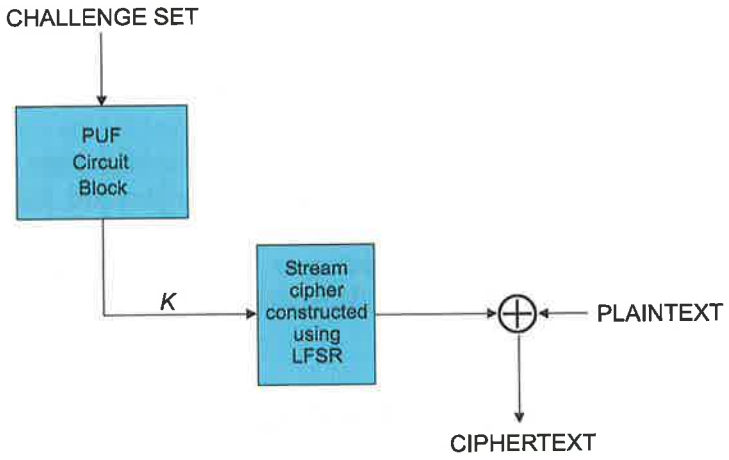

Figure 11.20 Tag implementation of a stream cipher performing an encryption operation.

Achieving a secure communication channel between a tag and an interrogator using a PUF for the secure storage of a secret key and a stream cipher as a fast and efficient source of a key stream is discussed below.

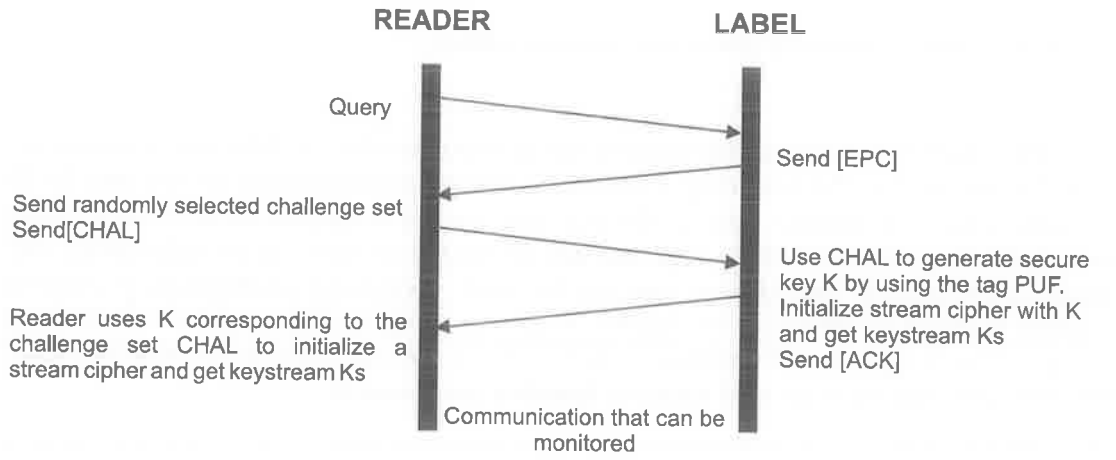

Figure 11.21 Communication protocol for achieving a secure communication channel.

The method assumes that an interrogator has uniquely identified the tag using its EPC and obtained a *CHAL* list from a secure database corresponding to the tag EPC. The interrogator then initiates the establishment of a secure channel by signalling to the tag and once the tag acknowledges the interrogators request, the interrogator transmits the *CHAL* list. The tag then uses the *CHAL* list to generate a key *K* which is used to initialise the stream cipher and to obtain a keystream *Ks* which can be used to encrypt the forward link as illustrated in Figure 11.20. The interrogator is able to generate the same keystream for decryption as the reader can obtain the secret key *K* corresponding to the *CHAL* set as the tag's PUF was characterised using the *CHAL* list prior to its deployment. An outline of the communication protocol is given in Figure 11.21.

11.4.2 Tag and Reader Authentication (Mutual Authentication)

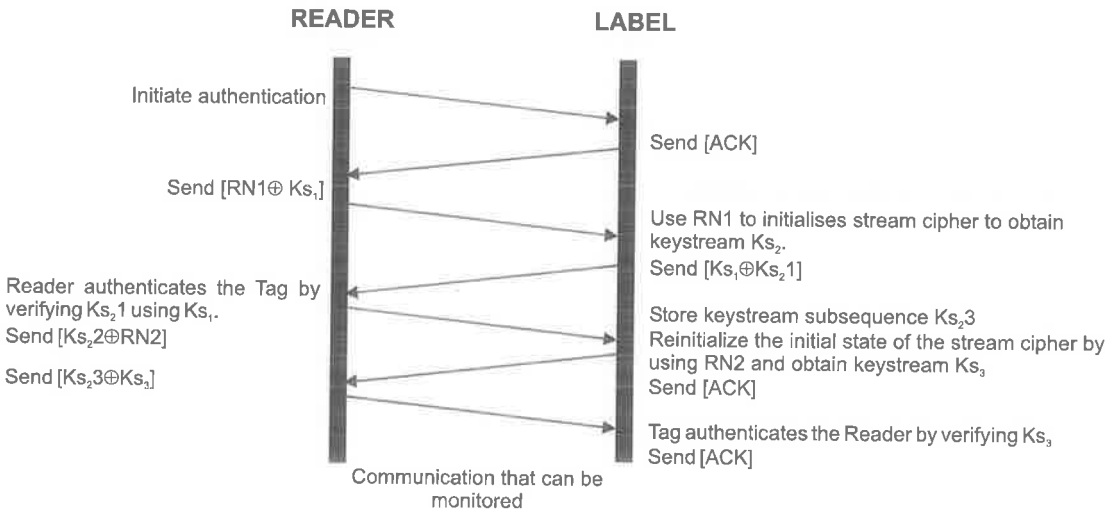

Figure 11.22 Protocol for mutual authentication after establishing a secure communication channel.

The Section 11.4.1 above discussed a mechanism for achieving confidentiality by allowing the establishment of a secure channel. However, a simple extension to the communication protocol using the existing hardware required for achieving a secure channel can provide mutual authentication of a tag and a reader.

The mutual authentication algorithm is based on using the stream cipher as a means of generating a one time pad. In the method outlined in Figure 11.22 an interrogator initiates the authentication after the establishment of a secure communication channel. Once the tag acknowledges the request, the interrogator sends a randomly generated number *RN1* encrypted using the keystream *Ks* generated previously (indicated as Ks_1 in Figure 11.22). A legitimate tag is able to obtain *RN1* and then use it to reinitialize the stream cipher to generate the new keystream Ks_2. In order to make the scheme more efficient the tag uses the previously generated keystream bits Ks_1 and encrypts them with an identical length sequence of bits Ks_21 from the keystream Ks_2 generated from the reinitialised stream cipher. The interrogator is then able to authenticate the tag since only a legitimate tag would have been able to produce the keystream Ks_1 to produce Ks_2.

If the tag is legitimate, the reader transmits an encrypted random number *RN2* to the tag. In the event the tag is not authentic the reader will pretend to complete the authentication by transmitting a nonce to the tag. Thus an adversary who does not have access to the reader output has no indication of whether the authentication attempt was successful.

The tag then reinitialises the stream cipher using *RN2* and sends an acknowledgement to the reader. The reader uses the previously generated keystream Ks_2 to obtain the bit sequence Ks_23 and encrypts it with the keystream Ks_3 generated from the reinitialised stream cipher to enable the tag to authenticate the reader based on the shared secret key generated using the stream cipher.

11.4.3 Evaluation

The two schemes outlined above can be evaluated together, and the results are detailed in Table 11.4. Using the knapsack generator as a stream cipher provides the most cost effective solution with a total gate cost of 2416. However, performance may still be a problem and increasing performance by storing the challenges in memory will cost an additional 6144 gates. Thus the total security engine will cost at least 8560 gates. This is still far less than that required for a heavyweight encryption scheme, delivered at a much greater performance benchmark (312 mutual authentications per second).

Table 11.4 Security mechanism evaluation.

Achieved Security Objectives	Confidentiality
	Tag authentication
	Reader authentication
Cost and Performance	PUF block: 856 gates *Stream cipher* Shrinking generator with 64 bit LFSRs and a 16 bit buffer: 1730 gates. Knapsack generator with 64 bit LFSRs and a 64 bit adder (1 HA + 63 FA) = 1560 gates. Memory cost of storing 64, 64 bit challenges: 6144 gates.
	Requires online access to secure database with *CHAL* lists or offline authentication may be performed using a local cache.
	Requires creating profiles of chips prior to deployment.
	Considering that the challenges are stored in memory Initialising the PUF will take an amount of time dictated by the tag memory access time for 64, 64 bit *CHALs*. Time to complete the transmission process using 128 bit random numbers: 3.2 ms, hence 312 tags per second can be authenticated on average. *Considering that the challenges are not stored in memory* Initializing the PUF:32 ms. Time to complete the transmission process using 128 bit random numbers: 3.2 ms. Hence 28 tags per second can be authenticated on average. *Using the stream cipher* There will be a small delay in initialising the stream ciphers but this will be in the time order of 10s of clock cycles and can be ignored. *Using a knapsack generator* On average 32, 64 bit additions need to occur prior to producing a 64 bit keystream and provided the adders are clocked with the worst case timing constraints in mind, 32 additions will take 62 clock cycles where the clock cycle length will be based on the worst case carry propagation path. However such a delay is only experienced at the initialising stage as the 64 bit keystream will buffer the time taken for the next output from the knapsack generator. Even if a 4 MHz tag oscillator is assumed the additions take negligible time and the bottle neck is still the data transmission times.
	Power consumption is not expected to be greater than that required for writing to EEPROM memory. Refer to Section 11.2.3.5 for a detailed discussion on LFSR power consumption.

11.4.4 Practical Issues

The first instance use of LFSR is in the form of a synchronous stream cipher as the key stream is generated by initialising the LFSR with the response from the PUF and thus has the advantage that a single bit error will only cause a single bit of the plaintext to be corrupted after decryption. However, subsequent use of RNs transmitted from the tag to initialize the stream cipher may cause a failure in the authentication process due to undetected errors in the transmission from the reader.

There is the added overhead of requesting and obtaining the *CHAL* list from a secure database and transmitting the challenges to the RFID IC. This process could be avoided by storing the *CHAL* list on the tag, as the discovering the *CHAL* list by way of a physical attack can not reveal the response a PUF circuit on an IC but as shown in Table 11.4 this is an expensive option.

The mechanism outlined above will require a number of proprietary commands as permitted by the C1G2 protocol. ***WRITECHL*** is one such command described in Section 11.3.1.10. However, the mutual authentication scheme will require further two commands: ***SENDENC(<flag>,<data>)*** and ***RESENC(<data>)***. The reader command ***SENDENC*** will be used to send encrypted text to the tag, where the *flag* value will be used to denote the authentication step. The ***RESENC*** command will be used by the tag to respond with the encrypted ciphertext to the reader. If either party detects a fraudulent tag or an unauthorised reader, the authentication process will be continued by the legitimate party, but the encrypted data will then be replaced by irrelevant bit sequences.

11.4.5 Possible Attacks

Similarly to the methods outlined in Section 11.3, the vulnerabilities and weaknesses of the PUF circuit also applies to the schemes described here and they are discussed in Section 11.3.1.11. In addition the vulnerabilities of the stream ciphers discussed in Section 11.2.3.3 and Section 11.2.3.4 also apply to the generators. However, it should be noted that as a result of the protocol used, the stream ciphers can not be subjected to a known plaintext attack, and attacks based on a known plaintext attack can be disregarded. As discussed previously, the existing vulnerabilities of these generators can be prevented with careful implementation.

11.4.6 Conclusions

The combination of a PUF circuit block with a stream cipher has created a practicable and a powerful solution capable of delivering both an authentication service and an encrypted communication channel. The mechanism is suitable for both Class I and Class II tags, especially Class II tags requiring an encrypted communication channel.

The performance of the knapsack generator may be increased by using a carry save adder instead of the ripple carry adder, but this will only be achieved at almost double the cost (in terms of the gate equivalent cost) of the ripple carry adder.

11.5 Anonymity and Untraceability

An RFID label implementing the C1G2 protocol will scroll out its EPC when queried after being singulated by any transceiver implementing the C1G2 air interface protocol. This unique identity carried by the RFID label poses various security threats and privacy violations illuminated in Section 9.3. Thus, it is important to control access to a label's EPC, or to allow an RFID label to respond with a non-identifying response as a way of concealing its unique identity to unauthorised readers. It is possible to avoid sending a tag's EPC which allows the tag, the associated object and perhaps even the person in possession with the tag to be identified. The following sections will consider two methods to achieve anonymity and untraceability.

11.5.1 Pseudonyms

The schemes presented in this section are based on the idea of a tag using randomly changing pseudonyms during the identification process as a way of addressing the privacy concerns outlined in Section 9.3.7. All pseudonym schemes generally rely on two means for changing the identifier on a tag. The mechanisms are based on where the pseudonym is generated. There can be two types of pseudonym updates; reader or backend database generated pseudonym updates or tag generated pseudonym updates. The security mechanisms presented in the sections below based on re-encryption and randomly varying object identifiers use simple lightweight protocols for pseudonym updates generated by backend databases or readers.

11.5.2 Re-encryption

The idea of re-encrypting was discussed in Section 9.9.8. Despite the resource limitations of an RFID label, it is possible to allow a low cost RFID tag to become a party to a computationally intensive security mechanism provided that the designer is able to transfer the computationally demanding aspects to a backend system, such as the reader itself, or a backend network of computers which may act as a proxy to the security mechanism. In essence it is a transfer of complexity out of the chip onto a shared resource with greater capability to execute a computationally intensive task in a timely manner. While this will essentially reduce the ability to perform offline operations it does allow tag cost to be kept to a minimum. Clearly pushing complexity further up the EPC Network reduces the ability to execute security services offline; an unfortunate compromise that must be made in view of tag cost constraints.

The following security mechanism is based in the idea of transferring complexity out of the tag silicon, the ideas behind re-encryption and a method used for establishing a secure communication channel outlined in Section 11.4. Re-encryption offers a novel perspective on achieving the privacy objectives of anonymity and location privacy (refer to Section 9.6). The scheme is described below.

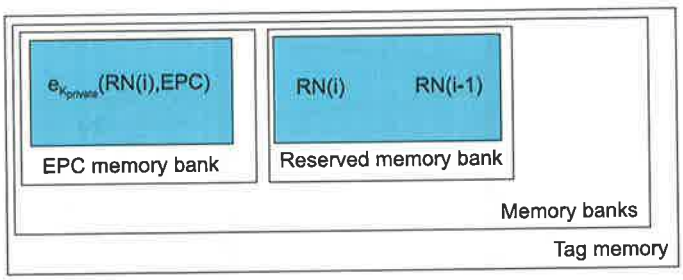

Figure 11.23 Tag memory contents in a typical implementation of a Class I tag.

Instead of storing the EPC in a write once format, it is possible, for instance, for a retailer in the supply chain to store an encrypted version of a tag EPC XORed with a nonce or a random number *RN(i)* on the tag where the EPC is now [EPC⊕*RN(i)*]. The XORing of the EPC ensures that the encrypted EPC appears different each time the encrypted version of the EPC is updated on the tag. However for simplicity of the following discussion the [EPC⊕*RN(i)*] operation is assumed to be implicit in the mention of an EPC.

The scheme also requires the storage and the transmission of a random number *RN(i)* so that the actual EPC can be obtained from [EPC⊕*RN(i)*] and an initial random number *RN(i-1)* on the tag as illustrated in Figure 11.23. The primary storage of the *RN*s needs to be performed in a secure environment such as an electromagnetically shielded room or a Faraday cage. Then it is possible for the tag owners to use their own private keys to encrypt the tag identifiers prior to their use within a facility such as at a supermarket store.

The encryption may be performed using a secret key that is known solely to the retailer or the person having ownership of the tags. Thus, when a label is requested to scroll out its EPC, it will scroll out an encrypted version of the EPC, which to a third party will appear as a stream of random bits. The protocol is illustrated in Figure 11.24.

The encrypted EPC alone will prevent profiling as the information obtained by eavesdropping does not reveal any information regarding the object to which the tag is attached as an adversary is unable to obtain the EPC without the private key $K_{private}$. However, the label still emits a predictable response and thus untraceability is not achieved.

On identification of the label, the reader has the option of storing back an encrypted version of the EPC padded with a new random number *RN(i+1)* used to perform the [EPC⊕*RN(i+1)*] operation. The protocol is outlined in Figure 11.24. Hence authorised readers may randomly, or on every occasion the tag is read, update the encrypted version of the EPC by storing a new encrypted version of the EPC.

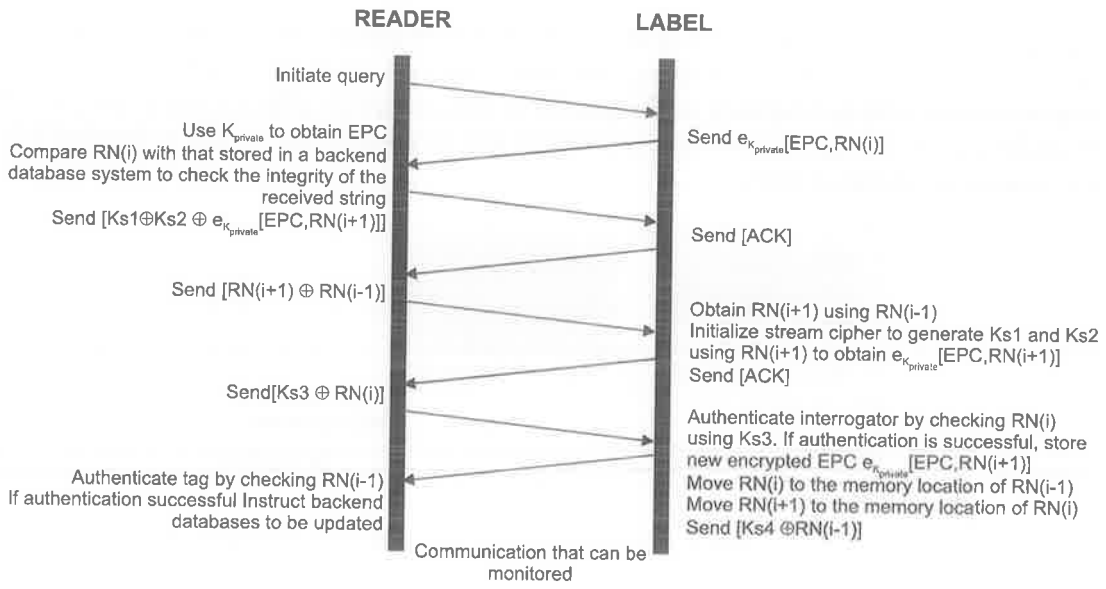

Figure 11.24 Communication protocol for the re-encryption scheme.

The above mechanism has the added advantage that the *RN(i)* padded to the EPC can be used as a message authentication code to ensure the integrity of the received tag response to a query by comparing the *RN(i)* obtained after decrypting the transmission $e_{Kprivate}[EPC, RN(i)]$ and also to authenticate the tag as being legitimate. It should be noted here that it is possible for the encrypted data transmitted from the tag to contain product specific information such as an EPAC that will extend the tag authentication scheme to a product authentication scheme. The use of EPAC is discussed in Section 11.6.1.

Alternatively, it is possible to use the protocol in Section 11.4 after the tag has transmitted its encrypted tag identifier. Then the $e_{Kprivate}[EPC, RN(i+1)]$ can be transmitted from the reader to the tag using the encrypted communication channel established after a mutual authentication process. The difference with the protocols in Figure 11.21 and Figure 11.22 with that given in Figure 11.24 is the removal of the overhead of using a PUF to initialize the stream cipher by transmitting the initialisation vector using a one time pad. Since the initialisation of the stream cipher is carried out only once, the performance will be increased by eliminating the re-initialisation time and the delay time of the stream cipher before a key stream of adequate length is available.

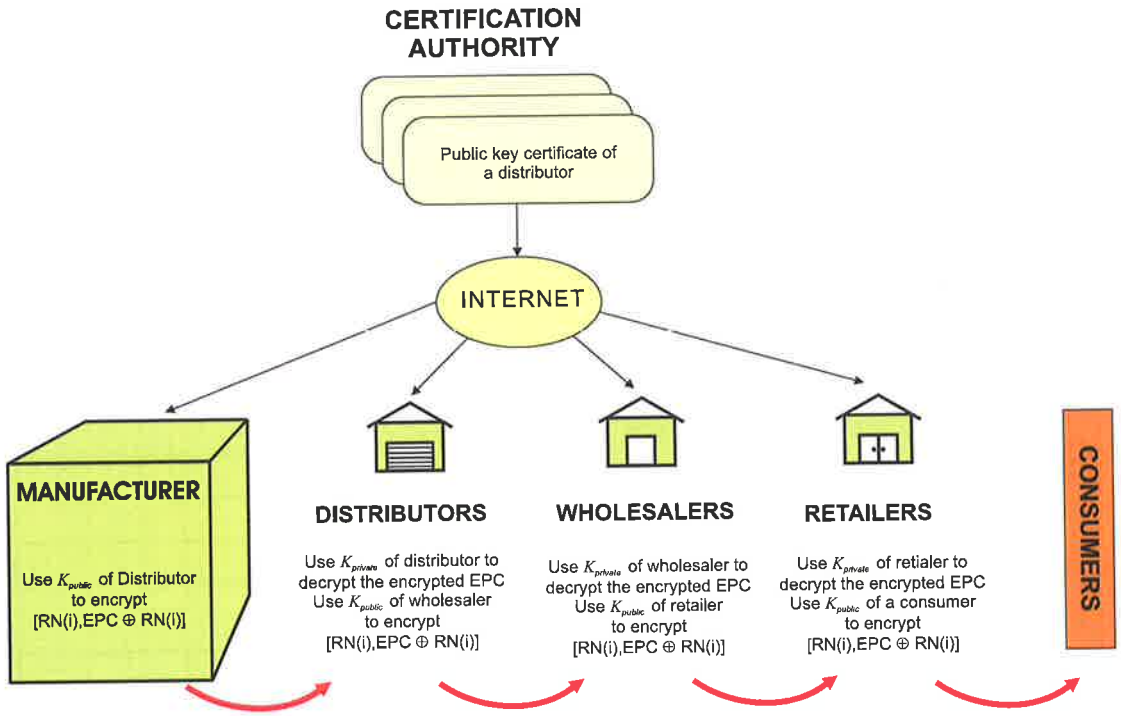

Figure 11.25 Tag data security infrastructure based on a PKI.

An attractive feature of the above approach is that it can be used throughout the supply chain at various stages by setting up a tag data security infrastructure based around a PKI (public key infrastructure) to allow the benefits of tag or product authentications, anonymity and untraceability to be available, at any time and anywhere along the supply chain all the way into the living quarters of consumers. Implementing such an infrastructure will require careful standards and agreements between various parties in the supply chain. The idea is illustrated in Figure 11.25. Considering a simple supply chain model given in Figure 3.8, the manufacture would use the public key of a distributor to encrypt the EPC prior to the transfer of goods to the supplier. The public key can be obtained from a certificate authority, with which the distributor's public key is published. The distributor is then able to decrypt the encrypted EPC using the distributor's private key and prior to transportation of the goods to a retailer, the EPCs of the products going to a particular retailer can be encrypted using that particular retailers public key. It is still possible for a party along the supply chain or even a consumer to then use their own private key to encrypt the EPC for both greater efficiency and greater security while the tagged item is in their ownership.

11.5.2.1 Evaluation

The scheme outlined above is evaluated and discussed in detailed in Table 11.5. Re-encryption, while at the cost of online requirements and tag initialisation requirements is capable of delivering adequate performance, using strong cryptographic primitives with

minimal cost consequences to a tag. Unlike the mechanisms discussed previously in Section 11.4, the present mechanism is able to satisfy a majority of the security objectives and all of the privacy objectives considered necessary in Table 9.3 and Table 9.5 respectively.

The total implementation cost of the mechanism on an IC is estimated to be 2328 gates (using a knapsack generator), however this might be higher in practice due to additional registers and the changes required to the finite state machine (refer to Figure 9.1) to implement the new commands necessary to execute the protocol and recover from a sudden power loss.

Table 11.5 Security mechanism evaluation

Achieved Security Objectives	Message content security
	Tag authentication
	Reader authentication
	Confidentiality
Achieved Privacy Objectives	Anonymity
	Untraceability
Cost and Performance Objectives	*Gate equivalent cost estimate based on the memory storage required for the keys* C1G2 tags already have the necessary hardware for XOR operations, string comparisons, CRC generation, registers for temporary storage of operands, and memory for the storage of the EPC. Using a 96 bit EPC, a 128 bit RN value (adequate for initialising two 64 bit LFSRs, as in the cases of a shrinking generator and also adequate for initialising both the initial state and the connection polynomials of a knapsack generator of length 64 bits) with a 128 bit private key where the encrypted message will be the same size as the key size (as is the case with the AES block cipher with a 128 bit key). Memory cost for the encrypted EPC: 384 gates Memory cost for *RN(i-1)* and *RN(i)*: 384 gates Cost of a shrinking generator: 1730 gates Cost of a knapsack generator: 1560 gates
	Real time authentication requires access to secure backend databases with *RN* values. However if real time authentication is not required and all that is required is a pseudonym change, where back end databases can be consulted at a later time for both the update procedure and the authentication process, the protocol can be executed without the need for online resources.
	Tags must be subjected to an initialisation phase prior to deployment in an electromagnetically secure environment for the initial storage of the at least the *RN* values. However this will be a one-time cost and can be carried out at the RFID chip verification phase.

	Neglecting network delays and computation time for the new encrypted version of a tag's EPC, the greatest delays will result from the time required for transmitting data between tags and readers. There will also be a small delay in initialising the stream ciphers but this will be in the time order of 10s of clock cycles and can be ignored. Estimated time to compete the protocol: approximately 5 ms Hence the number of tags that can be read, authenticated and pseudonyms updated: 200 tags This is a best case scenario and in reality the string comparisons and the calculation of the encrypted EPC will reduce the estimated performance.
	The most power consuming operation is the operation required to write two strings to the EEPROM and thus the mechanism will not violate power constraints assumed in Table 10.1. Refer to Section 11.2.3.5 for a detailed discussion on LFSR power consumption.

11.5.2.2 Practical Issues

The implementation of the scheme outlined in Figure 11.25 requires the creation of a PKI whereby parties can publish their public keys signed by a trusted third party. A simple system overview of a manufacturer in possession of a list of public keys associated with a product's destination party and a retailer using his own symmetric key encryption scheme is illustrated in Figure 11.26.

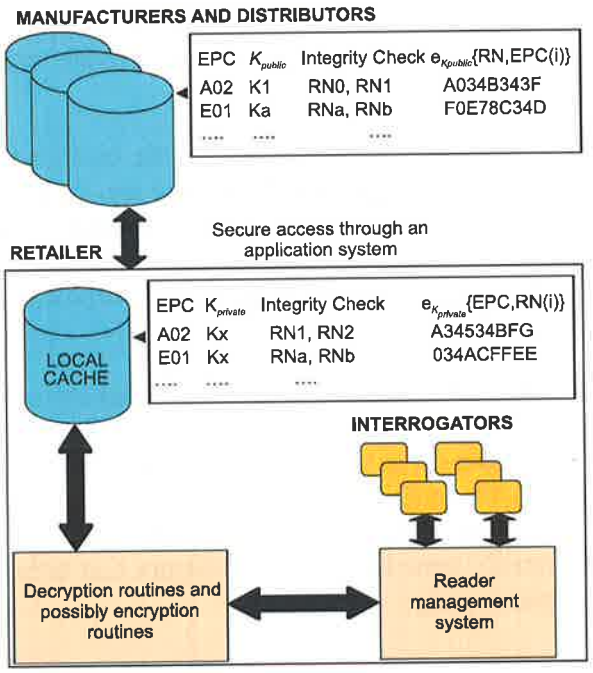

Figure 11.26 An overview of an implementation of an RFID system based on re-encryption.

During the journey of a product through the supply chain where a party along the supply chain is in ownership of that product, that party then has the option of using a symmetric key primitive to encrypt the EPCs to both improve performance by speeding up the encryption process and to utilise the increased security provided by symmetric keys in relation to public keys of similar size.

It has also not been discussed if the present mechanisms can be accommodated by the recently ratified C1G2 protocol. Tags will initially need to be placed in a locked state where the query command will initiate the execution of the protocol outlined, or a modified version of the existing query command is required to signal the finite state machine to execute the protocol in Figure 11.24. The mechanism outlined above will also require two proprietary commands as permitted by the C1G2 protocol: ***SENDENC(<flag>,<data>)*** and ***RESENC(<data>)***. The reader command ***SENDENC*** will be used to send encrypted text to the tag, where the *flag* value will be used to denote the protocol step. The ***RESENC*** command will be used by the tag to respond with the encrypted ciphertext to the reader. If either party detects a fraudulent tag or an unauthorised reader, the protocol will be continued by the legitimate party, but the encrypted data will then be replaced by irrelevant bit sequences. The commands mentioned above were also used and described in Section 11.4.4.

11.5.2.3 Possible Attacks

The security of the above mechanism relies on the difficulty of predicting the output of the stream cipher given only the ciphertext. Both, the knapsack generator and the shrinking generator have been found to be secure against known ciphertext only attacks, even if the connection polynomials are known. The stream ciphers used have been discussed in detail in Section 11.2.3.

Figure 11.27 provides a protocol verification schematic outlining the execution of the protocol over three consecutive interrogations of the same tag. At each stage the tag memory contents and also the possible information that can be obtained by a passive eavesdropper that can eavesdrop on both the forward and the backward communication channel is indicated to show that the information collected is not sufficient to defeat the security mechanism.

The information that is available to an eavesdropper provides an insight into looking at the weaknesses of the protocol to a passive or a malicious adversary who may be mobile or stationary. Clearly the protocol may be interfered with by a disruptive adversary conducting a man-in-the middle attack but this is a difficult proposition given the difficulty of capturing and altering the messages between a tag and reader during the rapid transition of messages and the adversary would also have to ensure that either the tag or the reader never receives the unaltered message.

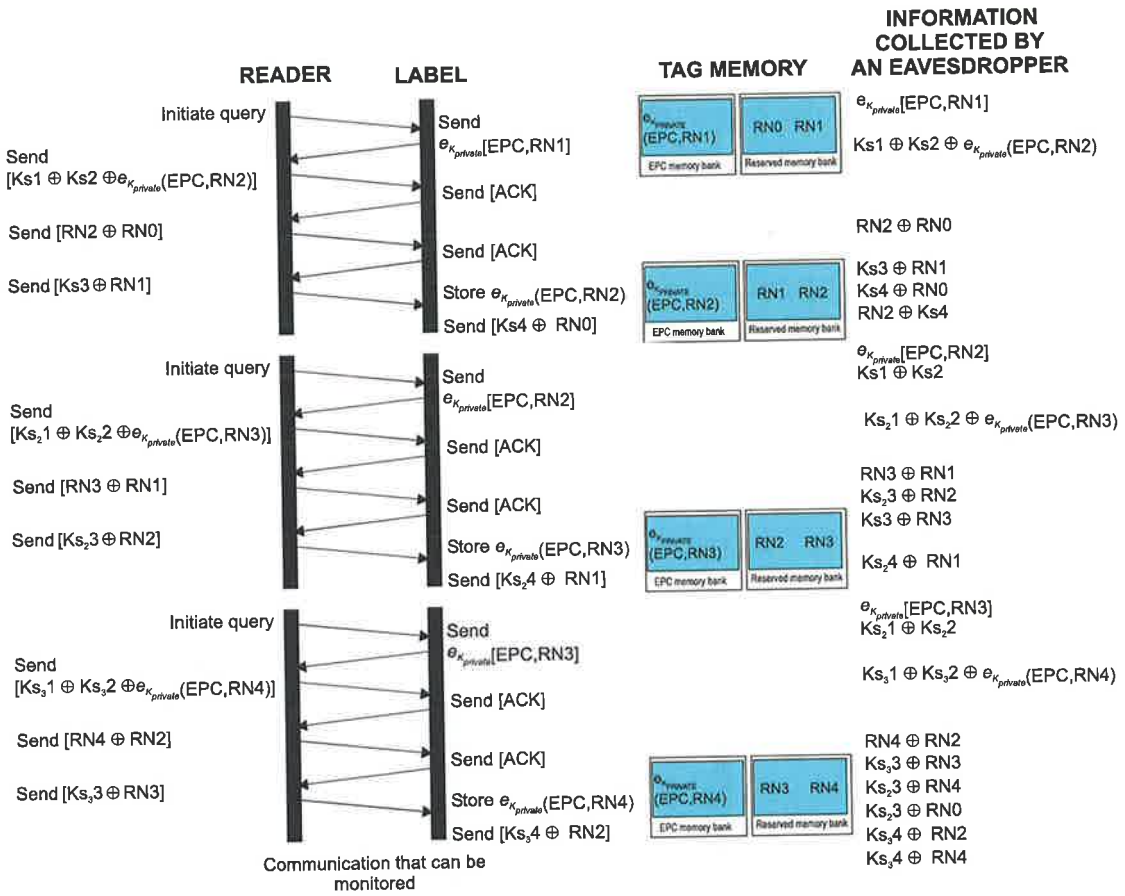

Figure 11.27 Verification of the re-encryption protocol.

An adversary with access to specialised equipment may subject a tag to a physical attack to obtain the *RN* values. This will only allow a single tag to be replaced with another tag, with identical memory contents but such an attack provides no useful information to defeating the security mechanism of other tags. However, a proximity mobile adversary will then be able to decipher all future communications between a tag and a reader. Preventing a physical attack requires the secure storage of keys on the tag as discussed in the protocol presented in Section 11.4.

The use of forward secrecy, that is the use of ephemeral keys, *Ks*, which are generated during each communication session, to be used only once, invalidates any information that the adversary is able to obtain about previous keys or any other information. Even if an adversary is able to recover a previous keystream, that is a Ks_i, that information is not useful in the next interrogation session. The security of the scheme relies on the ability of the label owners to keep the encryption key secret and thus the mechanism is vulnerable to an adversary who is a temporary or a permanent insider.

11.5.3 Randomly Varying Object Identifiers

An alternative to the re-encryption scheme outlined above for providing privacy protection and authentication, is a scheme based on the concept of using a pool of completely random EPCs is discussed below. Here the EPC number no longer has an information bearing structure and it is a random number which only acts as a temporary pointer to the actual EPC. The relationship created between the random EPC and the true EPC is securely stored in a backend database. Thus, it is possible for RFID labeled items of sensitive nature (such that required for use in supply chain logistic operations of the defence department) to be labeled with a randomly generated EPC.

The scheme requires the backend systems to manage a large pool of random numbers and to be able to search through such a large pool in a short time. This database will also need to be able to perform concurrent updates and searches while possibly being distributed in nature.

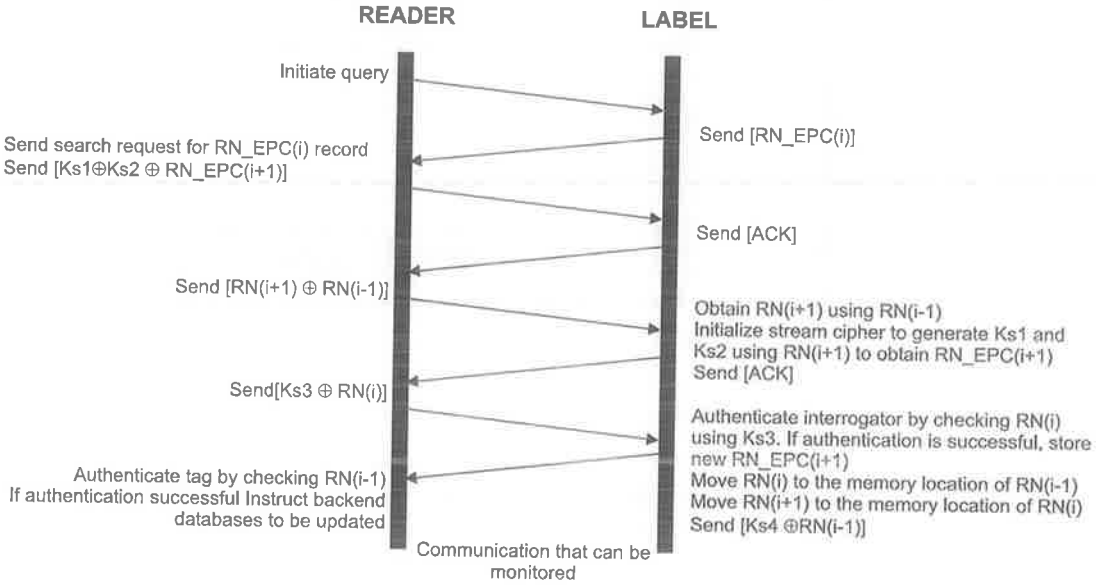

Figure 11.28 Protocol based on using randomly varying object identifiers.

It is possible to use an encrypted version of the random EPCs as done in the re-encryption scheme outlined in Section 11.5.2 to add another layer of security. Although using a random EPC will prevent an adversary from obtaining any useful information in the event the encryption scheme is compromised, encrypting the EPC data will be at the cost of performing the encrypting and decrypting operations, on the reader or by way of a trusted third party (proxy). It should also be mentioned here that it is possible for an encrypted version of the random EPC to contain product specific information such as an EPAC (Electronic Product Authentication Code) that will extend the tag authentication scheme to a product authentication scheme. The use of EPAC is discussed in Section 11.6.1.

However, the use of random EPCs eliminates the need to encrypt the EPC to hide the information bearing nature of the EPC while preventing the tag from emanating a predictable response to a query by an unauthorised reader. The protocol is detailed in Figure 11.28 and is similar to that given in Figure 11.24.

11.5.3.1 Evaluation

The randomly varying object identifier scheme outlined above is evaluated for its suitability for low cost RFID in Table 11.5. Similarly to the mechanism discussed in Section 11.5.2 the present mechanism is able to satisfy a majority of the security objectives and all of the privacy objectives considered necessary in Table 9.3 and Table 9.5 respectively.

Table 11.6 Evaluation of the randomly varying object identification scheme.

Achieved Security Objectives	Message content security
	Tag authentication
	Reader authentication
	Confidentiality
Achieved Privacy Objectives	Anonymity
	Untraceability
Cost and Performance Objectives	*Gate equivalent cost estimate based on the memory storage required for the keys* C1G2 tags already have the necessary hardware for XOR operations, string comparisons, CRC generation, registers for temporary storage of variables, and memory for the storage of the EPC. *Assume*: Using a 96 bit RN_EPC and a 128 bit RN value (adequate for initialising two 64 bit LFSRs, as in the cases of a shrinking generator and also adequate for initialising both the initial state and the connection polynomials of a knapsack generator of length 64 bits). Memory cost for *RN(i-1)* and *RN(i)*: 384 gates Cost of a shrinking generator: 1730 gates Cost of a knapsack generator: 1560 gates
	Real time authentication requires access to secure backend databases with *RN_EPC* values. If real time authentication is not required and all that is required is a pseudonym change, where back end databases can be consulted at a later time for both the update and the authentication process, then local databases caching a list of available random EPCs for future use, will greatly speed up the protocol and provide a method of offline authentication. Refer to Section 11.5.2.2 for database cost considerations.

	Tags must be subjected to an initialisation phase prior to deployment in an electromagnetically secure environment for the initial storage of the *RN* values.
	Neglecting network delays and computation time for new encrypted versions of a tag's EPC, the greatest delays will result from the time required for transmitting data between tags and readers. The small delay in initialising the stream ciphers in the time order of 10s of clock cycles can be ignored. Estimated time to compete the protocol: approximately 3.5 ms Hence the number of tags that can be read, authenticated and pseudonyms updated: 285 This is a best case scenario and in reality the string comparisons and the calculation of the encrypted EPC will reduce the estimated performance.
	The most power consuming operation is the operation required to write two strings to the EEPROM and thus the mechanism will not violate power constraints assumed in Table 10.1. Refer to Section 11.2.3.5 for a detailed discussion on LFSR power consumption.

An interesting result of transferring tag complexity to backend systems (with a complex data structure capable of concurrent updates, and efficient search algorithms), the randomly varying object identification techniques has reduced the security related tag costs from 2328 gates for the re-encryption protocol in Section 11.5.2 to 1944 gates in the current scheme by the removing the overhead from encrypting the EPC related information.

11.5.3.2 Practical Issues

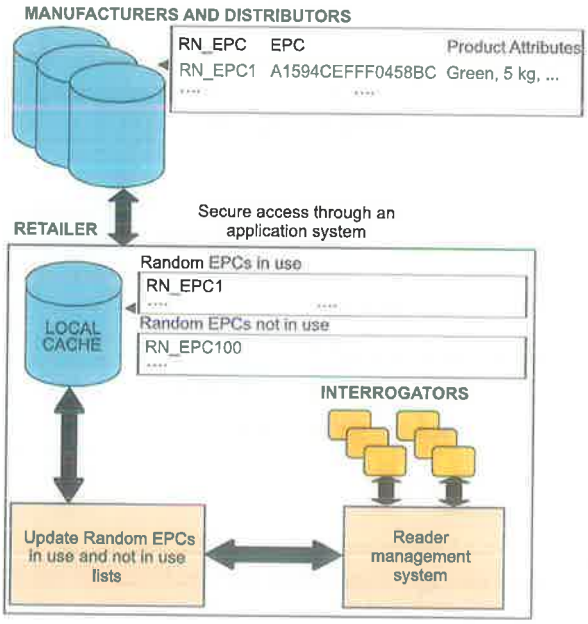

Figure 11.29 An overview an RFID system based on randomly varying object identifiers.

The scheme outlined above requires rapid access to databases over network infrastructure and secure databases maintaining records for each individual EPC as indicated in Figure 11.29. The management of such a database is not a significant hurdle. A data structure such as a self-balancing binary search tree implementation in the form of a red-black binary tree [216] with the ability to perform concurrent updates and search operations will perform efficient search, insert and delete operations. Infrastructural issues are important but can generally be addressed with greater investment to build networks of adequate bandwidth to mange the network delays. Currently Internet traffic network delays are in the rage of 10s of milliseconds [232] but these delays can be improved to the order of microseconds using fibre optic based infrastructure.

TPC (Transactions Processing Performance Council [224]) data provide a guide to assessing performance and cost criteria to evaluate the feasibility of a system implementation. TPC is a non-profit organisation founded to define transaction processing and database benchmarks. The term transactions has a broad meaning, however the TPC benchmarks define a transaction as it is referred to commonly in the business world. Thus a typical transaction defined by the TPC will encompass updating a database system for the purpose of inventory control, banking or the purchase of goods. TPC benchmarks generally measure transaction processing and database performance in terms of the number of transactions a given system and database can perform as transactions per second (tpc) or transactions per minute (tpm) TPC pricing and performance metrics are widely used by the industry to estimate IT infrastructure cost required to provide adequate system performance [225].

The TPC-C performance benchmark [226] for OLTP (on-line transaction processing) is a suitable benchmark to establish the cost of installing a clustered or a non clustered server capable of delivering the database processing times required to make randomly varying object identification practicable. The price performance number obtained for a non clustered system configuration capable of over 4 million transactions per minute, or 15 microseconds per transaction has a cost estimate that comprises of the cost of establishing and maintaining a system, including the cost of software, backup storage and three years maintenance cost, is approximately 15 million Australian dollars. While complexity is pushed further back up the EPC Network to reduce the cost of tags and readers, a significant investment will be required to establish and run the required backend systems. However this is not an ongoing cost and a significant portion of the cost will be one-time items (such as network infrastructure costs). In addition, usage over time (that is an accumulating number of transactions over the life time of the system), will reduce the cost per transaction to a level much less than that estimated in the TPC-C benchmark figures.

Reading a tag is clearly no more complicated than current implementations and requires no special commands. However a special query command or a modified version of the query command is required to signal the finite state machine to execute the protocol in Figure 11.28. The execution of the protocol following a query can be performed using the two additional commands outlined in Section 11.5.2.2.

11.5.3.3 Possible Attacks

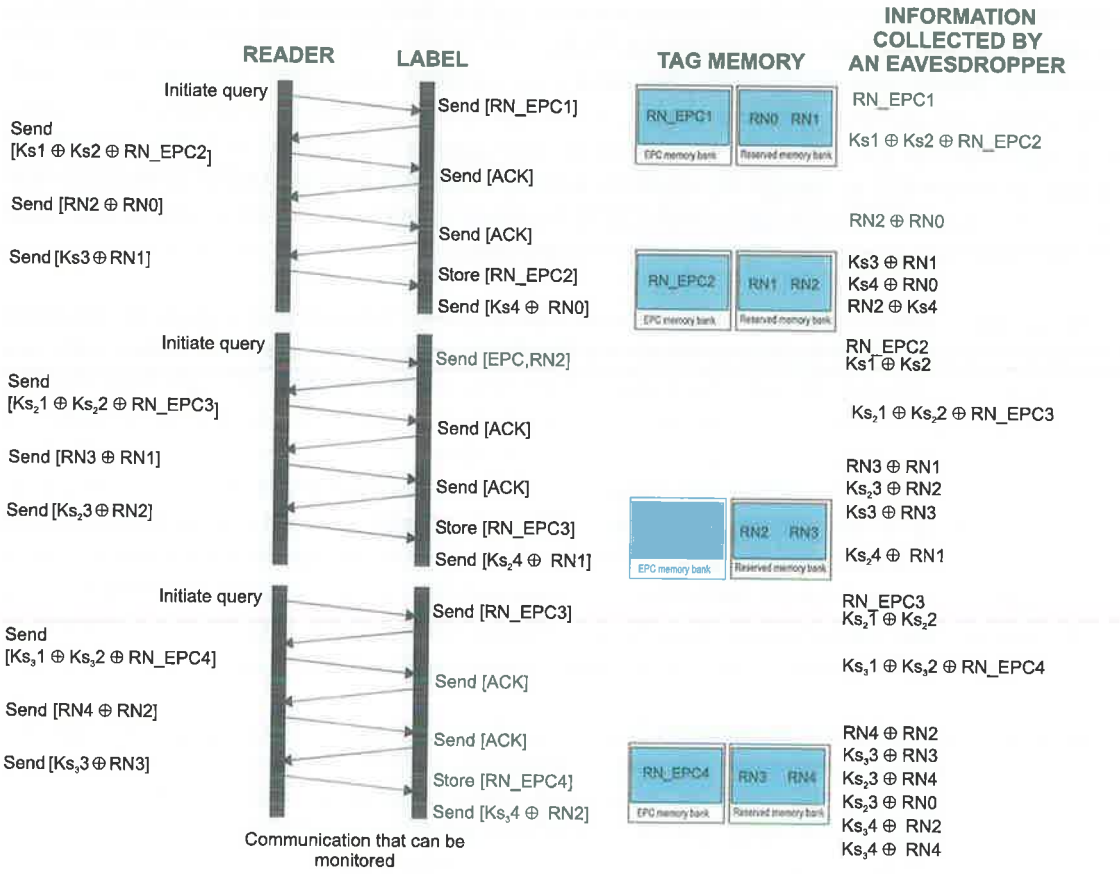

Figure 11.30 Verification of the randomly varying object identifier protocol.

The use of random tag identifiers provides anonymity by altering the tag response to a query command and thus never transmitting a predictable response. Figure 11.30 provides an outline of the execution of the protocol to query a tag on three consecutive occasions. The tag memory contents at the start and the end of the protocol is also indicated. The vulnerabilities of the system are identical to those discussed previously in Section 11.5.2.3 under the re-encryption based mechanism. Clearly given a mobile adversary who may be passive or malicious and who is able to collect all the information indicated in Figure 11.30, still has the task of breaking the stream cipher given only the ciphertext.

11.6 Anonymity, Untraceability, and Product Authentication

Section 11.5 outlined a scheme for providing anonymity and untraceability, while Section 11.3 outlined methods of authentication. The following scheme is aimed at combining the previous solutions to provide, in addition, a product authentication service.

Authentication implies the establishment of a tag's legitimacy by the definition in Section 9.5.3. However in a supply chain logistics environment, use of authentication services to establish the authenticity of a tag and hence the legitimacy of the article to which it is attached is not sufficient to guarantee the genuineness of the article; though authentication of a tag does eliminate cloning of tags. The absence of a method to ascertain the genuineness of goods may be a special concern in the secondary market for goods, and in the processing of returned items.

11.6.1 Product Authentication

The absence of a method to physically or electronically bind the tag identity to a product identity in existing RFID deployments implies that an authentication of a tag does not necessarily guarantee the authenticity of the object to which the tag is attached. The genuine article may be replaced with a counterfeited article or counterfeited goods fitted with stolen tags.

There are a variety of existing techniques for product authentication, based on optical technologies (watermarks, holograms, micro printing and biochemical technology [185]). These techniques are not without their list of associated problems. All the technologies above have static markers that are generally applied on a uniform scale to a single class of products. Biochemical marker tests provide the ability to detect markers but they do not generally quantify the marker, thus leaving open avenues of counterfeiting by dilution, while most optical technologies no longer present an adequate deterrent due to the reduction in the cost of producing watermarks and holograms.

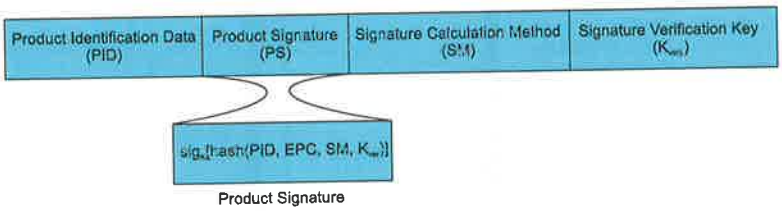

Figure 11.31 Electronic Product Authentication Code (EPAC).

It is into this environment that the following proposal introduces an electronic maker. Each tag attached to a product will contain an Electronic Product Authentication Code (EPAC) illustrated in Figure 11.31 and the various data fields are explained below.

Product Identifier (PI)

It is a unique identifier that characterises the product using verifiable, measurable or observable product specific information determined by the product manufacturer or the retailer. The product identifier may consist of, for instance, weight of the product, measurable physical characteristics such as the dielectric constant, or conductance, size

and shape of the product or an electronic copy of a signature printed or embossed on the product.

Product Signature (PS)

The signature value allows the confirmation of the integrity of the EPAC and also allows a third party to authenticate the identity of the signatory (such as a manufacture or a retailer). As illustrated in Figure 11.31 the signature value is calculated by hashing the PID, EPC, SM, and K_{ver}. This will allow the creation of a message digest to reduce the size of the bit string that needs to be transmitted as the product signature. Hashing the data will produce savings in transmission time as well as memory storage costs. Various signature methods such as RSA-PSS, DSA, ECDSA and ElGamal signature scheme can be used and unkeyed hash functions such as SHA-1 (produces a 160 bit hash value) or MD5 (produces a 128 bit hash value) can be used [78, and 79].

Signature Calculation Method (SM)

The binary sequence in this field will indicate the digital signature method and the type of hash function used to calculate the message digest.

Signature Verification Key (K_{ver})

This is the public key of the party that produced the product signature. The signature verification key can be used to verify the product signature.

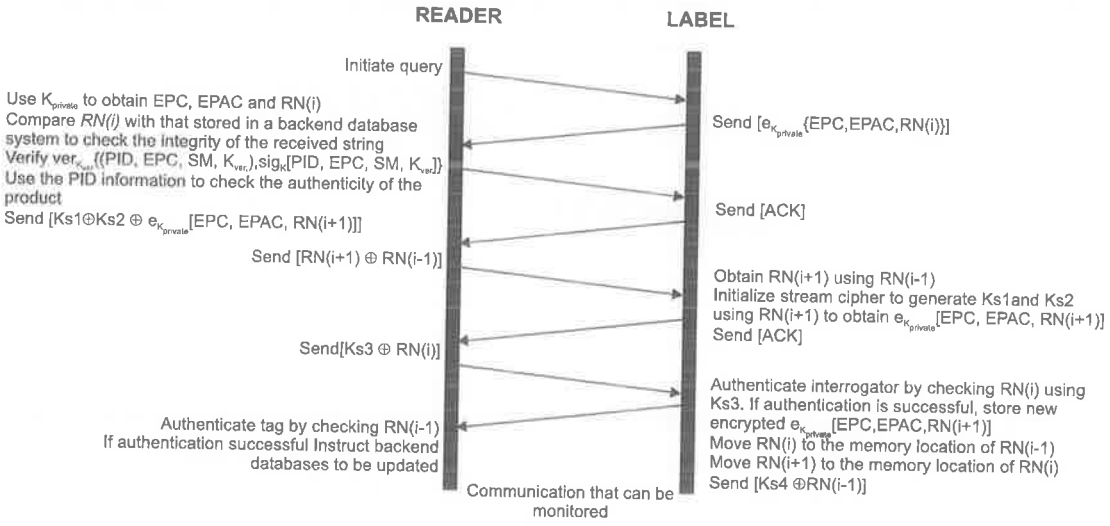

Figure 11.32 Protocol for product authentication.

Assuming that the product signing takes place at the manufacturer, any strong cryptographic signing algorithm with a reasonable digital signature size may be used for the process where the signing method used can be indicated as part of the EPAC along with the verification key, K_{ver}. The protocol for product authentication is illustrated in Figure 11.32.

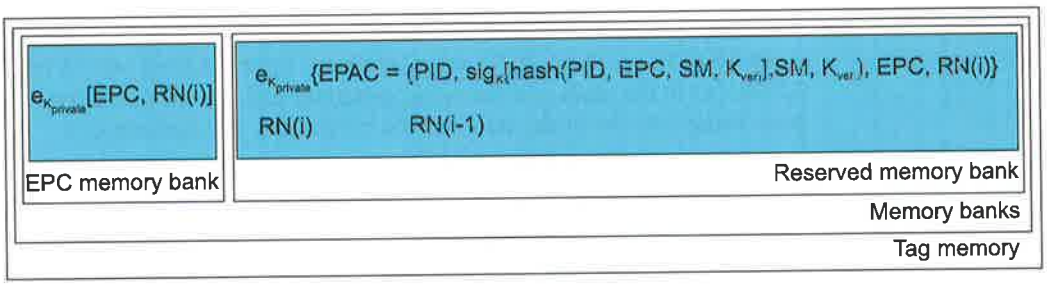

Figure 11.33 Tag memory contents with EPAC information.

The data that needs to be stored on the tag is shown in Figure 11.33. Similarly to the re-encryption scheme in Section 11.5.2 the EPC stored in the tag memory is a result of the EPC$\oplus$*RN(i)*, and *RN(i)* is transmitted along with the EPC. This ensures a random variation in the tag identifier. However for simplicity of the following discussion the [EPC$\oplus$*RN(i)*] operation is assumed to be implicit in the mention of an EPC. Once a tag transmits its tag identifier as indicated in Figure 11.32, a reader can decrypt the received information to obtain the tag EPC, EPAC and the *RN(i)* used in the XORing operation of the tags actual EPC. Although not mentioned previously, the EPC memory bank can also contain an encrypted version of the EPC without the EPAC data. This will allow a reader who does not wish to authenticate the product to execute the protocol based on the re-encryption scheme outlined in Figure 11.24.

11.6.1.1 Evaluation

Table 11.7 provides an evaluation of the product authentication mechanism discussed above. Examination of Table 11.7 reveals that there is a significant penalty in hardware costs (storing a product authentication code costs 1536 gates) and time required for transmitting the 962 bit long string from the tag to the reader and from the reader to the tag. The total hardware cost is less than 4000 gate equivalents. However, there is a serious performance limitation. This performance limitation may not be a hindrance in practice as product authentication may require a lengthy measurement or visual examination process.

Table 11.7 Evaluation of the product authentication protocol.

Achieved Security Objectives	Message content security
	Tag authentication
	Reader authentication
	Product authentication
	Confidentiality
Achieved Privacy Objectives	Anonymity
	Untraceability

Cost and Performance Objectives	*Gate equivalent cost estimate assumptions*: A 96 bit EPC, a 128 bit *RN(i)* value (adequate for initialising two 64 bit LFSRs, as in the cases of a shrinking generator and also adequate for initialising both the initial state and the connection polynomials of a knapsack generator of length 64 bits) A 128 bit private key for encrypting the EPC and EPAC. A 256 bit PID, and assuming that ECDSA is used to create the digital signature with a key size of 160 bits as recommended in FIPS 186-2 [227] which generates a digital signature of size 320 bits using the SHA-1 hash algorithm, which produces a 160 bit message digest. An ACM of 2 bits Verification key of size 160 bits and *RN(i)* of 128 bits Memory cost for encrypted EPC: 384 gates Memory cost for encrypted EPC and EPAC: 1536 gates Memory cost for *RN(i-1)* and *RN(i)*: 384 gates Cost of a shrinking generator: 1730 gates Cost of a knapsack generator: 1560 gates Total cost (using only the encrypted EPC and EPAC along with a knapsack generator): 3288 gates
	Real time authentication requires access to secure backend databases with *RN* values. However if real time authentication is not required and all that is required is a pseudonym change, where back end databases can be consulted at a later time for both the update procedure and the authentication process, the protocol can be executed without the need for online resources.
	Tags are required to undergo an initialisation phase prior to deployment in an electromagnetically secure environment where the initial *RN(i)* values can be imprinted in memory.
	Neglecting network delays and computation time for new encrypted versions of a tag's EPC and EPAC, the greatest delays will result from the time required for transmitting the 962 bit long string consisting of the EPAC between tags and readers. There will also be a small delay in initialising the stream cipher but this will be in the time order of 10s of clock cycles and can be ignored. Estimated time to complete the protocol: approximately 11 ms Hence the number of tags that can be read, authenticated and pseudonyms updated: 88 This is a best case scenario and in reality the string comparisons and the calculation of the encrypted EPC on the reader side will reduce the estimated performance.
	The most power consuming operation is the operation required to write three strings to the EEPROM and thus the mechanism will not violate power constraints assumed in Table 10.1. Refer to Section 11.2.3.5 for a detailed discussion on LFSR power consumption.

11.6.1.2 Practical Issues

Storing the EPAC on the tag is an expensive option both in terms of memory storage costs and transmission costs, but it does allow the offline authentication of the product. Alternatively it is possible to store the EPAC on a backend database, along with other product related data pointed to by the EPC. This will reduce tag complexity and reduce the bottlenecks produced by transmission times accumulated during the product authentication protocol.

It is also possible to store a portion of the EPAC such as the PID on the tag, and the rest of the data on a secure database pointed to by the tag EPC. This will greatly reduce product authentication times by reducing the data load transmitted during the protocol. If the PID data needs to be verified, it can be achieved by opting to retrieve the remaining EPAC data from the secure database.

As discussed in Section 11.5.2.2 tags will initially need to be placed in a locked state where the query command will initiate the execution of the protocol outlined or a modified version of the existing query command is required to signal the tag's finite state machine to execute the protocol in Figure 11.32. A modified query command can also signal if the current interrogation round requires the tag to participate in a product authentication round or if the query will only result in the update of the tag identifier based on the encrypted version of the EPC stored in the EPC memory bank shown in Figure 11.33. The mechanism outlined above will also require two propriety commands as discussed in Section 11.5.2.2.

11.6.1.3 Possible Attacks

The use of re-encryption provides anonymity by altering the tag identifier using a randomly generated number. Thus tags never transmit a predictable response. Figure 11.30 provides an outline of the execution of the re-encryption protocol, without the added product identification service, to query a tag on three consecutive occasions. The vulnerabilities of the system are identical to those discussed previously in Section 11.5.2.3 under the re-encryption based mechanism.

11.7 Acknowledgements

The work presented in Section 11.3.1.3 and 11.3.1.4 above has used previous work published by others in relation to IPUFs. While the concept of a PUF is not a novel idea, the use of PUFs in a low cost RFID system in the manner described in Section 11.3.1.3 and 11.3.1.4 by the author is novel and was first published in [142]. The author would like to thank and acknowledge the work of Daihyun Lim and Prof. Srinivas Devadas in the

developments of IPUFs. The author would also like to thank Srinivas Devadas and Daihyun Lim for kindly providing a number of IPUF chips for testing and assessment purposes to aid the work presented in Section 11.3.1.9.

11.8 Conclusion

The chapter has used lightweight hardware and lightweight protocols to address various vulnerabilities identified in Chapter 9, as strong cryptographic solutions are too area or power hungry to satisfy the limitations of RFID systems and much of the encryption hardware available is for smart card technology. Even though the solutions can be applied directly to RFID, the main obstacle is that smart card processors are much more powerful than a typical RFID label. Thus, the solutions are not portable to an RFID platform if we expect the cost of the secure labels to remain below the 5 cents target value.

The solutions presented have recognised that the resource limitation of low cost labels require the consideration of simplicity at the tag silicon level provided by small one time pads, which involve one or more small shared secrets between a label and an interrogator. Such methods required the use of shielded electromagnetic communications between the label and the reader system to store secret information at an initialisation phase.

The solutions presented have concentrated on the simple concepts of removing label IC complexity, and using the abundant resources available to the reader and application systems of an RFID system to counterbalance the resource limited nature of RFID labels.

Solutions presented to overcome privacy concerns in Sections 11.5 and Section 11.6 address profiling and, tracking and surveillance. However it should be noted here that issues concerning privacy are also public policy issues and require a combination of security mechanisms and properly formulated public policy.

The security mechanisms presented has been evaluated using the criteria developed in Chapter 10 to appraise their suitability for low cost RFID applications.

It is evident that RFID privacy and security are challenging areas of research that have lead to a blossoming new cryptographic paradigm called lightweight cryptography. There are two specific areas of research (lightweight hardware and lightweight protocols) which will greatly benefit low cost RFID security and privacy and the outcome of this research will be the widespread adoption of this technology.

It is important to note that the level of security and privacy will depend on the application. It is evident that there is no universal solution but a collection of solutions suited to different applications based on compromises and on security services required.

An important consideration that is often overlooked is the ability for a cryptographic system to use a piece of hardware repeatedly to result in a more secure encryption engine. Most modern UHF RFID chips use on board oscillators with frequencies over 1 MHz. Thus within

the operational timing constraints imposed as a result of US regulations, it is conceivable to allow a tag to expend around 400,000 clock cycles during a 400 millisecond period. Thus, it may be possible to redesign hardware for existing cryptographic primitives to exploit this unique scenario. However, this will be at the compromise of tag reading speeds. In addition a security mechanism that is capable of leveraging existing hardware on the tag will also reduce the cost of implementation; such a possibility may be found by using the hardware used to calculate the CRC (cyclic redundancy checks) on the tags.

A common requirement highlighted in the proposed solutions is the use of randomness to achieve unpredictability and confusion. Thus a carefully constructed source of randomness is essential to the implementation of the schemes outlined in this chapter. The following chapter will evaluate a suitably configured IPUF, as a pertinent source for generating random numbers both cost effectively and rapidly by interrogators for use in the security mechanisms described in this chapter.

Chapter 12

HARDWARE BASED RANDOM NUMBER GENERATOR

Random number generators are used in cryptographic operations involving generating random keys, and random challenges in a challenge response protocol. While pseudo random number generators based on computational complexity are widely used for most of cryptographic applications and probabilistic simulations, the generation of true random numbers based on physical randomness is required to guarantee the advanced security of cryptographic systems. This is an important aspect in both the implementation of UHF C1G2 Air interface protocol and the security schemes outlined in Chapter 10 as readers are required to generate random numbers on the fly. This chapter presents a method to exploit manufacturing variations, metastability, and thermal noise in integrated circuits to generate random numbers. The metastability based physical random number generator provides a compact and low-power solution which can be fabricated using standard IC manufacturing processes. Experimental results show that the generated random bits pass standard randomness tests successfully. The operation of the proposed scheme is robust against environmental changes since it can be re-calibrated to new environmental conditions such as temperature and power supply voltage.

12.1 Introduction

Random number generators are important in a number of modelling and simulation applications. However, the most significant is their application in cryptography. The wide array of cryptographic applications employ secret keys that must be generated using a random process to ensure the security of the cryptographic system. Numerous cryptographic protocols also require random or pseudorandom inputs such as in the generation of digital signatures, or the generation of challenges in a challenge-response protocol [78]. Random numbers are also used in the selection of winning numbers for lotteries and the picking of premium bonds as in the United Kingdom. Cryptographic needs and large Monte Carlo computations continue to advance the development and research into truly random number generators. In RFID systems a fast and efficient random number generator is both a necessary and an important component in the implementation of the C1G2 air interface protocol. This chapter will provide the details of a random number generator suitable for implementing on an RFID reader for the provision of randomly generated number for use in communication protocols and security mechanisms outlined in Chapter 11.

Random number generators can be broadly categorised as follows:

1. Pseudorandom number generators
2. Physical random number generators.

The above types of generators are more specifically referred to as random bit generators when they are used to produce a stream of binary numbers. This stream can be subdivided to form blocks of random numbers of required block sizes such as 32 bits, 128 bits, or 1024 bits.

Pseudorandom number generators are based on a computational algorithm that receives a random sequence of length l as an input and outputs a binary sequence of length $x >> l$ which has the appearance of being random [78]. Such a generator employs the existence of a one-way function f that is based on the complexity of computations to make it irreversible. There are various algorithms for pseudorandom number generation (PRBG); ANSI X9.17 and FIPS 186 are examples of such generators [78]. The output of a PRBG is not random, but completely deterministic, while the number of possible output sequences is at most $2^l / 2^x$, of all possible bit sequences of length x.

While the output sequences of pseudo random number generators appear to be random, there is the possibility for an adversary to predict random sequences by developing an equivalent algorithm or simply duplicating the generator hardware. To achieve unpredictable randomness, we can exploit the non-deterministic randomness in physical phenomena such as the decay of radioactive isotopes and laser scattering patterns through non-homogeneous materials to generate random numbers. Physical random number generators based on this physical randomness can be useful since there may not be an equivalent algorithm to simulate and predict the physical phenomena. However, exploiting this random source to produce a bit sequence that is both statistically independent and unbiased is not an effortless

task. Random bit generators based on natural sources of randomness are exposed to environmental variations that can be sensitive to the generation of random sequences. Therefore, most physical random number generators should employ post processing units to compensate for the environmental variation and statistical defects of output sequences as shown in Figure 12.1.

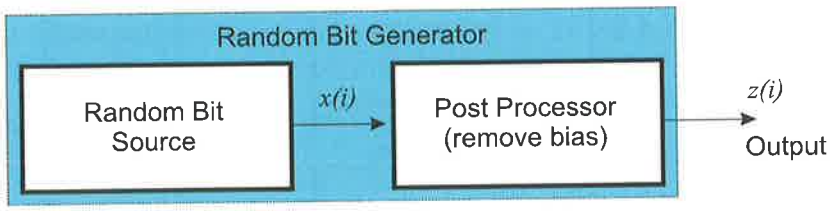

Figure 12.1 A random bit generator with a post processor.

12.2 Sources of Randomness

Random sources can be constructed from dedicated hardware devices (Hardware based generators); such as oscillators with considerable jitter. A number of hardware random number generators can be found in [186], [192], and [193]. Random sources may also be extracted from software procedures using the platform on which the generator is implemented (Software based generators). For instance, event timing may provide sources for SWG, such as mouse clicks, key strokes, size of input and output buffers, or network access.

Software based generators (SWG) are not based on very ideal sources, and it is difficult to properly evaluate and assess the robustness of the sources in regards to observations and possible manipulations. Thus software based generators use a combination of sources to obtain protection from one of its sources being manipulated. Raw bits generated from SWGs most often need to be heavily processed before a random sequence is obtained.

Hardware based generators (HWGs) have a number of advantages over software generators. Generally HWGs can be implemented using common integrated technologies and they can be fabricated into tamper resistant devices to prevent an adversary from performing observations or manipulating the generator. Hardware based generators are also faster, and are capable of producing generators with high throughput. Commonly used sources for hardware generators are physical phenomena such as thermal noise, shot noise, avalanche noise, phase noise, cosmic radiation and atmospheric noise.

In this article we present an assessment of the possibility of using metastability and thermal noise as a source for a hardware random number generator. The HWG presented in this chapter is fully integrable using standard CMOS technology.

12.3 Metastability

When a signal violates a device's signal setup and hold timing requirements of a latch, the output from the device becomes unstable [194]. In this case the observed output from a RS latch can be either high or low or it can even oscillate (see Figure 12.2). This widely undesirable phenomenon is known as metastability [195 and 196] and circuit designers try to avoid this metastability as the final state of the device is unpredictable. A latch is such a bistable device that can enter into a metastable state.

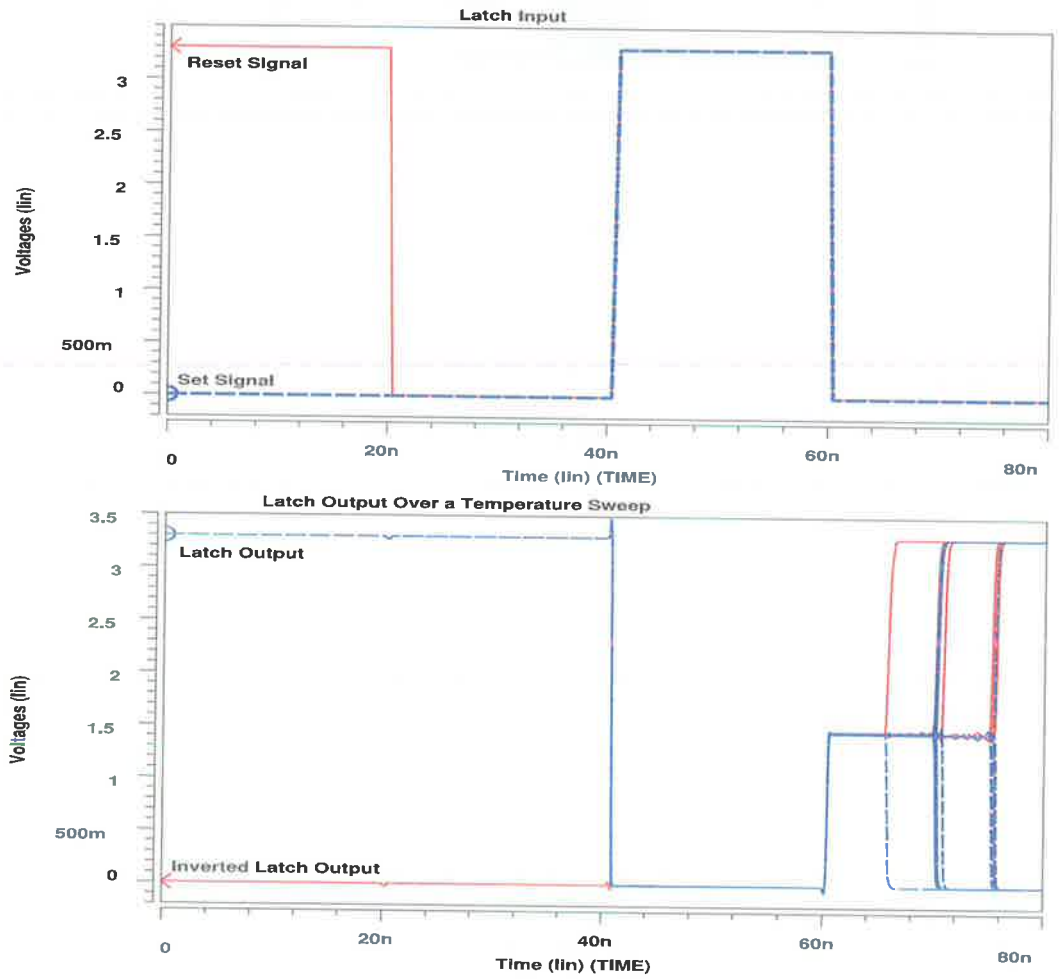

Figure 12.2 RS Latch under a metastable condition: The oscillations of the latch output can be seen on the simulation. The final stable output value of the latch varies with the operating temperature. This indicates the role played by thermal noise in determining the final output value of the latch.

Figure 12.2 shows a HSPICE simulation of a metastable condition in an RS latch as the temperature is swept from -25 to 125°C. The simulation results show that during the time period from 60 ns to 80 ns the latch is in a metastable condition where the output of the latch enters a state where it is neither a logic one nor a zero. Since the final output value can not be predicted due to thermal noise, this metastability provides a source of randomness that can be used to construct a simple and efficient physical random number generator. There have been attempts to use this metastability as a source of randomness [197, 198, and 199]. However, propagation delay variation by environmental changes such as temperature and power supply voltages makes the device difficult to keep in the metastable condition. The proposal presented in this chapter is an alternative method to maintain the generator circuit in a metastable condition to produce random sequences in a practical range of environmental changes.

12.4 Random Number Generator Design

The IPUF introduced in Section 11.2.4 based on the manufacturing variation in ICs provides a suitable solution to create and keep the metastability on a recurring basis. The observation of IPUF results reveals that for certain challenges, the setup and hold time violation of an arbiter (D-latch) leads to unpredictable responses as the arbiter enters into a metastable condition. It is possible to identify and exploit the metastable challenges to cause the D-latch to enter into a metastable state to produce random responses from the I-PUF. The transformed IPUF will be referred to as IPUF Random Number Generators (PUFRNGs) throughout this chapter.

Prior to proceeding further, it should be mentioned here that in the work presented below in this chapter, the IPUF circuits were designed by Marteen van Dijk, Daihyun Lim and Srinivas Devadas while the testing strategy was formulated by the author. The fabricated circuits were supplied by Prof. Srinivas Devadas and Daihyun Lim while the testing and the evaluation of the test results were carried out by the author.

12.4.1 Circuit Implementation

The block diagram in Figure 12.3 depicts the structure of a PUFRNG circuit which is based on the arbiter-based IPUF in [175, 200, and 201] and discussed in Section 11.2.4.1. The details of the switch component in Figure 12.3 are given in Figure 11.7. The circuit accepts an n bit challenge $b_0, b_2, b_3, \ldots, b_n$ to form two delay paths in 2^n different configurations. In order to generate a response bit, two delay paths are excited simultaneously to allow the transitions to race against each other. The arbiter block at the end of the delay paths determines which rising edge arrives first and sets its output to 0 or 1. The actual implementation of arbiter-based PUFs uses 64 bit challenges. For some challenges, the delays in the two paths are approximately identical. When two transitions violate the setup time of the arbiter, the arbiter becomes metastable and generates random responses.

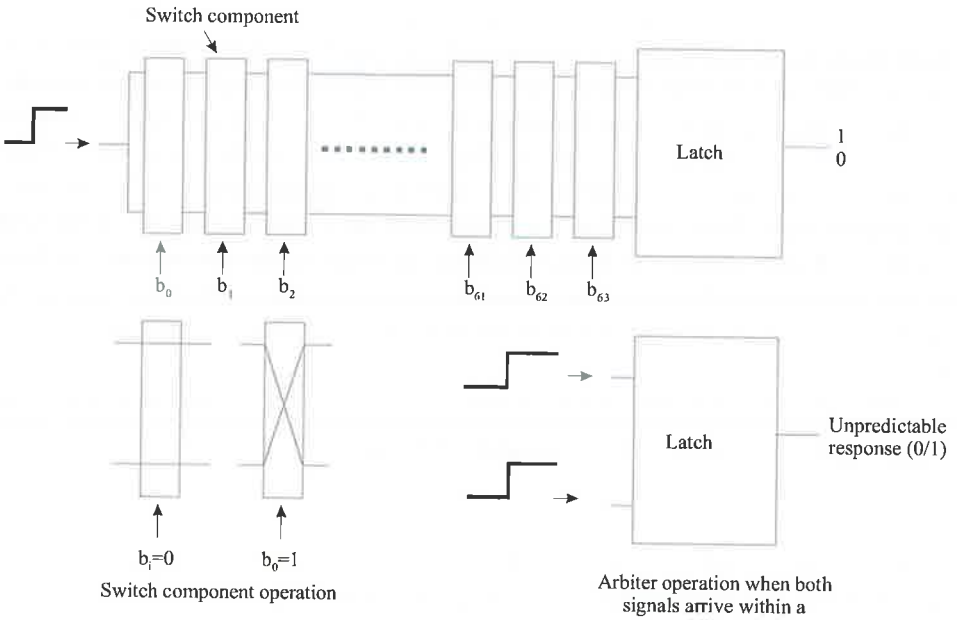

Figure 12.3 PUFRNG based on an arbiter-based PUF circuit [201].

12.4.2 Design Analysis

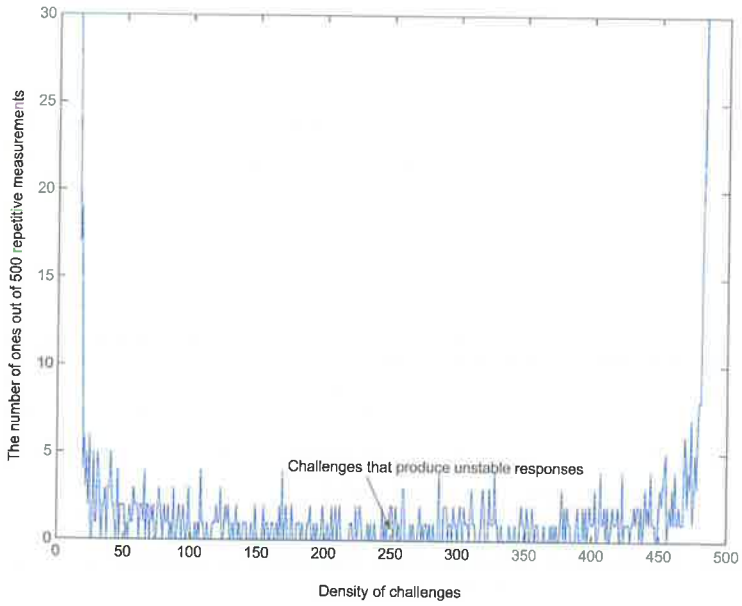

Figure 12.4 The density function of the random variable k, where k is the number of 1's out of 200 repetitive measurements.

The PUFRNG exploits the metastability of D-latch outputs caused by approximate identical timing of data and gate inputs. The output of the latch is largely determined by thermal noise. Figure 12.4 shows the probability density of the random variable k, which is the number of 1s in 200 repeated measurements for a given random challenge. In the middle of the density function, there exist the challenges whose responses consist of approximately 50% logic ones and 50% logic zeros. Within a temperature tolerance of ±5°C from the operating temperature, the responses from these challenges can be used to generate a random bit stream.

The responses from the PUFRNG are sensitive to environmental conditions such as temperature and power supply voltage. In addition fabrication process variations will also influence the responses obtained from one PUFRNG to another for the same challenge. A challenge that generates unreliable responses may not generate unreliable responses if environmental conditions change beyond the tolerance level of 5°C from the original temperature. Hence, each time a PUF is used as a source of randomness, a number of random challenges must be tested to select a challenge that produces unstable responses.

From experimental results, approximately 10 challenges out of 10,000 challenges (0.1%) produce unstable responses in a given environmental condition. Based on the performance of PUF circuits, it takes 0.5 seconds to test the randomness of 10,000 challenges by 1000 repeated measurements [175]. Hence it is possible to complete the initialization of a PUF random number generator within a second.

12.4.3 Increasing the Dynamic Range of Operation

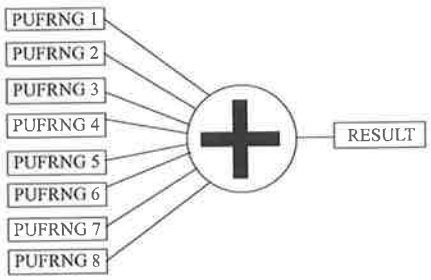

Figure 12.5 Using eight PUFRNGs to compensate for variation in operating temperature.

The generator is sensitive to the power supply voltage and the temperature of the surrounding environment. However problems caused by operational voltage changes can be minimised by the fabrication of a voltage regulator on the PUFRNG. In the testing stage of the PUFRNG a simple mechanism was used to allow the generator to function correctly over a temperature range of 80°C by using eight different PUFRNG circuits each calibrated at intervals of approximately 10°C, in view of the fact that each PUFRNG provided a temperature tolerance of ±5°C. The experiments were run in an oven with thermostat control to provide cyclic temperature changes for the experimental collection of the random bit stream. The output bit stream was a result of an XOR operation on each of the individual bit streams (as indicated in Figure 12.5).

The following section will examine the nature of the system used for generating random numbers and the randomness of generated bit sequence to evaluate the quality of randomness from the PUFRNG.

12.5 Evaluation of the Generator

In a true random number generator the probability of producing either a 1 or a 0 should be ½ where each bit is generated independently of any other bit in the bit stream. Hence it should not be possible to predict the value of a given bit with a probability greater than ½. These conditions form the framework of an ideal random number generator.

12.5.1 Chaos Theory (Dynamic System Analysis)

It is possible to use chaos theory, a division of nonlinear system analysis, to analyse the complex system used for random number generation to investigate if the system exhibits behaviour that is neither random nor periodic. The chaos theory allows the characterization of the PUFRNG as a random, probabilistic (or chaotic) and deterministic system. Furthermore the analysis can be used to discover underlying behaviour patterns, system information and dynamical system models that may render the generator deterministic without uncovering the laws and equations governing the dynamics of a given system.

A technique for the analysis of chaotic data is based on Taken's Embedding Theorem [202 and 203], which allows the reconstruction of the phase space of the system dynamics. The reconstruction of the phase space can be performed from a finite time series of observed random numbers (observation of a single variable). This method relies on the appropriate selection of the delay time and the embedding dimension.

A phase space is a collection of possible states of a dynamical system. The elements of a phase space represent possible states of the system [202]. Implicit in the notion of phase space is that a particular state in phase space specifies the system completely; it is all the information needed to have complete knowledge of the immediate future of the system. Thus the phase space of the planar pendulum is two-dimensional, consisting of the position (angle) and velocity. According to Newton, specification of these two variables uniquely determines the subsequent motion of the pendulum.

Chaotic systems are deterministic and the exact system state can be expressed as

$$\mathbf{X}(t) = \big((\mathbf{x}(t), \mathbf{x}(t-\tau), \mathbf{x}(t-2\tau), \ldots, \mathbf{x}(t-(k-1)\tau)\big). \tag{12.1}$$

where t is a scalar index for the data series and τ is the interval of observations.

Let $\mathbf{F}: \Re^k \rightarrow \Re^k$ be the nonlinear function governing the system. Then the future state of the system at time $t+\tau$ can be determined, such that

$$\mathbf{x}(t+\tau) = \mathbf{F}\big(\mathbf{X}(t)\big) + p(t) \tag{12.2}$$

There is relatively small, zero mean, probabilistic component $p(t)$ added since real world systems are not completely deterministic. This term accounts for the random effects.

12.5.1.1 Attractors

An attractor is simply a state into which a system settles (thus dissipation is needed). Hence over time a dissipative dynamical system may settle into an attractor. The attractor may be a point, a closed path or a complex object on a phase space plot. The attractor is a geometric representation of (12.1) for some large value of t where the effects of the transient have dissipated.

However a formal definition of an attractor is a set in the phase space that has a neighbourhood in which every point stays nearby and approaches the attractor as time tends to infinity. For the general chaotic system given by (12.2), the k-dimensional system will have a non-intersecting attractor with a bounded path of infinite length. However this attractor will be encapsulated in a finite k-dimensional volume.

12.5.1.2 Phase Space Reconstruction

Employing Taken's theorem to reconstruct the phase space requires the determination of the delay τ and the embedding dimension d. Choice of τ should provide low correlations between adjacent elements in the embedded vector so that the original data series is not reiterated. The popular average mutual information algorithms can be used to evaluate the lag τ. Hence the first minimum of $I(\tau)$ (average mutual information function) which measures the average amount of information (bits) shared by two measurements is used as the lag τ.

Furthermore, the correct dimension d unfolds the attractor from the time series. The false nearest neighbour [202] algorithm can be used to evaluate the embedding dimension d. However [204] proposed a method for determining a good embedding dimension using the ideas behind false nearest neighbour algorithm. In [204], Cao proposed two metrics, $E_1(d)$ and $E_2(d)$. Here, $E_1(d)$ is used to discover a good embedding dimension, while $E_2(d)$ is used to determine whether the original data series is random. A suitable embedding dimension is given by the value of d where $E_1(d)$ stops changing. While random signals will exhibit an $E_2(d)$ that is close to unity for all values of d, chaotic signal will have $E_2(d)$ values that are less than unity for small values of dimension d.

12.5.2 Statistical Testing

Randomness is a property that can be characterised and described in terms of probability. The outcome of statistical tests applied to a random number sequence can be thus described in probabilistic terms. There is an array of statistical tests available to test the randomness of random and pseudorandom number generators. Even though these statistical tests do not provide definite results, it is possible to interpret these results with care and caution to determine the randomness of a generator. The general rule of thumb is "more tests the better". The generator bit stream was subjected to a battery of statistical tests for randomness used by The National Institute of Standards and Technology (NIST; an agency of the U.S. Commerce Department's Technology Administration [205]). A more detailed discussion of the tests and their interpretations can be found in [206]. It is however important to note that the test suite is suitable for identifying "deviations of binary sequences" from randomness. However factors contributing to these deviations are numerous and it is possible to expect a certain number of failures from a particular generator.

12.5.2.1 Hypothesis Testing

Each NIST statistical test assesses a binary sequence to establish whether there is significant evidence to suggest that the null hypothesis (H_0) should be rejected in favour of the alternative hypothesis. Here the null hypothesis H_0 is that the sequence being tested is random; while the alternative hypothesis H_1, is that the sequence being tested is not random. Thus for each applied test a decision is made to accept or reject the null hypothesis based on statistical evidence.

A statistical hypothesis, commonly denoted as H_0 is an assertion about a distribution of one or more random variables [207]. This assertion can then be tested using a method based on the observations made on the random variables. The test can provide evidence to support or reject the hypothesis H_0.

Table 12.1 Consequences of accepting and rejecting a hypothesis.

Correct Result	Test Result Decision	
	Accept H_0	Reject H_0
H_0 True	Correct Decision	Type I error
H_1 True	Type II error	Correct

Each test statistic obtained for each individual test is used to calculate a P-value that indicates the strength of the evidence against the null hypothesis. Thus for each test, the P-value is the probability that a perfect random number generator would have produced a sequence that is less random than the tested sequence, given the particular non-randomness being gauged by that particular test. Hence to reject the null hypothesis H_o (signalling the failure a test) at a 95% confidence level would require a P-value < 0.05. The possible

outcomes of the conclusions from a statistical test are outlined in Table 12.1. It is clear from Table 12.1 that *P*-value only asses the relative incidence of Type I errors. Hence, it is important to note here that the test only provides a measure of the strength of the evidence provided by the data against the hypothesis and that the deduction derived from the test is not irrefutable but rather probabilistic.

12.5.2.2 Statistical Test Suite

The statistical test suite employed (as detailed in [206]) is briefly described in Table 12.2.

Table 12.2 Description of the tests used from the NIST test suite

No	Test	Description
1	The frequency test	The test aims to determine whether the proportion of 1's and 0's in a given sequence is that expected from for a random sequence [78].
2	Frequency block test	Similar to the above test but the focus is now the M-bit blocks within a given sequence. The block size used was 128 bits [78].
3	The runs test	The purpose of the test is to determine whether the number of runs (either 0's or 1's) of various lengths in the given sequence is as expected from a random sequence [208].
4	Test for the longest run of once in a block	The test examines if the length of the longest run of ones within the given sequence is consistent with the length of the longest run of ones that would be expected in a random sequence.
5	The binary matrix rank test	The purpose of the test is to discover linear dependence among fixed length substrings of the original sequence [209 and.210].
6	The discrete fourier transform test	This test is used to examine the peak heights in the Discrete Fourier Transform of a given sequence. The test is able to depict periodic features in the tested sequence by examining the number of peaks exceeding the 95% peak height threshold value. The number of peaks exceeding this threshold peak height is less than 5% for a random sequence [211].
7	The non-overlapping template matching test	This test is designed to search for the number of occurrences of pre-specified bit patterns. The test is aimed at detecting generators that produce large occurrences of a certain aperiodic bit pattern. The size of the template used was 9 bits in length, which resulted in a total of 148 templates being applied to each of the sequences. The results from this test are similar to having applied 148 different tests on the sequence of numbers provided [212].

8	The overlapping template matching test	Similar to the above test, with the exception that once a pattern is found the search window is now advanced only one bit instead of advancing the window to the end of the pattern as performed in the non-overlapping template matching test [212].
9	The serial test	The purpose of this test is to establish whether $2m$, m-bit, overlapping patterns occurs as many times as expected from a random sequence. In a random sequence every m bit pattern has the same probability of appearing as every other m-bit pattern [213].
10	The approximate entropy test	The purpose of the entropy test is to compare the frequency of overlapping bit patterns of two consecutive lengths of m and m+1 bits with that expected from a random sequence [214].
11	The cumulative sums test	The cumulative sums (Cusum) test determines whether the cumulative sum of partial sequences occurring in a given bit string is that expected from a random sequence. The cumulative sum may be considered as a random walk and thus for a random sequence the deviation from the random walk should be near zero. This test is performed once going forward in the sequence and then going in the reverse direction [78].

12.6 Analysis and Interpretation of the Test Results

Fabricated PUF generators mounted on a circuit board and interfaced to a PC using a JTAG interface provided the experimental set-up. Each PUF generator was initialised using 10,000 random challenges to select challenges that produced an unstable responses at different temperatures (10, 20, 30, 40, 50, 55, 60, 70 °C). The challenge was repeatedly used to obtain 4.5 million random bits after post processing of the output sequence.

12.6.1 Post Processing

In order to make the stream of bits emanating from the PUFRNG uniformly distributed, it was necessary to pass the bits through an entropy distillation process [78]. Post processing is required to remove bias from an original bit stream at the expense of reducing the overall size of the original bit stream. The method adopted is detailed in [78] and is that of von Neumann [215], and it involves the parsing of bits generated from the random number generator in pairs, and then transforming them according to the scheme outlined in Table 12.3. This method resulted in a typical reduction in the original PUFRNG bit stream in the range of 65-75%. However this is only a general estimate as it is a variable parameter that tends to depend on the environmental conditions such at the device operating temperature.

Table 12.3 Post processing transformations. The original bit stream from the PUFRNG is obtained as non-overlapping pairs (input bits). The corresponding new output is then depicted in the 'output bit' column, where 'Ignore' indicates that the bits a discarded.

Input bits	Output bit (transformation)
10	1
01	0
11	Ignore (indicates that nothing is appended to the post processed bit stream)
00	Ignore (indicates that nothing is appended to the post processed bit stream)

12.6.2 System Analysis

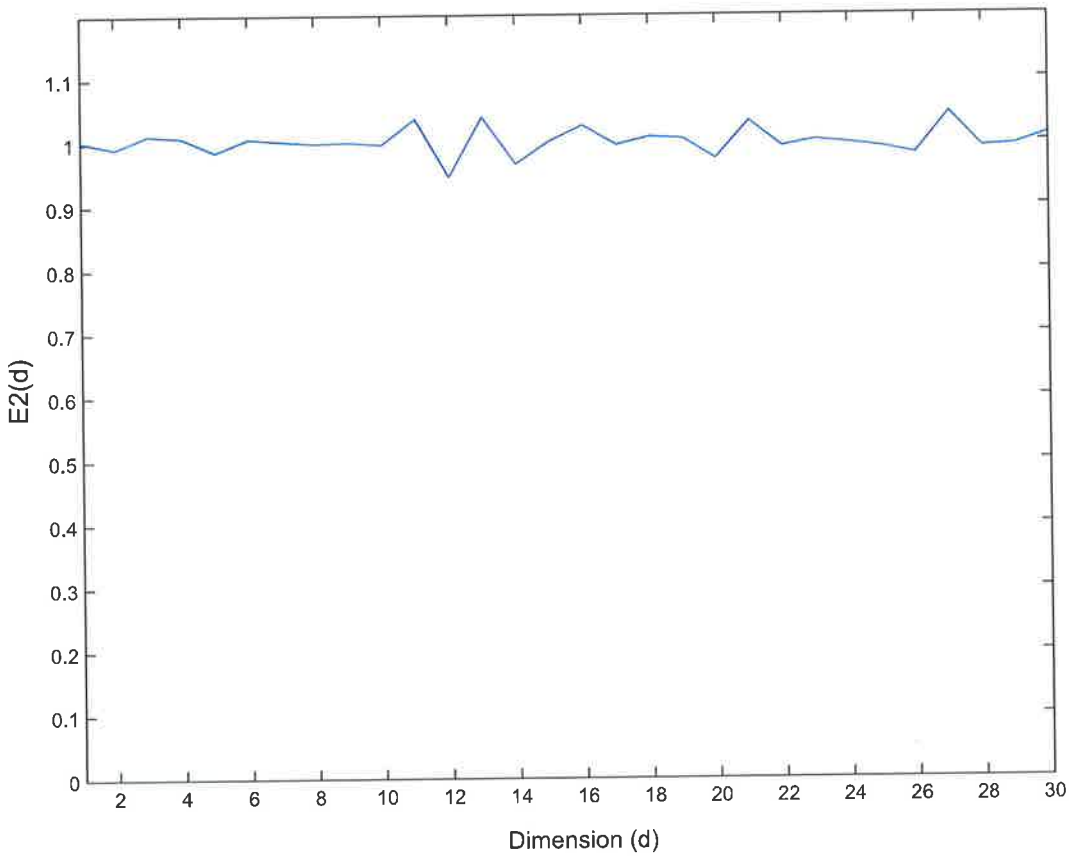

Figure 12.6 Plot of *E2(d)* metric for the series of 32 bit random numbers.

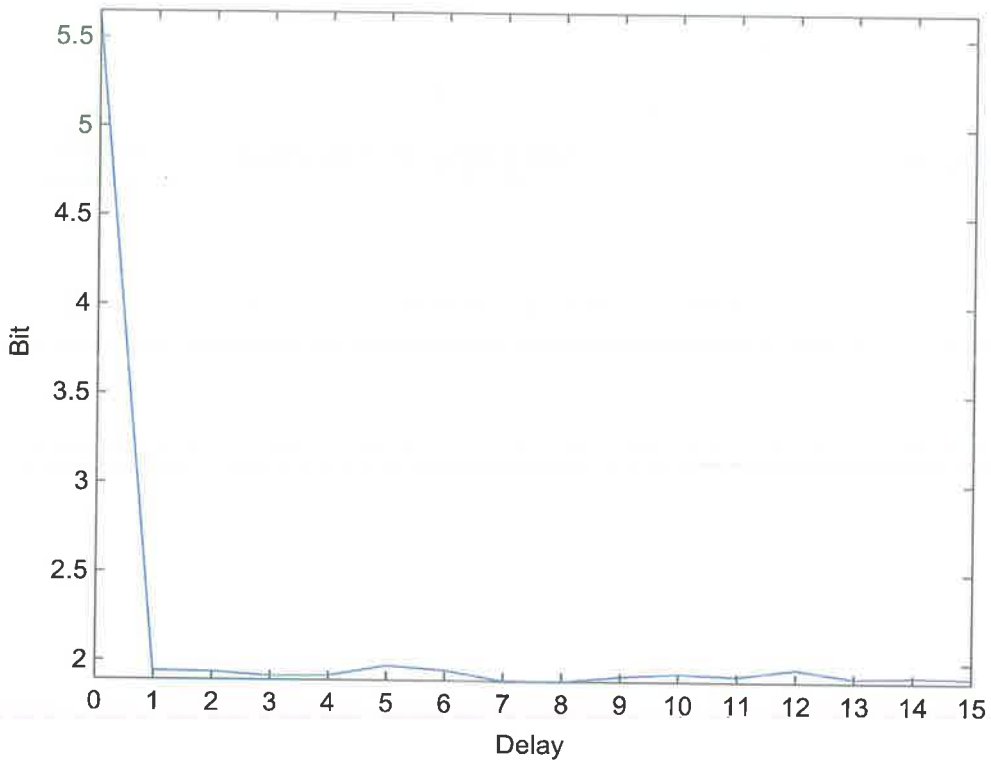

Figure 12.7 Mutual information function for the random data.

The Discrete Fourier Transform spectrum in Figure 12.10 shows a broadband spectrum. Both chaotic and random systems will exhibit such a spectrum. However Section 12.5.1 described the features of a chaotic system and the reconstruction of a phase space. This section shows the results from the analysis. All the calculations depicted here were performed using the TSTool add-on program for Matlab. The bit stream subdivided into 32 bit blocks provided the random numbers for the following analysis.

Cao's [204] $E_2(d)$ metric applied to identify whether the number sequence is random is shown in Figure 12.6. Here, $E_2(d)$ remains at unity for all values of d, suggesting a random system and hence the system is not chaotic. Nevertheless, this can be shown graphically for a two and a three dimensional phase space reconstruction by using the calculated lag. Figure 12.7 shows the average mutual information plot, $I(\tau)$ to support the computation of lag time. The first minimum of the display is at lag one. This will be used as the delay coordinate in the phase space reconstruction. Figure 12.8 and Figure 12.9 depicts the two dimensional and three dimensional phase space reconstruction respectively. The plot in Figure 12.8 does not show an attractor, however strange attractor of a square shape is only an artefact of the use of 32 bit blocks to generate random numbers. Similarly the plot in Figure 12.9 does not depict an attractor; however the space of the plot is limited by the use of 32 bits of the random bit stream to generate data.

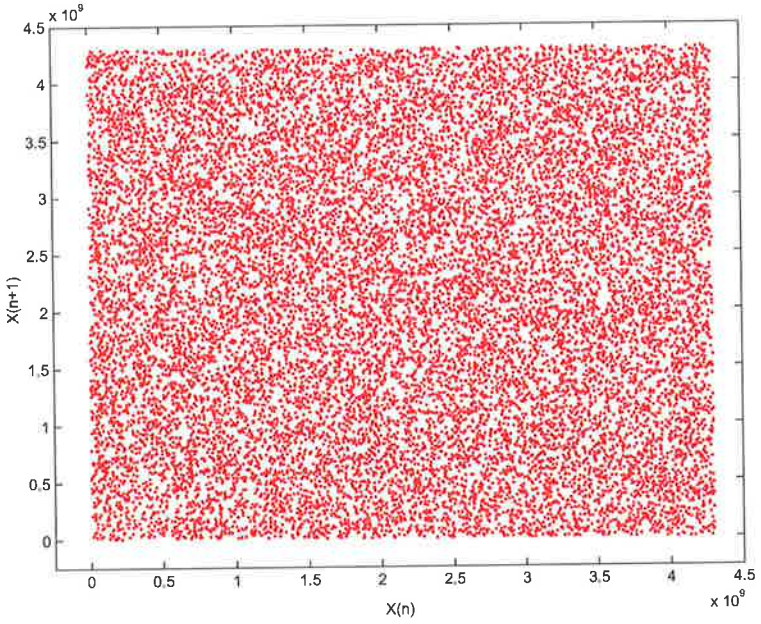

Figure 12.8 2-D phase space plot of the random numbers using a delay of one estimated from the average mutual information algorithm.

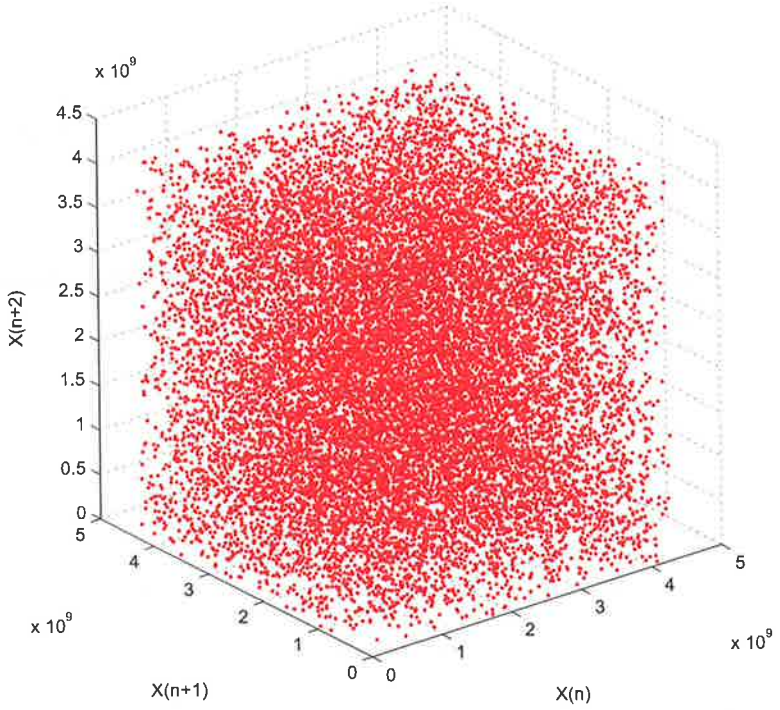

Figure 12.9 3-D phase space plot of the random numbers using a delay variation of one estimated from the average mutual information algorithm.

12.6.3 Statistical Testing

The following sections will summarise the results obtained from the statistical test suite along with a description of the parameters used in testing.

12.6.3.1 Parameters Used in the Test Suite

Table 12.4 provides the test suite specific parameters for all the parameterised tests.

Table 12.4 Test parameters used in the NIST test suite for analysis of the generator output.

No.	Test	Parameter value
2	Frequency Block Test	The block size used was 128 bits.
7	The Non-overlapping Template Matching Test	Template length used was 9 bits.
8	The Overlapping Template Matching Test	Template length used was 9 bits.
9	The Serial Test	Block length used was 16 bits.
10	The Approximate Entropy Test	Block length used was 10 bits.

12.6.3.2 Evaluation of Test Results

Table 12.5 provides a summary of the tests conducted and the results of those tests. A test result depicting a PASS indicates that the generator results have met the test criteria.

Table 12.5 Test result summary.

No.	Test	Description
1	The Frequency Tests	PASS
2	Frequency Block Test	PASS
3	The Runs Test	PASS
4	Test for the Longest-Run-of-Once in a Block	PASS
5	The Binary Matrix Rank Test	PASS
6	The Discrete Fourier Transform Test	PASS
7	The Non-overlapping Template Matching	PASS (except four templates failed)
8	The Overlapping Template Matching Test	PASS
9	The Serial Test	PASS
10	The Approximate Entropy Test	PASS
11	The Cumulative Sums Test	PASS

The guidelines in [206] can be used to interpret the test results. Table 12.5 gives a summary of the number of sequences that passed each of the tests performed using a *P*-value of 0.01 (that is $\alpha = 0.01$) as the significance level to reject or accept the null hypothesis. A pass in a test indicates that there is no significant evidence to reject the null hypothesis and thus the sequence can be considered to be random.

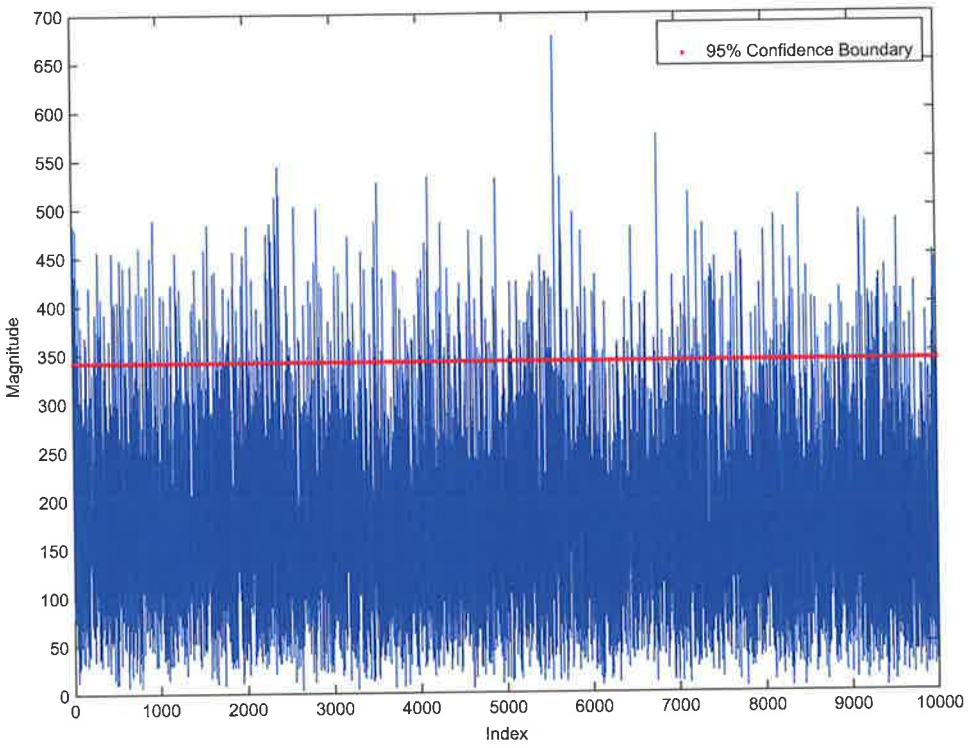

Figure 12.10 Discrete Fourier Transform test results of the random bit sequence.

Figure 12.10 shows the Discrete Fourier Transform of the random bit sequence. The sequence has a broadband spectrum and there is no significant frequency component in the spectrum. Measurement results show that only 4.5% of the spectral lines are above the 95% confidence boundary. Hence less than 5% of the peaks are below the 95% confidence level. The number of peaks exceeding this confidence level is less than 5% for a random sequence [206 and 211].

The test results from the individual tests do not indicate a deviation from randomness. However, two recommended approaches by NIST to interpret the empirical results can be used to investigate the validity of the null hypothesis. The method adopted by NIST includes the following.

1. The examination of the proportion of sequences that pass a given statistic test
2. The estimation of the distribution of *P*-values to ensure that they are uniformly distributed

If either of the above evaluations fails, the null hypothesis can be rejected. However, further testing on the generator should be performed to ensure that the conclusion was not due to a statistical irregularity.

12.6.3.3 Proportion of Sequences Passing a Test

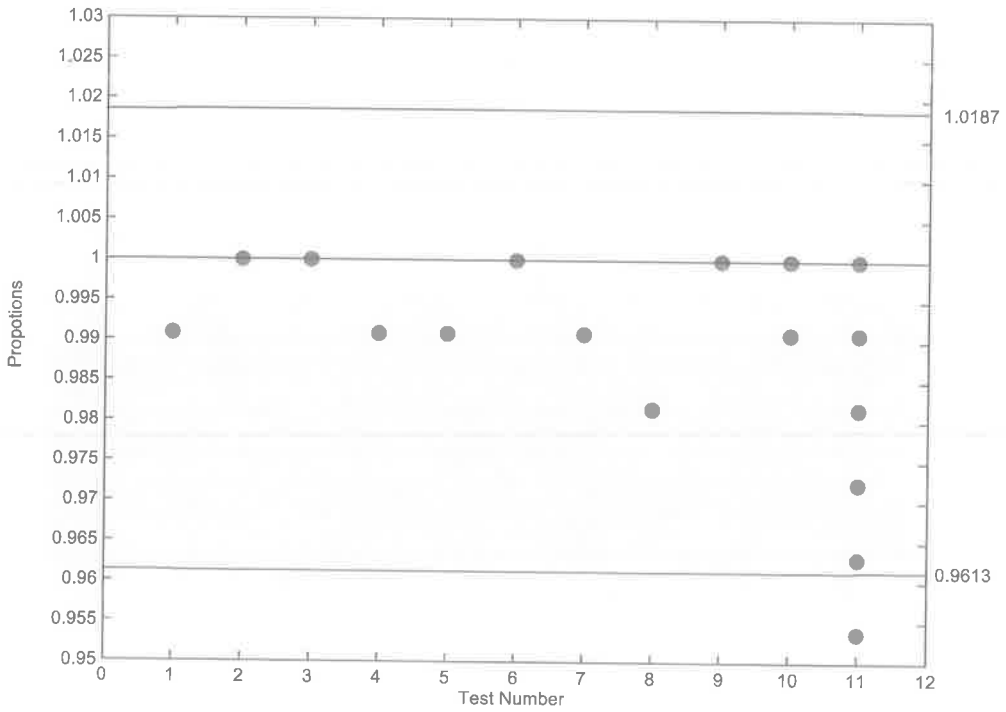

Figure 12.11 Proportion of sequences passing each test based on their *P*-value.

The empirical results can be used to calculate the proportion of sequences that pass a given test. The range of acceptable proportions is given in [206] and evaluated using the confidence interval defined as

$$\hat{p} \pm 3\sqrt{\frac{\hat{p}(1-\hat{p})}{m}}, \tag{12.3}$$

where $\hat{p} = 1 - \alpha$ and m is the number of sequences tested.

Figure 12.11 is a graph of the proportion of sequences that passed each test, with the confidence intervals marked with dotted lines. It can be seen that there is an outlier below the lower limit of the confidence interval due to four templates from the template matching test failing to pass the expected proportion of passing sequences.

12.6.3.4 Uniform Distribution of *P*-values

Evaluation of the uniformity of the distribution of *P*-values is discussed in detail in [206]. A *P*-value calculated on the distribution of *P*-values of a statistical test can be used to accept the distribution of *P*-values as being uniform or non-uniform.

An χ^2 test and a determination of a *P*-value that corresponds to the goodness-of-fit distributional test of the P-values can be used to evaluate the so called '*P*-value of *P*-values'. Thus using ten bins to analyse the distribution of P-values the χ^2 statistic is given by [28]:

$$\chi^2 = \sum_{i=1}^{10} \frac{(F_i - s/10)^2}{s/10} \tag{12.4}$$

where F_i is the number of *P*-values in the bin i of Figure 16 and s is the sample size. The *P*-value of *P*-values is then given by the complemented incomplete gamma function:

$$1 - \Gamma(a, z) \tag{12.5}$$

where $a = 9/2$ and $z = x^2/2$ [206]. The resulting calculation yields a mean x^2 value for the distributions in Figure 16.

Table 12.6 Results evaluating the uniform distribution of *P*-values.

No.	Test	Uniformity of *P*-value distribution
1	The Frequency Tests	PASS
2	Frequency Block Test	PASS
3	The Runs Test	PASS
4	Test for the Longest-Run-of-Once in a Block	PASS
5	The Binary Matrix Rank Test	PASS
6	The Discrete Fourier Transform Test	PASS
7	The Non-overlapping Template Matching Test	PASS
8	The Overlapping Template Matching Test	PASS
9	The Serial Test	PASS
10	The Approximate Entropy Test	PASS
11	The Cumulative Sums Test	PASS

A significance level of $\alpha = 0.0001$ was used to assess the uniformity. Thus a *P*-value calculated on the distribution of *P*-values ≥ 0.0001 was considered to have a uniform distribution [206]. Table 12.6 summarises the assessment performed on the *P*-values obtained for each statistical test.

Figure 12.12 below is a plot of the *P*-value distributions for the test conducted. The graphed data show that all the distributions are approximately uniform.

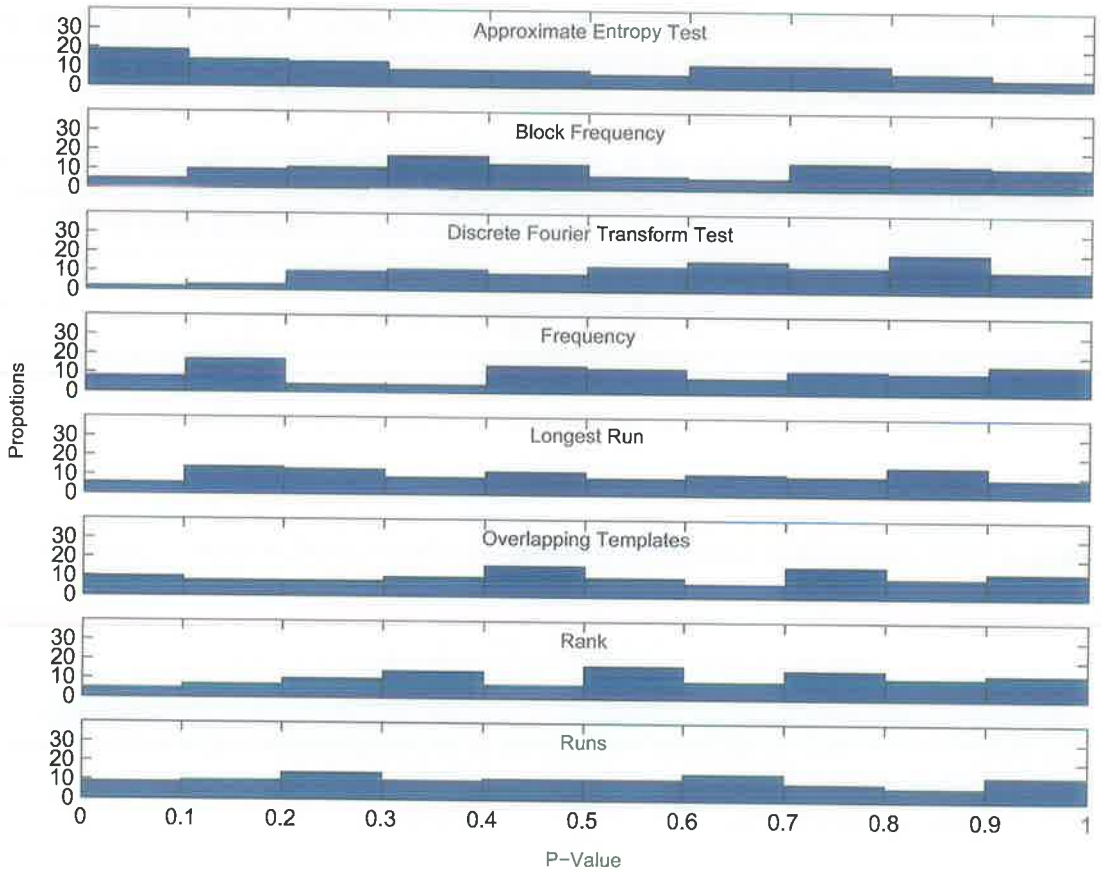

Figure 12.12 Histogram of the *P*-value distributions resulting from applying the eleven statistical tests from Table 12.2 to 108 bit stream of length 38912 bits.

12.7 Acknowledgements

The author would like to thank and acknowledge the work of Daihyun Lim and Srinivas Devadas in the area of PUFs and their collaborations. The author would also like to thank Srinivas Devadas and Daihyun Lim for kindly providing a number of PUF chips for conducting the analysis presented in this chapter.

12.8 Conclusion

The system analysis showed that the random number generator based on an arbiter based PUF circuit is not deterministic but rather random. The analysis of the generated bit stream showed that the generator successfully passed the eleven tests used from the NIST test suite. This proves the quality of the randomness of the bit stream from a PUFRNG under varying environmental conditions.

The PUFRNG provides a cost effective solution to produce millions of random bits in a very short time by fabricating a number of PUFRNGs on a single IC. This generator can be easily constructed using standard digital gates and layout tools. However, the PUFRNG requires some overhead such as post processing and calibration prior to its use.

Nevertheless the ability to calibrate the PUFRNG very rapidly can be an advantage. This unique ability of the PUF random number generator allows the generator to adapt to external influences and to fine-tune the generator for greater performance. Compared to other physical random number generators, the PUF random number generator can be a compact and a low-power solution. A 64-stage PUF circuit costs less than 1000 gates and the circuit can be implemented using standard IC manufacturing processes. Additionally, various kinds of low power techniques such as sub-threshold logic design and multi-thresholds CMOS design can be utilised to reduce the power consumption to make it suitable for use in devices sensitive to low power consumption.

The effects of environmental conditions on the measurements obtained from a PUF are documented in [175], and the symmetrical nature of the circuit counter acts to reduce much of the variation provided otherwise. Changing temperature affects the propagation delays and the metastability window and thus the particular set of unstable response producing challenges may not all produce acceptable results if the operating temperature of the device changes drastically. The tolerance of performance and the calibration of PUFRNG to operate at a range of temperature provide compensation for varying conditions of temperature. However effects of power supply voltage still need to be investigated to discover practical performance boundaries such that the PUFRNGs do not need to be calibrated prior to its use on every occasion. Nevertheless it is possible to fabricate a voltage regulator onboard the PUFRNG to prevent effects from higher voltage variations, but it will not be able to counteract conditions induced by voltages below a calibrated power supply voltage.

Future work will also involve the investigations into the effects of voltage on the performance of the PUFRNG. It may also be possible to avoid calibration by evaluating the performance of the circuits under different environmental conditions of varying voltage. It is also left to investigate whether the generator throughput can be improved. Eliminating the need for post processing the output of the generator would one method of improving the throughput of the generator.

PART III: TURN-ON CIRCUITS

Chapter 13

TURN-ON CIRCUITS FOR ACTIVE LABELS

The focus thus far has been on low cost RFID technology used in Class I and Class II labels in the label class hierarchy outlined in Chapter 2. The consideration of the rest of this dissertation will be on battery powered labels; a set of higher class labels referred to as active labels. The battery, powering such active labels, must have very low current consumption in order to prolong the life of the battery. However due to circuit complexity or the desired range, the electronic circuits can drain the battery more rapidly than desired.

A turn on circuit is a type of receiver, which applies the battery power to an RFID transponder after receiving an appropriate trigger signal level. Turn-on circuits will allow active labels to operate at greater distances than those achieved with passive labels, and also allow a well controlled trigger field to be established for active labels. The primary goal, however, is to activate the label only when required, thus conserving power and extending the life of the label battery. Two solutions available for the development of a turn-on circuit use resonance in a label rectification circuit. This chapter presents the results of experiments conducted to evaluate the performance of the suggested turn on circuits and the design of a fully integrable turn-on circuit for higher class active RFID labels along with test results published in [69] and [70].

13.1 Introduction

The primary focus of this chapter is on active labels (Class III and Class IV labels) that will be equipped with a power source (such as a paper battery) to enable long range communication. The interrogation of active RFID labels will inevitably involve the development of a mechanism for turning on the labels as power conservation is an important factor that requires the labels to be turned off when not being interrogated. This situation will also be true for active sensors and sensor networks.

The turn-on requirements of active backscatter labels are different from independent reply generating labels. An active backscattering label will modulate the powering carrier or a sub-carrier to establish a communication link with the reader while using the battery to power the logic circuits of the label. However an active label that uses an independent source of power for generating a reply to a reader may also alternatively use the on-board battery to power the transmitter of the label. This distinction is more apparent in the range of operation of the label. A reply from a backscattering label is very weak, and under an RFID system operating under the US regulations for the ISM (Industrial, Scientific and Medical) band of 902-926 MHz (allowed transmit power in this band is 4W EIRP), a backscattered reply can only be correctly decoded in the range of tens of metres. Thus a turn-on circuit need only work within the range of tens of metres. However an active label with an independent source of power for reply generation will work in the range of several hundred metres.

Thus a turn-on circuit for this situation will need to be operated at a greater distance. The turn-on circuit presented in this chapter is for active backscattering labels and for active labels with an independent source of power for reply generation operating in the UHF ISM band (refer to Table 7.1 for a list of frequencies). In the following Sections the paper describes a practical evaluation of the concepts and a discussion of a detailed implementation and test results of the circuit.

13.2 Turn on circuits

The practical options for turn-on circuits are two fold:

- Rectifier circuits that can produce a voltage from an illuminating RF field of the order of 1V that can turn a CMOS transistor from fully off to fully on; or
- Rectifier circuits that can produce a voltage from an illuminating RF field of the order of 10mV that can be amplified in a sub-threshold current CMOS amplifier to a level in which a transistor can be turned from fully off to fully on.

For the production of a rectified output even to an open circuit load, a rectifying diode must experience a voltage across the junction capacitance of the order of, or greater than, the rectified output, and hence a minimum of reactive power must flow into and out of the junction capacitance. To service that reactive power, a resonant rectifier must be provided

and the power lost in that resonant rectifier must be provided by the available source power from the antenna. Circuits of this latter type are described in the following Sections of this chapter.

13.2.1 Evaluating Turn-On Circuit Concepts

A label antenna, that in this application is preferably inductive, and the rectifying circuit that is intended to produce from a UHF signal a rectifying voltage used for circuit turn-on, can be modelled as indicated in Figure 13.1. Here R_r represents the antenna radiation resistance, X_s represents the antenna reactance, X_l represents the reactance of the diode capacitance, X_B is the reactance of the reservoir capacitor that also serves as an RF bypass, R_l represents the loss in bringing reactive power into and out of the diode junction capacitance, and R_a is the ohmic loss contribution from the antenna.

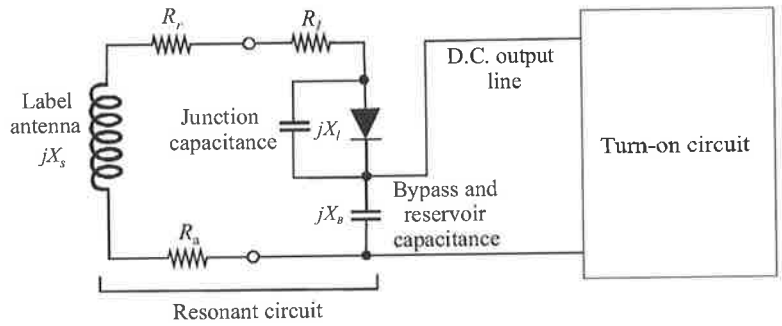

Figure 13.1 Label rectification circuit.

The antenna ohmic losses can be ignored if the antenna is not too small, and the antenna construction in a good design can be a slot antenna containing a significant amount of copper. In addition the series combination of the impedance jX_l and jX_B will be approximately equivalent to that provided by the diode junction capacitance, as the reservoir capacitor has a relatively larger capacitance of the order of 100 pF. It is assumed that no d.c. power is removed from the diode. Shaping the antenna and its connection points appropriately, allows an impedance match between R_r and R_l to be achieved.

Determining experimentally the minimum power required to produce one volt across the reservoir capacitor of the label circuit requires care. The procedure involves the following steps.

1. Selecting a suitable diode.
2. Setting up an impedance matching circuit.
3. Setting up an RF rejection circuit.
4. Minimising damping caused by radiation.

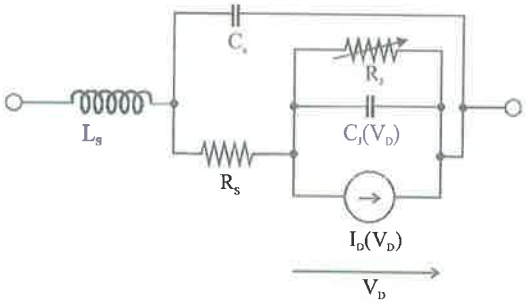

Figure 13.2 High frequency diode model.

A model of a Schottky diode at high frequencies is presented in Figure 13.2 where R_s is the parasitic series resistance of the diode, L_s is the series package inductance and C_s is the package capacitance, where the capacitance C_j which depend on the bias voltage V_d is the voltage across the junction. Current CMOS manufacturing techniques can produce small Schottky diodes with junction capacitances (diode depletion layer capacitance) ranging from 0.1 pF to 1 pF. However a Schottky junction is relatively delicate and sensitive to excessive RF power. RFID applications may work in poorly controlled environments where high power many cause the diode to burn out. Hence in an application it is important to use power limiters to protect the sensitive Schottky diode.

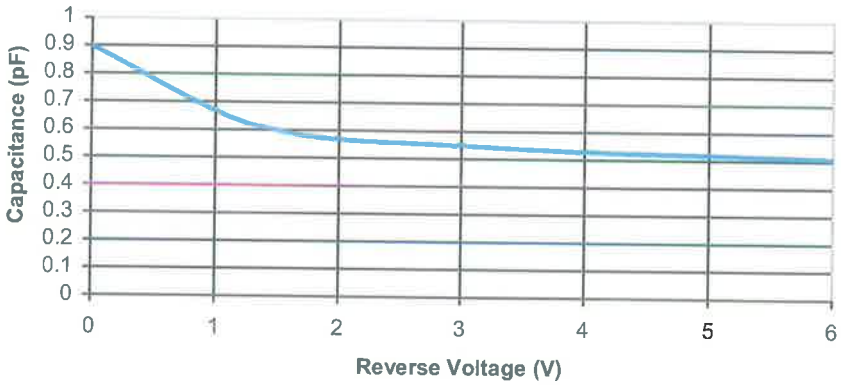

Figure 13.3 Variation of the diode junction capacitance as a function of the reverse biasing voltage [71].

The experiments described below utilised the Hewlett Packard 5082-2835 Schottky diodes [71] (a surface mounting version of the same diodes are available as HSMS-2820). These are good candidates for the application due to the range of its capacitance (1 pF – 0.5 pF) and low cost. Figure 13.3 provides a plot of the variation of the junction capacitance as a result of the reverse biasing voltage across the depletion region.

In addition, consideration must be given to practical matters regarding the construction of the circuit and the measurement of the voltage developed across the reservoir capacitor at various RF energies fed from the antenna. The circuit design must minimise damping caused by measuring instruments, connectors and radiation from physical structures such as the inductor used to represent the antenna or the leads connecting the electronic components. The issues of "hot wires", and radiating connections and components are serious at high frequencies. Thus the construction of the experimental circuit should employ low loss and low series inductance capacitors, small coil inductors, and short connections with adequate shielding provided by a metal box as shown in Figure 13.4. The shielding from the metal box serves to reflect radiated energy back into the circuit, and so reduce losses from the radiation mechanism. It also produces a small and unimportant change to the inductance and capacitance properties.

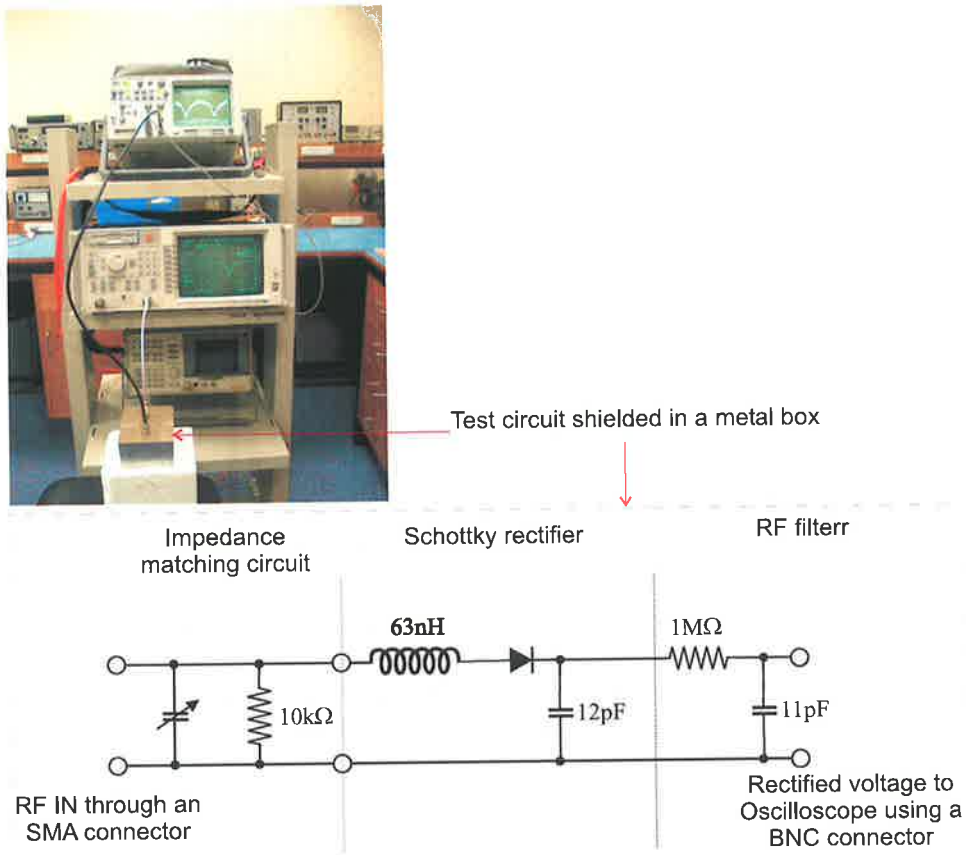

Figure 13.4 Instrumental arrangement and the schematic of the circuit used to conduct the turn on circuit experiments.

The experimental arrangement used to conduct the turn-on circuit experiments and a schematic of the test circuit for diode rectification studies is provided in Figure 13.4. As the measurement method uses a network analyser, the RF input port of the measurement circuit

requires an impedance matching network with a capacity for external adjustment. This is provided by an adjustable trimmer capacitor in series with stray inductance of the capacitor connections. The inductor is used to represent an antenna element which fulfils the function of resonating the diode junction capacitance. Measuring the output voltage from the reservoir capacitor requires very good filtering to remove all the RF content in order to minimise radiation from the output connection before the voltage is measured across the reservoir capacitor. The simple RF filter achieves this function.

The impedance matching network utilised employs a capacitor in parallel with a large (10 KΩ) resistor (to provide a DC path across the capacitor for the case where no such path exists in the source). The impedance of the matching circuit was measured with the Schottky rectifier and the RF filter circuits identified in Figure 13.4 disconnected.

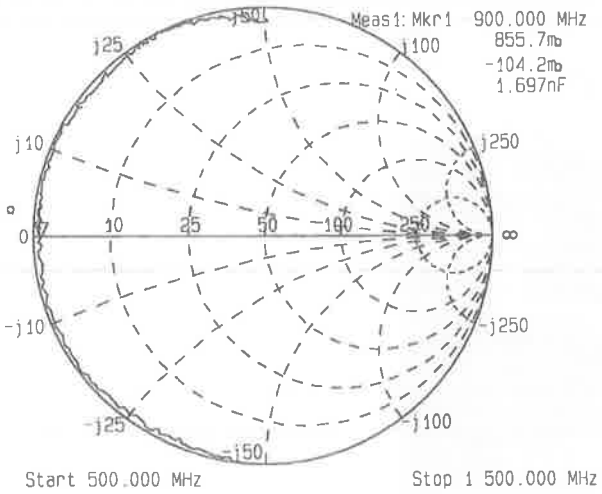

Figure 13.5 Impedance properties of the matching circuit.

The impedance of the matching circuit over a bandwidth of 1 GHz is given in the Smith chart provided in Figure 13.5. The impedance appears close to the periphery of the Smith chart implying a lossless termination and from the nature of the chart it is clear that the impedance below 900 MHz is capacitive while above 900 MHz is inductive.

However this is not true for all values of the trimmer capacitor, which can be adjusted to obtain a wider range of capacitive or inductive impedances as indicated in Figure 13.6 and Figure 13.7. A point of resonance for the setting that produced Figure 13.5 can be observed at 900MHz but at resonance the impedance of the transmission cable is mismatched to the impedance of the resonant circuit. The fact that the impedance plot is at the left edge of the Smith's chart is indicative of a detuned open circuit and hence indicates the matching network to be a series resonant circuit.

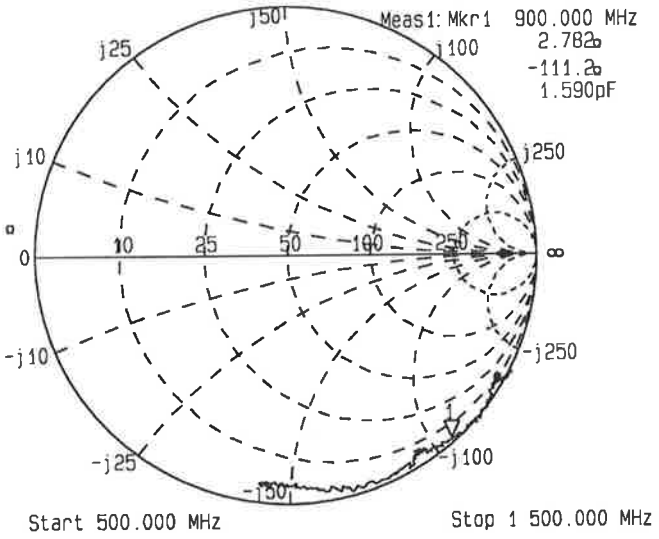

Figure 13.6 Smith chart of the impedance matching network showing impedance values with the trimmer capacitor set to its minimum value.

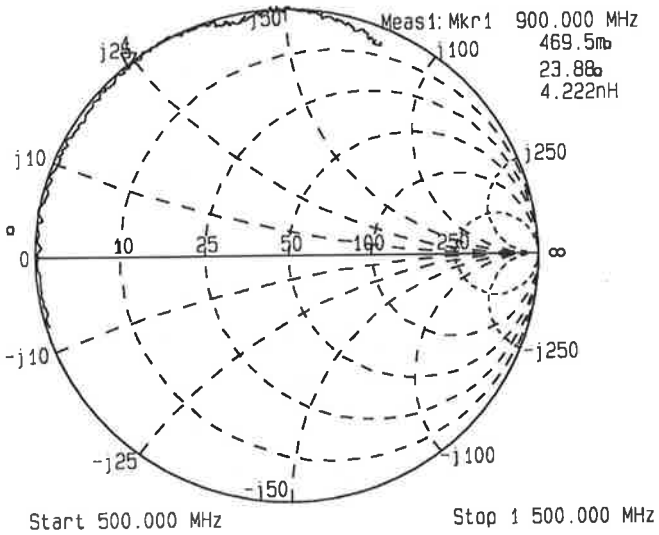

Figure 13.7 Smiths chart of the impedance matching network showing impedance values with the trimmer capacitor set to its maximum value.

It is important to gauge the Q of this circuit as it provides an indication of the lossless nature of the impedance matching network. This is an important consideration as we require most of the RF energy to be localised in the diode resonance. Estimating the Q requires a measure of the inductance, L and capacitance, C of the series resonant circuit model. The reactance graph in Figure 13.8 can be used to obtain the C and L parameters of the resonant circuit. The capacitance is found to be about 5.0 pF and the inductance is found to be about 6.36 nH.

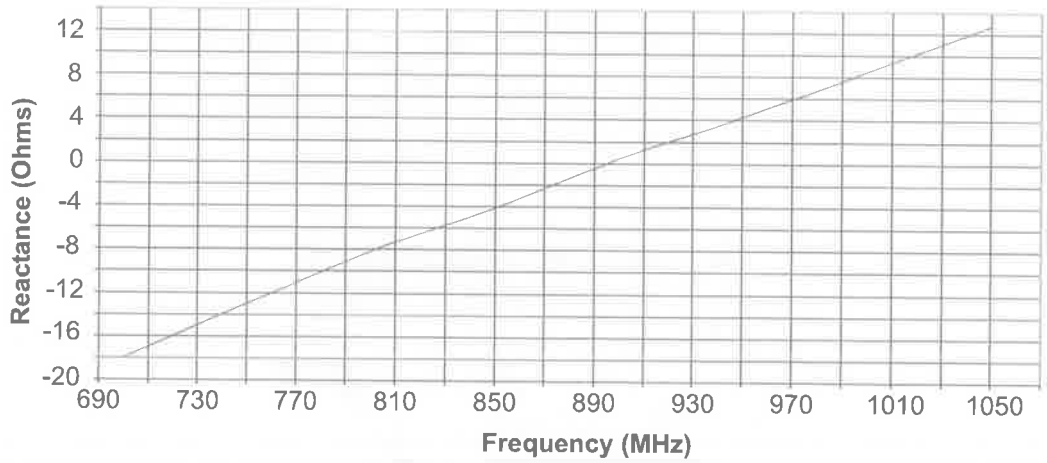

Figure 13.8 Impedance matching circuit input reactance obtained using a network analyser sweeping across a frequency range of 500MHz to 1500MHz.

The resistance of the series resonant circuit is much less than 50 Ω since the resulting Smith chart is at the periphery for a broad range of frequencies; this point can also be observed by examining the dynamic resistance on the Smith chart (refer to Figure 13.6 and Figure 13.7). The quality factor of this resonance is in excess of 35. Hence the impedance matching network is relatively broadband in relation to the quality factors expected in the diode resonance and low loss.

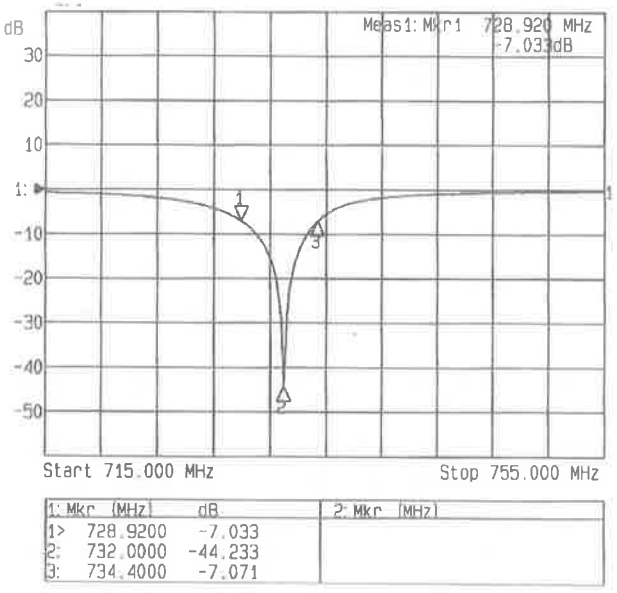

Figure 13.9 Low power Q: the return loss plot obtained with the network analyser output power level set to -35dBm indicates a low power Q of approximately 130.

The return loss plots provided in Figure 13.9 with the diode connected (that is with the Schottky rectifier and the RF filter circuit in Figure 13.4 connected) can be utilised to obtain the low power Q of the diode resonance (by tuning the matching network to obtain a deep dip on the return loss curve), while the plot in Figure 13.10 provides a return loss curve under high power and thus can be used to obtain the high power behaviour of the diode as the source frequency is swept. The non-linearity of the circuit response at high power shown in Figure 13.10 precludes a meaningful definition of high power Q.

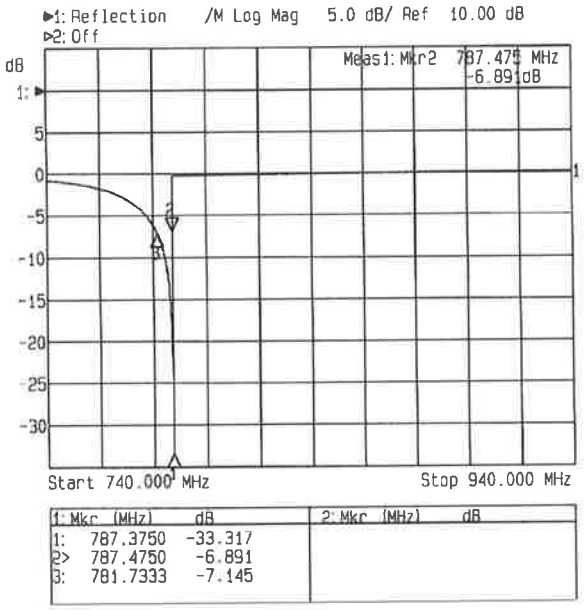

Figure 13.10 High power Q: the return loss plot obtained with the network analyser output power level set to -19 dBm depicting the non-linearity of the circuit response.

13.2.2 Turn on Range Estimation for a Zero Power Turn on Circuit

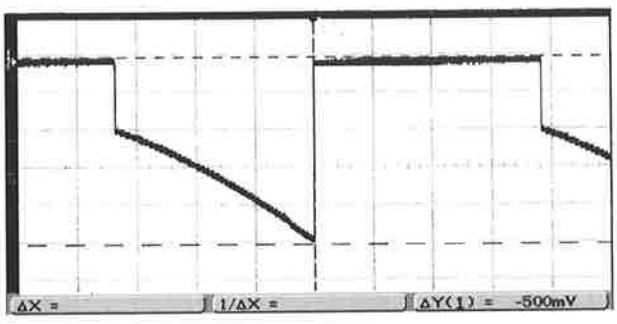

Figure 13.11 A DC voltage of 500 mV obtained across the reservoir capacitor using an oscilloscope with an input impedance of 1 MΩ and a capacitance of 4 pF.

To obtain a d.c. output of 1 volt as indicated in Figure 13.11 while sweeping across a 26 MHz frequency range, the minimum UHF input power required from the 50 Ω source provided by the network analyser was –46.20 dBW. The resulting return loss curve presented in Figure 13.12 shows that over 90% of the incident power at 825 MHz is feeding into the diode resonance.

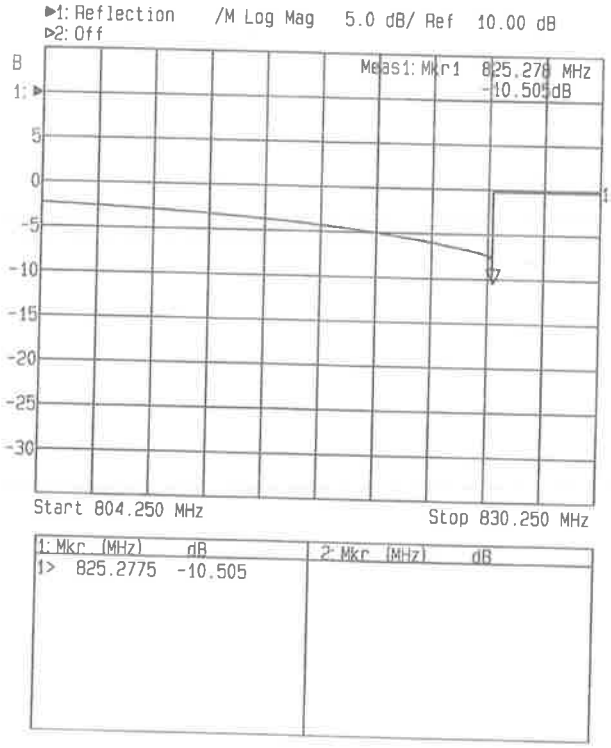

Figure 13.12 The return loss plot with the network analyser output RF power fixed to -16.20 dBm.

Calculation, using standard far field antenna formulae, of the range at which, for favourably oriented antennas, a reader with antenna gain of 6 dBi and output power of 1W, (as is allowed under the US regulations [45]) providing an available source power of -46.2 dBW from a tag antenna of gain 1.5, gives a range of 12.6 m.

However, an interesting and an important phenomenon can be observed when the signal sweep bandwidth at high power is reduced. Unless the sweep begins at a frequency that is somewhere near the low power resonance frequency, and follows upward in frequency as the diode develops voltage and begins to raise its resonant frequency, the full diode output will not be obtained. Zero power turn-on circuits have a limited range of operation due to the large voltage needed. However zero power turn-on circuits are useful for extending the battery life of battery assisted backscatter labels.

13.2.3 Turn on Range Estimation for a Low-Power Turn on Circuit

An alternative means that can still exploit the diode resonance in a turn-on circuit is to compare a small d.c. voltage developed across a diode with an internal reference voltage, and to activate a switch when the rectified voltage exceeds the internal value. In order for the low power turn-on circuit to be useful the current drain in its "off" state must be low with respect to the self discharge current of the battery. Unlike the previous turn on circuit the present design is triggered by a small d.c. voltage, rectified and amplified by diode resonance where the minimum value will be dictated by rectified RF noise [69 and 70].

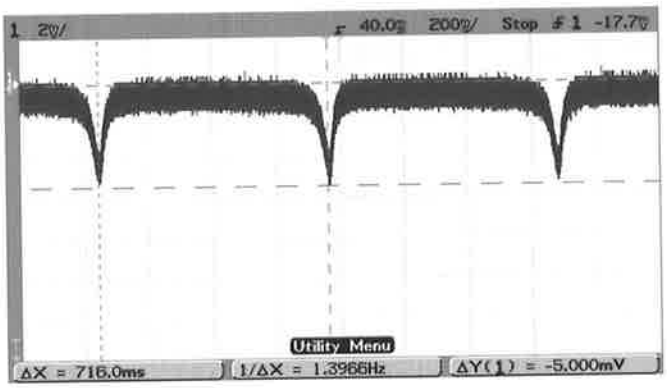

Figure 13.13 A DC voltage of 5 mV obtained across the reservoir capacitor using an oscilloscope with an input impedance of 1 MΩ and a capacitance of 4 pF.

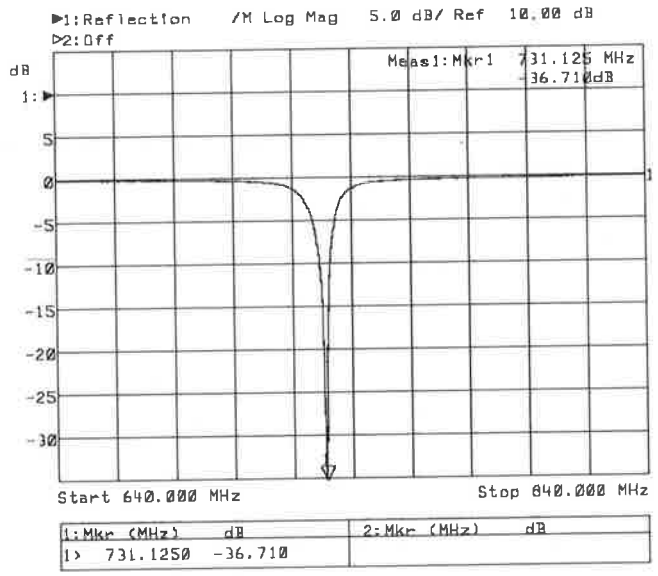

Figure 13.14 The return loss plot with the network analyser output RF power fixed to -43 dBm.

Experimental evidence has proved that a minimum RF power of -46 dBm is required at resonance to obtain a 5 mV d.c. output from the reservoir capacitor. Figure 13.13 shows the voltage output from the reservoir capacitor while Figure 13.14 shows the return loss plot at resonance.

Calculation, using standard far field antenna formulae, of the range at which, for favourably oriented antennas, a reader with antenna gain of 6 dB and output power of 1W will provide this available source power from a tag antenna of gain 1.5 gives a range of 390 m. The operation range makes these turn-on circuits suitable for independent reply generating active labels.

13.3 Design and Implementation

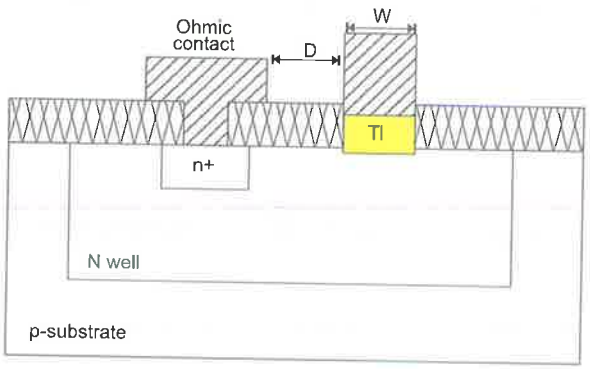

Figure 13.15 Cross-sectional view of the Schottky barrier diode. Here the Schottky diode contact width is *W*, and the separation between the Schottky contact and the Ohmic contact is *D* [70].

While it is possible to construct circuit topologies that utilize the diode resonance to turn a battery powered RFID label "on", the primary concern is the design and implementation of a Schottky diode with characteristics that are similar to those studied in HSMS-2820. In order for the circuit design to be cost effective it requires the diode design to be fabricated using a standard CMOS process. The possibilities of achieving such a goal are explored in [72 and 73].

It should be noted here that the Schottky diodes built on a standard CMOS process and used by the author to validate the experimental results presented in Section 13.2 were fabricated by Mr. David Hall but the concept to use them in the turn-on circuit tests described below and the experiments performed below are the work of the author.

The layout of the CMOS diode used in the turn-on circuits is outlined in Figure 13.15 while Figure 13.16 gives a voltage and current plot of the diode extracted jointly by Hall and the author.

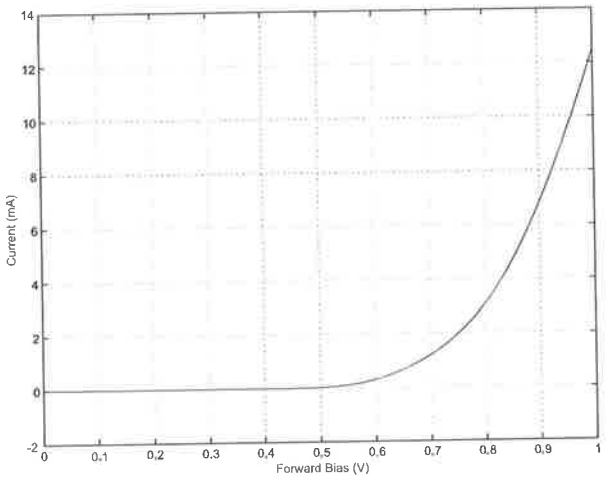

Figure 13.16 Measured IV curve of the CMOS Schottky barrier diode [70].

The diodes were fabricated using a standard 0.5 μm digital process. The Schottky diode was made by placing metal over a minimum size (0.5 μm x 0.5 μm) contact cut to a region of minimum size (3.5 μm x 2.4 μm) n-well [70]. The capacitance from the n-well to p-substrate was 4.6 fF. The diode had a dc voltage drop of 0.56 V at 427 μA, and an RF sensitivity of 14.23 mV/μW at 915 MHz when developing a rectified dc voltage of 4 mV. As a comparison, an HP HSMS-2820 Schottky diode had an RF sensitivity of 40 mV/μW, under similar conditions. At 915 MHz the diode had a series impedance of $21.10 + j173.66\ \Omega$ when the rectified d.c. output was 4 mV [70]. Diode sensitivity measurements were obtained using a network analyser and a double stub tuner to achieve a match to the diode input impedance measured using a network analyser. The rectified voltage was measured using an Agilent digital oscilloscope.

13.3.1 Zero Power Turn-On Circuit

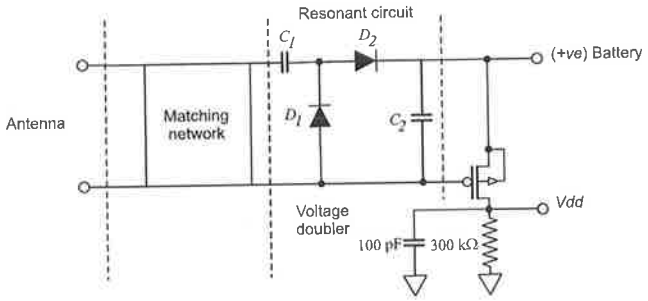

Figure 13.17 Turn on circuit implementation.

The proposed turn-on circuit topology in Figure 13.17 is adequate and cost effective for a backscattering active label. In this proposal a p-channel FET is used as a switch to control the power supply to a label's control circuits when triggered by incident RF radiation of a particular frequency on the label antenna. In the circuit outlined in Figure 13.17 the reactive power flowing through the junction capacitance of the Schottky diodes fabricated using standard CMOS technology, amplified by the quality factor of the Schottky diode resonance, is utilised to turn a p-channel FET from an off state to an on state. Figure 13.17 gives the general concept of the turn-on circuit while Figure 13.18 gives an implementation used in an HSPICE simulation of the circuit. In the circuit used for simulations (Figure 13.18), the antenna was modelled as a voltage source with an impedance of 5Ω. Again, the inductors are required to match the input impedance of the rectifier structure to the 5 Ω test input but the dependence of the junction capacitance on the biasing voltage was not modelled. Hence the simulations did not require the input RF signal to be swept over a 26 MHz frequency band.

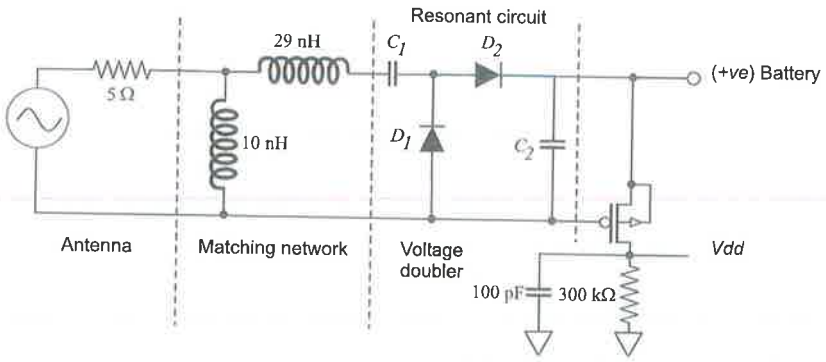

Figure 13.18 Turn on circuit implementation used in an HSPICE simulation.

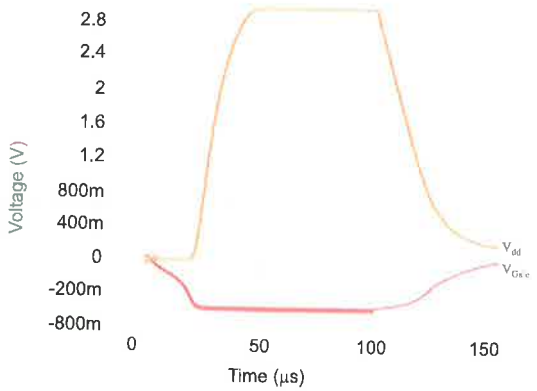

Figure 13.19 Simulated results for the turn on circuit implementation using a HSPICE diode model for the fabricated CMOS Schottky diode.

Simulation results (refer to Figure 13.19) indicate that the turn on circuit in Figure 13.18 performs adequately at a minimum available source power of -47 dBW at the resonant frequency of the turn on circuit (915 MHz). This result correlates with the measurement results performed previously in Section 13.2.2 with HP Schottky diodes. However the simulated results indicate a lower power requirement to generate the required turn on voltage in comparison to the measured results in Section 13.2.2. This is due to the elimination of losses in the experimental circuit and the use of a voltage doubler instead of a single rectification diode. Figure 13.19 shows the simulated output for a 915 MHz, -47 dBW pulse input to indicate the circuit switching off after removal of the excitation. The simple circuit requires the excitation for the time of transmission, but could be triggered by an RF pulse by latching the battery switch on for the time of transmission. While Figure 13.20 shows the simulated return loss curve of the turn on circuit, where the deep dip in excess of -30 dB indicates the resonance frequency of the turn on circuit.

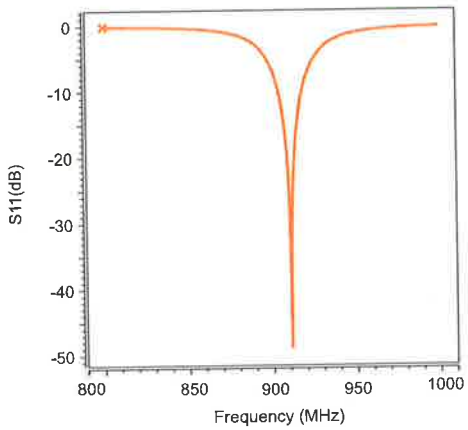

Figure 13.20 Simulated return loss curve of the zero power turn on circuit.

Calculations using standard far field antenna formulae of the range at which, for favourably oriented antennas operating at a frequency of 915 MHz, with antenna gain of 6 dBi and output power of 1W (as is allowed under FCC regulation for the UHF ISM band), providing an available source power of -47 dBW from a label antenna of gain 1.5, give a range of approximately 13m. Operation at this power level would thus enable a backscattering tag to be turned on and off in the range of 10 m under ETSI or FCC jurisdiction.

The zero power turn on is a practicable design that can be fabricated using a standard CMOS process. Simulation results and tests performed using discrete components show that the diode fabricated is adequate to obtain a desired level of performance for a zero power turn on circuit. However the complete fabrication of a zero power turn on circuit was not available for testing at this stage.

However for a zero power turn-on circuit to be of practicable significance the current drain of MOSFET must be less than the self discharge current of the battery. The zero power turn-on circuit presents a current drain of less then 0.1% of the self discharge current of a typical 3.5 V, 750 mA battery.

13.3.2 Low Power Turn-On Circuit

It is also possible to build a turn-on circuit based on achieving a small d.c. voltage across the reservoir capacitor [69 and 70] that exploits diode resonance investigated in Section 13.2.3. Such a low power turn-on circuit only requires a small d.c. voltage (of around 10 mV d.c.) to be developed by a Schottky diode rectifier, which is then compared with an internal reference voltage using a voltage comparator circuit. Hence the internal circuits of the RFID label are turned on when the rectified voltage exceeds the internal reference value. However unlike the previous turn-on circuit implementation the minimum turn-on voltage value will be dictated by rectified noise from other users of the frequency band and the intended operational temperature range [69 and 70].

Simulation results presented in [70] has shown that a rectified voltage of 4.2 mV d.c. output from the reservoir capacitor was achieved with a minimum RF power of –65.3 dBW at resonance. Calculations, using standard far field antenna formulae, of the range at which, for favourably oriented antennas operating at a frequency of 915 MHz, with antenna gain of 6 dBi and output power of 1W (as is allowed under FCC regulations for the UHF ISM band [45]), providing an available source power of -65.3 dBW from a label antenna of gain 1.5 give a range of 117.4 m. The long operation range of the low power turn-on circuit presents itself as a candidate solution for turning on independent reply generating labels.

Similar to a zero power turn-on circuit, a low power turn-on circuit is only useful when the current drain from the MOSFET switch [70] in its "off" state is lower with respect to the self discharge current of the battery on board the label. The low power turn-on circuit presented in [70] presents a current drain of less than 5% of the self discharge currents of a typical 3.5V, 750mA battery.

13.4 Acknowledgements

The solutions presented and analysed in this chapter was carried out collaboratively with David Hall. The Schottky diode used in the analysis was designed and fabricated by David Hall. The author would also like to thank Zheng Zhu and Peter H. Cole for their collaborations.

13.5 Conclusions

The development of active labels and sensors will eventually involve the incorporation of turn-on circuits. This chapter has presented some concepts and a number of ways in which they can be exploited.

The concept provided for a zero power turn-on circuit in Section 13.2.1 involves the design of a turn-on circuit that functions by sweeping the excitation across a UHF bandwidth. The concept of using a trigger RF wave and diode resonance is a practicable alternative and it is

illustrated through performance measurements taken in a scenario modelling a far field, and through range predictions under favourable conditions based on that scenario. The concept outlined and tested were used to design, analyse and simulate a zero power turn-on circuit based on a Schottky diode fabricated on a standard CMOS process. The simulations were performed at single frequency and no frequency sweeping was necessary as the diode junction capacitace of the simulation model was not dependent on the applied voltage across the diode junction. Hence it was not necessary to "chase" the resonace by sweeping across a wide frequency band. Experiments conducted in Section 13.2.2 have shown 1 V can be generated at an adequate distance from a trigger field creation source, by sweep through a 26 MHz frequency band, the frequency band of operation in the United States (refer to Table 6.1). Future work should consider reducing the sweep frequency range by sweeping from a high frequency value to a low frequency value so that the frequency sweep is able to meet the diode resonance instead of following the diode resonance as the junction capacitance of the diode varies with the applied voltage (refer to Figure 13.3).

It has been shown that the turning-on range can be increased by using a low power turn-on circuit triggered by a small d.c voltage obtained by rectifying and amplifying (by diode resonance) a received RF signal, as investigated in Section 13.2.1. Here the selected minimum voltage value for the turn-on circuit is determined by RF noise contributed by other users of the frequency band.

The turn-on circuits presented can be fabricated on a single poly, single metal CMOS process, allowing easy incorporation into existing transponder designs. At UHF frequencies the rectifier needs a CMOS process supporting Schottky diodes. Test results show that the practical realisation of the above concept in active labels is a possibility.

A turn-on circuit is permanently powered, however for it to be effective the current drain of the turn on circuit must be low with respect to the self discharge current of the battery. The low power turn-on circuit presented in the paper presents a current drain of less than 5% of the self discharge currents of a typical 3.5 V 750 mAh lithium battery while the zero power turn-on circuit presents a current drain of less than 0.1%.

A turn-on circuit may also be conceived as a field sensor. The following chapter will look at a MEMS based turn-on circuit for an active RFID label that can be used as a magnetic field sensor suitable for both a turn-on circuit for conserving the on-board battery of an active RFID label and theft detection.

Chapter 14

AN APPLICATION OF A MEMS BASED TURN-ON CIRCUIT

In the proliferation of RFID technology anti-theft labels are continuing to evolve. Often these active labels are employed for the tagging of expensive goods, with aim of both tracking and preventing the theft of the item. The battery, powering such active labels, must have very low internal and external current drain in order to prolong the life of the battery while being in a state of functionality to signal a theft of the labeled item. However due to circuit complexity or the desired operating range the electronics may drain the battery more rapidly than desired and the label may not last the shelf life of the product.

The theft detection mechanism presented in this chapter conserves power and thus prolongs the battery life of an active anti-theft label. A solution available for the development of such a theft detection circuit uses electroacoustic energy conversion using a MEMS device on a label IC to provide a high sensitivity turn-on circuit which acts as a field sensor. This chapter presents the results of an analysis conducted and published in [217] to evaluate the performance and the capabilities of such a theft detection circuit.

14.1 Introduction

The primary focus of this chapter is on active labels (Class III and Class IV labels) that will be equipped with a power source (such as a paper battery). In the category of active labels the most common objective is to obtain a long range in a battery-assisted backscatter label. However other types of active labels may not use backscatter but instead use a battery for powering and transmitting requirements. This chapter examines and presents a practicable solution for theft detection of high value tagged items by the implementation of a "screaming corridor", described later in Section 14.2. The solution presented is capable of conserving battery power and thus extending the lifetime of the battery while providing a high level of security in theft detection by use of high performance long lifetime theft detection labels.

14.2 Theft Detection Circuit

There are number of practical options for a theft detection circuit for a battery powered RFID label. Although very low quiescent power theft detection circuits are possible, the requirement of increasing the lifetime of RFID labels (perhaps due to longer shelf life of the labeled product) demands a zero quiescent power theft detection circuit. The proposed theft detection circuit is a zero power turn-on circuit for active RFID labels that will rely on generating a voltage of the order of 1V that can turn a CMOS transistor from fully off to fully on when triggered by a low frequency large volume magnetic field.

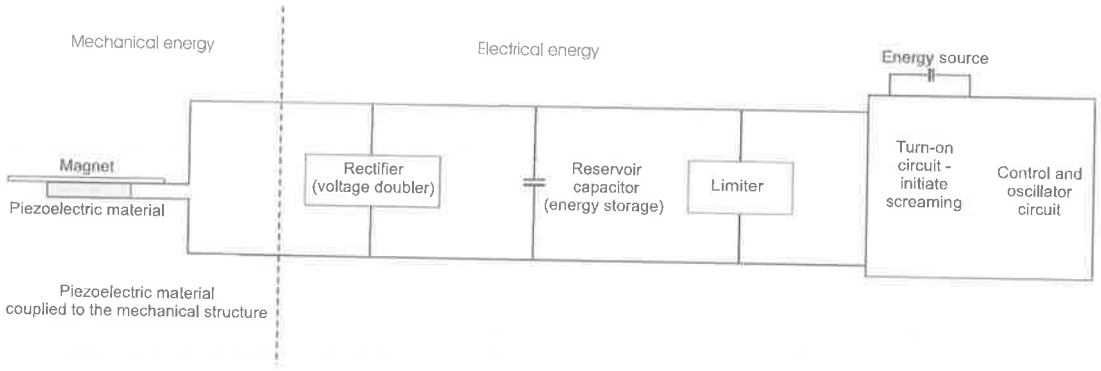

Figure 14.1 Components of the theft detection system.

The proposed approach utilizes a piezoelectric material to convert mechanical energy provided by an oscillating magnet into electric power as shown in Figure 14.1. There are four essential components described in Figure 14.1. The piezoelectric material coupled to a magnet (mechanical structure) is used to convert mechanical stress placed in the slab of piezoelectric material due to the forced oscillations induced on the magnet from an oscillating magnetic field, to electric power. The oscillating voltage is then rectified to

provide a reliable 1 V to turn on the switch operated by a FET (refer to Figure 14.8 for turn-on circuit schematic) to a theft alert circuit which will cause the label to "scream" by alerting interrogators of the theft and also provide a beacon for tracking the item. The theft detection process is illustrated in Figure 14.2. The low frequency large volume magnetic field provides the trigger for the MEMS circuit. Such a field can be setup in and around the vicinity of a large corridor exit to turn the MEMS theft circuitry "on" when a thief attempts to flee with stolen goods.

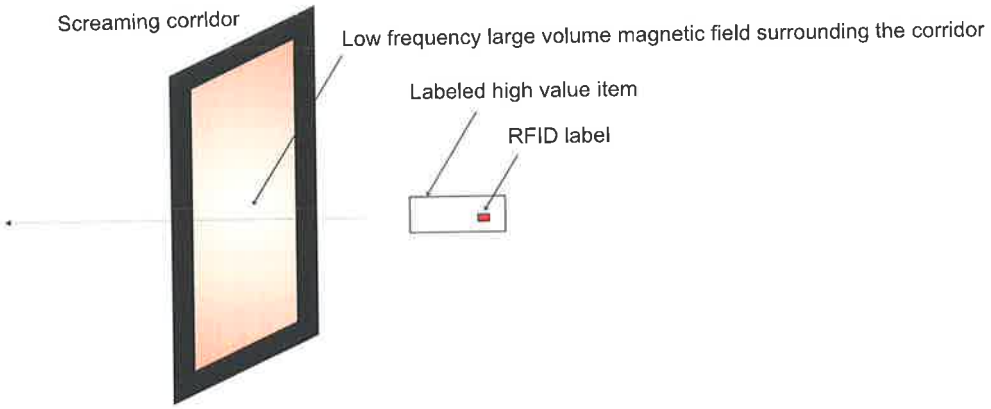

Figure 14.2 The "screaming corridor".

14.3 Magnetic-Electroacoustic Energy Conversion System

A piezoelectric material acts as a transformer element between mechanical energy and electrical energy. When an external mechanical force applied to a piezoelectric material strains the element, the unit cells of the crystal shift and realign. This results in the development of an electrostatic potential between certain opposing faces of the element. The relationship between the applied force and the resulting electric charge is dependent upon a number of characteristics inherent to the material as well as its size, shape and mechanical distortion.

The concept outlined here depends on being able to generate a large volume low frequency magnetic field in the vicinity of the active labels to harness the required mechanical energy such that the oscillating magnetic field will influence the magnet to oscillate with sufficient energy to generate through the piezoelectric effect a voltage V that is adequate to turn on a FET.

The intended frequency of operation of the turn-on circuit should allow the creation of a large volume magnetic field without exceeding the electromagnetic regulations regarding radiation (field confinement without undue radiation). In addition the intended frequency of operation should not consume excessive power to create large volume interrogation fields, and the field created should not be easily screened. It is also important that the frequency be

high enough such that ambient electromagnetic fields do not cause the turn on circuit to become operational unintentionally. All these considerations point to the use of a frequency around the LF electromagnetic spectrum. FCC electromagnetic compatibility regulations part 15, section 15.209 has defined the use of the LF electromagnetic spectrum for general use while CEPT/ETSI has an unlicensed band from 119-135 kHz, among other bands in the LF spectrum. Considering that the frequency range below 135 kHz is unlicensed and is available for general use around the world and taking note of the aforementioned considerations, 130 kHz was considered for creating a large volume, low frequency, magnetic field to power the turn-on circuit. It will be important to consider later the practically achievable magnetic field levels and any regulatory limits on them.

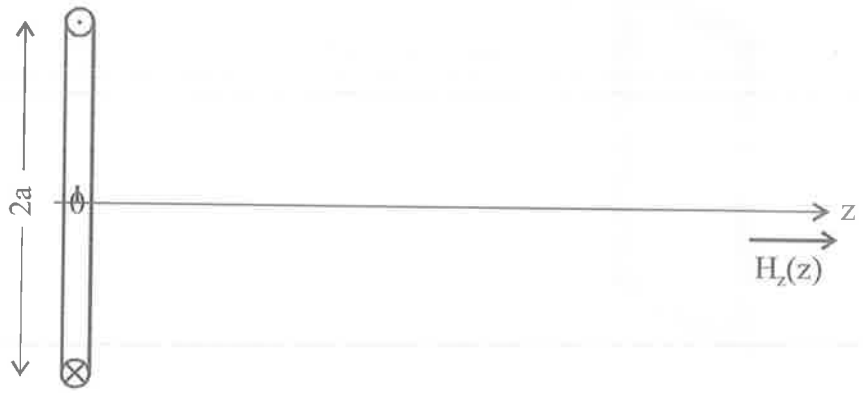

Figure 14.3 Geometry of the large coil.

Consider, as shown in Figure 14.3, a single turn circular planar coil of diameter d, used to create at a distance z a magnetic field $H_z(z)$, at a frequency f when there is r.m.s current I flowing through the coil. Then the magnetic field at a distance z is given by (14.1). Figure 14.4 shows the variation of the magnetic field strength as a function of the normalised distance from the coil. Clearly, creating a large volume magnetic field requires the use of a large loop structure. The following analysis will consider such a structure.

$$H_z(z) = \frac{Ia^2}{2\left(a^2 + z^2\right)^{3/2}}. \tag{14.1}$$

For the geometry considered in Figure 14.3, the total reactive power flowing per unit volume at a point where the magnetic field described by a real r.m.s. phasor H is given by

$$W_H = \omega\mu_0 |\mathrm{H}|^2. \tag{14.2}$$

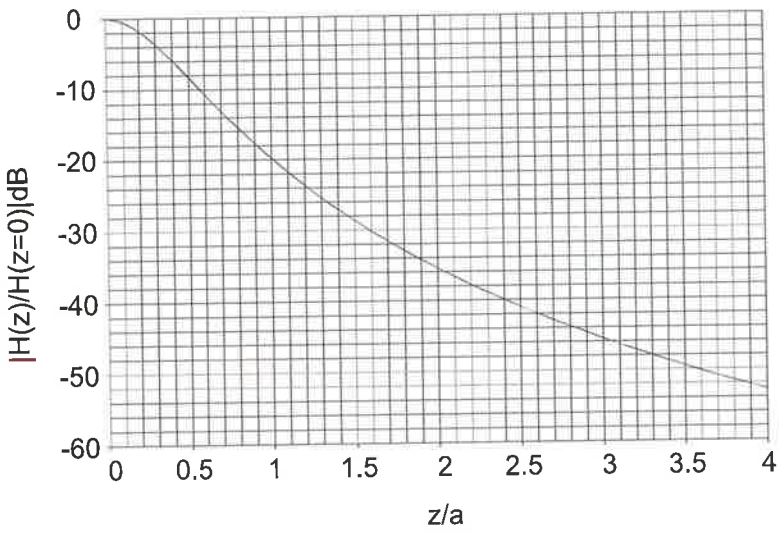

Figure 14.4 Relative magnetic field strength at a normalised distance (z/a) from a circular coil.

In the situation where skin currents flow on surface of the coil conductor the self-inductance is given by

$$L = \frac{\mu_0 D}{2}\left[\log_e\left(\frac{8D}{d}\right) - 2\right]. \tag{14.3}$$

The power P dissipated by an r.m.s current I flowing in the coil is given by (14.4), where R is the resistance of the coil or is an effective resistance that lakes into account all losses in the resonant circuit formed by the coil and a tuning capacitor normally placed in parallel with a lossless capacitor.

$$P = I^2 R \tag{14.4}$$

The quality factor of resonance can be obtained using the resistance of the coil wire. At a frequency of 130 kHz the skin depth effects must be taken into consideration. The skin depth of copper at 130 kHz is 183 μm. Considering a practicable geometry for the coil the following parameters outlined in Table 14.1 can be used to evaluate the reactive power density at distance z from the centre of the coil (refer to Figure 14.3).

Table 14.1 Field generating coil configuration.

Description	Symbol	Value
Coil diameter	$D(2a)$	3 m
Diameter of the coil wire	d	10 mm
Power supplied	P	50 W
Frequency of operation	f	130 kHz

A coil constructed with the dimensions outlined in Table 14.1 will have an inductance L of 10.9011 μH while the effective quality factor of the coil is expected to be around 60. (The calculated quality using the resistance of the coil based on its skin depth is a much higher value but in a practical implementation there are additional losses from tuning capacitors). The evaluations based on the following sections of this chapter will assume a magnetic field strength created by a pair of coils with 50 W of power supplied to each coil.

14.4 Analysis

Electroacoustic energy conversion systems have traditionally been modelled using mass-spring damper systems [218 and 219]. However we have used coupling relations to analyse the feasibility of the proposal.

14.4.1 Electroacoustic Energy Conversion

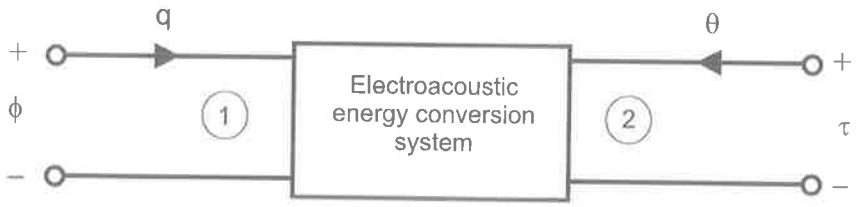

Figure 14.5 Representation of an electroacoustic conversion system [66].

The energy conversion system outlined in Figure 14.5 can be formulated using a pair of electrical terminals and a pair of mechanical terminals as indicated in Figure 14.5. The quantities represented by ϕ and q are the instantaneous applied voltage and charge that has entered the electrical port, and τ and θ are the instantaneous applied torque and angular displacement at the mechanical port. The electromechanical system can thus be described by the relationship given in (14.5) [66].

$$\begin{bmatrix} q \\ \theta \end{bmatrix} = \begin{bmatrix} C_{11P} & C_{12P} \\ C_{21P} & C_{22P} \end{bmatrix} \begin{bmatrix} \phi \\ \tau \end{bmatrix} \tag{14.5}$$

In the above matrix, C_{11P} is the input capacitance at the electrical port with zero torque applied to the mechanical port. C_{11P} for the slab piezoelectric material can be calculated in the thin parallel plate approximation by

$$C_{11P} = \frac{\varepsilon^T \varepsilon_0 w l_p}{h}. \tag{14.6}$$

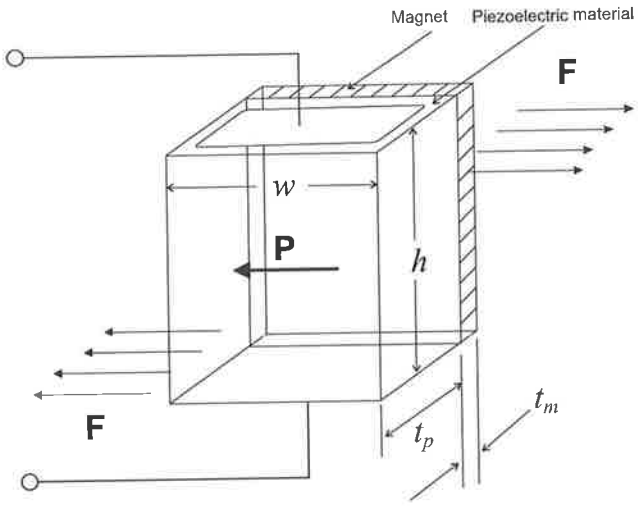

Figure 14.6 Magnet coupled to the piezoelectric material. Here **F** is the shearing force applied by the magnet and **P** is the direction of polarization of the piezoelectric material.

Here ε^T is the relative dielectric constant under constant stress, t_p is the thickness of the piezoelectric material (as well as the width of the electrode plates) while h is the height of the piezoelectric block and w is the width of the electrode plates or the piezoelectric material (refer to Figure 14.6). C_{22P} is the compliance of the piezoelectric structure given that the piezoelectric material compliance is s_P^E.

$$C_{22P} = \frac{s_P^E}{hwt_p} \tag{14.7}$$

The off diagonal element C_{21P} is the angular displacement at port 2 when a voltage is applied to port 1 and no torque is applied to port 2. C_{12P} gives the charge entering port 1 when it is short circuit and a torque is applied to port 2. The elements C_{21P} and C_{12P} are a result of the piezoelectric effect and these elements will be equal as an expression of reciprocity.

14.4.2 Electrical Power

A FET will require a turn on voltage of 1 V across the gate to turn it from an "off" state to an "on" state. This voltage is generated by the electroacoustic energy conversion system and rectified by the voltage doubler shown in Figure 14.1. Hence the voltage generated at the electrical port of the electroacoustic converter will have to serve an electrical load of approximately 1 pF and a series resistance of around 10 Ω presented by a Schottky diode used in the rectification structure. At the operational frequency range of 902-228 MHz, under FCC regulations, the electrical load is primarily the junction capacitance of the Schottky diode. Figure 14.7 shows the electroacoustic energy conversion system considering its electrical load.

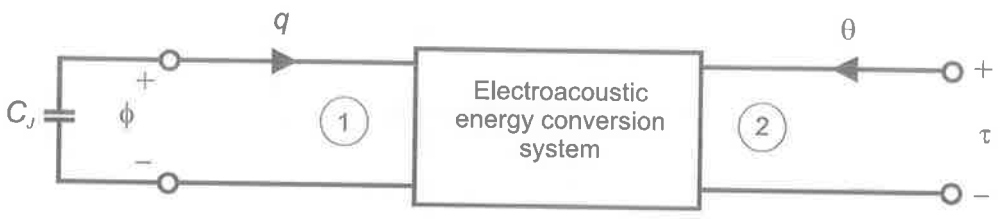

Figure 14.7 Representation of the electroacoustic energy conversion system with an electrical load.

The electroacoustic converter characterised above, connected to an external load of C_J will have a charge q developed at the port 1 when a torque τ is applied to port 2. The charge q is given by

$$q = -C_J \phi . \tag{14.8}$$

Solving the matrix given in (14.5) for ϕ gives

$$\phi = -\frac{C_{12P}\tau_P}{C_{11P} + C_J} . \tag{14.9}$$

Considering the dimensionless ratio $r = C_J/C_{11P}$

$$\phi = -\frac{C_{12P}\tau_P}{C_{11P}(1+r)} \tag{14.10}$$

Solving the matrix given in (14.5) for θ and using (14.10) gives

$$\theta = -\frac{C_{12P}C_{21P}\tau_P}{C_{11P}(1+r)} + C_{22}\tau_P \tag{14.11}$$

We can define k^2 as [220]

$$k^2 = \frac{C_{12P}C_{21P}}{C_{11P}C_{22P}} . \tag{14.13}$$

For a piezoelectric material k^2 is described as the electromechanical coupling factor of the material. For a piezoelectric material the coupling factor can be simply described as given in (14.12).

$$k^2 = \frac{\text{Mechanical energy converted to stored electrical energy}}{\text{Mechanical energy input}} \tag{14.12}$$

However for a structure in which we have uniform stress, strain, electric field and polarization it can be proved that the electromechanical coupling factor for a piezoelectric material is identical to that of the piezoelectric structure. Hence

$$\theta = \left[1 - \frac{k^2}{1+r}\right] C_{22P} \tau_P . \tag{14.14}$$

Thus we have an effective compliance for the capacitive loaded electroacoustic converter given by (14.15), where C_{22P} is the open circuit compliance at the electrical port.

$$C_{22eff} = \left[1 - \frac{k^2}{1+r}\right] C_{22P} \tag{14.15}$$

Observing that θ is proportional to τ, the stored energy for a final torque τ applied to the mechanical port is given as

$$E_{MP} = \frac{1}{2}\left(1 - \frac{k^2}{1+r}\right) C_{22P} \tau^2 . \tag{14.16}$$

The energy E_M expended at the mechanical port will appear as stored energy in the capacitor C_J, hence the electrical energy E_E at the electrical port is given as

$$E_{EP} = \frac{1}{2} C_J \left(\frac{C_{12P}^2}{C_{11P}^2 (1+r)^2} \right) \tau^2 . \tag{14.17}$$

Equation (14.18) can be obtained by using (14.16) and (14.17)

$$\frac{E_{EP}}{E_{MP}} = \frac{rk^2}{(1+r)^2 - k^2(1+r)} . \tag{14.18}$$

Equation (14.18) gives the effective electromechanical coupling factor k_{eff}^2 for the piezoelectric structure as

$$k_{eff}^2 = \frac{rk^2}{(1+r)^2 - k^2(1+r)} . \tag{14.19}$$

14.4.3 Mechanical Power

In a practical realisation of the turn-on circuit the mechanical stress on the structure is provided by the torque exerted on the structure in Figure 14.6 by the oscillatory magnetic field. If the magnitude of the total toque on the composite structure is represented by the r.m.s phasor T_S and is provided by the torque exerted by the field on the magnet, then

$$T_S = \mu_0 v M H . \tag{14.20}$$

Where ρ is the material density, v is the volume of the magnet and M is the remnant magnetisation. It should be noted that (14.20) assumes that the direction of the magnetisation M is orthogonal to the magnetic field H.

The mechanical stress on the structure is provided by a permanent magnet excited by an oscillating magnetic field where the resulting energy E_{MS} of the structure flowing into the stiffness of the structure can be evaluated as

$$E_{MS} = \left|\mathrm{T_S}\right|^2 C_{22S}\,. \tag{14.21}$$

C_{22S} is the total compliance of the structure at the mechanical port. C_{22S} is calculated as the harmonic mean of the effective compliance of the piezoelectric material C_{22eff} and the compliance C_{22M} of the magnetic structure. Thus, C_{22S} can be calculated as given in (14.22) where C_{22M} is given by (14.23), while s_M^E is the compliance of the magnetic material.

$$C_{22S} = \frac{1}{\left(\dfrac{1}{C_{22M}} + \dfrac{1}{C_{22eff}}\right)} \text{ where} \tag{14.22}$$

$$C_{22M} = \frac{s_M^E}{hwt_m} \tag{14.23}$$

From the total mechanical stress placed on the structure, only the mechanical stress on the piezoelectric material will result in transferring electrical energy E_{MP} into the stiffness of the piezoelectric. That electrical energy is given as

$$E_{MP} = \left|\mathrm{T_P}\right|^2 C_{22eff} \text{ where } \mathrm{T_P} = \mathrm{T_S}\left(\frac{C_{22S}}{C_{22eff}}\right). \tag{14.24}$$

14.4.4 Mechanical Resonance

The section above evaluated the energy power flow into the stiffness of the piezoelectric material. It can be viewed that the piezoelectric material is placed under a shearing stress as the magnet inclines to oscillate at the frequency of the surrounding magnetic field. However, at the mechanical resonance frequency of the combined structure with a mechanical quality factor Q_m the voltage developed at the electrical port is multiplied by the quality factor of mechanical resonance Q_m. Thus the voltage at the electrical port is increased by a factor of Q_m. Hence the energy E_{EPR} stored in C_J at the electrical port is given by (14.25).

$$E_{EPR} = k_{eff}^2 Q_m^2 E_{MP} \tag{14.25}$$

Using equation (14.24),

$$E_{EPR} = k_{eff}^2\, Q_m^2 \left|\mathrm{T_P}\right|^2 C_{22eff}\,. \tag{14.26}$$

Hence a magnet occupying a volume v, and a magnetisation constant M will provide the following energy flow into the capacitor C_J at the electrical port.

$$E_{EPR} = k_{eff}^2 Q_m^2 (Mv\mu_0)^2 |\mathrm{H}|^2 C_{22S}^2 \frac{1}{C_{22eff}} \tag{14.27}$$

14.4.5 Zero Power Turn-On Requirements

The r.m.s voltage V_{EPR} can be calculated as given in (14.28) using the relationship between the energy stored in a capacitor to an r.m.s voltage V_{EPR} applied across the terminals of that capacitor.

$$V_{EPR} = \sqrt{\frac{2E_{ERP}}{C_J}}. \tag{14.28}$$

For the production of a rectified output even to an open circuit load, a rectifying diode must experience a voltage across the junction capacitance of the order of or greater than the rectified output. Thus, the r.m.s voltage V_{TO} (V_{EPR}) available to turn on a FET can be calculated as follows

$$V_{TO} = \sqrt{2k_{eff}^2 Q_m^2 (Mv\mu_0)^2 |\mathrm{H}|^2 C_{22S}^2 \frac{1}{C_J C_{22eff}}}. \tag{14.29}$$

In order for the theft detection circuit to activate V_{TO} must be about 1 V. It should be noted here that in the event that the Schottky barrier diode junction capacitance is comparable to C_{11P}

$$k_{eff}^2 = \frac{k^2}{4 - 2k^2}. \tag{14.30}$$

14.5 Practical Evaluation

Figure 14.8 outlines a possible schematic for a turn on circuit based on a piezoelectric source. It should be noted here that the power lost in the rectification process will not be considered here, however it is sufficient to say that the turn-on voltage calculated is an absolute minimum turn on requirement. Nevertheless several circuits simulated showed that a voltage doubling rectifier used in the turn-on circuit is the optimal design that minimizes the diode losses.

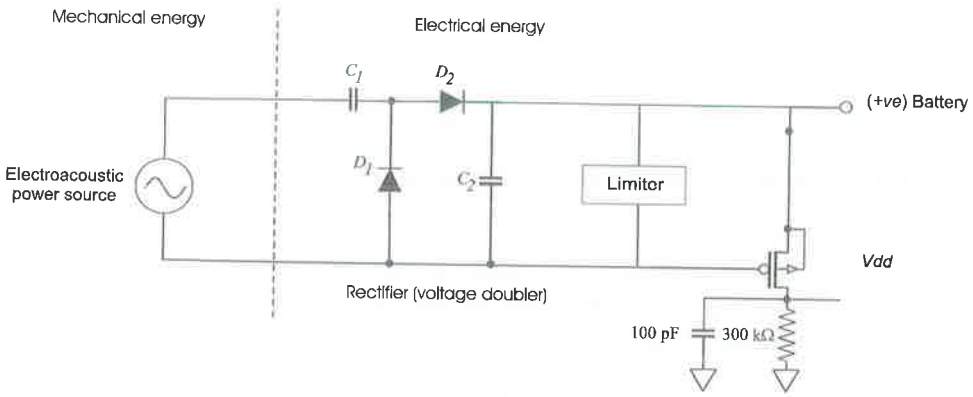

Figure 14.8 MEMS based theft detection turn on circuit schematic.

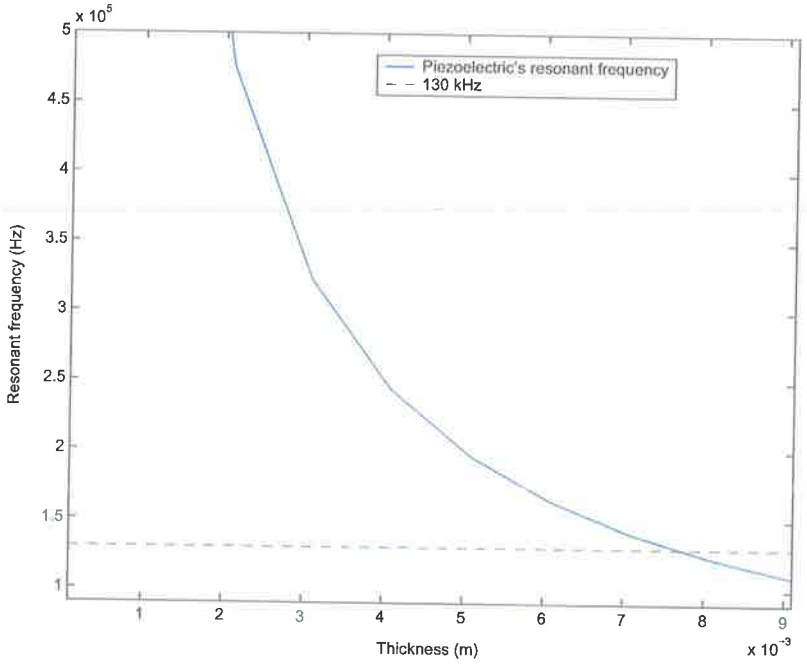

Figure 14.9 Resonant frequency of a PZT ceramic as a function of its thickness.

The piezoelectric material chosen for practical evaluation purposes is PZT, largely due to its higher coupling coefficient. It can be seen from (14.19) that a large k^2 is a desired characteristic. There are a variety of different PZT ceramics available with a diversity of characteristics while some ceramics can be custom made with desired characteristics [221 and 222]. However for the evaluation of the theft detection circuit a shear coupling coefficient, k_{15} of 0.69 and a frequency constant N_{15} of 1000 Hzm is used. Figure 14.9 shows the variation of the resonant frequency for a piezoelectric under a shearing stress. In order to ensure the efficient transfer of the energy density in the magnetic field to electrical energy at the output port of the piezoelectric material it is important to operate the piezoelectric

material at the resonant frequency of the structure. Ideally for the application under consideration the mechanical resonance frequency should be based around a point between the minimum impedance frequency (series resonance frequency) and the maximum impedance frequency (parallel resonance frequency) of the piezoelectric material. From Figure 14.9 it can be noted that a piezoelectric material of height h = 7 mm - 8 mm is required for the construction of the piezoelectric component if a mechanical resonance frequency of 130 kHz is to be achieved. The bulk compliance of PZT, s_{44} (under constant electric field) used was 30×10^{-12} m^2N^{-1} [221 and 222] while the relative permittivity was considered to be 200.

The magnetic material considered for the bar magnet should suitably have a high residual magnetisation M while occupying the largest possible volume. The latter is illustrated by both equations (14.27) and (14.29). Using a range of PtCo (Platinum and Cobalt) based alloys a residual magnetization M of 400,000 Am^{-1} is achievable. The bulk compliance of PtCo used in the evaluations was 10×10^{-12} m^2N^{-1}. While the compliance of the magnetic material is taken into account in the following simulations, its inertia is not taken into account. An assumption is made that the self resonant frequency of the magnetic structure is close to that of the PZT material. The simulated results can be improved in future by examining that the self resonant frequency of the magnetic material to ensure that it is indeed in the vicinity of 130,000 ± 100 Hz or there is for the composite structure a new and known mechanical resonance frequency in this region.

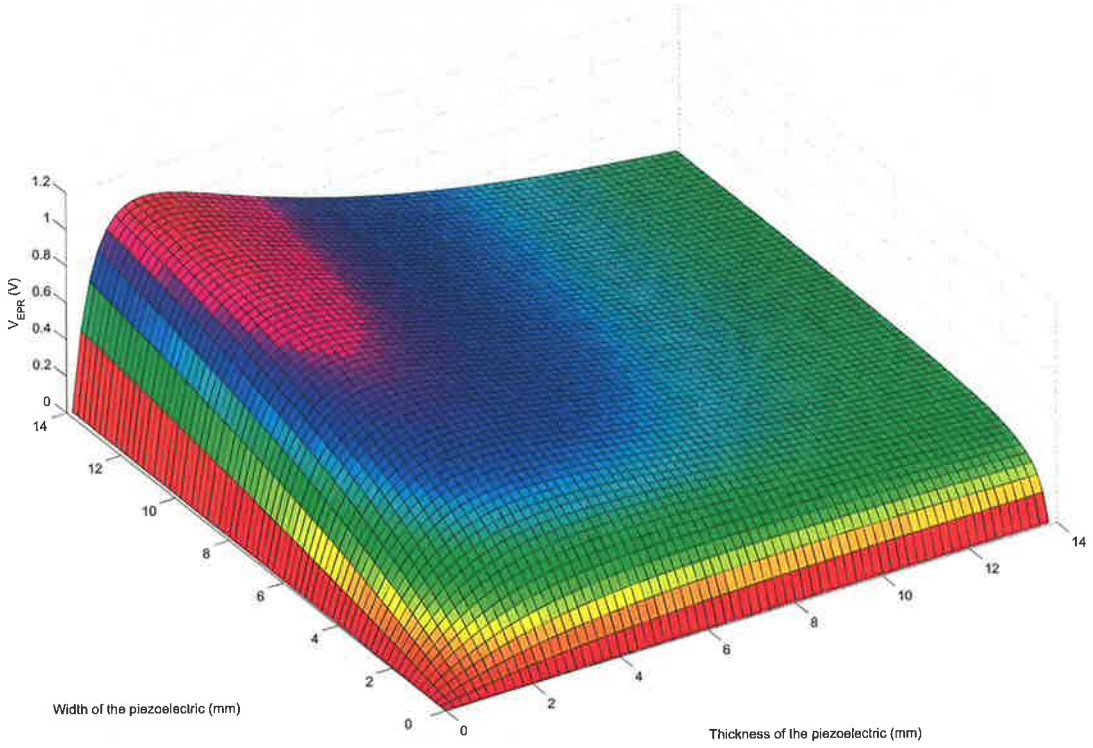

Figure 14.10 Effect of the piezoelectric structure dimensions on V_{EPR} at z = 2 metres from the "screaming corridor".

Figure 14.10 gives the variation in V_{EPR} at a distance of 2 metres from the loop structure generating the magnetic field. Here the effect of the width and the thickness of the piezoelectric are investigated. However the height of the piezoelectric was fixed to ensure mechanical resonance while the thickness of the magnet (t_m) was fixed to 3 mm and the width of the magnet was set to equal that of the piezoelectric.

As can be seen from Figure 14.10, increasing the PZT thickness has a gradually increasing effect on the voltage V_{EPR} that can be extracted from the electrical port. The thickness of the piezoelectric has a significant detrimental effect on performance beyond 3 mm. It can be seen that V_{EPR} is also a function of the width of the material. However there is an optimum value for the thickness of the piezoelectric to be used for a given width of the material. It can be observed from Figure 14.10 that the optimum thickness of the PZT for a given width of material is in the range of 2 mm – 3 mm.

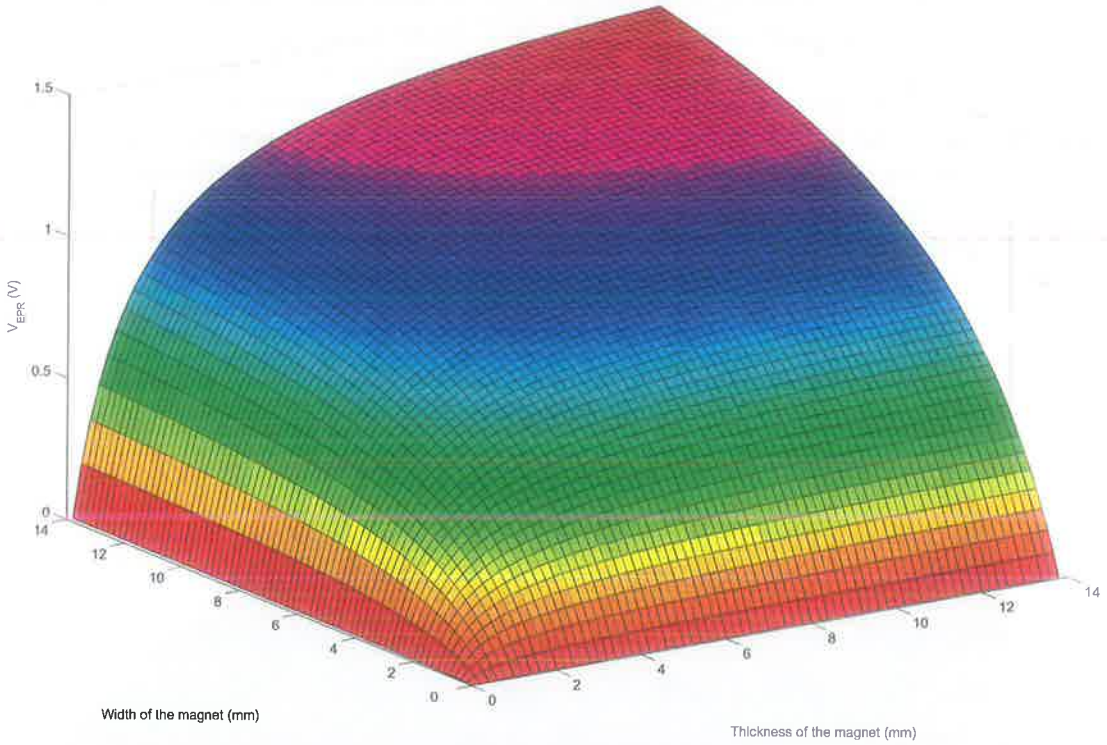

Figure 14.11 Effect of the magnetic structure dimensions on V_{EPR} at z = 2 metres from the "screaming corridor".

Figure 14.11 investigates the electrical power as functions of the width and thickness of the magnetic structure used in the application. In the simulation conducted the width of the magnet was set to be the equivalent to that of the PZT denoted by w in Figure 14.6. It is evident from Figure 14.11 that V_{EPR} is a strong function of the width of the magnetic structure as well as the thickness. This is as a result of the increase in the torque produced on the piezoelectric material by an increasing quantity of magnetic moment.

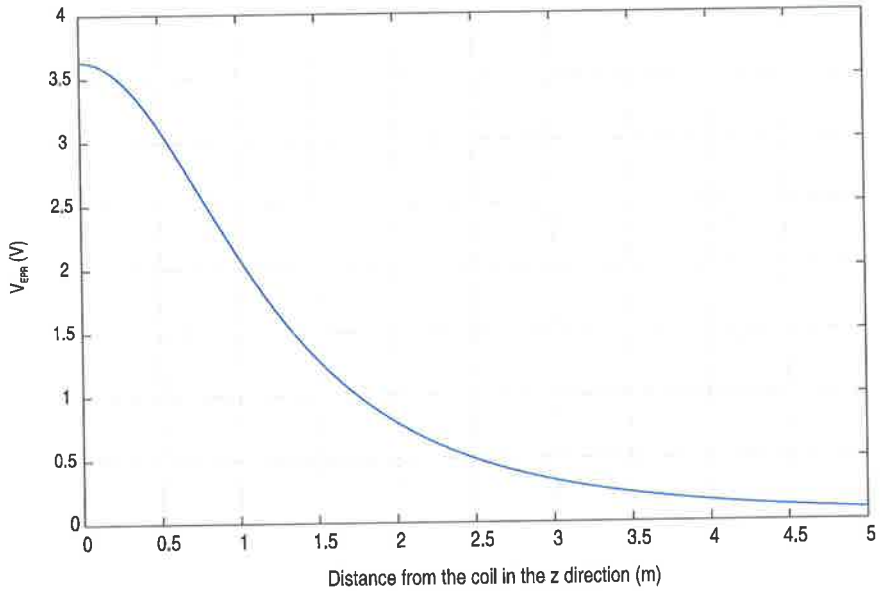

Figure 14.12 Turn-on range of the theft detection tag (measured from a "screaming corridor").

Based on the previous analysis, Figure 14.12 shows the $V_{EPR}(V_{TO})$ generated along a possible "screaming corridor". From the graph in Figure 14.12 a minimum operational distance of approximately 1.8 metres can be obtained for using the MEMS structure under consideration with the dimensions, w = 5 mm, t_p = 2.5 mm , t_m = 2 mm, and h = 7.5 mm to obtain an operating range of around 2 metres from the centre of the coil.

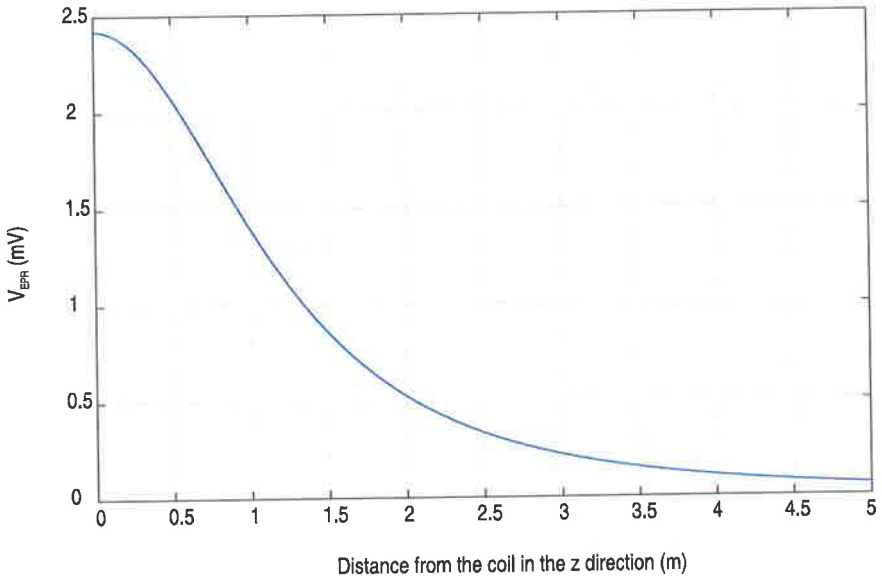

Figure 14.13 Turn-on range of the theft detection tag (measured from a "screaming corridor").

The mechanical Q of a piezoelectric material is very high; hence the MEMS device has a very narrow resonance bandwidth of around 100 Hz. This high Q_m yields a very high sensitivity result for the theft detection circuit. When the frequency of the magnetic field is reduced below the mechanical resonance frequency the significant advantage gained by the high Q_m will diminish rapidly. As a result it will greatly reduce the power transfer ratios and will eliminate the magnification of the energy at the electrical port. Figure 14.13 indicates the expected voltage from the device if mechanical resonance did not occur. In conjunction with the fact that the operating frequency is 130 kHz, the latter reality prevents other stray vibrations and fields from generating enough voltage to turn-on the theft detection circuits of the label to raise a false alarm.

The importance of Q_m factor to achieved the desired result emphasises the need to operate close to the resonance frequency of the structure and the need avoid raising a false alarm implies that the resonant frequency is set high enough to avoid the effects of stray fields. However this may not always be practicable as it will depend largely on selecting suitable magnetic and piezoelectric materials with desirable characteristics to optimise the energy at the electrical port. In addition the cost of the structure will eventually determine a practical implementation in an active theft detection label.

14.6 Acknowledgements

I would like to thank and acknowledge the work of Prof. Peter H. Cole in the formulation of the analysis presented in this section, as the analysis presented above would not have been possible without his guidance.

14.7 Conclusions

There is an increasing demand for high performance anti-theft RFID labels. The development of active RFID labels has helped overcome performance barriers of passive labels and provided useful solutions to preventing theft of high value products. The problem of designing a highly sensitive theft detection circuit that does not drain the limited power supply of an active label but instead works to prolong the lifetime of the battery powered label while at the same time providing security implemented through a "screaming corridor" has been examined in this Chapter.

The MEMS device developed exploits a combination of magnetic and piezoelectric effects for the generation of a turn on voltage to a theft detection circuitry. Unlike more traditional analysis of electromechanical systems, the analysis presented has used coupling relationships to investigate the extraction of mechanical energy from a magnetic field and investigated the feasibility of operating a trigger system based on a combination of magnetomechanical and piezoelectric effects within acceptable material volumes in a practical exciting field. The conclusion that sufficient energy transfer is possible at a distance of approximately 2 m from a "screaming corridor" has been established. The above analysis also addressed the issue of

adjusting material proportions to achieve resonance in the desired LF frequency range (100-135 kHz).

Future work will involve the investigation of the use of flexural mode to generate turn-on voltages and the study of the interplay between the electrode capacitance and the piezoelectric capacitance. The present size of the MEMS device needs to be further optimised in size and shape. Future work will also involve the simulation of the mechanical structure using tools provided by the ANSYS software package [233] to confirm the results obtained from analytical methods.

Appendix A

LIST OF FORMULAE AND SPICE MODEL

A.1 Inductance Calculations

Planar Circular Coil

The self inductance of a single-turn circular coil of diameter D made from wire of diameter d is, when the currents flow on the surface, given by

$$L=\frac{\mu_o D}{2}\left[\log_e\left(\frac{8D}{d}\right)-2\right]$$

An empirical but useful formula for the self inductance of a thin wire solenoidal coil of N turns wound over a length of l on a former of diameter of $2r$ as shown in the diagram below is

$$L=\frac{\mu_0 \pi r^2 N^2}{l+0.9r}.$$

Twin Wire Line

The self inductance L of a twin-wire line in which the conductors have diameter d and separation s is given by

$$L = \frac{\mu_o}{\pi} \text{arc cosh}\left(\frac{s}{d}\right)$$

$$\approx \frac{\mu_o}{\pi} \log\left(\frac{2s}{d}\right) \qquad when\ s >> d$$

A.2 Axial Field of a Circular Coil

In the magnetostatic approximation, the field at a point at a distance z along the axis of a single turn circular coil of radius a carrying a current I is given by

$$H_z(0,0,z) = \frac{Ia^2}{2\left(a^2 + z^2\right)^{\frac{3}{2}}}$$

A.3 Skin Effect

Skin depth δ in a metal at an angular frequency ω is given by

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu\sigma}}$$

The surface resistivity R_s per square due to skin effect is

$$R_s = \frac{1}{\delta\sigma} = \sqrt{\frac{\omega\mu}{2\sigma}}$$

and the wave impedance η at the surface is

$$\eta = (1+j)R_s$$

A.4 Radiation Resistances

Electric Dipole

The radiation resistance of a short electric dipole of length L, operating at a frequency for which the free space propagation constant has a magnitude β, is given in Ohms by

$$R_r = 20(\beta L)^2$$

Magnetic Dipole

The radiation resistance of a small current loop of radius a, operating at a frequency for which the free space propagation constant has magnitude β, is given in Ohms by

$$R_r = 20\pi^2 (\beta a)^4$$

Small loops of other shapes but the same area have the same radiation resistance.

A.5 SBD SPICE Model

```
% ################################################################
%
% SPICE model used for simulating the SBD used in Chapter 13
% turn-on circuit simulations
%
% ################################################################

.MODEL TurnOnSBD d
+ LEVEL=1
+ BV=9
+ CJO=0.7e-12
+ EG=0.69
+ IBV=10e-4
+ IS=2.2e-8
+ N=1.08
+ RS=5
+ PB=0.56
+ XTI=2
+ MJ=0.5
*
```

BIBLIOGRAPHY

1. Royal Air Force website, http://www.raf.mod.uk/history/line1940.html.

2. R. Ames, *Perspectives on radiofrequency identification*, Van Nostrand Reinhold, New York, 1990.

3. GS1 Home page, http://www.gs1us.org/.

4. A. L. Haberman, *Twenty-five years behind bars*, Harvard University Wertheim Publications, 2001.

5. J. Woodland, and B. silver, US Patent No. 2,612,994, Oct. 1952.

6. K. Finkenzeller, *RFID Handbook: Radio Frequency Identification Fundamentals and Applications*. John Wiley & Sons, New York, 1999.

7. Auto-ID Labs home page: http://www.autoidlabs.org.

8. International Organisation for Standardisation, "SI units and recommendations for the use of their multiples and of certain other units". International Standard ISO 1000 (1992).

9. P. H. Cole, D. M. Hall, M. Loukine, and C. D. Werner, "Fundamental constraints on RF tagging systems", in *Proceedings of the fourth annual wireless symposium and exhibition,* Santa Clara, pp. 294-303, February 1995.

10. K. Eshraghian, P. H. Cole, and A. K. Roy, "Electromagnetic coupling in subharmonic transponders", *Journal of Electrical and Electronic Engineering*, vol. 2, pp. 28-35,1982.

11. T. A., Scharfeld, "An analysis of the fundamental constraints on low cost passive radiofrequency identification system design", Masters thesis, Massachusetts Institute of Technology, August 2001.

12. D. L. Brock, "The Electronic Product Code – A Naming Scheme for Physical Objects", *Technical Report MIT-AUTOID-WH-002*, Auto-ID Center, January 2001. http://www.autoidlabs.org/researcharchive/.

13. EPCglobal Inc. home page, http://www.epcglobalinc.org.

14. EPCglobal Inc: "EPC Generation 1 Tag Data Standard Version 1.1 Rev. 1.27", (June 2005).

15. Sarma, S., and Engels, D. W., "On the Future of RFID Tags and Protocols", *Technical Report*, 2003. http://www.autoidcenter.org/research.

16. EPCglobal Inc., Specification for RFID air interface, http://www.epcglobalinc.org/standards_technology/EPCglobal2UHFRFIDProtocolV109122005.pdf.

17 Personal discussion with P. H. Cole and D. W. Engels

18. B. Roshan, and J. Leary, *802.11 Wireless LAN Fundamentals,* Cisco press, 2003.

19. B. Bing, *Broadband Wireless Access*, Boston Kluwer Academic Publishers, 2000

20. D. W. Engels, and S. E. Sarma, "The reader collision problem", IEEE International Conf. On systems, man and cybernetics, vol. 3, 2002, pp. 6-13.

21. B. Jamali, "High performance RFID systems", PhD Thesis, University of Adelaide, 2005.

22. Sarma, S., Towards The 5c Tag, Technical Report MIT-AUTOID-WH-006, 2001. http://www.autoidcenter.org/research/MIT-AUTOID-WH-006.pdf.

23. Hall, D., Senior Design Engineer, TAGSYS, Australia. Personal Conversation, July 2004.

24. RFID Journal news article, EM Micro Readies New RFID Chip, March 2003. http://www.rfidjournal.com/article/articleview/350/1/1.

25. Takaragi, T., Usami, M., Imura, R., Itsuki, R., and Satoh, T., "An Ultra Small Individual Recognition Security Chip", *IEEE Micro*, November-December 2001.

26. Auto-ID Labs home page: http://www.autoidlabs.org.

27. D. C Ranasinghe, K. Seong, M. Leng, D. W. Engels, and P. H. Cole, "A Distributed Architecture for a Ubiquitous Item Identification Network", *Smart Object Systems Workshop*, Japan, September 2005.

28. T. Erl, *Service-Oriented Architecture: Concepts, Technology, and Design*, Prentice Hall, 2004.

29. Auto-ID Center: Auto-ID ALE Engine specification 1.0, (Work in Progress), Oct. 2004.

30. Auto-ID Center: Auto-ID Object Name Service (ONS) 1.0, (Work in Progress), Oct. 2004.

31. P. Cole, and D. W. Engels, "Auto-ID – 21st century supply chain technology", *Proc. of the AEEMA Cleaner Greener Smarter Conference* (invited paper), Melbourne , 2002.

32. B. Atkinson, et al., "Web Services Security (WS-Security) Specification version 1.0.05", http://www-128.ibm.com/developerworks/webservices/library/ws-secure/, 2005.

33. IBM Corporation and Microsoft Corporation, "Security in a Web Services World: A proposed Architecture and Roadmap", white paper, http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwssecur/html/securitywhitepaper.asp, April 2002.

34. M. Harrison, "EPC Information Service (EPCIS)", *Auto-ID Labs Workshop*, Zurich 2004.

35. H. L. Lee, V. Padmanabhan, and S. Whang, "Information distortion in a supply chain: the bullwhip effect", *Management Science*, vol. 43(4), pp. 546-558, 1997.

36. G. Chappell, D. Durdan, G. Gilbert, L. Ginsburg, J. Smith and J. Tobolski, "Auto-ID in the Box: The value of Auto-ID Technology in Retail Stores", Auto-ID Centre White Paper, retrieved on 02 January 2006 http://www.autoidlabs.org/whitepapers/acn-autoid-bc006.pdf.

37. P. H. Cole, D. C. Ranasinghe, B. Jamali, "Coupling relations in RFID systems", Auto-ID Center white paper, June 2003.

38. M. N. O. Sadiku, *Elements of Electromagnetics*, 3rd ed., Oxford University Press, 2004.

39. P. H. Cole, D. C. Ranasinghe, B. Jamali, "Coupling relations in RFID systems II: practical performance measurements", Auto-ID Center workshop, June 2003.

40. P. H. Cole, "A study of factors affecting the design of EPC antennas and readers for supermarket shelves", Auto-ID Center workshop, October 2003.

41. W. L. Stutzman and, G. A. Thiele, *Antenna Theory and Design*, 2nd ed., John Wiley and Sons, Inc., New York, 1988.

42. D. K. Cheng, *Field and wave electromagnetics*, 2nd ed., Addison-Wesly Publishing, New York, 1989.

43. J. D. Kraus, R. J. Marhefka, *Antennas – For all applications*, 3rd ed., McGraw-Hill, New York, 2002.

44. ETSI, European Telecommunications Standards Institute, ETSI EN 302 208-1 V1.1.1 (2004-09), http://www.etsi.org/, 2006.

45. FCC Regulations, Title 47, Telecommunications, Chapter 1, Part 15, Radio frequency devices, http://www.fcc.gov, 2005.

46. G. H. Brown, and O. M. Woodward, "Experimentally determined radiation characteristics of conical and triangular antennas", *RCA Review*, pp. 425- 452, 1952.

47. J. D. Kraus, *Electromagnetics*, Third Edition, McGraw-Hill Series in Electrical Engineering, 1984.

48. C. A. Balanis, *Antenna theory: analysis and design*, John Wiley and Sons, New York, 1996.

49. Greg Pope, Michael Y. Loukine, David M. Hall, and Peter H. Cole "Innovative Systems Design for 13.56 MHz RFID", *Proceedings of the First Annual Wireless and portable Design Conference*, pp. 240, Sep. 15-18, 1997.

50. Peter H. Cole, and David M. Hall, "Metal Screened Electronic Labelling System", PCTWO0005675, PCT/AU99/00587.

51. S.-Y. Chen and P. Hsu, "CPW-fed folded-slot antenna for 5.8 GHz RFID tags," *Electronic Letters*, vol. 24, pp. 1516–1517, Nov. 2004.

52. M. Hirvonen, P. Pursula, K. Jaakkola, and K. Laukkanen, "Planar inverted-F antenna for radio frequency identification," *Electronic Letters*, vol. 40, pp. 848–850, Jul. 2004.

53. Q. Xianming and Y. Ning, "A folded dipole antenna for RFID," *Proc. IEEE Antennas and Propagation Soc. Int. Symp.*, vol. 1, pp. 97–100, Jun. 2004.

54. K. S. Leong, M. L. Ng, D. M. Hall and P. H. Cole, "A small passive UHF RFID tag for livestock identification", *IEEE 2005 International symposium on Microwave, antenna, propagation and EMC technologies for wireless Communications*, vol. 1, pp. 67-70, August 2005.

55. STMicorelectronics, XRA00, "UHF, EPCgloabal Class 1b, contactless memory chip 96 bit ePC with inventory and kill Function", Technical report, 2005.

56. Imping RFID Technology Series, "The RFID antenna: maximum power transfer", Technical report, 2005.

57. Alien Technologies, RFID tags, "ALC-140-xx RFID transponder IC", http://www.alientechnology.com/products/rfid_tags.php, date accessed January 2005.

58. M. B. Eunni, "A novel planar microstrip antenna design for UHF RFID", Master's thesis, A. M. A. College of engineering, Kancheepuram, Madras University, May 2004.

59. H. A. Wheeler, "Fundamental limitations of small antennas", *Proc. IRE*, vol. 35, pp. 1479-1484, Dec. 1947.

60. L. J. Chu, "Physical limitations of omni-directional antennas", *J. Appl. Phys.*, vol. 19, Dec. 1948, pp. 1163-11175.

61. J. S. McLean, "A Re-Examination of the Fundamental Limits on the Radiation Q of Electrically Small Antennas", *IEEE Trans. Antennas Propagat.*, vol. 44, no. 5, pp.672-676, May 1996.

62. H. A. Wheeler, "Small Antennas", *IEEE Trans. Antennas and Propagat.*, vol. AP-23, July 1975, pp. 462-469.

63. D. C. Ranasinghe, K. S. Leong, M. L. Ng, and P. H. Cole, "Small UHF RFID label antenna design and limitations", *IEEE International workshop on antenna technology: Small antennas and novel metamaterials*, March, 2006.

64. R. M. Fano, "Theoretical Limitations on the Broadband Matching of Arbitrary Impedances", *J. Franklin Inst.*, vol. 249, pp. 57-83, Jan. 1950.

65. R. H. Clarke, D. Twede, J. R. Tazelaar, K. K. Boyer, "Radio frequency identification (RFID) performance: the effect of tag orientation and packaging contents", *Packaging Technology and Science*, vol. 19(1), pp. 45 – 54, Nov. 2005.

66. Cole, P. H. "Coupling and quality factors in RFID", *Design, Characterisation and Packaging for MEMS and Microelectronics*, Paul D. Franzon, Editor, *Proceedings of SPIE*, vol. 4593, pp. 1-11, 2001.

67. D. C. Ranasinghe, and P. H. Cole, "Extending coupling volume theory", *IEEE International Workshop on Antenna Technology: Small Antennas and Novel Metamaterials*, March, 2006.

68. Federal Communication Authority (FCC) Regulations Part 15, 2005, http://www.fcc.gov.

69. D. M. Hall, and P.H. Cole, "A fully integrable turn-on circuit for RFID transponders," *Wireless and Portable Design Conference*, Burlington, Massachusetts, pp. 66-71, September 1997.

70. D. Hall, D. C. Ranasinghe, B. Jamali, P. H. Cole, "Turn-on circuits based on standard CMOS technology for active RFID labels" in *VLSI Circuits and Systems II, Proceedings of SPIE*, vol. 5837, p. 310-320, June 2005.

71. Agilent Technologies, "Schottky barrier diode for general purpose applications", in Technical data sheet, 2003, http://www.agilent.com.

72. B. Jamali, P. H. Cole, D. C. Ranasinghe, and Z. Zhu, "Design and optimisation of Schottky diodes in CMOS technology with application to passive RFID systems", *Smart Structures Devices and Systems II, Proceedings of SPIE*, vol. 5649, pp. 323-331, Feb 2005.

73. B. Jamali, D. C. Ranasinghe, P. H. Cole, "Design and optimisation of power rectifiers for passive RFID systems in monolithic CMOS", *Proceedings of SPIE*, vol. 6035, pp. 366-376, Dec. 2005.

74. Verichip corporation home page, http://www.4verichip.com/, date accessed Sept. 2004.

75. ITU, International Telecommunication Union, http://www.itu.int/home/index.html, 2006.

76. RFID Privacy and corporate data, *RFID Journal*, 2 June 2003, http://www.rfidjournal.com, accessed on August 2005

77. Gobioff, H., Smith, S., Tygar, J. D., and Yee, B., "Smart cards in hostile environments", *2nd USENIX Workshop on Electronic Commerce*, 1996.

78. Menezes, A., Van Oorchot, P. and Vanstone, S., *Handbook of Applied Cryptography*, CRC Press, 1996.

79. B. Schnier, *Applied Cryptography Protocols: Algorithms, and Source Code in C*, John Wiley & Sons, Inc, New York, 1994.

80. D. R. Stinson, *Cryptography Theory and Practice*, CRC Press, 1995.

81. R. A. Mollin, *Introduction to Cryptography*, Chapman & Hall/CRC, London, 2001.

82. SmartCode Debuts Smallest Chip, http://www.rfidjournal.com/article/articleprint/764/-1/1/, 23rd January, 2004.

83. Sarma, S., and Engels, D. W., RFID Systems, "Security & Privacy Implications", *Auto-ID Center white paper*, Feb 2003. http://www.autoidlabs.org/researcharchive/.

84. S. Bono, M. Green, A. Stubblefield, A. Juels, A. Rubin and M. Szydlo, "Security analysis of a cryptographically-enabled RFID Device", *Proceedings of 14th USENIX Security Symposium*, pp 1-16.

85. J. Westhues. "Hacking the prox card", *RFID: Applications, Security and Privacy*, Addison-Wesley, pp. 291-300, 2005.

86. K. Albrecht, "Chipping workers poses huge security risks", February 2006, http://www.freemarketnews.com/Analysis/139/3812/2006-02-15.asp?wid=139&nid=3812,

87. Z. Ker and A. Wool, "Picking virtual pockets using relay attacks on contactless smartcard systems", *Proceedings IEEE/CreateNet SecureComm*, pp. 47-58, 2005.

85. A. Juels, "RFID Security and Privacy: A research Survey", RSA Laboratories, September 2005.

86. M. R. Rieback, R. Crispo, and A. S. Tanenbaum, "Is your cat infected with a computer virus?", *Fourth IEEE International Conference on Pervasive Computing and Communications (percom)*, pp. 169-179, 2006.

87. Y. Oren, and A. Shamir, "Power analysis of RFID Tags", accessed on March 2006, http://www.wisdom.weizmann.ac.il/~yossio/rfid/.

88. G. Avoine and P.Oeschlin, "RFID traceability: a multilayer problem", Financial Cryptography, 2005.

89. Weigart, S.H., "Physical security devices for computer subsystems: a survey of attacks and defences". *Workshop on Cryptographic Hardware and Embedded Systems, LNCS*, vol. 1965, pages 302-317.

90. Anderson, R, and Kuhn, M., "Low cost attacks on tamper resistant devices", *International Workshop on Security Protocols, LNCS*, 1997.

91. E. Bovenlander, Invited talk on smartcard security, Eurocrypt 97.

92 Boycott Benetton web site, accessed Dec. 2005, http:/www.boycottbenetton.com.

93. M. Benetton, "Benetton explains RFID privacy flap". RFID Journal, 23 June 2004, http://www.rfidjournal.com/article/articleview/471/1/1/.

94. M. Roberti, "Analysis: RFID and Wal-Mart", September 2003, http://www.cioinsight.com/article2/0,1540,1455103,00.asp.

95. A. Jha, "Tesco tests spy chip technology", The guardian, July 19, 2003, http://www.guardian.co.uk/uk_news/story/0,3604,1001211,00.html.

96. Spychips web site, http://www.spychips.com, date accessed Dec. 2005.

97. J Collins. Marks & Spencer expands RFID retail trial, RFID Journal, 10 February 2004.

98. D. Molnar and D. Wagner, "Privacy and security in library RFID: Issues, practice, and architectures", B. Pfitzmann and P. McDaniel, editors, *ACM Conference on Communications Security*, pp. 210-219. ACM Press, 2004.

99. "RFID Upgrades Gets Goods to Iraq", *RFID Journal*, 23 July 2004.

100. F. Stajano, and R. Anderson, "The resurrecting duckling: security issues for ad-hoc wireless networks", *International Workshop on Security Protocols, LNCS*, vol. 1796, pp 172-194, 1999.

101. R. Clarke, "Introduction to data surveillance and information privacy and definition of terms", August 1997, http://www.anu.edu.au/people/Roger.Clarke/DV/ Intro.html#Id, date accessed January 2006.

102. B. Subirana, and M. Bain, *Towards Legal Programming of Software Agents. Research Monograph*, Kluwer. 2004.

103. Commonwealth Freedom of information Act 1982, Australia.

104. Commonwealth Privacy Act 1988, Australia.

105. Electronic Privacy Information Centre, EPIC web site, accessed March 2004, http://www.epic.org.

106. A. Beresford, and F. Stajano, "Location Privacy in Pervasive Computing", Pervasive computing, January-March 2003.

107. A. Juels, and R. Pappu, "Squealing euros: privacy protection in RFID Enabled banknotes", *Financial Cryptography*, 2002.

108. S. A. Weis, S. E. Sarma, R. L. Rivest, and D. W. Engels, "Security and privacy aspects of low-cost radio frequency identification systems", *Security in Pervasive Computing*, 2003.

109. A. Juels, and S. A. Weis, "Defining Strong Privacy for RFID", RSA Laboratories, 2006.

110. J. M. Rabaey, A. Chandrakasan and B. Nikolic, *Digital integrated circuits - A design perspective*, 2nd Edition, Prentice Hall, New Jersey, 2003.

111. J. Rabaey, and M. Pedram, *Low-Power Design Methodologies*, Kulwer Academic Publishers, 1996.

112. G. Avoine, "Adversary model for radio frequency identification", Technical Report, Security and Cryptography Laboratory, Swiss Federal Institute of Technology; Lausanne; 2005.

113. Y. Nohara, S. Inoue, K. Baba, and H. Yasuura, "Quantitative evaluation of unlinkable ID matching schemes", *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*; ACM Press, Pages 55-60, 2005.

114. M. Aigner, "Crypto implementations for RFID tags", presentation, Graz University of Technology.

115. S. Wolfram, *A New Kind of Science*, 2nd edition, Wolfram Media, 2002.

116. S. Wolfram, "Cryptography with cellular automata", *Advances in Cryptology: Crypto '85 Proceedings, LNCS*, vol. 218, pp 429-432, 1986.

117. J. Daemen, R. Govaerts, and J. Vandewalle, "Hash functions based on block ciphers: a synthetic approach", Advances in Cryptology, LNCS. Springer-Verlag, 1991.

118. A. Sen, C. Shaw, D. R. Chowdhuri, N. Ganguly, and P. P. Chaudhuri, "Cellular automata based cryptosystem (CAC)", *LNCS*, vol 2513, pp 303-314,2002.

119. S. R. Blackburn, S. Murphy, and K. G. Paterson, "Comments on "theory and applications of cellular automata in cryptography"", *IEEE Transactions on Computers*, vol. 46, no. 5, pp 637-638, May 1997.

120. P. H. Bardell, "Analysis of Cellular Automata Used as Pseudorandom Pattern Generators", Proceedings of 1990 International Text Conference, pp. 762-768.

121. C. H. Meyer and W. L. Tuchman, "Pseudo-random codes can be cracked", Electronic Design, vol. 23. Nov. 1972.

122. C. H. Meyer and W. L. Tuchman, "Design considerations of cryptography", Proceedings of the NCC, vol. 42, Montvale, N.J. AFIPS Press, pp.594-597, Nov. 1972.

123. J. Hoffstein, J. Pipher, and J.H. Silverman, NTRU: A Ring-Based Public Key Cryptosystem, In *Proceedings of ANTS III*, Portland, June 1998.

124. D. Micciancio, "The hardness of the closest vector problem with pre-processing", *IEEE Transactions on Information Theory*, vol. 47, no 3, pages 1212-1215, March 2001.

125. O. Goldreich, S. Goldwasser, and S. Halvei, "Public-key cryptosystems from lattice reductions problems", MIT LCS, 1996.

126. R. J. McEliece, "A public key cryptosystem based on algebraic coding theory", *JPL Pasadena*, 1978.

127. NTRU web site, accessed August 2003, http:www.ntru.com/products/genuid.html

128. D. Wheeler, and R. Needham, "TEA, a Tiny Encryption Algorithm", Computer Laboratory, Cambridge University, England, 1994. Accessed at http://www.ftp.cl.cam.uk/ftp/papers/djw-rmn/djw-rmn-tea.html in June 1995

129. F. Stnadaert, G. Piret, N. Gershenfeld, and J. Quisquater, "SEA: A scalable encryption algorithm for small embedded applications", *CARDIS 2006, LNCS* 3928, pp. 222-236, 2006.

130. A. Juels, and R. Pappu, "Squealing euros: privacy protection in RFID-enabled banknotes", *Financial Cryptography, LNCS*, vol. 2742, pp. 103-121, 2003.

131. M. Aigner, and M. Feldhofer "Secure symmetric authentication for RFID tags", *Telecommunications and Mobile Computing TCMC2005*, March 8th-9th, 2005.

132. M. Feldhofer, S. Dominikus, J. Wolkerstorfer, "Strong authentication for RFID systems using the AES algorithm", *Proceedings of the Workshop on Cryptographic Hardware and Embedded Systems, LNCS*, vol. 3156, ISBN 3-540-22666-4, pp. 357-370, Springer Verlag, 2004.

133. J.Wolkerstorfer, "Is elliptic-curve cryptography suitable to secure RFID tags?", *Workshop on RFID and Light-Weight Cryptography*, Graz (Austria), 2005.

134. F. Martin, A. Manfred, D. Sandra, "An application of RFID tags using secure symmetric authentication", *Proceedings of 1st International Workshop on Privacy and Trust in Pervasive and Ubiquitous Computing*, pp. 43–49, ISBN 960-531-179-8, Santorini Island, Greece, July 14, 2005.

135. S. Tillich, J. Großschädl, "Accelerating AES using instruction set extensions for elliptic curve cryptography", *Proceedings of Computational Science and Its Applications, LNCS*, vol. 3481, pp. 665–675, May 9-12, 2005.

136. L. Batina and J. Guajardo and T. Kerins and N. Mentens and P. Tuyls and I. Verbauwhede, "An Elliptic Curve Processor Suitable For RFID-Tags, *Cryptology ePrint Archive*, Report 2006/227, 2006, http://eprint.iacr.org/

137. N. J. Hopper and M. Blum, "Secure human Identification, Protocols", *LNCS*, vol 2248 pp. 52, 2001.

138. A. Juels and S. Weis. "Authenticating pervasive devices with human protocols", *Advances in Cryptology, Crypto 2005, LNCS,* vol. 3621, pp. 293-308, 2005.

139. J. Katz, J. S. Shin, "Parallel and Concurrent Security of the HB and HB+ Protocols", *Eurocrypt* 2006.

140. T. Dimitriou, "A Lightweight RFID Protocol to Protect Against Traceability and Cloning Attacks", *Proceedings of IEEE Conference on Security and Privacy for Emerging Areas in Communication Networks - SECURECOMM*, 2005.

141. S. Pramuthu, "HB and related lightweight authentication protocols for secure RFID tag/reader authentication", *CollECTeR Europe Conference*, Basel, Switzerland.

142. D. Ranasinghe, D. W. Engels, P. H. Cole, "Security and Privacy Solutions for Low Cost RFID Systems", Proc. of the 2004 Intelligent Sensors, Sensor Networks & Information Processing Conference, Melbourne, Australia. pp. 337-342, 14-17 December, 2004.

143. A. Juels, "Minimalist cryptography for low cost RFID tags", *LNCS* vol. 3352, pp.149-164, Springer-Verlag, 2001.

144. Szewczykowski, United States Patent, Patent number 5818021,Date of patent Oct. 6 1998.

145. Cole, P. H., Secure Data Tagging Systems, In International Patent Application, Applicant TagSys Australia Pty. Ltd, Patent number PCT/AU02/01671,Date Feb. 10 2003.

146. H. Chabanne, and G. Avoine, "Noisy Cryptographic Protocols for low RFID tags", *Workshop on RFID lightweight Crypto*, 2005.

147. C. Castelluccia and G. Avoine, "Noisy tags: a pretty good key exchange protocol for RFID tags", *International Conference on Smart Card Research and Advanced Applications (CARDIS'06)*, Spain, April 2006.

148. A. Juels, R.L. Rivest and Mszydlo "The blocker tag: selective blocking of RFID tags for consumer privacy", V. Atluri editor, *8th ACM conference on Computer and Communications Security*, pp. 103-111., ACM Press, 2003.

149. Jules 2005, T Daniels, M. Mina, S. Russell, "A small fingerprinting paradigm for physical layer security in conventional and sensor networks", in *IEEE/CreateNet Secure Comm*, 2005 to appear.

150. G. P. Hancke, M. G. Kuhn, "An RFID distance bounding protocol", *IEEE SecureComm*, Athens, 2005.

151 Tsudik, G. "YA-TRAP: Yet