Far Field RF Power Extraction Circuits and Systems

by

Soumyajit Mandal

B.Tech Electronics and Electrical Communication Engineering
Indian Institute of Technology, 2002

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

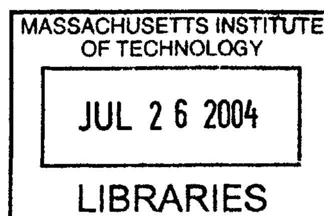
June 2004

© Massachusetts Institute of Technology 2004. All rights reserved.

Author ....
Department of Electrical Engineering and Computer Science
May 12, 2004

Certified by...
Rahul Sarpeshkar
Associate Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students


MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUL 2 6 2004
LIBRARIES

BARKER

Far Field RF Power Extraction Circuits and Systems

by

Soumyajit Mandal

Submitted to the Department of Electrical Engineering and Computer Science
on May 12, 2004, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering

Abstract

In this thesis, I describe efficient methods for extracting DC power from electromagnetic radiation. This will become an important necessity for a number of applications involving remotely powered devices, such as Radio Frequency Identification (RFID) tags and bionic implants. I first investigate the problem abstractly, allowing theoretical bounds on system performance to be derived. Next I devise circuit, antenna and impedance matching network design strategies to efficiently approach these theoretical bounds. Finally, I use these strategies to create an experimental power extraction system that collects RF power at low electromagnetic field strengths. This system enables a substantial increase in the operating range of remotely powered devices.

Thesis Supervisor: Rahul Sarpeshkar
Title: Associate Professor of Electrical Engineering

Acknowledgments

First of all, I would like to thank my advisor, Prof. Rahul Sarpeshkar. Rahul has been unfailingly supportive and encouraging throughout this project. I couldn't have done it without him. Thanks also go out to my academic advisor, Prof. John Wyatt, for his support.

Many thanks to Prof. Neil Gershenfeld for letting me use his group's RF test equipment, Prof. Scott Manalis for permission to use the wire bonder, and Prof. Joe Paradiso for access to the reflow solder system.

Next a word for all the people in the Analog VLSI & Biological Systems group, especially Chris Salthouse, Keng-Hoong Wee, Maziar Tavakoli-Dastjerdi, Micah O'Halloran, Ji-Jon Sit, Michael Baker & Heemin Yang. You guys rock! Thanks also to our group's long suffering secretary, Rose Englund, and to Ben Vigoda, Manu Prakash and Peter Russo for helping out at various crucial junctures.

Special mention for the nice people in Symbol Technologies, especially Jerry Swartz, Raj Bridgelall, Mark Duron & Dick Knadle, for believing in this project.

A big round of applause to my friends all over the world. You guys make life worth living. A special word here for Kalyanjit Ghosh and Siddharth Ray, two people who have made graduate life memorable.

I would like to acknowledge my wonderful family, especially my parents. What can I say? Thanks for everything. All I am today is because of you.

And, before I get too sentimental, here's a list of other things to be thankful about:

- Google
- IEEE Xplore
- Matlab®
- LaTeX, WinEdt, MiKTeX
- Mapquest

- Agilent, Keithley, Fluke...
- MOSIS
- Dragon Boat Club of Boston
- Hot coffee, clean air, the Charles river...

Contents

1 Introduction **23**

1.1 Background 23

1.2 Motivation and Context 24

1.3 Backscatter RFID Systems 25

2 System Design **27**

2.1 Introduction 27

2.2 Global Considerations 27

2.2.1 Robustness 27

2.2.2 Cost 28

2.2.3 Performance 29

2.2.4 Physical Size 30

2.3 Internal Considerations 31

2.3.1 Impedance Matching 32

2.3.2 The Bode-Fano Criterion 36

2.3.3 The L-Match 45

2.3.4 Efficiency 47

2.3.5 Packaging & Assembly 54

2.4 RFID Protocol Issues 59

2.5 System Testing 60

2.5.1 Test Setup 61

2.5.2 Test Results 61

3 **Rectifier Design** **69**
3.1 Introduction 69
3.2 Common Rectifier Structures 70
3.3 Schottky Based Rectifiers 70
3.4 The Four Transistor Cell 72
3.4.1 Cascading Rectifier Cells 75
3.4.2 Startup Behavior 77
3.4.3 Threshold Voltage 79
3.4.4 Body Effects 82
3.5 Floating Gate Structures 85
3.5.1 Fowler-Nordheim Tunneling 85
3.5.2 Rectifier Structures 90
3.6 Performance Optimization 94
3.6.1 Layout Issues 94
3.6.2 Circuit Optimization 116
3.6.3 Input Impedance Measurement 127

4 **Antenna Design** **131**
4.1 Introduction 131
4.2 Transmission Lines 131
4.2.1 Coplanar Strips Line 132
4.2.2 Parallel Strips Line 133
4.3 Transmit Antennas 135
4.3.1 Typical Transmit Antennas 135
4.3.2 Final Design 135
4.4 Receive Antennas 139
4.4.1 Typical Receive Antennas 139
4.4.2 Final Design 141
4.5 Antenna Tuning Networks 147
4.5.1 Algorithm 148

4.5.2 Current Reference 149

5 Summary and Conclusions 155

5.1 Summary 155

5.2 Future Work 156

A Non-standard CMOS Devices 159

A.1 Schottky Diodes 159

A.1.1 Type 1: metal to n-well 161

A.1.2 Type 2: metal to epitaxial p-substrate 166

A.1.3 DC Modeling 168

A.1.4 RF Performance 173

A.1.5 Matching across dies and wafers 178

A.2 Solar Cells 180

A.2.1 Properties of Solar Cells 180

A.2.2 CMOS Solar Cells 184

B The Mathematics of F-N Tunneling 191

B.1 Introduction 191

B.2 Tunneling Current 191

B.3 Tunneling Into Floating Nodes 193

List of Figures

2-1 The impedance matching problem. The antenna impedance on the left has to be matched to the load impedance on the right for maximum power transfer. 33

2-2 Illustrating the derivation of the Bode-Fano impedance matching criterion. A parallel R-C load with a quality factor Q_L is shown. 37

2-3 Reflection coefficient characteristics of impedance matching networks of various orders. 39

2-4 Plot of the function $f\left(|\Gamma_m|\right)$ in (2.21). 42

2-5 Plot of the function $g\left(Q_L\right)$ for various values of the fractional bandwidth $\frac{\omega_0}{B}$. 43

2-6 The simple L-match. For maximum power transfer, $X_A = -X_C$ and $R_A = R_S$. 46

2-7 Typical shapes of the rectifier function $F(\bullet)$ for different values of I_L. $I_{L1} < I_{L2} < I_{L3}$. 49

2-8 Theoretical efficiency curves for parameters in text and C_S varying from 0.4 pF to 2 pF. 51

2-9 Theoretical maximum power up range curves for parameters in text and C_S varying from 0.4 pF to 2 pF. 52

2-10 Theoretical power up threshold curves for parameters in text and C_S varying from 0.4 pF to 2 pF. 53

2-11 Circuit used for efficiency calculations, with R_s being a series representation of rectifier losses. The load resistance $R_L = \frac{V_L}{I_L}$ is transformed into an additional loss component R_{eff} at the rectifier input. 54

2-12 Circuit used for efficiency calculations, with R_p being a parallel representation of rectifier losses and R_{eff} representing power delivered to the load. 55
2-13 A fairly generic lumped single section package model showing three neighboring pins. Non nearest-neighbor coupling between pins has been ignored. 56
2-14 Typical RF feed scenario. The input RF signal is fed into the rectifier through the package. Z_{in} includes all package capacitances. 59
2-15 Setup used for testing the power extraction system. 61
2-16 Photograph of completed power extraction system. 62
2-17 Measured frequency response of the tag for different available power levels P_A and $I_L = 2\ \mu$A. 64
2-18 Measured load curves (varying load current I_L) of the tag for different available power levels P_A. 65
2-19 Measured output voltage of the tag versus available power level P_A for different load currents I_L. 66
2-20 Measured power conversion efficiency η of the tag for different available power levels P_A. 67
2-21 Measured transient response of the tag to amplitude modulation of the carrier. The modulation frequency was 100 Hz and modulation depth was M % 68

3-1 N-stage capacitive voltage multiplier circuit using diode doubler stages. 71
3-2 Simulated rectification curves (output DC voltage V_L versus RF input voltage V_{in}) for Schottky diode rectifiers with different values of N (the number of stages). $I_L = 2\ \mu$A. 72
3-3 The four transistor cell using MOS devices. Body (bulk) terminals are not shown. 74
3-4 Typical steady state waveforms produced by the four transistor switching cell for (a) square wave and (b) sinusoidal inputs. 75

3-5 The switching cell configured as a rectifier. NMOS bulk terminals are tied to the chip substrate, which is taken to be the ground terminal. . 76

3-6 A rectifier formed by cascading three switching cells in series. For power extraction applications V_{RF} is generated by an antenna and impedance matching network. 76

3-7 Parasitic body-source and body-drain junction diodes for (a) NMOS and (b) PMOS transistors. 77

3-8 The four transistor cell with MOS parasitic diodes included (a) actual and (b) simplified. A n-well CMOS process is assumed. 78

3-9 Typical transient startup waveforms produced by the four transistor switching cell for sinusoidal inputs. 79

3-10 Effect of changing the PMOS and NMOS threshold voltages (at zero V_{BS}) VTP_0 and VTN_0 on the output DC voltage of a five stage rectifier (simulated). The input RF amplitude was 1 V at 900 MHz and the load current was 2 μA. 81

3-11 The four transistor cell using MOS devices in a n-well CMOS process. NMOS bulk terminals are tied to the chip substrate, which is taken to be the ground terminal. 83

3-12 Body bias effect on rectifier performance for different load currents I_L. Solid lines show the output voltage when NMOS bulks are grounded ($V_{SB} \geq 0$). Dashed lines show the output when NMOS bulks are tied to sources ($V_{SB} = 0$, no body effects). 84

3-13 Simulated rectification curves (output DC voltage V_L versus RF input voltage V_{in}) for rectifiers with different values of N (the number of stages). $I_L = 2$ μA. 86

3-14 Energy band diagrams for F-N tunneling in metal-oxide-semiconductor structures for (a) zero bias, (b) negative bias and (c) positive bias on the metal electrode. 88

3-15 Canonical floating gate transistor structures. (a) NMOS and (b) PMOS. 89

3-16 Floating gate transistor structures suitable for use in rectifiers to shift transistor threshold voltages. (a) NMOS and (b) PMOS. 90
3-17 The four transistor cell using floating gate MOS transistors. Body (bulk) terminals are not shown 91
3-18 Experimental variation of output voltage V_L with change in threshold voltages ΔV_{TN} and $\Delta |V_{TP}|$ for a five stage floating gate rectifier. . . 92
3-19 Simulated rectification curves (output DC voltage V_L versus RF input voltage V_{in}) for floating gate rectifiers with different values of N (the number of stages). $I_L = 2$ μA, $C_P = 0.3$ pF. 93
3-20 Alternative representations of a reactive element with finite Q. (a) Series and (b) Parallel. 95
3-21 Transistor geometry with multiple gate fingers. Only four fingers are shown; in general there can be N fingers. 97
3-22 Single L-R-C section used to model interconnect between terminals 1 and 2. 98
3-23 RC transmission line model of an interconnect of length l open circuited at one end. 99
3-24 Contact finger geometry for reducing well resistance. The well has dimensions $W \times L$ and is divided by the fingers into n rectangles. . . 101
3-25 Generic model of an integrated circuit capacitor. 103
3-26 Cross sectional geometry of integrated parallel plate capacitors (a poly1-poly2 capacitor is shown). Drawing is not to scale. 104
3-27 Simplified model of an integrated circuit parallel plate capacitor. Terminal 2 is the bottom plate. 105
3-28 View of parallel plate capacitor geometry. The top plate is assumed to be lossless, and the bottom plate has a sheet resistance ρ. 106
3-29 Illustrating the geometry of the circular parallel plate capacitor problem. (a) Polar coordinates and (b) a differential sector element. . . . 107
3-30 Tapered 1D RC transmission line representation of the parallel plate capacitor problem. $R(0) \to \infty$ and $C(0) \to 0$. 108

3-31 Calculated Q of capacitors at 900 MHz for different values of ρ. C_a=900 aF/μm^2, corresponding to poly1-poly2 capacitors in AMI's 0.5 μm process. 110

3-32 Calculated resistance of capacitors at 900 MHz for different values of ρ. C_a=900 aF/μm^2, corresponding to poly1-poly2 capacitors in AMI's 0.5 μm process. 111

3-33 AC (charge storage) models for (a) diodes, (b) PMOS transistors and (c) NMOS transistors, showing the various parasitic capacitors. PMOS transistors are assumed to be fabricated inside n-wells. 113

3-34 Circuit model used to find optimum pump capacitance C_P when the RF signal V_{in} is fed to the (a) top and (b) bottom plate of C_P. 119

3-35 Calculated effect of pump capacitor C_P sizing on rectifier input voltage V_G. Values have been normalized. 120

3-36 (a) A single voltage doubler stage and (b) its model at zero input amplitude. 122

3-37 Rectifier figure of merit for Schottky diode based voltage multipliers with different package capacitances C_{pack} for C_P = 0.25 pF, V_L = 1.5 V and I_L = 2 μA. 124

3-38 Figure of merit for floating gate based switching rectifiers with different package capacitances C_{pack} for C_P = 0.30 pF, V_L = 1.5 V and I_L = 2 μA. Plot normalized w.r.t. Figure 3-37. 125

3-39 Figure of merit for switching rectifiers with different package capacitances C_{pack} for C_P = 2 pF, V_L = 1.5 V and I_L = 2 μA. Plot normalized w.r.t. Figure 3-37. 126

3-40 Simulated rectification curves of a three stage floating gate rectifier for various load currents I_L. 127

3-41 Typical measurement scenario. The device under test can only be accessed through a test fixture. 128

3-42 Measured input reflection coefficient Γ_{in} of a three stage rectifier in a LCC package after model based de-embedding for different input power levels. 129

4-1 Cross section of coplanar strips (CPS) transmission line geometry. The dielectric substrate is infinitely wide and has a thickness h (which may also be infinite). 132

4-2 Calculated characteristic impedance Z_0 of coplanar strips transmission line. 133

4-3 Cross section of parallel strips transmission line geometry. The dielectric substrate is infinitely wide and has a thickness d. 134

4-4 Differential (parallel strips line) to single ended (back to back microstrip lines) impedance conversion. 134

4-5 Structure of the planar Yagi antenna with parallel strips line feed. Metal 1 and Metal 2 are on opposite sides of a thin dielectric layer. . 136

4-6 Measured and simulated reflection coefficients (in dB) of the Yagi antenna into a 50 Ω load. 138

4-7 Coupled resonator impedance matching networks. The two circuits are duals of each other. 142

4-8 Examples of lumped impedance inverters (a) K-type and (b) J-type. . 143

4-9 Equivalent circuit of the rectifier (chip), antenna and impedance matching network. 144

4-10 Physical structure of the final tag antenna. The rectifier chip is fed differentially from ports 1 and 2. 145

4-11 Simulated input impedance $R_A + jX_A$ of the tag antenna with matching network, matched to $R_{in} - jX_{in}$, the input impedance of the chip. . 146

4-12 Simulated reflection coefficient magnitude $|S_{11}|$ of the tag antenna when connected to the chip input impedance for R_{in} = 4, 7 and 10 Ω. 148

4-13 CMOS current reference. The startup circuit is shown in gray. 151

4-14 Measured output current I_{ref} of the current reference versus power supply voltage V_{DD}. Results from five different chips are shown. . . . 153

4-15 Measured startup time T_{start} of the current reference versus power supply voltage V_{DD}. 154

A-1 Energy band diagram of a metal / n-type semiconductor junction at equilibium. 160

A-2 Energy band diagram of a metal/ p-type semiconductor junction at equilibrium. 161

A-3 Cross section of metal / n-well Schottky diode. The device is fingered in order to reduce series resistance. 162

A-4 Layout of a metal / n-well Schottky diode with 10 contacts. 163

A-5 Measured DC I-V characteristics of metal / n-well Schottky diodes with 10, 20, 80 and 320 contacts. Wells were connected to the chip substrate (ground). 164

A-6 Simple circuit models for integrated circuit diodes. (a) metal / n-well type and (b) metal / p-substrate type. 165

A-7 Measured DC I-V characteristics of metal / n-well Schottky diodes with 10, 20, 80 and 320 contacts. Wells were unconnected (floating). . 166

A-8 Measured DC I-V characteristics of metal / n-well Schottky diodes. Each diode has 320 contacts but different values of d, the minimum spacing between Schottky and Ohmic well contacts. 167

A-9 Cross section of metal / p-epitaxial layer Schottky diode. The device is fingered in order to reduce series resistance. 168

A-10 Measured DC I-V characteristics of metal / p-substrate Schottky diodes with 10, 20, 80 and 320 contacts. 169

A-11 Modeling DC I-V characteristics of metal / n-well Schottky diodes using a single diode (ignoring series resistance). 171

A-12 Modeling DC I-V characteristics of metal / p-substrate Schottky diodes using a single diode (ignoring series resistance). 172

A-13 Improved circuit models for integrated circuit Schottky diodes. (a) metal / n-well type and (b) metal / p-substrate type. 173
A-14 Modeling DC I-V characteristics of metal / n-well Schottky diodes using two back to back diodes (ignoring series resistance). 174
A-15 Modeling DC I-V characteristics of metal / p-substrate Schottky diodes using two back to back diodes (ignoring series resistance). 175
A-16 Approximate zero bias RF models of integrated Schottky diodes. (a) metal / n-well type and (b) metal / p-epitaxial type. 176
A-17 Experimental rectification curve (at 2 MHz) of a voltage doubler implemented with Schottky diodes compared with that predicted by simulation. Both diodes were of the metal / n-well type. 178
A-18 Experimental rectification curve (at 2 MHz) of a voltage doubler implemented with Schottky diodes compared with that predicted by simulation. The freewheeling diode was of the metal / p-substrate type. . 179
A-19 Measured DC I-V characteristics of metal / n-well Schottky diodes with 20 and 320 contacts for 7 different dies on the same wafer. Data was current limited to 1 mA. 180
A-20 Measured DC I-V characteristics of metal / p-substrate Schottky diodes with 20 and 320 contacts for 7 different dies on the same wafer. . . . 181
A-21 Measured DC I-V characteristics of metal / n-well Schottky diodes with 10, 20, 80 and 320 contacts for two different process runs (wafers) 182
A-22 Schematic of a generic solar cell. I_L is the load current and V_{DC} is the output voltage. 183
A-23 Theoretical I-V curves of a p-n junction exposed to light. 184
A-24 Measured open circuit voltage obtained from a eight diode CMOS solar cell as a function of the incident illuminance. 185
A-25 Possible phototransducing structures in a vanilla CMOS process (a) n^+ active to p^- substrate, (b) n-well to p^- substrate, (c) p^+ active - n-well - p^- substrate. 186
A-26 Optical absorption length $L(\lambda)$ for silicon at room temperature (300 K).187

A-27 Equivalent circuit of a CMOS solar cell (a) using transistors and (b) using diodes and photocurrents. 188

A-28 Measured load curves of a eight diode CMOS solar cell for different illuminance levels. Each diode was on a separate chip. Absolute illuminance values should be treated as approximate. 190

List of Tables

2.1 Test setup parameters 63

4.1 Printed circuit board parameters 132

A.1 Calculated Schottky diode parameters 170

A.2 Typical CMOS process parameters 186

Chapter 1

Introduction

In this introductory chapter I describe some of the background for this thesis. In other words, I try to establish a context for the research which is discussed in later chapters.

1.1 Background

The problem of powering up autonomous systems has two possible solutions. One solution is transmitting power to the system remotely. This usually involves RF energy, though one can envision systems where laser light, for example, can be used instead. RFID systems fall into this category. The other possibility is energy harvesting, where the system utilizes natural energy sources in its environment to power up. Such energy sources can take a wide variety of forms, such as ambient RF signals [18], light, thermal gradients, mechanical vibrations [31] and convection flows.

Remotely powered devices have become increasingly important over the past few years. One of the most important applications is for Radio Frequency Identification (RFID). RFID technology looks set to have far reaching implications for the global supply chain [16]. Other applications include bionic implants and low power sensor networks. These devices must include a power extraction system that can extract enough DC power from incident electromagnetic radiation for the device to function. In this context, I should now define what I mean by *far field* power extraction systems.

In antenna design, far field refers to distances from the antenna r which are large enough such that there is negligible reactive energy storage in the neighborhood. The only RF energy is in the form of a propagating electromagnetic wave, the power density of which drops off as $\frac{1}{r^2}$. Since reactive fields drop off faster than $\frac{1}{r^2}$, far field conditions are always satisfied for large enough values of r. Precisely what this value should be is however doubtful, and at least 10 different definitions exist. All these values are $\sim \lambda$, the electromagnetic wavelength in the medium. I shall therefore assume $r >$ 'a few λ' as a suitably vague definition of far field. More importantly for power extraction applications, far field conditions imply low power densities which have to be collected efficiently in order to power up circuitry. By a far field power extraction system, one therefore means something that can efficiently power up a load at distances more than a few wavelengths away from a RF source. At UHF frequencies, the maximum power up range for such systems is currently limited in practice to a few meters.

1.2 Motivation and Context

In this thesis, I have concentrated on the first step in getting any remotely powered system to work: powering it up. This is also frequently the hardest step when the incident power levels are low. Rectification, i.e., converting bidirectional power (like that in RF or mechanical vibrations) into unidirectional DC power becomes extremely difficult at low power levels. Fundamentally, this is because rectification, like any other frequency conversion, is a nonlinear operation, and all physical systems and devices look linear for small signal amplitudes. Several design strategies exist for minimizing the power consumption of active circuits. In contrast, the designer of low power rectifiers very soon runs into basic physical limitations. This thesis examines these limitations analytically and using simulations. Any performance predictions are then verified against experiments. It is my belief that my results have implications for a wide range of remotely powered systems.

1.3 Backscatter RFID Systems

The constraints and tradeoffs involved in RFID system design have been analyzed previously [33]. The commonest RFID systems use passive (batteryless, or remotely powered) tags. Tags communicate with the reader using backscatter modulation. In this technique, the tag transmits data by changing (modulating) its reflection coefficient to incoming radiation from the reader, i.e., modulating its scattering/radar cross section. The reflection coefficient is either varied in magnitude, producing Amplitude Shift Keying (ASK), or in phase, producing Phase Shift Keying (PSK), or both. It has been shown [22] that PSK is more energy efficient for this application. Varying the reflection coefficient modulates the amount of power reflected back to the reader, thereby transmitting information. Because the tag does not actually have to produce and broadcast any RF in this method, it can be designed to consume very little power.

Communication in the other direction (i.e., between reader and tag) is simpler. The reader can use any modulation scheme to transmit information to the tag. ASK, PSK and Pulse Width Modulation (PWM) are in common use. The tag can now demodulate the modulated RF carrier using standard techniques to recover reader commands. A finite state machine present on the tag determines the appropriate backscatter response when reader commands are fed into it.

Chapter 2

System Design

2.1 Introduction

In this chapter, I examine the system level issues, first at the level of general qualitative characteristics, and then more quantitatively. I end the chapter with a discussion of the testing strategies used to evaluate system performance after design and fabrication, and present test results from a prototype power extraction system.

2.2 Global Considerations

This section discusses, in a qualitative way, several of the characteristics one wants the power extraction system to have. Simultaneously, I also develop general design philosophies and techniques which allows these system requirements to be fulfilled effectively.

2.2.1 Robustness

Our power extraction system has to be robust to a wide range of environmental conditions if it is to be used for RFID tags which will be part of the global supply chain. System performance has to be maintained over wide ranges of temperature and humidity. In addition, the system has to work when attached to different types of

packaging material and subjected to bending and other abuse. System performance has to be verified at all these corners.

As designers, an effective way to enhance robustness is *simplicity*. The KISS (Keep It Simple, Stupid) philosophy says that when other factors are equal, a simple, ingenious system is more likely to work under adverse conditions than a complicated one, simply because the complicated system has more modes of failure. The KISS principle is broadly applicable to engineering design, but with caveats. Complexity without functional benefits is valueless, but ingenious use of a certain amount of complexity is frequently desirable. Any system must possess a minimum level of complexity in order to be robust and efficient in a noisy environment. This is why biological systems have evolved to use feedback and adaptation.

Coming back to our problem, the discussion above certainly does not exclude the use of adaptation and other active methods to enhance system robustness. It means that one has to be clever; functional enhancements have to be achieved without significantly increasing system complexity.

2.2.2 Cost

Another reason for keeping the system design as simple as possible is to keep costs down. To make RFID tags attractive to retailers, they should be as cheap as possible (ideally, less than a few cents each). The high volumes envisioned for tags means that even 1 cent less per tag translates into millions of dollars saved per year. The cost of a RFID tag is primarily set by the embedded integrated circuit, the cost of which scales as the die size (i.e., the cost/unit area of active circuitry is approximately constant). A typical figure of 7 cents/mm^2 of silicon area has been quoted. This means that the die area occupied should be as small as possible, placing heavy penalties on increased circuit complexity. This also ties in well with the KISS philosophy mentioned previously.

However, the cost per unit area is not really fixed, but varies with the IC fabrication process. In addition to minimizing area, one therefore also has to minimize the cost/unit area by using as cheap a process as possible. This basically limits one

to vanilla (standard, unmodified) CMOS processes with large feature sizes (unit area cost rises sharply as the minimum feature size scales downwards). On the other hand, available circuit complexity/unit area and high frequency performance both worsen rapidly as feature size increases. The optimum for many RFID applications today seems to lie around a minimum feature size of $\sim 0.5\ \mu m$ [22]. An alternative, which may prove to be important for the future, is flexible printed circuit technology.

2.2.3 Performance

Among the main performance criteria for RFID tags are read range and read speed. Read range is the maximum distance at which the tag can be rad by the tag reader. For many applications, such as in warehouses, the read range should be maximized to reduce reader density. For a successful read, two things have to happen. First the tag has to power up that distance from the reader (i.e., at the resultant RF power level). Secondly, the reader must be sensitive enough to pick up backscatter modulation from the tag with adequate Signal to Noise Ratio (SNR) for reasonably low Bit Error Rate (BER) communication between the two. Maximizing read range thus involves both tag and reader design. In this thesis I concentrate on the tag.

In free space (i.e., with no environmental effects and far away from the source) the RF power density (W/m^2) drops off as $\frac{1}{r^2}$, where r is the distance from the RF source (reader). This just follows from a power conservation argument. The radiated power is spread (non uniformly, since isotropic radiators do not exist) over the surface of a sphere with radius r. This area is $4\pi r^2$, so the power density $\propto \frac{1}{r^2}$. For multipath situations with environmental reflections and losses, the drop off exponent is n is greater than 2 and situation dependent. If one assumes free space propagation, the read range is $\propto \frac{1}{\sqrt{P_{th}}}$, where P_{th} is the power up threshold of the tag and is defined as the minimum incident power density at which the tag can respond to reader commands. In the multipath case, the dependence of read range on P_{th} is even weaker. Thus, decreasing P_{th} by a factor of 10, even in the best (free space) case, only increases the read range by a factor of about 3. The lesson from this is that increasing read range is difficult and that any increase one gets in read range is

reasonable cause for celebration. Even if the additional read range is not needed in a particular application, the lower P_{th} means that the reader can reduce its transmitted power and still power the tag up at a given distance. This is extremely desirable since it reduces reader-to-reader interference.

Another issue is read speed, which is essentially the number of tags the reader can read per second. One obviously wants to maximize this quantity. This puts constraints on the overall RFID system. Assume that the desired read speed is N tags per second. If each tag stores B bits of information, the information transmission rate needed between the tag and the reader is BN bits/second. Since the tag usually transmits using a binary modulation scheme (binary amplitude or phase shift keying), the minimum clock frequency of the tag is $f_{clk} \approx 2BN$ Hz. For most digital systems, the power consumption is dominated by switching power loss, which is given by

$$P_{sw} = \alpha C_L V_{DD} \Delta V f_{clk} \tag{2.1}$$

where α is a constant called the node transition activity factor, C_L is the load capacitance, V_{DD} is the power supply voltage and ΔV is the voltage swing. One sees that $P_{sw} \propto f_{sw}$, i.e., ultimately $P_{sw} \propto N$, the read speed. Thus tags which can be read faster consume more power. This makes sense intuitively.

The power up threshold P_{th} has a component which scales with P_{sw}. Thus there is positive correlation between P_{sw} and P_{th}. In general, tags which consume more power have higher power up thresholds. This means that P_{th} should increase as the read rate and/or the information stored on the tag increases. For very low read rates when $P_{sw} \approx 0$, P_{th} is limited by physical and practical constraints. Some of these are explored in this thesis.

2.2.4 Physical Size

RFID tags should be as small and inconspicuous as possible. The smaller the tag, the more easily can it be placed on small retail items. However, the physical size of antennas scale as the free space electromagnetic wavelength. At UHF (900 MHz),

the free space wavelength is about 30 cm, and most of the tag area is occupied by the antenna. Since the receive antenna on the tag still has to be physically small enough to fit on most packages, it's size will necessarily be limited to a small fraction of a wavelength. The theory of small antennas was first developed by Wheeler [39]. Wheeler's theory predicts that the antenna Q is inversely proportional to its effective volume. In other words, a small antenna is an inefficient radiator. Most of its energy is stored in reactive near fields and is not radiated, which accounts for the high Q. So there is a lower limit on the antenna (and hence the tag) size at a given operating frequency. As shall be shown, the high Q of small antennas also makes impedance matching to a load difficult.

Since most of the tag area is occupied by the antenna, the physical size of far field tags also scales as the wavelength, i.e., inversely as the frequency of operation. Higher frequency tags are smaller. However, as I show in the next section, practically all aspects of tag performance, including read range, cost and efficiency, degrade steeply as the operating frequency is increased. UHF tags operating around 900 MHz were proposed as a practical compromise to this problem - small enough to fit on most objects of interest, and at low enough frequencies to be cheap and provide adequate read range and efficiency for many applications.

Much attention has recently been devoted to fractal antennas as a way to reduce antenna (and hence tag) area without sacrificing radiation performance. However, it has been shown that, for small antennas, fractal geometries do not provide significantly better performance than simpler geometries like meander lines [5]. Indeed, in the small antenna limit, the radiation resistance is found to be independent of geometry. Hence I do not discuss fractal antennas further in this thesis.

2.3 Internal Considerations

This section analyzes system level performance quantitatively and comes up with theoretical performance bounds. The discussion is kept as general as possible to allow the applicability of the principles discussed to a wide range of frequencies and

power levels.

2.3.1 Impedance Matching

The impedance matching problem is now examined mathematically. Assume the electromagnetic field intensity in the neighborhood of our tag is P_{rad} W/m^2. Assuming free space propagation conditions (i.e., no environmental reflections or losses) P_{rad} is given by

$$P_{rad} = G_t \frac{P_t}{4\pi r^2} \tag{2.2}$$

where P_t is the transmitted power in Watts, r is the distance between the tag and the transmitter (tag reader) antenna and G_t is the gain of the transmitter antenna, usually quoted in dBi, i.e., dB power gain over an isotropic radiator with unity gain. G_t varies with direction, i.e., in (r, θ, ϕ) co-ordinates with $G_t \equiv G_t(\theta, \phi)$. The power P_{ant} collected by the tag antenna at its terminals is given by

$$P_{ant} = A_{eff} P_{rad} \tag{2.3}$$

where A_{eff} is the effective area of the tag antenna. A fundamental theorem, valid for all antennas in reciprocal media, says that in the far field (i.e., when the incident radiation locally looks like a plane wave) A_{eff} is given by

$$A_{eff} = G_r \frac{\lambda^2}{4\pi} \tag{2.4}$$

where G_r is the tag antenna's gain over an isotropic radiator, and λ is the electromagnetic wavelength in the medium. Of the power collected, only half is actually available for delivery to a load (the maximum power transfer theorem). This *available power* P_A is thus given by

$$P_A = G_r P_{rad} \frac{\lambda^2}{8\pi} \tag{2.5}$$

In (2.5), both G_r and P_{rad} are functions of (θ, ϕ), so the available power depends

on the relative orientation of the reader and tag antennas. Now consider a load which is a reactive element with a quality factor Q_L. The real part of the load impedance consumes the real power delivered by the antenna. A parallel representation for the load is now chosen (see Figure 2-1). Clearly, P_{diss}, the power dissipated in the load, is

$$P_{diss} = \frac{V_{in}^2}{2R_p} \tag{2.6}$$

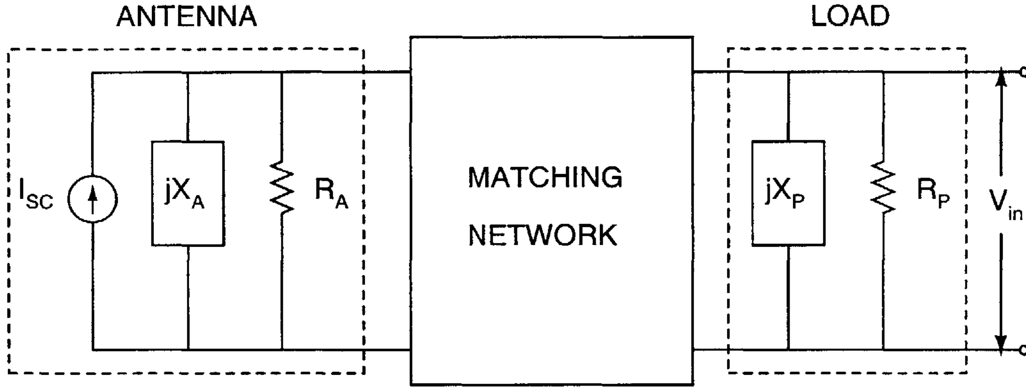


Figure 2-1: The impedance matching problem. The antenna impedance on the left has to be matched to the load impedance on the right for maximum power transfer.

where V_{in} is the amplitude of the (assumed sinusoidal) voltage across the load (in this case, the input of the rectifier), and the load resistance R_p incorporates both actual power delivered to the load and unwanted losses in the power extraction system. The maximum possible value of P_{diss} for a given electromagnetic field intensity is clearly the available power P_A. This corresponds to an antenna impedance matched to the load. From (2.5) and (2.6), the load voltage V_{in} attains a maximum in this case and is given by

$$V_{in,max} = \sqrt{\frac{G_r R_p P_{rad} \lambda^2}{4\pi}} \tag{2.7}$$

It is more convenient in this case to work with a series representation $R_s + jX_s$ of the load. To transform between the series and parallel forms, first note that $R_p = Q_L X_p$ by definition. In the high Q case, i.e., $Q_L^2 \gg 1$, we then have $X_s \approx X_p$

and $R_s \approx \frac{R_P}{Q_L^2}$. Note that the results which follow do not critically depend on this assumption (which is true for most cases of practical interest) but it does make the formulas simpler by eliminating R_p from (2.7). One then has

$$Q_L = \frac{X_s}{\frac{R_P}{Q_L^2}} \Rightarrow R_P = Q_L X_s \tag{2.8}$$

Substituting (2.8) into (2.7), one gets

$$V_{in,max} = \sqrt{\frac{G_r P_{rad} Q_L X_s \lambda^2}{4\pi}} \tag{2.9}$$

For example, when the reactive part of the load consists of a capacitance C_s (which is approximately true for practical power extraction systems), one has $X_s = \frac{1}{\omega C_s}$, so (2.9) becomes

$$V_{in,max} = \sqrt{\frac{G_r P_{rad} Q_L \lambda^2}{4\pi\omega C_s}} \tag{2.10}$$

Using the fact that $\lambda = \frac{2\pi c}{\omega}$, where c is the speed of light in the medium, (2.11) can be rewritten as

$$V_{in,max} = c\sqrt{\frac{\pi G_r P_{rad} Q_L}{\omega^3 C_s}} \tag{2.11}$$

Substituting (2.2) in (2.11), one finally gets the maximum available input voltage to the rectifier in terms of the source power P_t as

$$V_{in,max} = \frac{c}{2r}\sqrt{\frac{G_r G_t P_t Q_L}{\omega^3 C_s}} \tag{2.12}$$

The formula above is true for free space propagation conditions. In the more general multipath case, power density drops off with distance from the transmitter as $\frac{1}{r^n}$, where $n > 2$ is known as the propagation exponent. In this scenario, r in (2.12) is replaced by $r^{\frac{n}{2}}$. Typical values of n in urban environments range from 3 to 5. In free space, $n = 2$.

Since any rectifier one employs to extract power will be a nonlinear device which

will have a linear 'dead zone' at low values of V_{in} (where it will be very inefficient), it is in one's interest to maximize V_{in} for a given P_{rad}. To do this, one cannot increase G_t too much, since the reader has no *a priori* information about the spatial location of the tag. Thus the reader antenna's gain is limited to a maximum of about 6 dBi (dB gain over an isotropic radiator). For similar reasons (i.e., the tag does not know where the reader is, either), the tag antenna must be nearly isotropic. In practice, G_r is limited to a maximum of about 2 dBi. One clearly want to operate at low frequencies, but am limited by physical area constraints of the tag. In other words, the frequency ω is fixed by my application (and the FCC). For obvious reasons, one cannot change c. What remains is to maximize Q_L and minimize C_s. Hence I shall place considerable stress on increasing the quality factor and minimizing the capacitance of the rectifier input impedance. However, increasing Q_L also makes it more difficult to impedance match the antenna to the rectifier (load) over some required bandwidth (see Section 2.3.2).

Note that no form was assumed for the matching network used to impedance match the rectifier with the antenna. It does not matter for this derivation. The job of the matching network is to present a conjugately matched impedance to the antenna over some frequency range. This also corresponds, as has been shown, to maximum load voltage V_L and maximum energy stored in the load reactance. For example, when this is a capacitance C_s, the average energy stored in the capacitor with voltage V_C across it, $\frac{1}{4}C_sV_L^2$, is maximized when V_L is maximized, because $V_L = -j\omega\frac{V_C}{Q_L} + V_C = V_C\left(1 - \frac{j\omega}{Q_L}\right)$.

When one says 'matched over a bandwidth', one means that the reflection coefficient looking into the load is less than some specified value over a range of frequencies $\Delta\omega$. This defines the network Q to be $\frac{\omega_0}{\Delta\omega}$, where ω_0 is the center frequency. How does one design the matching network to maximize this? The first thing to realize is that, for a sufficiently complicated passive network, the *minimum* impedance matching bandwidth of the network (i.e., the highest network Q) has in general no relationship to the Q's of its constituent elements, the source or the load [21]. However, perfect impedance matching (zero reflection coefficient) can only be achieved for zero band-

width, i.e., at a finite number of frequencies, and the *maximum* impedance matching bandwidth of the network is governed by a criterion known as the Bode-Fano criterion, which is described below.

2.3.2 The Bode-Fano Criterion

Practically, for robustness, one would like as broadband an impedance match as possible between the antenna and the rectifier, i.e., one would like the network Q to be low. Also, since network elements are never lossless, the more complicated the network, the higher the losses within it and the more physical space it occupies. Thus one wants simple, broadband matching networks between specified source and load impedances. Unfortunately, such networks do not exist. In general, the more complicated the network, the better the matching bandwidth that can be achieved. Ultimately, however, fundamental limits exist that place hard upper bounds on achievable impedance matching bandwidth (i.e., the lowest network Q) to an arbitrary load with a lossless matching network at some specified mismatch level. These limits can be reached only in the distributed case, when there are an infinite number of differential circuit elements in the matching network. However, one can approach it as closely as desired by using finite matching networks of increasing complexity.

The problem is set up as follows. A source impedance Z_0 is to be matched to a load impedance Z_L through a lossless matching network E. By matched, one means that the power transferred to the load from the source is maximized. For any loads and sources with reactive (frequency dependent) components, i.e., whenever the load and source impedances are not pure resistances, the achievable impedance matching bandwidth at some specified mismatch level is limited. Bode and Fano were the first to solve the so-called single matching problem, where the source impedance Z_0 is purely resistive. The maximum possible impedance matching bandwidth (i.e., lowest network Q) to a load with a given Q_L in this case is given by the classical Bode-Fano criterion [29, 15]. The more complicated double matching problem, where the source impedance is arbitrary, is more difficult to analyze [8]. Source impedances of antennas are in general not purely resistive, but we can still apply the Bode-Fano criterion to

our impedance matching problem. This is because the antenna impedances are largely controllable by the designer (see the last paragraph of this section).

For definiteness, consider the single matching case shown in Figure 2-2 (this is the case originally analyzed by Bode). The purely resistive source impedance Z_0 is to be matched to a complex load with quality factor Q_L using a lossless matching network (using only inductors and capacitors). I The load is considered to be a parallel R-C combination, with $Q_L = \omega RC$, but similar expressions will occur for other simple R-C and R-L combinations. The parallel R-C case is particularly useful in this case since it is a good model of the input impedance of integrated circuits at frequencies of interest.

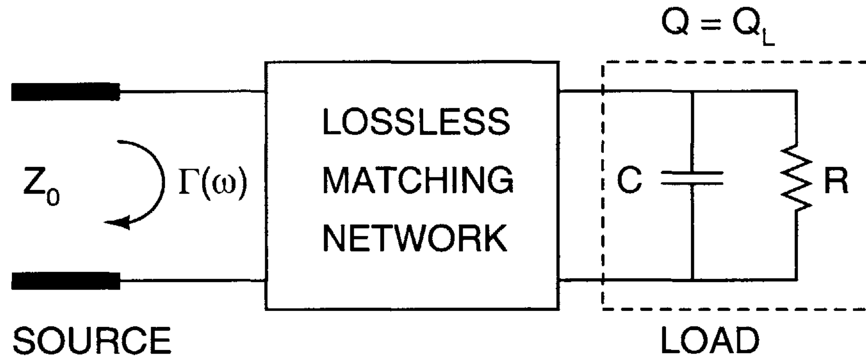


Figure 2-2: Illustrating the derivation of the Bode-Fano impedance matching criterion. A parallel R-C load with a quality factor Q_L is shown.

Defining the reflection coefficient looking from the source as $\Gamma(\omega)$, Bode and Fano showed that one must have

$$\int_0^\infty \ln\left(\frac{1}{|\Gamma(\omega)|}\right) d\omega \leq \frac{\pi}{RC} \tag{2.13}$$

Since $0 < |\Gamma| < 1$, one can divide (2.13) by -1, take the absolute value of both sides, and get

$$\left|\int_0^\infty \ln\left(|\Gamma(\omega)|\right) d\omega\right| \geq \frac{\pi\omega_0}{Q_L} \tag{2.14}$$

where ω_0 is the center frequency of the matching bandwidth of interest. This sub-

stitution only works in the narrowband case, i.e., when Q_L is almost constant over the bandwidth of interest. This is a reasonable approximation that will be assumed in all that follows. (2.14) expresses the fundamental gain-bandwidth-like constraint of this problem. For a given load, if one wants a 'high gain' matching network (one that realizes a very low Γ) one can only realize it over a narrow bandwidth.

One's objective here is to maximize the range of ω where $|\Gamma(\omega)| \leq |\Gamma_m|$, where $|\Gamma_m|$ is the maximum reflection coefficient magnitude one can tolerate. Since the area under the $\Gamma(w)$ curve is fixed, by (2.14), for a given load, the way to get maximum bandwidth is by having $|\Gamma(\omega)| = |\Gamma_m|$ over a bandwidth $\Delta\omega$, and be completely mismatched ($|\Gamma| = 1$) everywhere outside the bandwidth of interest. This ensures that $\ln(|\Gamma|) = 0$ everywhere outside one's desired bandwidth and contributes nothing to the integral in (2.14). This corresponds to the 'ideal matching network' case shown in Figure 2-3. In this case

$$\int_0^\infty \ln(|\Gamma(\omega)|)\, d\omega = \Delta\omega \ln(|\Gamma_m|) \tag{2.15}$$

giving one a theoretical upper bound on the achievable impedance matching bandwidth

$$\Delta\omega \leq \frac{\pi\omega_0}{Q_L} \frac{1}{\ln\left(\frac{1}{|\Gamma_m|}\right)} \tag{2.16}$$

Suppose one wants $|\Gamma_m| = 0.3$, which corresponds to 10 % of the incident power being reflected back to the source (in terms of s-parameters, $S_{11} = -10$ dB). Defining the Q of the impedance matching network as the 'network Q' $Q_n = \frac{\omega_0}{\Delta\omega}$, one gets, for the ideal matching network at the -10 dB level,

$$Q_n \geq 0.383 Q_L \tag{2.17}$$

Similarly, when $S_{11} = -20$ dB, i.e., $|\Gamma_m| = 0.1$ and only 1 % of the incident power is reflected, one gets

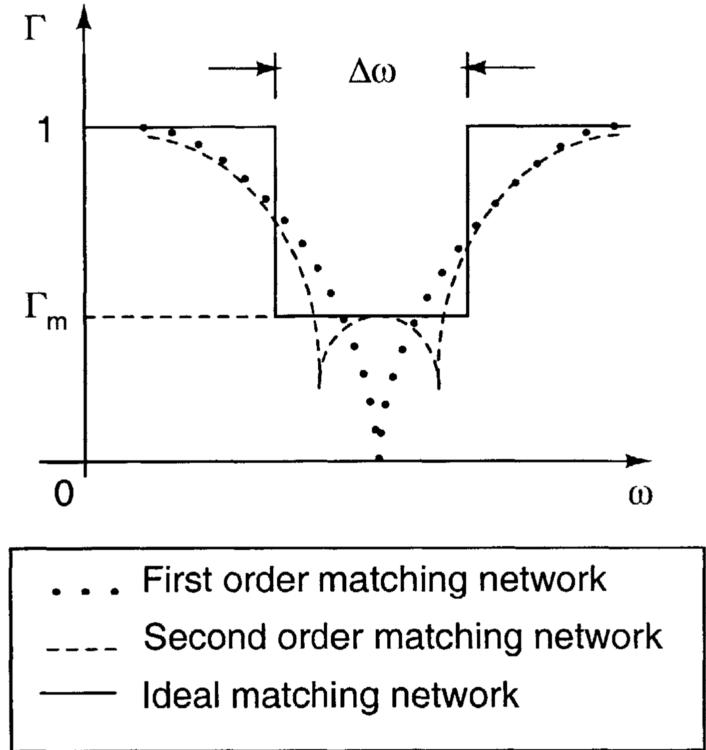


Figure 2-3: Reflection coefficient characteristics of impedance matching networks of various orders.

$$Q_n \geq 0.733 Q_L \tag{2.18}$$

One can make some observations based on what has been said so far. Since $\ln(0) \rightarrow -\infty$ and the integral in (2.13) is finite, one cannot have a perfect match, where $\Gamma = 0$, unless $\Delta\omega = 0$, i.e., only at a finite number of frequencies. Secondly, the higher the Q of the load, the harder is it to match over a wide bandwidth.

The design of impedance matching networks is closely allied to that of passive filters. The impedance matching problem translates to a filter design problem, where the ideal matching network (which reaches the maximum allowable Q_n as defined by (2.16)) is a brick wall bandpass filter. As one knows, such a filter can only be realized by using an infinite number of elements. However, it be approached as closely as one wants by making the network progressively more complex.

The commonest matching network design procedure uses coupled resonators to create bandpass filters from lowpass filter prototypes [29]. Any of the classical filter types - Butterworth, Bessel, Chebyshev or Elliptic - can be used. If a certain amount of passband ripple can be tolerated, a Chebyshev filter is often a good choice. First and second order matching network responses are shown in Figure 2-3. Note that the second order response has substantially more bandwidth. The bandwidth increases further for higher order networks. However, it rarely pays to use filter orders (number of resonators) higher than third or fourth, since the incremental bandwidth gain from adding another filter stage decreases rapidly with an increasing number of stages. In this thesis, I use second order matching networks, which increase bandwidth by factors of 2 to 3 over first order networks and can get quite close to the Bode-Fano limit without adding much circuit complexity.

I now examine what constraints are put on a power extraction system by its bandwidth requirements. Defining the required operating bandwidth as $B \equiv \Delta\omega$, one can find the theoretical upper bound on input voltage available to a rectifier (or any other load). To do this, one assumes that B is defined by a maximum allowable reflection coefficient magnitude $|\Gamma_m|$, and write (2.16) as

$$Q_L \leq \frac{\pi}{\ln\left(\frac{1}{|\Gamma_m|}\right)} \frac{\omega_0}{B} \tag{2.19}$$

where ω_0 denotes the center frequency. The equality in (2.19) holds when the impedance matching network is ideal in the sense of Figure 2-3. One can now substitute (2.19) in (2.12). However, (2.12) was derived assuming perfect impedance matching, i.e., $\Gamma_m = 0$). Recognizing that in general only $\left(1 - |\Gamma_m|^2\right)$ of the available power is actually delivered to the load, (2.12) is modified to

$$V_{in,max} = \frac{c}{2r}\sqrt{\frac{\left(1 - |\Gamma_m|^2\right) G_r G_t P_t Q_L}{\omega_0^3 C_s}} \tag{2.20}$$

One can finally substitute (2.19) in (2.20) to get the desired upper bound on the input amplitude V_{in}

$$V_{in} \leq \frac{c}{2\omega_0 r}\sqrt{\frac{G_r G_t P_t}{BC_s}\frac{\pi\left(1-|\Gamma_m|^2\right)}{\ln\left(\frac{1}{|\Gamma_m|}\right)}} \tag{2.21}$$

Figure 2-4 plots the function $f\left(|\Gamma_m|\right) = \frac{\left(1-|\Gamma_m|^2\right)}{\ln\left(\frac{1}{|\Gamma_m|}\right)}$ for $0 < |\Gamma_m|$. This is a very interesting plot. One sees that the function increases monotonically with $|\Gamma_m|$ from $f(0) = 0$ to $f(1) = 2$. At first glance, this is completely counterintuitive. As one increases the impedance mismatch, the maximum available voltage at the input terminals actually increases. The win in $V_{in,max}$, after taking the square root, is by about a factor of 2.2 as $|\Gamma_m|$ varies from 0.1 to 1. The catch is that as $|\Gamma_m|$ increases, Q_L must dramatically increase in order to achieve the equality in (2.19) and thus allow the upper bound of (2.21) to be reached.

In other words, if Q_L (the quality factor of the rectifier input impedance) is large, it is possible to simultaneously increase bandwidth B and rectifier input voltage $V_{in,max}$ by allowing $|\Gamma_m|$ to increase in-band. The process of increasing Q_L cannot be carried on indefinitely in practice, since the power available at the rectifier input steadily decreases as $|\Gamma_m|$ increases. The maximum possible Q_L is set by the loading effect of the rectifier itself on the input (normally parasitic loss components dominate Q_L: see Section 2.3.4). This corresponds to the ideal case of no parasitic losses and a system efficiency of 100 %. In this case, one gets $|\Gamma_m| \to 1$, $f\left(|\Gamma_m|\right) = 2$ and

$$V_{in} \leq \frac{1.25c}{\omega_0 r}\sqrt{\frac{G_r G_t P_t}{BC_s}} \tag{2.22}$$

As described earlier, a typical definition of $|\Gamma_m|$ used in practical systems is 0.3 ($S_{11} = -10$ dB). If Q_L is large this is not optimal, and $|\Gamma_m|$ should be allowed to increase in order to maximize $V_{in,max}$. Nevertheless, I consider this situation so that it can be compared with the ultimate limit of (2.22). Here one can substitute $|\Gamma_m| = 0.3$ into (2.21) to get

$$V_{in} \leq \frac{0.77c}{\omega_0 r}\sqrt{\frac{G_r G_t P_t}{BC_s}} \tag{2.23}$$

Thus one loses by a factor of about 0.62 in V_{in} as compared to (2.22). Other

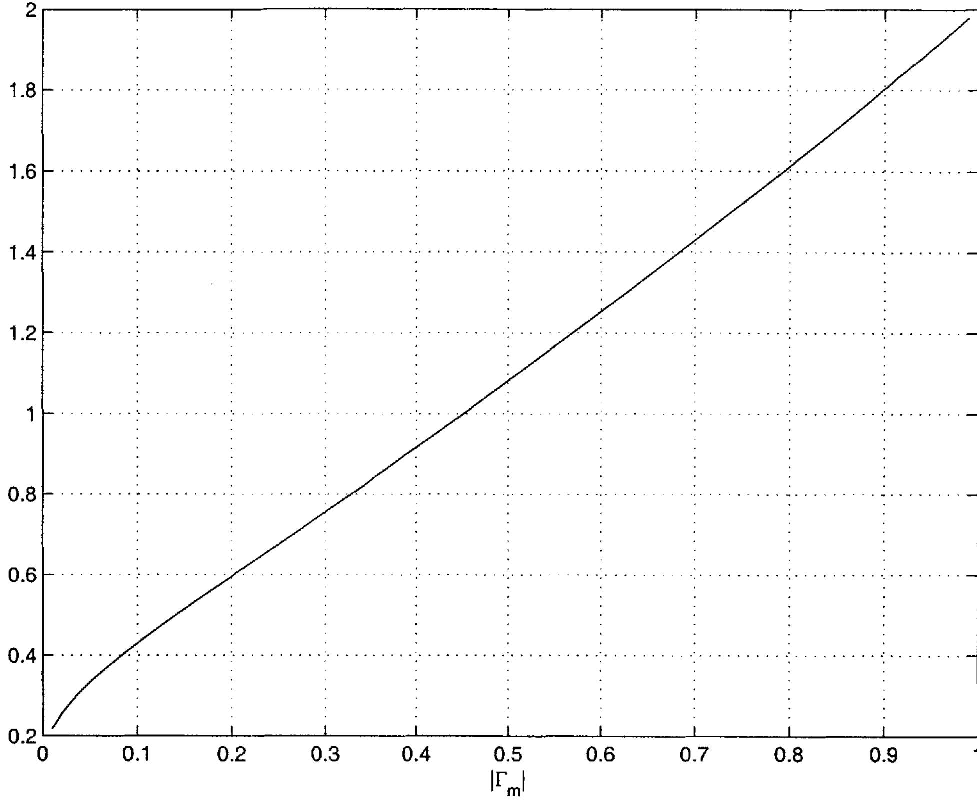


Figure 2-4: Plot of the function $f\left(|\Gamma_m|\right)$ in (2.21).

values of $|\Gamma_m|$ change the coefficient in (2.23) from 0.77. In practice, it usually lies within the range 0.5 - 1. Thus one sees that the operating bandwidth B is important in determining the performance of power extraction systems. Broadband power extraction systems are more difficult to build efficiently than narrowband ones.

The relationship between Q_L, bandwidth B and input voltage V_{in} is made clearer if one writes (2.19) as

$$|\Gamma_m|^2 \geq \exp\left(\frac{-2\pi\omega_0}{BQ_L}\right) \tag{2.24}$$

Substituting (2.24) in (2.20), one gets another form of the upper bound

$$V_{in} \leq \frac{c}{2r}\sqrt{\frac{G_r G_t P_t Q_L \left(1 - \exp\left(\frac{-2\pi\omega_0}{BQ_L}\right)\right)}{\omega_0^3 C_s}} \tag{2.25}$$

The function $g(Q_L) = Q_L\left(1 - \exp\left(\frac{-2\pi\omega_0}{BQ_L}\right)\right)$ increases monotonically from 0 to $\frac{2\pi\omega_0}{B}$ as Q_L increases from 0 to ∞. Consider the infinite Q_L (no parasitic losses) limit. If one substitutes $g(\infty) = \frac{2\pi\omega_0}{B}$ in (2.25), one gets back (2.22) as expected. Figure 2-5 shows $g(Q_L)$ for various values of $\frac{\omega_0}{B}$. The function increases monotonically with Q_L, improving system performance, but eventually saturates at $\frac{2\pi\omega_0}{B}$. Thus one sees that V_{in} saturates to (2.22) for high values of Q_L, so that further increases in Q_L barely improve performance. To get V_{in} within 90 % of this infinite Q_L limit, one needs $g(Q_L) \approx 0.8$, i.e.,

$$Q_L \geq 13.51\frac{\omega_0}{B} \tag{2.26}$$

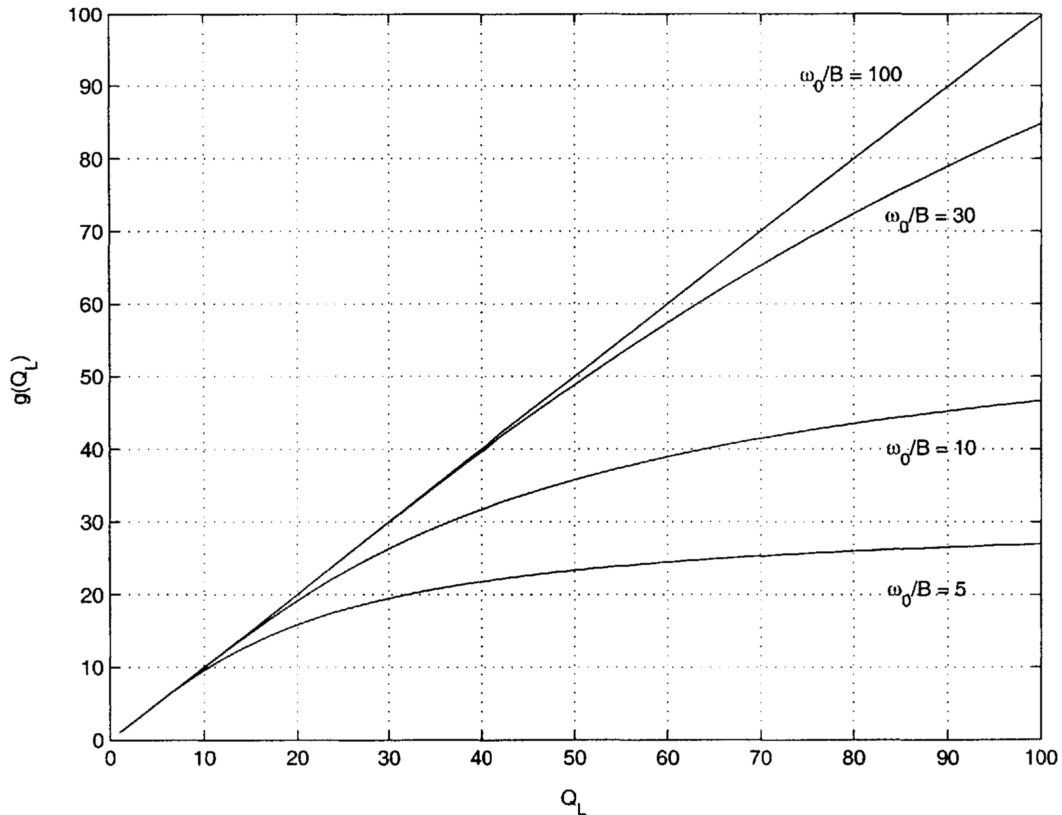


Figure 2-5: Plot of the function $g(Q_L)$ for various values of the fractional bandwidth $\frac{\omega_0}{B}$.

Thus, the Q_L required to achieve high efficiencies decreases as the fractional band-

width $\frac{B}{\omega_0}$ increases. The fractional bandwidths of typical UHF RFID applications are narrow enough to make the Q_L in (2.26) too high to reach. For example, for $B = 30$ MHz at $\omega_0 = 900$ MHz, the equality in (2.26) corresponds to $Q_L = 405$. The $|\Gamma_m| = 0.3$ case in (2.23) corresponds in this case to $Q_L = 78$, which is more reasonable. However, at lower frequencies Q_L can be much higher and very efficient power extraction becomes possible.

To obtain some physical insight into the problem, one notes that when R (i.e., Q_L) is high, it is easy to get high input voltage into the rectifier since $V_{in} \propto \sqrt{P_A R}$. Thus, a broadband impedance matching network that only achieves a poor match and reflects most of the input power can still produce a value of V_{in} high enough for the rectifier to power up the load. When $R\,(Q_L)$ is low, the broadband matching network achieves a better match (reflects back less power) and provides more power to the rectifier, while $\sqrt{P_A R}$ maintains reasonable values of V_{in}. The Bode-Fano criterion, because of its logarithmic dependence on $|\Gamma_m|$, predicts that the degradation in broadband impedance match as $R\,(Q_L)$ increases is gentle enough such that it is better to have a very poor impedance match and huge $R\,(Q_L)$ than to have a smaller $R\,(Q_L)$ and good broadband impedance match. This is true up to a point. Beyond this point, increases in $R\,(Q_L)$ result in $|\Gamma_m|$ increasing by just enough to saturate V_{in} to the maximum value given by (2.22).

In practice, parasitic losses limit the achievable R at the input for a given input capacitance C such that a maximum practical input voltage V_{in} is reached with a high but non-unity value of $|\Gamma_m|$. Furthermore, because of antenna and rectifier efficiencies, for a given impedance matching bandwidth, R cannot be indefinitely increased (i.e., there is an upper bound on $|\Gamma_m|$). There is a simple feedback loop interpretation of this. As R increases, $|\Gamma_m|$ increases by the Bode-Fano criterion to create a negative feedback loop on V_{in}. Initially with R low and $|\Gamma_m| \approx 0$, the loop gain is small so that V_{in} rises with R. When $R \to \infty$ and $|\Gamma_m| \approx 1$, the loop gain becomes high and pins V_{in} to its maximum value.

One also notes here that the applicability of the single matching problem to the power extraction scenario may not be immediately obvious. Since the antenna

(source) impedance is complex in general, the problem may seem to be of a double matching type. However, double matching only applies when *both* the source and load impedances are known and fixed, and one has to match them. In the power extraction case, one can design the reactive part of the antenna impedance (and to a certain extent, the resistive part too) to be whatever one wants. This flexibility is equivalent to saying that the reactive part of the antenna impedance can be considered part of the matching network. Thus the 'effective' source may be considered purely resistive, reducing the problem to the single matching situation.

2.3.3 The L-Match

It has been claimed that for many small antennas the simple L-section impedance matching network shown in Figure 2-6 can achieve the maximum efficiency that is possible given matching elements with finite Q [35]. The matching network should be composed of reactive elements having the opposite sign of reactance to that of the antenna. This ensures that the matching elements only exchange energy with the antenna and not amongst each other, thus minimizing their losses. For example, an inductive antenna should be matched using capacitors only.

Let one therefore examine the merits of the simple L-match. The main merit is simplicity. This is a first order impedance matching network. It requires no additional elements and occupies no extra area. The antenna is made inductive and resonates out the rectifier capacitance at frequencies of interest. The antenna (source) quality factor Q_A and rectifier (load) quality factor Q_L are equal. One might then naively think that the network quality factor is simply

$$Q_n = \frac{1}{2}Q_L = \frac{1}{2}Q_A = \frac{X_A}{2R_A} \tag{2.27}$$

However, (2.27) is *incorrect*. The real network Q (as measured by the impedance matching bandwidth) is much higher. The Q defined by (2.27) is the ratio of the imaginary and real parts of the impedance presented by the circuit to the load, and does not correspond to the network Q defined in Section 2.3.2. In reality, the

impedance matching bandwidth for Q_n is well short of the Bode-Fano limit. Typical values for $\frac{Q_n}{Q_L}$ are 1.5 at $|\Gamma_m| = -10$ dB and 5 at $|\Gamma_m| = -20$ dB, i.e., factors of 4 and 7 higher than the theoretical limits of (2.17) and (2.18). Apart from the high network Q, the main disadvantage also arises from simplicity. There are no adjustable parameters. Assuming that one has maximized Q_L to the best of one's abilities, the antenna impedance, Q_A and Q_n are all specified.

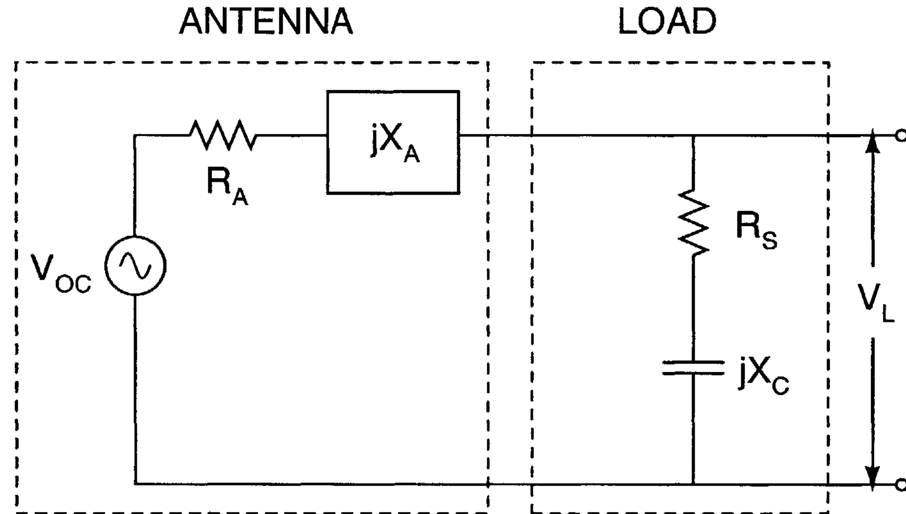


Figure 2-6: The simple L-match. For maximum power transfer, $X_A = -X_C$ and $R_A = R_S$.

I shall now examine whether the lack of flexibility in the L-match proves to be a problem for RFID tags at UHF. Since high Q_L is vital for rectifier efficiency, one wants Q_L to be as high as possible. This unfortunately decreases one's impedance matching bandwidth. At $Q_L = 50$, (practically, this is about the maximum one can expect to get at this frequency), designing a small antenna with $Q_A = Q_L$ to provide 1-2 dB of gain (all that the tag can afford), does not prove to be difficult. However, the network Q provided by the single stage L-match is then about 75 (at the -10 dB level), providing 12 MHz of bandwidth about 900 MHz. The allocated RFID bandwidths are 26 MHz (902-928 MHz) in the USA and 2 MHz (868-870 MHz) in Europe, so this amount of bandwidth is enough for Europe but not enough for the American market. To provide a bandwidth of 40 MHz, which is barely enough to

cover the American bandwidth with some error margin, $Q_L < 15$. One would also like a single tag that can span both European and American frequency bands. This needs $Q_n = 12$. The single L-stage only allows $Q_L = 8$ at the -10 dB level, whereas the Bode-Fano limit is $Q_L = 32$. Thus the L-match is quite undesirable and better matching networks, which approach the limit more closely, are needed. I shall use second-order matching networks for this purpose.

2.3.4 Efficiency

The power extraction efficiency η of the system is defined as the P_L, the DC power delivered to the load, divided by P_A, the RF power available at the antenna. Thus,

$$\eta = \frac{P_L}{P_A} \tag{2.28}$$

The available power is defined as the maximum power which can be delivered by the antenna. This happens when the antenna impedance is conjugately matched to the input impedance of the rectifier, possibly as seen through an impedance matching network i.e., $Z_A = Z_{in}^*$. Assuming free space propagation conditions, P_A is given by

$$P_A = 0.5 G_t G_r \left(\frac{\lambda}{4\pi r}\right)^2 P_t \tag{2.29}$$

where P_t is the transmitted power in Watts, G_t and G_r are the gains of the transmitting and receiving antennas, respectively, r is the distance between the transmitter and the receiver, and $\lambda = \frac{2\pi c}{\omega}$ is the electromagnetic wavelength in the medium. The power delivered to the load is simply

$$P_L = V_L I_L = \frac{V_L^2}{R_L} \tag{2.30}$$

where V_L and I_L are the DC load voltage and current, respectively, and the effective load resistance $R_L = \frac{V_L}{I_L}$. I define V_{in} to be the input voltage at the rectifier terminals. Since integrated circuits are forced to store energy primarily in capacitors, and because diodes and MOS transistors are voltage controlled devices, one wants the

amplitude of V_{in} to be as large as possible. Increasing V_{in} generally leads to dramatic improvements in rectifier efficiency (the 'dead zone' effect). This may be formalized by defining a rectifier function $F(\bullet)$, as follows

$$V_L = F\left(V_{in}, I_L\right) \tag{2.31}$$

The function F is process and circuit topology dependent. However, it's general shape remains the same across processes and rectifier topologies and is shown in Figure 2-7 for different values of I_L, the current drawn by the load. V_L increases monotonically with V_{in} when I_L is fixed, but the curves are not linear. V_L remains small until V_{in} reaches a value V_D, beyond which it increases rapidly. V_D is called the 'dead zone' of the rectifier and defined by noting that for $V_{in} < V_D$, the load voltage V_L is 'small' and the V_L - V_{in} rectification curves are almost flat. The dead zone is usually a weakly increasing function of the load current (logarithmic for a diode, for instance). The rectification curves shift downward as the load current increases. This is because of the increased voltage drop across the finite output resistance of the rectifier. In addition, the rectifier becomes more efficient and it's output resistance decreases as V_{in} decreases. This explains why the rectification curves approach each other for large values of V_{in}.

At large values of V_{in}, the curves become linear with slope K, which may be termed the no load rectifier voltage gain. In this regime the load voltage is given by $V_L = KV_{in} - R_{out}I_L$, where R_{out}, the output resistance of the rectifier, is input level dependent. One wants rectifiers with low values of V_D and high values of K. Unfortunately, this usually means a complex circuit, which will increase the input capacitance C_s and reduce the available input voltage $V_{in,max}$ at a given received power level (from (2.21)).

Finally, suppose one wants a certain load voltage V_{L0} in order to power up on chip circuitry. The input RF amplitude required to produce V_{L0} with a load current I_{Li} is called the rectifier threshold voltage V_{toi}. Figure 2-7 shows the threshold voltage for three different values of the load current. As we might expect, the threshold voltage

increases as the load current increases and V_{L0} is fixed. Good rectifiers have small values of C_s and V_{to}, i.e., they have low input capacitance and can supply the load even when V_{in} is small.

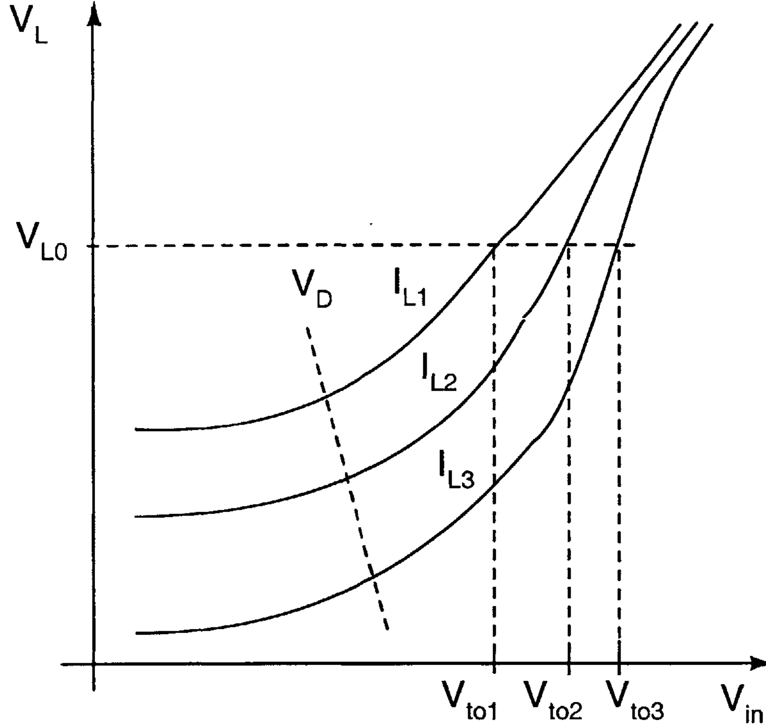


Figure 2-7: Typical shapes of the rectifier function $F(\bullet)$ for different values of I_L. $I_{L1} < I_{L2} < I_{L3}$.

The power delivered to the load can now be written in terms of $F(\bullet)$ as

$$P_L = I_L F\left(V_{in}, I_L\right) \tag{2.32}$$

Since, from (2.21), there exists an upper bound on V_{in} for a given transmitter power and other physical parameters, one can theoretically calculate various performance metrics of the system if the function $F(\bullet)$ is known (theoretically or experimentally). V_{to} is defined as the minimum value of V_{in} which allows the rectifier to supply a current I_L at a voltage V_{L0}. This allows one to find theoretical upper bounds on efficiency η, power up range r_{max} and power up threshold P_{th} for different values of C_s and V_{to} by simply using the previously derived equations for available input

voltage and power (equations (2.23) and (2.29)) and replacing V_{in} by V_{to}. The power up threshold is defined as the minimum power collected by the antenna for which the tag powers up, and is twice the power available to the tag at that range. It is better to use (2.23), which uses $|\Gamma_m| = 0.3$, instead of (2.22) as an upper bound since the values of Q_L required to approach (2.22) are difficult to achieve in most practical situations (see Figure 2-5).

As an example, let me take typical values for my power extraction system. I have $\frac{\omega}{2\pi} = 900$ MHz, $\frac{B}{2\pi} = 30$ MHz, $G_r = 1.5$ dBi, $G_t = 4.5$ dBi and $P_t = 1$ W (this corresponds to the maximum allowable transmitted power in this frequency band). Suppose one needs V_L to be at least 1.5 V at $I_L = 2$ μA for powering up on chip circuitry, i.e., $V_{L0} = F(V_{to}, 2\mu A) = 1.5$ V. Figures 2-8, 2-9 and 2-10 then show the theoretical efficiency, power up range and power up threshold, respectively, as a function of V_{to} with C_s varying from 0.4 pF to 2 pF assuming that the Bode-Fano limit has been reached with $|\Gamma_m| = 0.3$.

One sees that P_{th} is a strongly increasing function of both V_{to} and C_s. In addition, η and r_{max} both increase rapidly as V_{to} and C_s decrease. The lesson is thus the same as before: the rectifier needs to have both low threshold voltage V_{to} and low input capacitance C_s. These are usually mutually conflicting requirements, and must be traded off optimally to get the best design.

It is illustrative to demonstrate an alternative method of looking at the efficiency. By using energy conservation, one can define the effective resistive load R_{eff} on the resonant input circuit due to power being delivered to the load (see Figure 2-11) as follows

$$\begin{aligned} \frac{V_L^2}{R_L} &= \frac{V_{in}^2}{2R_{eff}} \\ \Rightarrow R_{eff} &= \left(\frac{V_{in}}{V_L}\right)^2 \frac{R_L}{2} \end{aligned} \quad (2.33)$$

where $R_L = \frac{V_L}{I_L}$ is the load resistance. Because real rectifiers have dead zones, arbitrary values of R_{eff} are not achievable. Achievable values of R_{eff} are determined by the nonlinear, circuit dependent function $F(V_{in}, I_L)$.

In Figure 2-11, R_s represents losses in the rectifier. Losses occur in the devices,

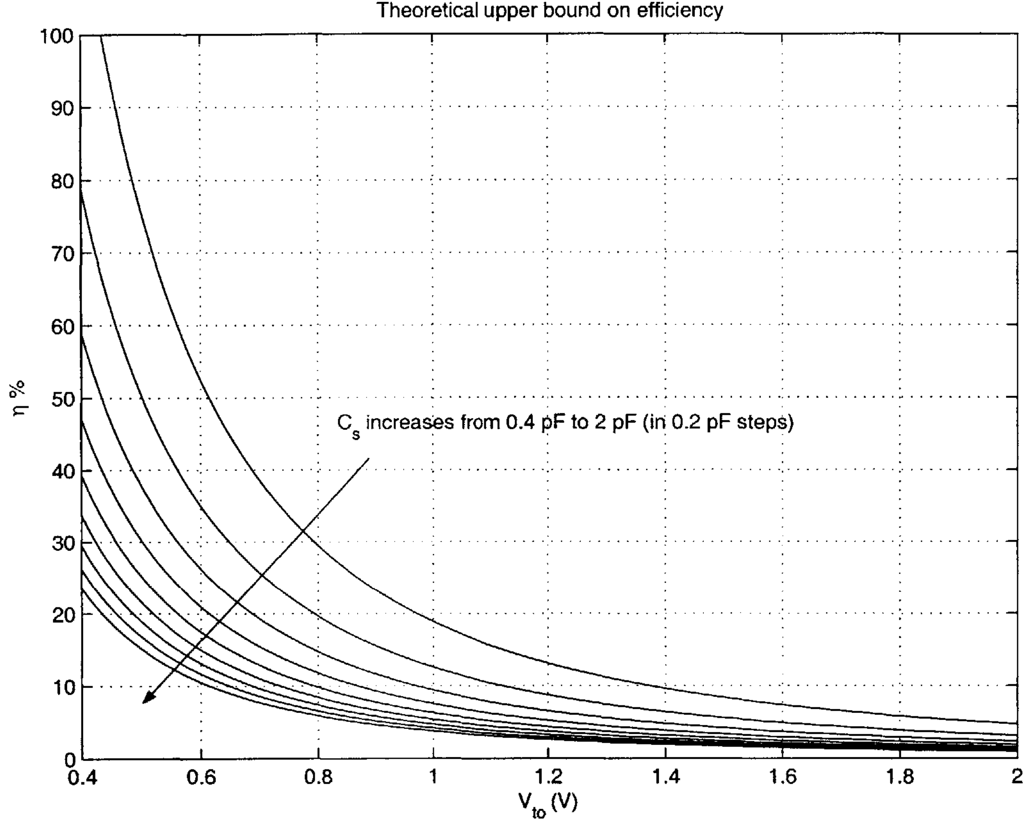


Figure 2-8: Theoretical efficiency curves for parameters in text and C_S varying from 0.4 pF to 2 pF.

interconnects and passive components. One can transform R_s into an equivalent parallel resistance R_p, where $R_p = Q_{L0}^2 R_s$, and $Q_{L0} = \frac{1}{\omega C_s R_s}$ is the input quality factor at no load, i.e., $R_{eff} \to \infty$, and $Q_{L0} \gg 1$.

In other words, R_p represents unwanted losses in the power extraction system, and R_{eff} represents wanted losses, i.e., useful power supplied to the load (see Figure 2-12). Since these resistances are connected across the same terminals, the efficiency η can now be simply written as

$$\eta = (1 - |\Gamma|) \frac{R_p}{R_{eff} + R_p} \tag{2.34}$$

where Γ is the input reflection coefficient. Efficiency is lowered by the factor $(1 - |\Gamma|)$ since only that fraction of the available power is actually delivered to the rectifier

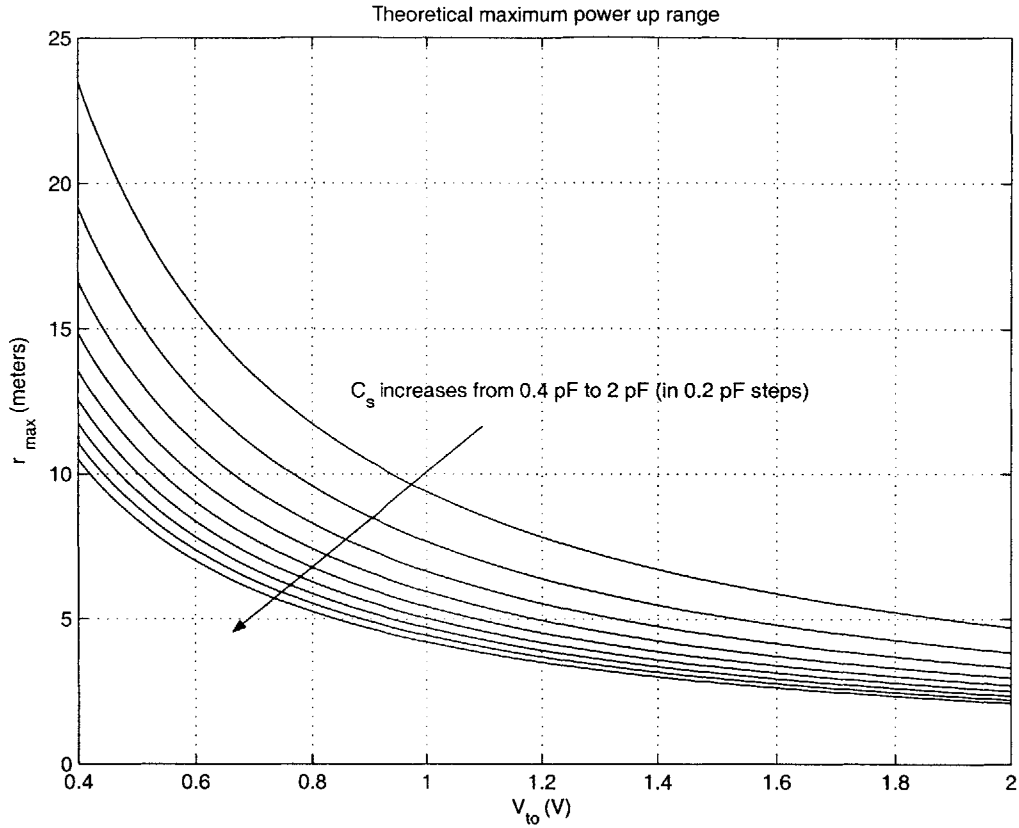


Figure 2-9: Theoretical maximum power up range curves for parameters in text and C_S varying from 0.4 pF to 2 pF.

(the rest is re-radiated by the antenna). Another way to write (2.34) is as follows. The loaded quality factor Q_L of the rectifier input is given by

$$Q_L = \omega C_s \frac{R_p R_{eff}}{R_p + R_{eff}} \tag{2.35}$$

From (2.34) and (2.35), one gets

$$Q_L = Q_{L0}\,(1 - \eta) \tag{2.36}$$

Since the power delivered to the load is $P_L = \frac{V_L^2}{R_L}$, $R_{eff} = \frac{V_{in}^2}{2P_L}$. From (2.34) and (2.35), one gets

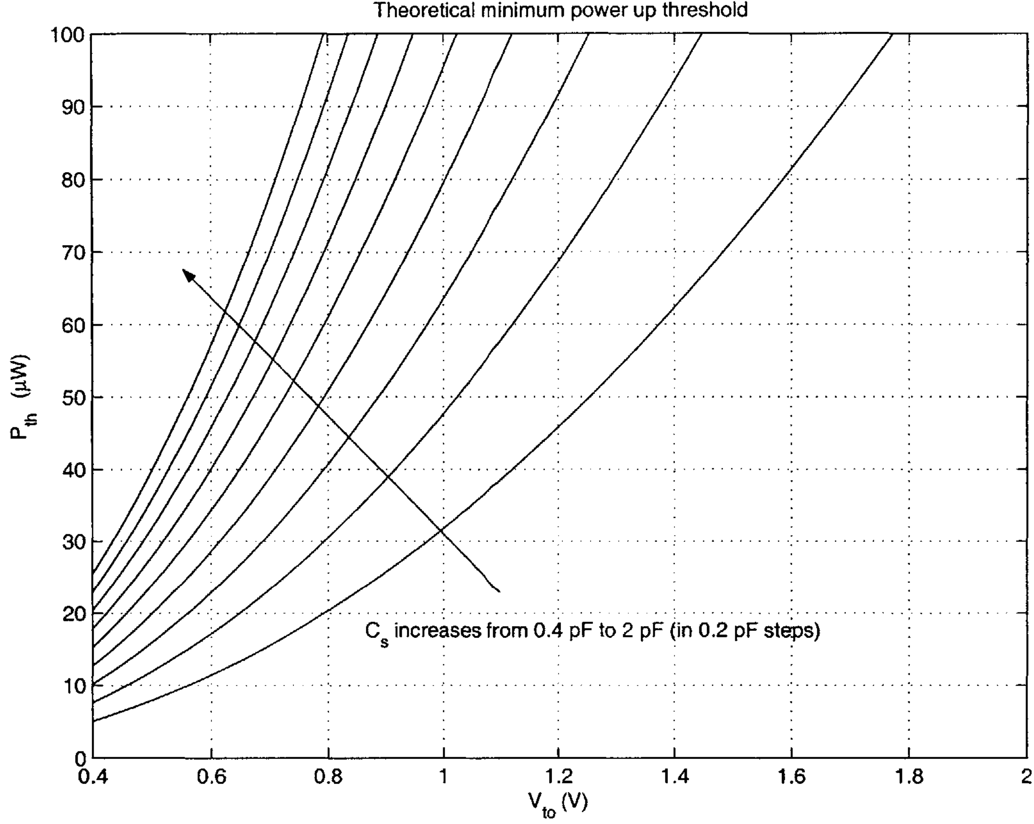


Figure 2-10: Theoretical power up threshold curves for parameters in text and C_S varying from 0.4 pF to 2 pF.

$$\eta = (1 - |\Gamma|)\frac{Q_L}{\omega C_s R_{eff}} \tag{2.37}$$

Finally, substituting for R_{eff} and using (2.36), one finds the efficiency to be

$$\eta = (1 - |\Gamma|)\frac{1}{1 + \frac{V_{in}^2}{2R_p P_L}} \tag{2.38}$$

This explicit expression for η is useful because it is automatically limited to a maximum value of 1, i.e., 100 %. It emphasizes the obvious point that parasitic losses must be minimized (i.e., R_p must be increased) in order to improve the efficiency. In Section 3.6.1, I describe several layout techniques that can be used for achieving this goal.

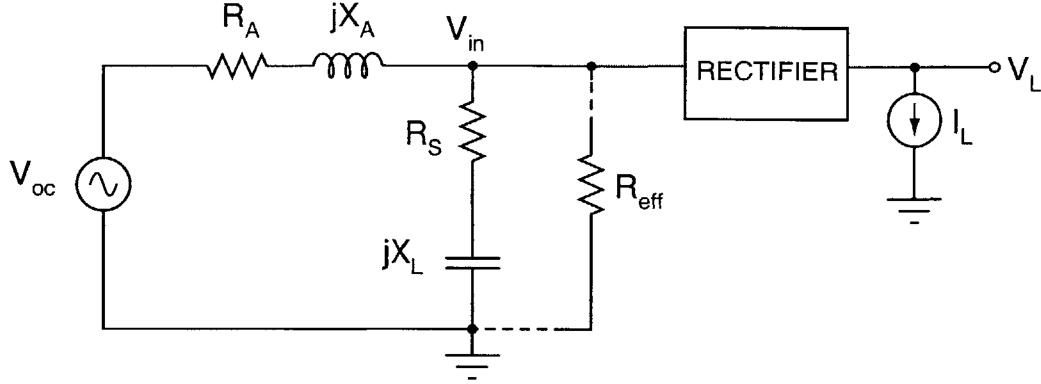


Figure 2-11: Circuit used for efficiency calculations, with R_s being a series representation of rectifier losses. The load resistance $R_L = \frac{V_L}{I_L}$ is transformed into an additional loss component R_{eff} at the rectifier input.

Finally, I work out an example of how to calculate the efficiency and evaluate effective resistances. Assume that $V_{to} = 0.5$ V and $P_L = 3$ μW, and all other system parameters are the same as previously stated. Using the theoretical efficiency curves of Figure 2-8 directly gives $\eta \approx 0.3$ for $C_s = 1$ pF. If one substitutes this value of η into (2.38) with $V_{in} = V_{to} = 0.5$ V, one gets $R_{eff} \approx 41.7$ kΩ and $R_p \approx 20.8$ kΩ for $V_L = 1.5$ V. At 900 MHz, this corresponds to $R_s \approx 1.5$ Ω, $Q_{L0} \approx 118$ and $Q_L \approx 80$. This value of Q_L is equal to the maximum allowed by the Bode-Fano criterion (equation (2.19) with $|\Gamma_m| = 0.3$). This is because Figure 2-8 assumes that the Bode-Fano limit has been reached with $|\Gamma_m| = 0.3$. In practice, the efficiency η will be somewhat lower than 0.3, since impedance matching networks that reach the Bode-Fano limit cannot be built with a finite number of elements. Failure to reach the limit may be considered to effectively decrease the value of Q_L.

2.3.5 Packaging & Assembly

The physical design of a power extraction system is at least as important as the circuit design. Among the important issues are circuit packaging and assembly. Unless the circuit is printed on the flexible substrate of the tag itself (an option which is currently unavailable for UHF tags, because flexible circuits which operate at UHF are not com-

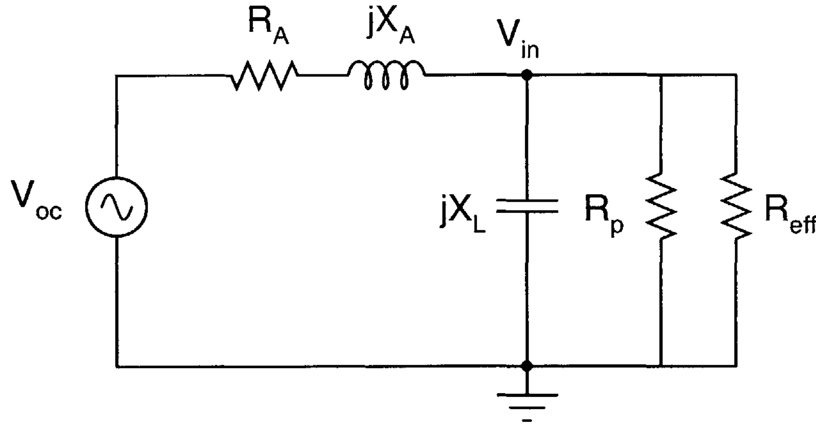


Figure 2-12: Circuit used for efficiency calculations, with R_p being a parallel representation of rectifier losses and R_{eff} representing power delivered to the load.

mercially available), it must be packaged and connected to the antenna. The package protects the circuit from the environment, but introduces unwanted inductances and capacitances (package parasitics) which degrade performance. Commercial tags minimize package effects by using flip chip bonding or fluidic self assembly of bare silicon dies into the tag. This eliminates package effects to a large extent. Since these techniques are expensive and generally proprietary, in this thesis I have limited myself to conventionally wire bonding the die into a surface mount package, which is then soldered onto the surface of the tag. This is not a serious problem since package effects can be measured and modeled; these can be used to evaluate system performance in their absence.

A simple lumped model of the package, which models each pin as a single *L-R-C* section, will prove to be sufficient for my purposes. L is primarily the inductance of the bonding wire. R accounts for ohmic losses in the bonding wire, while C accounts for stray parasitic capacitances to the body of the package (ground). This picture is complicated by the fact that nearby pins are coupled to each other. To account for this, mutual inductances and capacitances are introduced between adjacent pins (see Figure 2-13).

A multiplicity of formulae exist for the inductance and capacitance of various common wire configurations. The most accepted formula [25] for the self inductance

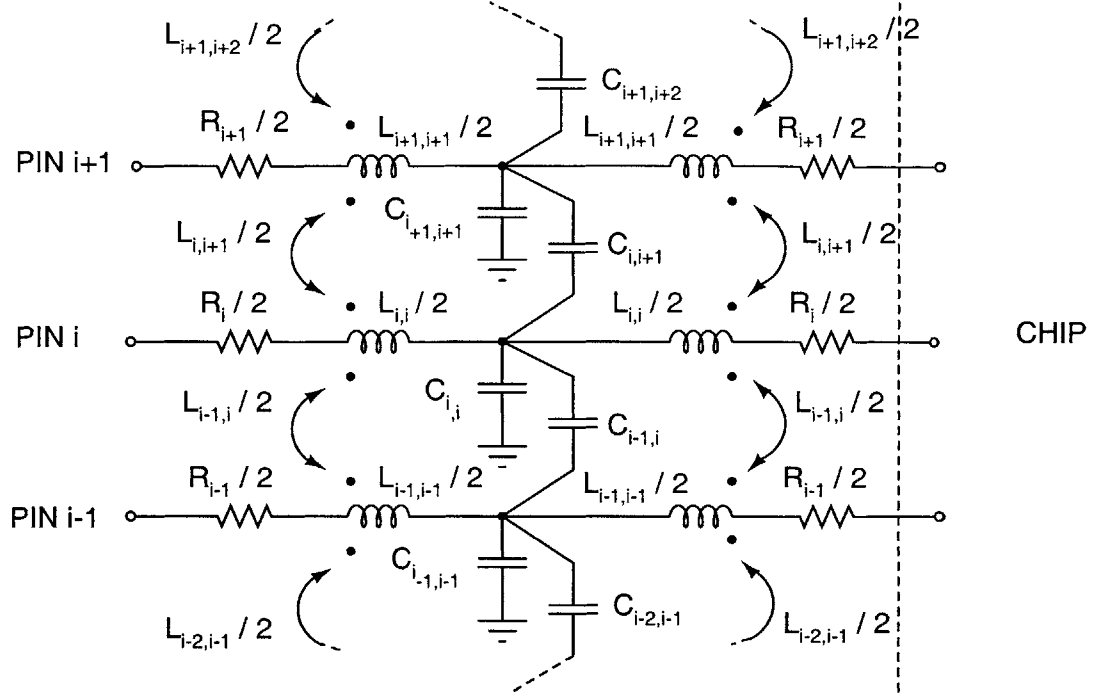


Figure 2-13: A fairly generic lumped single section package model showing three neighboring pins. Non nearest-neighbor coupling between pins has been ignored.

L of a long straight conductor in free space is

$$L \approx \frac{\mu_0 l}{2\pi}\left[\ln\left(\frac{4l}{d}\right) - \frac{3}{4}\right] \tag{2.39}$$

where d is the diameter of the wire, l is its length and $\mu_0 = 4\pi \times 10^{-7}$ is the permeability of free space. This formula assumes that the return current is infinitely far away. Technically, (2.39) is called the partial inductance of the wire in a Partial Element Equivalent Circuit (PEEC) model. Because of the weak dependence of $\frac{L}{l}$ on geometry, a handy rule of thumb is $\frac{L}{l} \approx 1$ nH/mm.

Adjacent bond wires introduce mutual capacitance and inductance. In the absence of a ground plane, the mutual capacitance of two parallel wires of length l and diameter d separated by a distance D in free space is given by

$$C_m \approx \frac{\pi \epsilon_0 l}{\ln\left(\frac{2D}{d}\right)} \tag{2.40}$$

where ϵ_0 is the permittivity of free space. The mutual inductance of the same configuration is given by [25]

$$M \approx \frac{\mu_0 l}{2\pi}\left[\ln\left(\frac{2l}{D}\right) - 1 + \frac{D}{l}\right] \tag{2.41}$$

To keep things simple, any non-nearest neighbor interactions between pins shall be ignored. This is not always a valid approximation. Because of the logarithmic falloff of M with distance, long range mutual inductances, in particular, can be significant.

For standard 1 mil (25 μm) diameter bond wire at UHF and microwave frequencies, skin depth effects dominate the resistance, since the skin depth δ in aluminum at 900 MHz is only 2.8 μm (see Section 3.6.1). Assuming current flows only within a skin depth of the surface of the wire, the resistance of bond wire of conductivity σ, diameter d and length l is given by

$$R \approx \frac{l}{\pi\delta\sigma d} \tag{2.42}$$

(2.42) gives a value of 0.11 Ω/mm for 1 mil diameter aluminum wire at 900 MHz. I shall now define a simpleminded lumped element package model that provides acceptable accuracy at UHF. Define i as the pin index. i numbers the pins of the package sequentially from some starting position, counter clockwise (see Figure 2-13). One then has

$$\begin{aligned}
L_{i,i} &= L_0 l_i \\
C_{i,i} &= C_0 w_i \\
R_i &= R_0 l_i \\
C_{i,i\pm1} &= C_1\left(\tfrac{l_i + l_{i\pm1}}{2}\right) \\
L_{i,i\pm1} &= L_1\left(\tfrac{l_i + l_{i\pm1}}{2}\right)
\end{aligned} \tag{2.43}$$

In (2.43), L_0 and R_0 denote the (frequency independent) per unit length inductance and resistance of the bond wire of length l_i. C_0 is a constant which depends on the package type. w_i is a package specific weighting function that takes into account capacitance variations between pins on the same package. L_1 and C_1 are

package dependent coupling inductance and capacitance coefficients which decrease with increasing spacing between pins. Finally, $L_{i,i}$, R_i and $C_{i,i}$ are the self inductance, resistance and self capacitance of pin i, and $L_{i,i\pm1}$ and $C_{i,i\pm1}$ are the mutual inductance and capacitance between pin i and its neighbors, pins $i \pm 1$.

To parameters of this model are obtained as follows. The formulas for mutual capacitance and resistance ((2.40) and (2.42), respectively) are used to get $C_1 = \frac{C_m}{l}$ and $R_0 = \frac{R}{l}$. C_0 is usually specified by the package manufacturer. To get L_0 and L_1, the self and mutual inductance formulas (2.39) and (2.41) are further simplified by assuming constant self or mutual inductance per unit length. This works because of the weak (logarithmic) dependence of $L_0 = \frac{L}{l}$ and $L_1 = \frac{M}{l}$ on the length l.

I now calculate typical values of the model parameters for a symmetric surface mount package (such as a leadless chip carrier or quad flat pack) at UHF. The package uses 1 mil diameter bond wire, the pitch (spacing between adjacent pins) is 40 mil (1 mm), and the typical bond wire length is 3 mm (i.e., $d = 1$ mil, $D = 40$ mil, $l = 3$ mm). Due to symmetry, all pins have similar capacitances, i.e., $w_i \approx 1$. From ((2.39), ((2.40), ((2.41) and ((2.42), we get $L_0 \approx 1$ nH/mm, $C_1 \approx 6.4$ fF/mm, $L_1 \approx 0.22$ nH/mm and $R_0 \approx 0.12$ Ω/mm. One can neglect C_1 in most cases since it so small. Finally, a typical (manufacturer specified) value of C_0 for such packages is 300 fF.

Consider the typical situation shown in Figure 2-14 where a pair of adjacent pins is used to feed a differential RF signal into the chip, with neighboring pins being at RF ground. The two pins and their bond wires are assumed to be identical. Each bond wire has a self inductance L and the mutual inductance between the two bond wires is M. Since the feed pins carry equal currents in opposite directions, M is negative (this is similar to what happens in a transmission line) and the inductance seen by the RF signal decreases from $2L$ to $2(L - |M|)$. A 20 % reduction is obtained for the typical values of $\frac{L}{l}$ and $\frac{M}{l}$ derived earlier. The percentage reduction increases as the bond wires grow longer, because $L_1 = \frac{M}{l}$ increases as l increases (equation (2.41)). In addition, it is found that C_1 is usually small enough to be neglected. In other words, the effects of capacitive coupling to the adjacent 'ground' pins are usually negligible.

As the current return path is moved further away (e.g., by connecting the RF

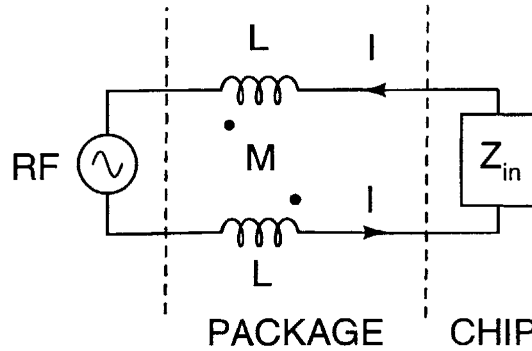


Figure 2-14: Typical RF feed scenario. The input RF signal is fed into the rectifier through the package. Z_{in} includes all package capacitances.

signal across non-adjacent pins), M remains negative but decreases in absolute value, increasing the total inductance. Bond wire inductance is unwanted in this application. One should thus connect the RF signal between adjacent pins and decrease D, the spacing between them, so that $|M|$ increases. Hence one favors packages with small pitch length.

2.4 RFID Protocol Issues

Protocols for passive backscatter RFID systems have recently been standardized and are in the process of being accepted globally. These protocols set the frequency ranges that can be used by RFID systems, set tag functionality requirements and define the communication protocols to be used between the reader and the tag. Like other protocols, they do not specify what the circuits or antennas on the tag have to be like in order to achieve the required functionality. The main factors to consider when designing the power extraction circuit which are protocol dependent are

1. The backscatter modulation scheme to be used for communicating with the reader (usually ASK or PSK), and the effect this has on the power collected and rectifier efficiency. PSK has been found to be more efficient [22].

2. The desired tag read rate, and consequently the maximum allowable power up time for the tag.

3. The allowable power supply droop during the interrogation phase. This directly impacts the size of the load capacitor needed at the rectifier output.

The commonest way to backscatter modulate is binary ASK, where the amount of power reflected to the reader is changed by changing the *magnitude* of the reflection coefficient Γ looking into the tag from the antenna. To do this, a resistive load is switched in parallel with the rectifier input, or the input is shorted out entirely. The extra resistive load burns power and lowers the Q of the input, degrading power collection and causing the power supply voltage to droop in either the 0 or the 1 state. In binary PSK, only the *phase* of Γ is modulated, usually with a varactor or switched capacitor. Ideally, this does not consume any extra power; the same power is collected in both modulation states. However, designing a low loss switch that does not degrade the Q of a switched capacitor often proves difficult. MOS varactors can be used instead.

Many protocols place strict limits on the power up time of a tag in order to maintain a high read rate. Often a continuous wave carrier signal is transmitted by the reader before querying the tag. This allows the reader to power up, and sets the maximum allowable power up time. Typical values for this time interval range from 50 μs to 500 μs. Finally, the tag protocols place limits on allowable frequency droop of the on-tag oscillator that sets the return data rate during the read out process. A typical value for this is 25 %. Since the oscillator frequency usually depends directly on the power supply voltage V_{DD}, this places limits on the minimum size of the load capacitor that holds V_{DD}.

2.5 System Testing

In this section I describe the procedures used to test the power extraction system, and present my test results. The design of the circuits and antennas used to obtain these results is described in the following chapters and constitutes the rest of this thesis.

2.5.1 Test Setup

The basic setup used for testing the power extraction system (tag) is shown in Figure 2-15. An RF source is used to generate a sine tone at the frequency of interest. This is then amplified by an external amplifier and fed into the transmit antenna. The tag is located a distance R away from the transmit antenna. The source meter is used to monitor the output DC voltage produced by the tag. It can also be used to sink the load current which has to be delivered by the tag. The test environment is not anechoic. It is a typical laboratory environment with shelves, tables and electronic test equipment surrounding the test setup. However, the path between the transmit antenna and the tag is kept clear and microwave absorbers are placed on the floor to approximate free space propagation conditions as much as possible. This approximation was verified to be accurate by actual measurement of the path loss using antennas with known gain.

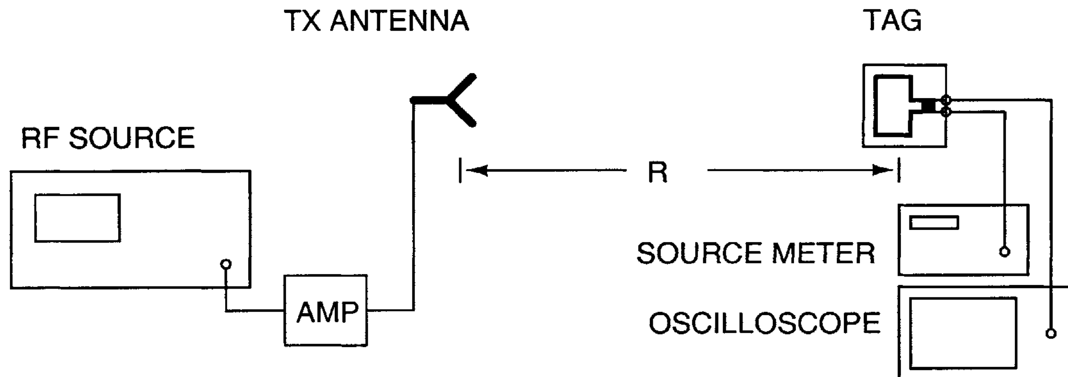


Figure 2-15: Setup used for testing the power extraction system.

2.5.2 Test Results

This section shows results obtained from testing a complete RF power extraction system. For collecting RF power the system uses the antenna shown in Figure 4-10. For supplying DC power to a load, the system uses the rectifier circuit shown in Figure 3-6 with three stages and $C_L = 2$ pF. The antenna was fabricated by a commercial printed circuit board vendor. The rectifier chip was fabricated by the

chip vendor MOSIS in a standard three metal, non-salicided CMOS process with a minimum feature size of 0.5 μm. The measured input impedance of the rectifier when packaged in a Leadless Chip Carrier (LCC) package was $Z_{in} \approx 6.5 + j175\ \Omega$ at 900 MHz, corresponding to an input capacitance $C_s \approx 1.0$ pF. A photograph of the completed tag is shown in Figure 2-16. The size of the tag is 2.9" $\times$ 2.0", most of which is occupied by the antenna.

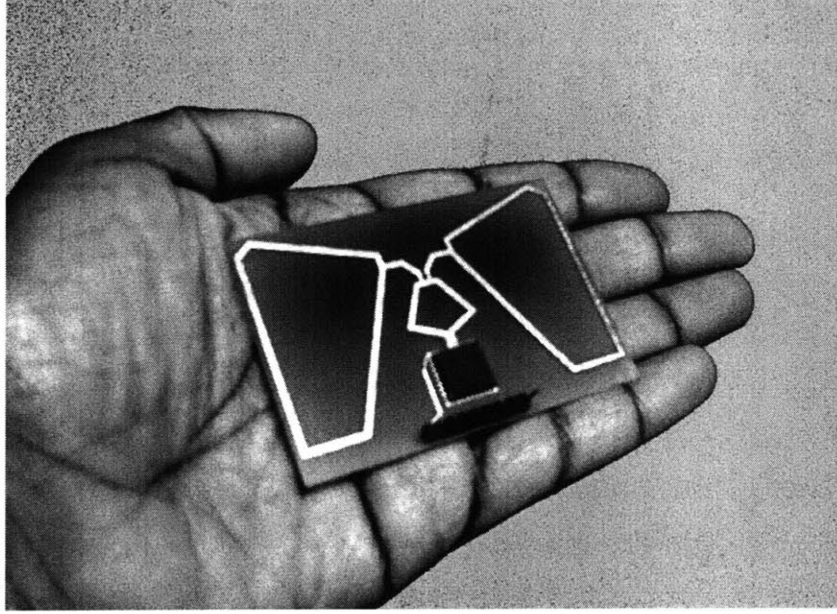


Figure 2-16: Photograph of completed power extraction system.

For characterizing the transmitter-tag RF link, I use the classical Friis equation, which is just a statement of power conservation in free space propagation conditions. According to this equation, P_{rec}, the maximum power received (in dB) by the tag at a distance r from the transmit antenna is given by

$$P_{rec} = P_s + G_A + G_t + G_r + X + 20 \log \left(\frac{\lambda}{4\pi r} \right) \tag{2.44}$$

where P_s is the power generated by the RF source (in dB), G_A is the gain of the amplifier in Figure 2-15 in dB, G_t and G_r are the gains of the transmit and receive (tag) antennas, respectively, in dBi, and X is a fudge factor which takes into account the departure of the test range from free space propagation conditions. The last term

represents the power loss due to free space propagation, and is called the path loss.

For my RF link measurements, one always spatially aligns the transmitter and the tag so that (2.44) applies. The tag is held aloft using a clamp and stand apparatus to eliminate near field effects as much as possible. Since the human body, being a big lossy dielectric mass, has pronounced effects on the received power, all measurements are made remotely using GPIB. One can use measured or simulated values of antenna gain for G_t and G_r. The amplifier gain G_A can be directly measured using a network analyzer. For estimating X, the transmit antenna and tag are removed and two antennas with known values of gain are placed at the same locations. The difference between the measured path loss between these antennas and the theoretical free space value is one's estimate for X. Once r (measured between the phase centers of the transmit and receive antennas) is known, one can use this estimate for X and find P_{rec}, the received power at the tag. The power actually available to the tag (assuming perfect impedance matching) is $P_A = \frac{P_{rec}}{2}$. Typical values of the parameters in (2.44) are shown in Table 2.5.2. G_t and G_A are measured values. G_r is simulated, and the path loss value is theoretical but experimentally verified (using the above procedure for estimating X) to within ± 0.5 dB.

Table 2.1: Test setup parameters

Parameter	**Symbol**	**Value**
Transmit antenna gain	G_t	4.4 dBi
Receive (tag) antenna gain	G_r	1.4 dBi
Amplifier gain	G_A	39 dB
Path loss (at 900 MHz)	-	-31.5 dB/m

Figure 2-17 shows the measured frequency response of the tag for P_A varying between 25 and 50 μW for a load current $I_L = 2$ μA. This value of I_L is typical for commercial RFID tags in this frequency range. One sees that the tag is almost correctly tuned for the United States RFID band (902-928 MHz). The operating frequency range is about 880-910 MHz. I define the power up threshold P_{th} for the tag to be the minimum available power level at which the output voltage is > 1 V at $I_L = 2$ μA over some defined bandwidth B. In my case, $B \approx 30$ MHz (the

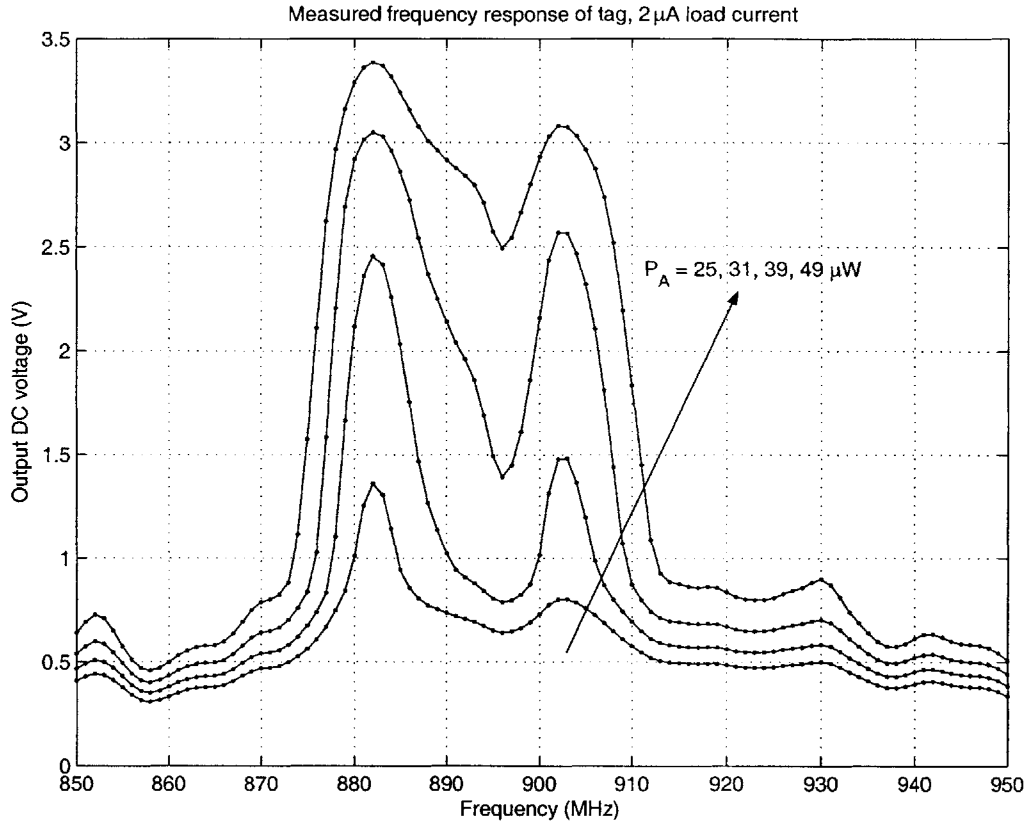


Figure 2-17: Measured frequency response of the tag for different available power levels P_A and $I_L = 2$ μA.

allocated RFID bandwidth in the United States is 26 MHz). From Figure 2-17, this gives a power up threshold of $P_{th} \approx 35$ μW (-14.5 dBm). The maximum allowable Equivalent Isotropic Radiated Power (EIRP) in this band is 4 W (36 dBm), i.e., the maximum value $P_s + G_A + G_t$ in (2.44) is 36 dBm. Assuming free space propagation (i.e., $X = 0$) this gives a maximum power up range of $r_{max} = 7.5$ meters (24' 7") for the tag. This level of performance compares well with commercial systems for the American market, which typically have $P_{th} \approx 40$ - 60 μW. Performance can be further improved by applying the layout and circuit optimization techniques described in this thesis (the rectifier on this tag was an early version which did not implement many of these optimizations).

Figure 2-18 shows measured load curves of the tag for different values of P_A. Load

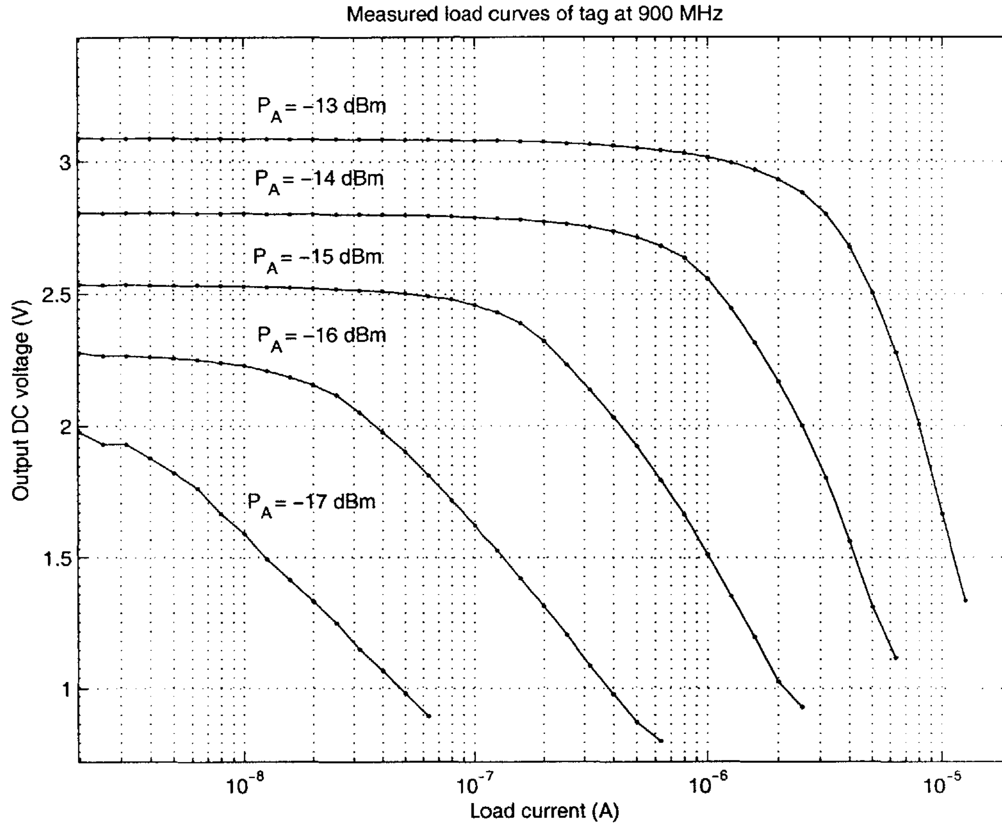


Figure 2-18: Measured load curves (varying load current I_L) of the tag for different available power levels P_A.

curves are obtained by varying the load current I_L drawn from the rectifier. Each curve corresponds to a certain value of rectifier output resistance R_{out}. R_{out} decreases as P_A increases. The load voltage is given by

$$V_L = V_0 - R_{out} I_L \tag{2.45}$$

where V_0 is the no load voltage, i.e., V_L when $I_L = 0$. The logarithmic scale in Figure 2-18 converts this linear relationship into the exponential load curves that were observed. The curves shift downward and to the left as P_A decreases. This is because decreasing P_A makes V_0 smaller and R_{out} larger.

Figure 2-19 shows the measured output voltage of the tag versus P_A for different load currents. Figures 2-18 and 2-19 are useful for predicting P_{th} for different load

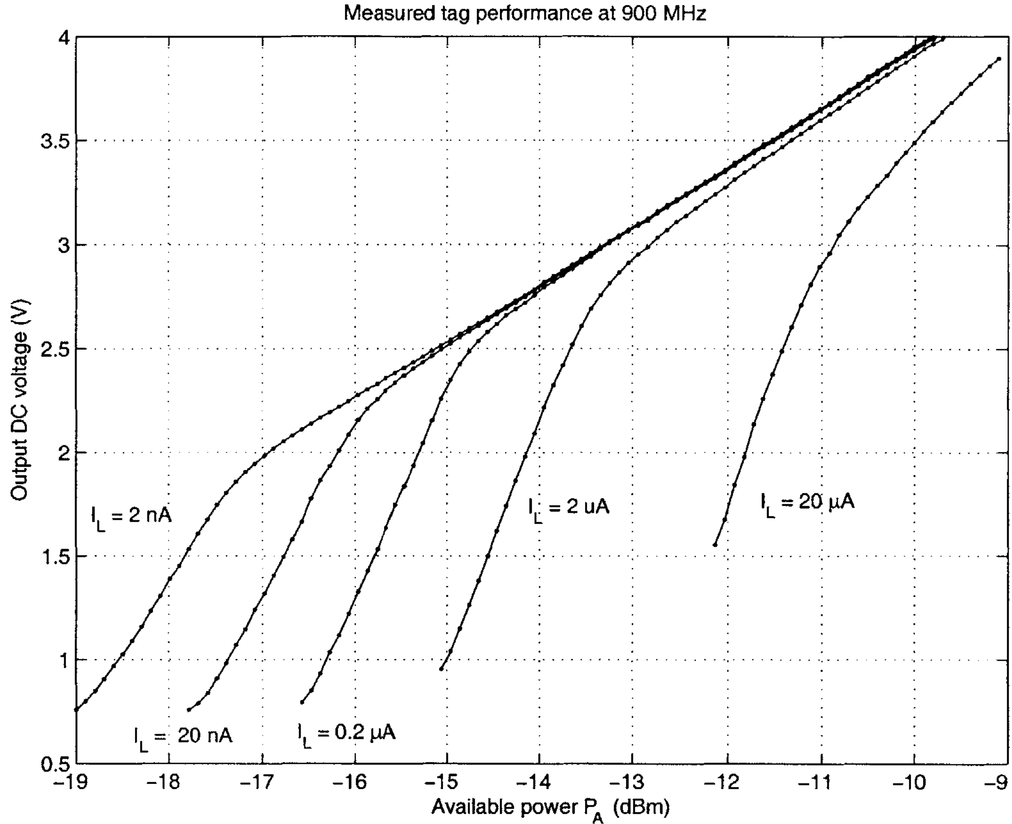


Figure 2-19: Measured output voltage of the tag versus available power level P_A for different load currents I_L.

currents. For example, if $I_L = 0.2$ μA and the threshold is set to 1 V, P_{th} = -16.3 dBm (23 μW). Thus P_{th} can be improved significantly if the power consumption of the tag circuits can be lowered.

Figure 2-20 shows measured DC to RF power conversion efficiencies η for the power extraction system for different values of P_A. The efficiency increases sharply as P_A increases, which is another way of saying that it is very difficult to efficiently extract DC from RF at low power levels.

Figure 2-21 shows the measured transient response of the tag's output voltage when the RF carrier was amplitude modulated at 100 Hz with a square wave having a 50 % duty cycle. Different modulation depths M were used, and the load current was 2 μA. All the signals had the same average power P_{av}, corresponding to $P_A = 31$

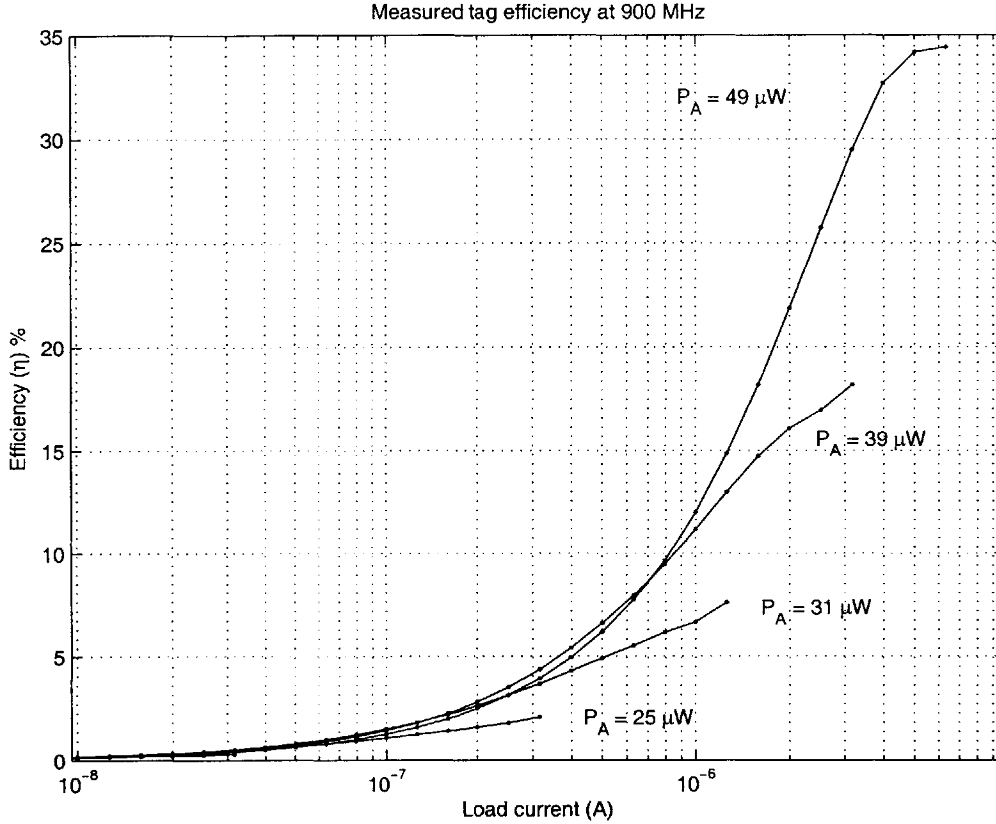


Figure 2-20: Measured power conversion efficiency η of the tag for different available power levels P_A.

μW. The maximum DC voltage increases as M increases because the peak power P_{max} of the amplitude modulated (AM) carrier increases. For AM signals with modulation depth M

$$P_{max} = P_{av}\frac{(1+M)^2}{1+M^2} \tag{2.46}$$

for $0 < M < 1$. The ratio $\frac{P_{max}}{P_{av}}$ increases from 1 to 2 as M increases from 0 to 1 (i.e., 0 to 100 %). One can see from Figure 2-21 that the power up time τ (defined as the 10 % to 90 % rise time of the output voltage) is a strong function of the received power. Assuming a simple exponential RC charging process, this means that the output resistance of the rectifier increases as the power level decreases (no big surprise). The load resistance at the rectifier output was $C_L = 20$ pF. The falling

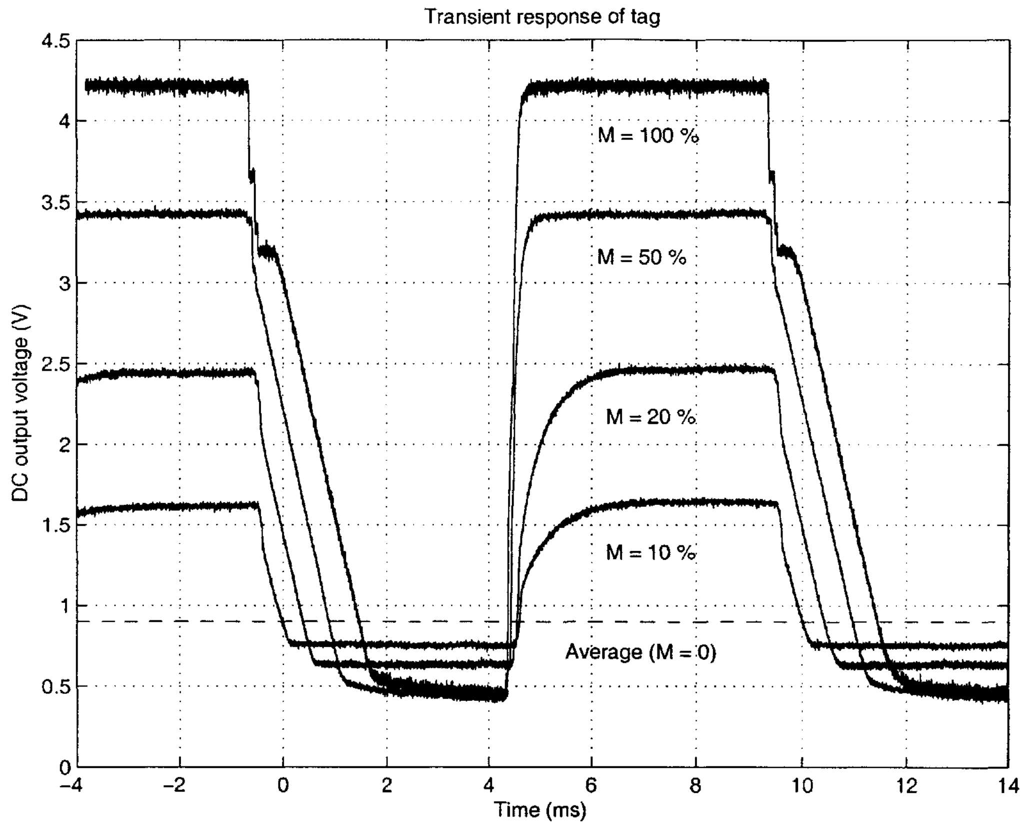


Figure 2-21: Measured transient response of the tag to amplitude modulation of the carrier. The modulation frequency was 100 Hz and modulation depth was M %

edges of the voltage waveforms show kinks. This is due to slew rate limitations of the output buffer used to test the tag and is not related to the rectifier itself.

In conclusion, a word of caution: the performance metrics shown so far, especially the maximum power up range, should be treated as upper bounds only. In real environments the power up range will be lower because of multipath effects, resonance frequency shifts, field nulls and so on. Experimentally the center frequency of the tag antenna was found to shift by ± 20 MHz or more depending on location in the room where tests were conducted. This variation is expected to become worse when the tags are attached to various packaging materials, further degrading the power up range. Hence adaptation of the tag's resonant frequency or other properties based on environmental conditions may be of interest for future designs.

Chapter 3

Rectifier Design

3.1 Introduction

All rectifier designs use a non-Ohmic device as the basic rectifying circuit element. The I-V characteristic of this element is asymmetric about the voltage axis and usually nonlinear. Historically, such structures have included point contact Schottky diodes used in crystal radio sets (see [32] for a description of Indian pioneer J.C. Bose's experiments with point contact rectifiers), vacuum diodes, thyratrons and mercury arc rectifiers (ignitrons). Today, semiconductor devices dominate in practically all rectifier applications, except at very high power levels, where ignitrons and vacuum diodes are still used. Semiconductor devices used as rectifiers include junction diodes, metal-semiconductor (Schottky) diodes, triacs, Silicon Controlled Rectifiers (SCR's) and so on, and are available in a wide range of sizes and power handling capabilities.

The applications of rectifiers have tended to follow two distinct streams. The first has been concentrated at high frequencies and low power levels, mostly for radio and microwave detectors. The other has been at high power levels and line frequencies (50 or 60 Hz), and constitutes traditional power electronics. For an excellent review of the development of power electronics, see [40]. The high frequency, high power and low frequency, low power regions of the power-frequency plane have fewer practical applications and have been less explored.

The work described in this thesis uses power electronics concepts and extends

them into the high frequency, low power 'radio' region, the application being the wireless transmission of power at RF frequencies.

3.2 Common Rectifier Structures

In this section, I first discuss common diode-based rectifier structures, and then go on to discuss some innovative MOS transistor based synchronous rectifiers which were used in this thesis for RF power extraction.

3.3 Schottky Based Rectifiers

Schottky diodes are widely used in rectifiers because their reverse saturation currents can often be made orders of magnitude higher than common semiconductor p-n junction diodes. This translates into lower forward voltage drop when supplying a given load current. In addition, Schottky diodes are largely majority carrier devices and do not have minority carrier charge storage. This leads to lower capacitance and better RF performance than p-n junction diodes. Integrated Schottky diode based voltage multipliers have been used for powering up transponder chips at UHF [22] and microwave [2] frequencies and in rectennas (rectifier-antenna combinations) [19].

Figure 3-1 shows a simple diode based capacitive voltage multiplier circuit. Such circuits have been used for many years in various rectifier and charge pump applications. This particular circuit was used in [22] for powering up a UHF RFID tag, and consists of N diode doubler stages in cascade. The RF input is fed in parallel to the all the stages through pump capacitors C_P. The DC outputs are developed across the load capacitors C_L and add up in series to produce the final load voltage V_L. V_L is approximately given by

$$V_L = 2N\left(V_{in} - V_{th}\right) \tag{3.1}$$

where V_{in} is the amplitude of the RF input, and V_{th} is the forward voltage drop of the diodes at the specified load current I_L. Ideally, each doubler stage produces a

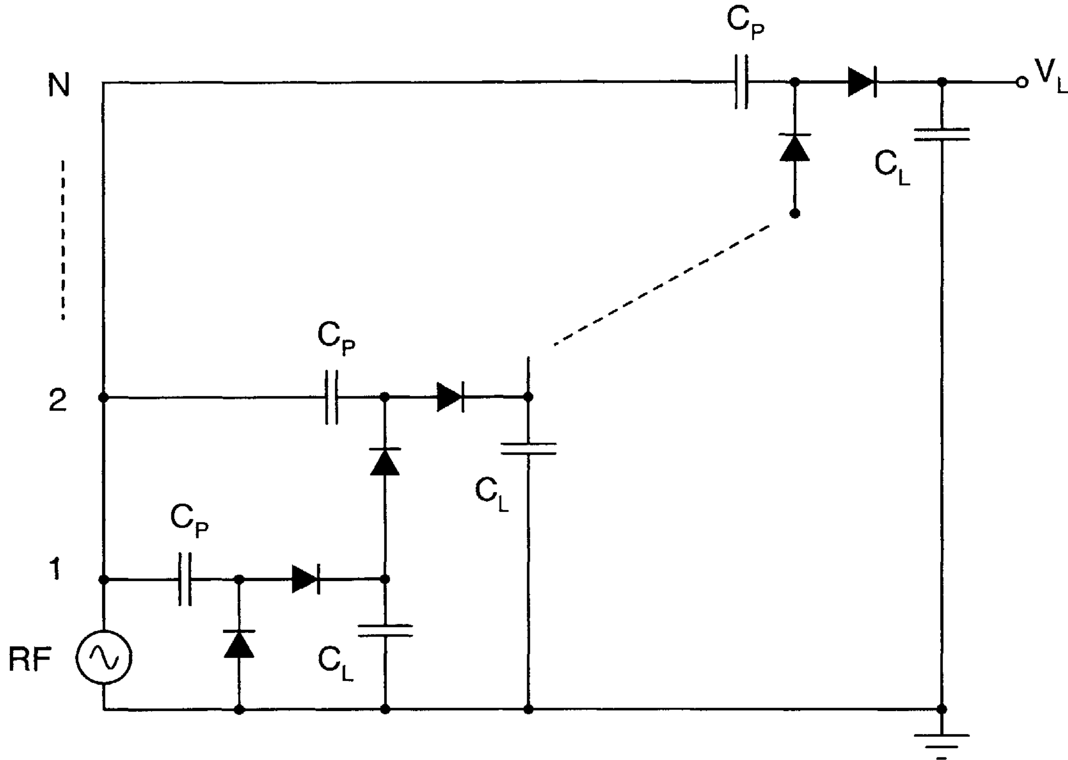


Figure 3-1: N-stage capacitive voltage multiplier circuit using diode doubler stages.

DC voltage $2(V_{in} - V_{th})$. To obtain operation at low values of V_{in}, Schottky (metal-semiconductor) diodes with low values of V_{th} are used. For this circuit (and in general for all diode based rectifiers), I want V_{th} to be as low as possible to improve performance. This is because reverse biased diodes conduct a constant current I_S. Ideally, I_S is constant with reverse bias and negligible compared to forward biased current levels. Thus decreasing V_{th} reduces the rectifier output resistance and does not appreciably increase the reverse conduction loss. One way to decrease V_{th} (i.e., increase I_S) for a given junction type is to increase the junction area. This cannot be carried on indefinitely because the junction capacitance is also proportional to the area of the junction. The optimum diode size for a particular circuit thus depends on frequency of operation and the load current level (see Section 3.6.2). Figure 3-2 shows the simulated rectification curves of the circuit in Figure 3-1 for various values of N when Schottky diodes developed within a standard CMOS process (see Appendix A.1) were

used. Rectification curves are defined as plots of the output DC voltage V_L of a rectifier versus the input RF amplitude V_{in} for a given load current I_L. The RF frequency was 900 MHz, $I_L = 2$ μA, $C_P = 250$ fF and $C_L = 1$ pF. The value of C_P and the diode sizes were optimized to give best performance at this frequency as described in Section 3.6.2.

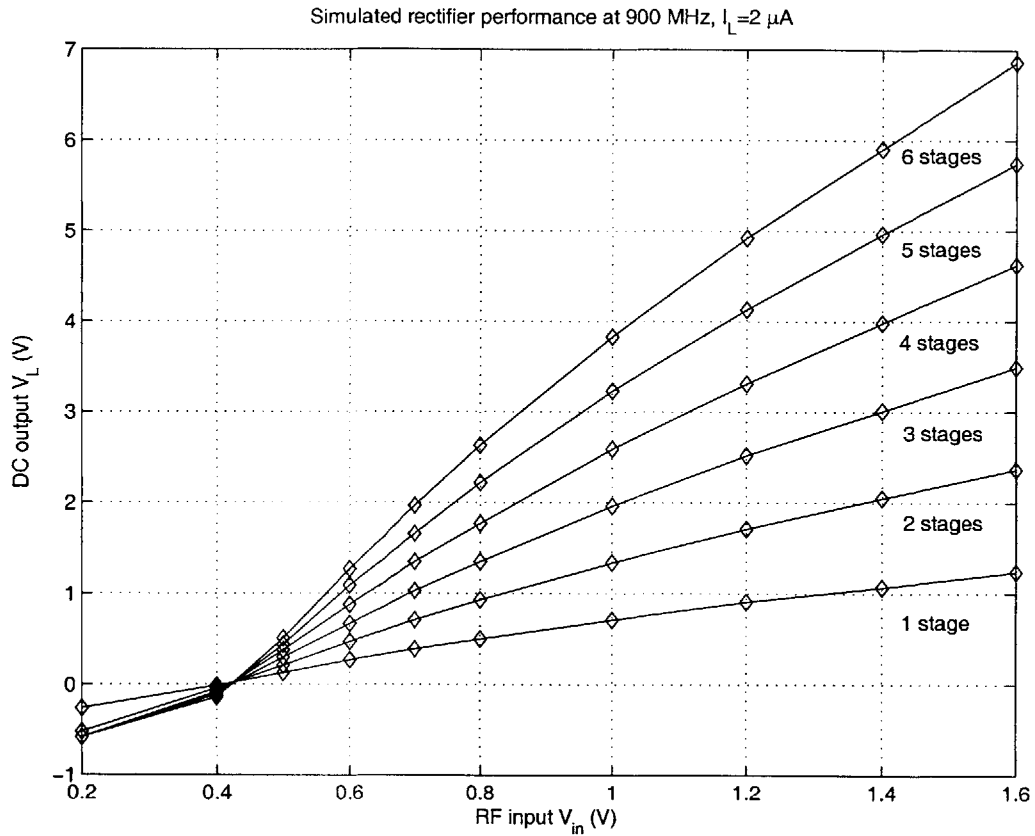


Figure 3-2: Simulated rectification curves (output DC voltage V_L versus RF input voltage V_{in}) for Schottky diode rectifiers with different values of N (the number of stages). $I_L = 2$ μA.

3.4 The Four Transistor Cell

Devices used in power electronics can be passive devices like diodes, or active devices like transistors and thyristors. Passive devices have two terminals and are not con-

trollable - their behavior is determined entirely by the voltages at the two terminals. Active devices have one or more control terminals which can be used to change the behavior of the device. The MOSFET gate and body (bulk) and the BJT base are examples. Thus the basic difference between transistors and diodes as rectifying devices is the presence of one or more control terminals. This is also true of other three and four terminal devices used as rectifiers in power electronics, such as thyristors, silicon controlled rectifiers and triacs. Synchronous rectifiers [23], which are more efficient than diode based rectifiers, need active devices. However, control terminals have an obvious disadvantage for self powered applications - what does one control the control terminals with? One possibility is to derive the control voltages from the terminal voltages, i.e., create a self-driven structure.

Essentially the only controllable device available to a designer in a normal CMOS process is the MOSFET. The MOSFET is a voltage controlled device that has two main regions of operation: linear (triode) and saturation. In the linear region, the device acts like a resistor whose resistance is controlled by V_G and V_B, the gate and bulk voltages. Drain and source terminals are symmetric and both control the current through the transistor. In saturation, the current through the MOSFET is primarily controlled by its gate-source and bulk-source voltages V_{GS} and V_{BS}. The drain voltage has much less effect. In other words, the device looks like a constant current source or sink to the drain. In the absence of active control circuitry, the gate and bulk voltages must be derived from the input RF signal itself.

Another thing to consider is that a standard CMOS process has two types of MOS transistors available - PMOS and NMOS, whose control voltages are complementary to each other. The most basic self-driven synchronous rectifier structure in this case is the four transistor cell shown in Figure 3-3. V_P and $\overline{V_P}$ are complementary (differential) AC signals, and the rectified DC voltage is $V_H - V_L$. Unlike many rectifier circuits, the direction of power flow in this structure cannot be reversed. This is because the transistors are driven by the AC input signals V_P and $\overline{V_P}$ themselves. In other words, placing a voltage source between V_H and V_L does not produce a voltage difference between V_P and $\overline{V_P}$. It has to be the other way around.

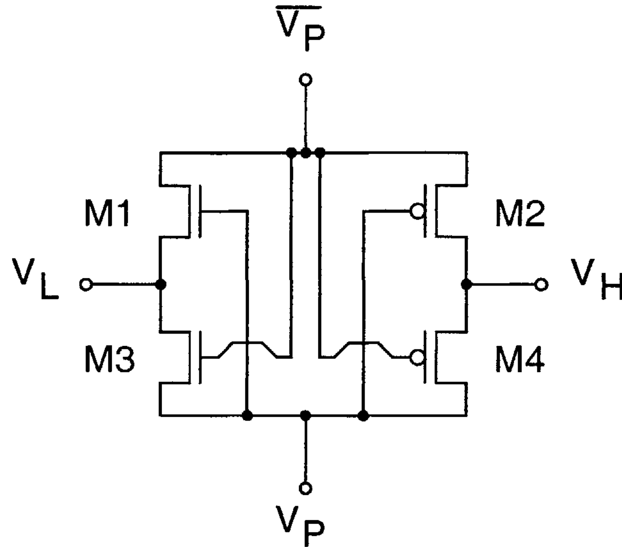


Figure 3-3: The four transistor cell using MOS devices. Body (bulk) terminals are not shown.

The operation of the four transistor cell is easy to understand if V_P and $\overline{V_P}$ are assumed to be square waves of large enough amplitude to turn the transistors on and off. The transistors then operate as switches. During half of the switching cycle, V_P is high and $\overline{V_P}$ is low. In this case $M1$ and $M4$ are on and $M2$ and $M3$ are off. Current flows into V_H through $M4$ and out of V_L through $M1$. During the other half of the cycle, $M1$ and $M4$ turn off and $M2$ and $M3$ are on, but the current flow at V_H and V_L has the same direction as before. Thus a DC voltage is developed across a load connected between V_H and V_L. Typical steady state waveforms are shown in Figure 3-4. In general, $V_{DC} = (V_H - V_L) = (2V_{RF} - V_{drop})$. The quantity V_{drop} represents losses due to switch resistance and reverse conduction. V_{drop} increases as the load current I_L increases and V_{RF} decreases. The behavior with a sinusoidal input is similar, but the efficiency will be lower (i.e., output resistance will be higher) since the devices only turn on for part of the input cycle.

The use of a single switching cell as a rectifier is shown in Figure 3-5. The RF signal from the antenna is matched to the input impedance of the rectifier using a matching network. The output voltage V_{DC} is developed across the load capacitor C_L. The energy stored in C_L is used to run the rest of the chip circuitry. Ground

represents the potential of the chip substrate.

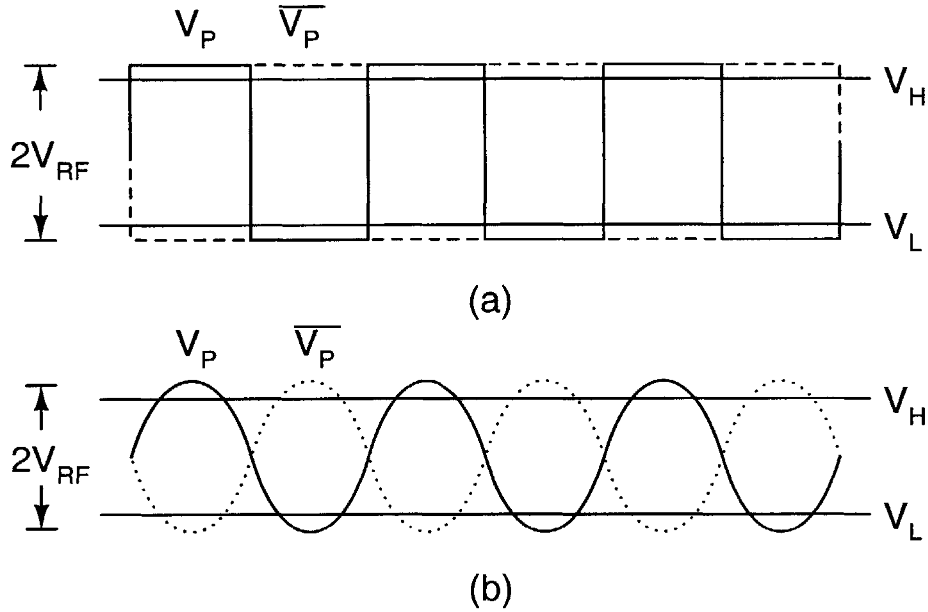


Figure 3-4: Typical steady state waveforms produced by the four transistor switching cell for (a) square wave and (b) sinusoidal inputs.

Assuming lossless switches, the maximum value of $(V_H - V_L)$ that can be obtained from a single four transistor cell is limited to $2V_{RF}$, the peak to peak amplitude of V_P and $\overline{V_P}$ (i.e., $V_{drop} = 0$). To obtain larger voltages, several cells can be cascaded in series. This is described in the next section.

3.4.1 Cascading Rectifier Cells

To obtain large DC voltages from a given input DC amplitude, several four transistor cells can be cascaded in series. Figure 3-6 shows the resultant circuit when three cells are used. In general, N stages can be cascaded in this way. The input RF signal V_{RF} is assumed to be purely differential (no defined common mode voltage). This leads to circuit startup issues which shall be discussed in the next section. V_P and $\overline{V_P}$ for the first stage are directly connected to V_{RF}. Succeeding stages are capacitively coupled to V_{RF} through C_P, allowing V_{DC} to build up at the output. Thus the circuit behaves as a charge pump voltage multiplier.

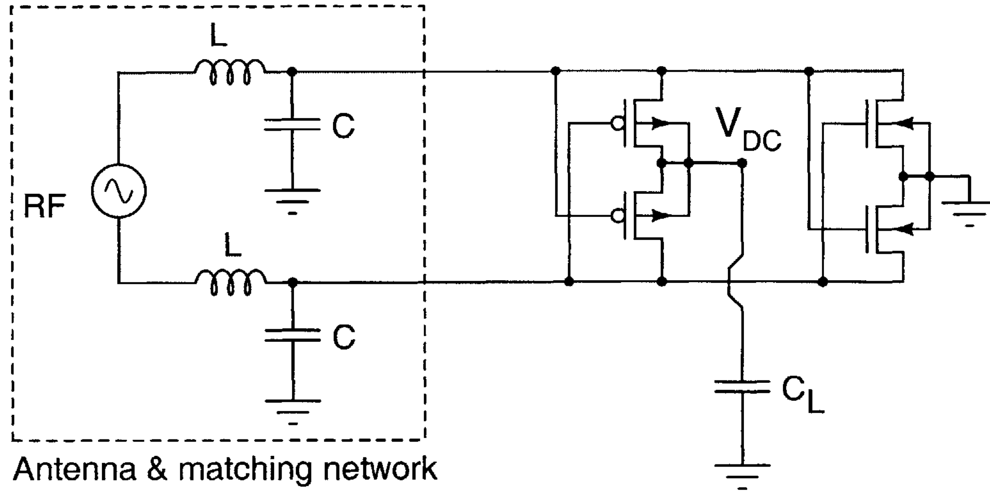


Figure 3-5: The switching cell configured as a rectifier. NMOS bulk terminals are tied to the chip substrate, which is taken to be the ground terminal.

For N stages, one expects the output DC voltage to be $V_{DC} = N\left(2V_{RF} - V_{drop}\right)$. However, in practical implementations, V_{DC} is generally found to be lower than expected. This is because V_{drop} is not constant and increases down the cascade of cells due to body bias effects (see Section 3.4.4).

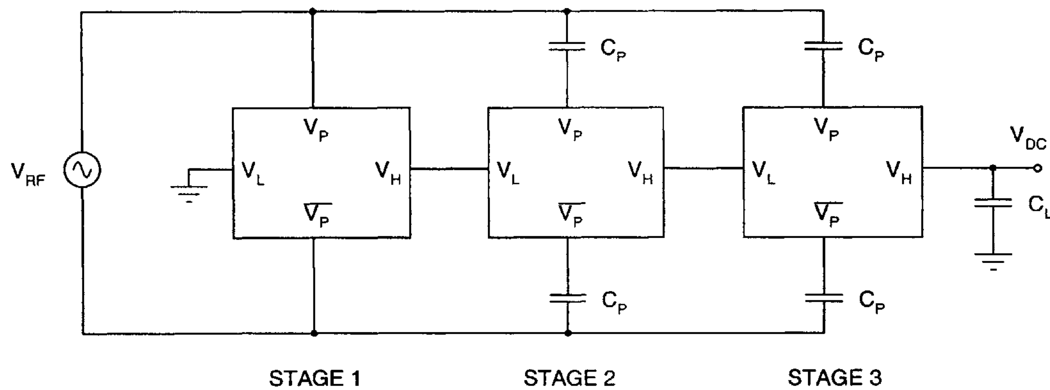


Figure 3-6: A rectifier formed by cascading three switching cells in series. For power extraction applications V_{RF} is generated by an antenna and impedance matching network.

The simple view of the operation of the four transistor cell (and structures derived from it) that has been described so far is complicated by several factors. I shall consider them in turn.

3.4.2 Startup Behavior

The startup behavior of the four transistor cell is interesting, Assume that initially all four nodes in Figure 3-3 - V_H, V_L, V_P and $\overline{V_P}$, are at the substrate potential (ground). When the RF signal is applied to V_P and $\overline{V_P}$, both nodes have a common mode voltage at ground, so the RF causes both nodes to swing above and below the substrate. If the RF amplitude is large enough, this will cause the parasitic body-source and body-drain junction diodes (see Figure 3-7) of the FETs to turn on. There are eight diodes to consider (two for each transistor in the cell). Refer to Figure 3-8(a). Assuming that $V_L \geq 0$, the two diodes connected to V_L are reverse biased and carry negligible current. Similarly, the two diodes connected to V_H are zero biased and can be ignored. This leaves one with the simplified circuit of Figure 3-8(b), which shows the four remaining diodes.

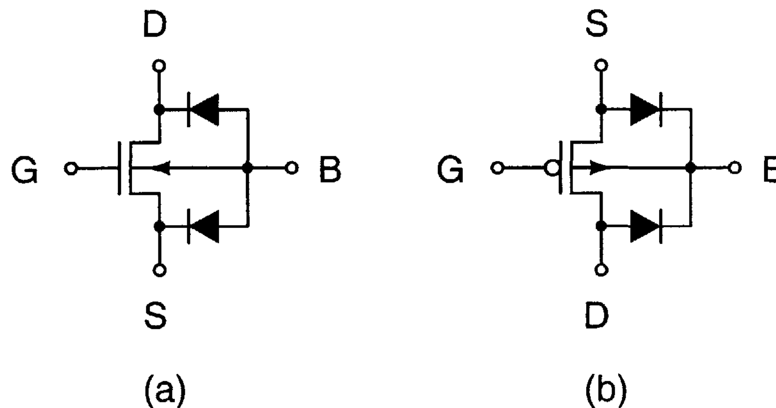


Figure 3-7: Parasitic body-source and body-drain junction diodes for (a) NMOS and (b) PMOS transistors.

One appealing feature of the four transistor cell is evident from Figure 3-8(b). The four parasitic diodes form a full bridge rectifier connected between V_P and $\overline{V_P}$. The output of this rectifier is between V_H and ground. The parasitic rectifier is operational during startup. As a result, the common mode voltage of nodes V_P and $\overline{V_P}$ increases to $V_{RF} - V_{th}$, where V_{RF} is the input RF voltage amplitude and V_{th} is the forward voltage drop of the parasitic diodes (about 0.6 - 0.7 V for CMOS processes). At this point the parasitic bridge turns off and the switching cell takes over. The resultant

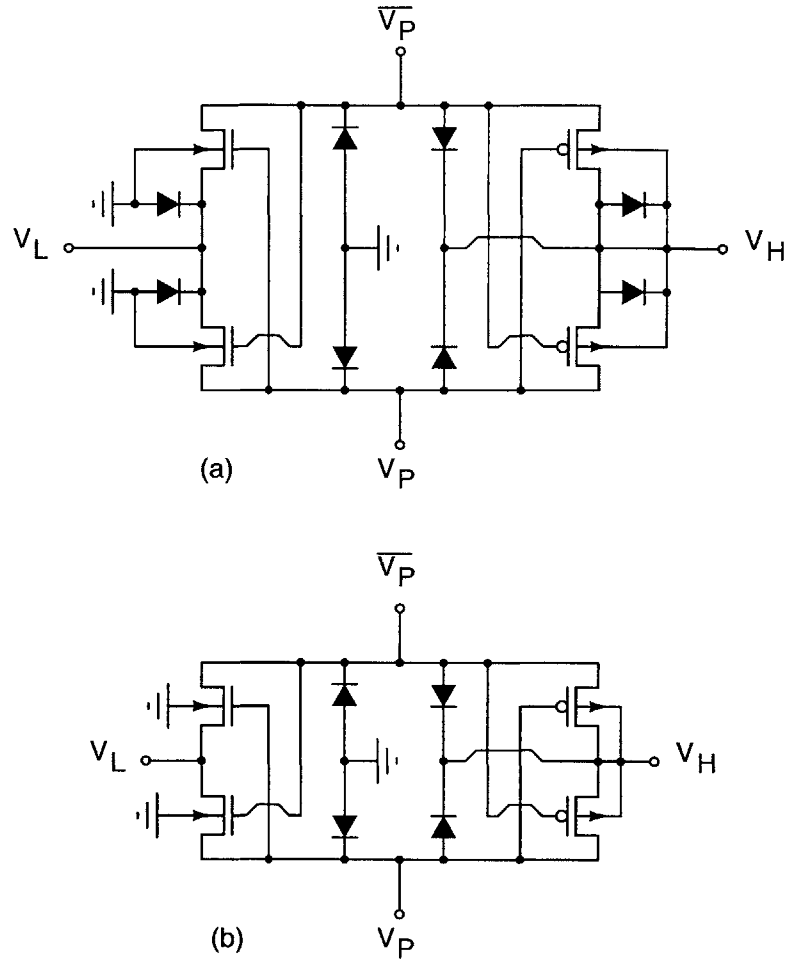


Figure 3-8: The four transistor cell with MOS parasitic diodes included (a) actual and (b) simplified. A n-well CMOS process is assumed.

waveforms are shown in Figure 3-9. The voltage at V_P and $\overline{V_P}$ in steady state swings below V_L by approximately $\frac{V_{drop}}{2}$. Normally, however, $V_{drop} \ll V_{th}$, so that V_P and $\overline{V_P}$ never swing below ground far enough for the parasitic diodes to turn on again even if $V_L = 0$ (grounded). The parasitic bridge thus remains off during normal operating conditions. It is only of interest during startup.

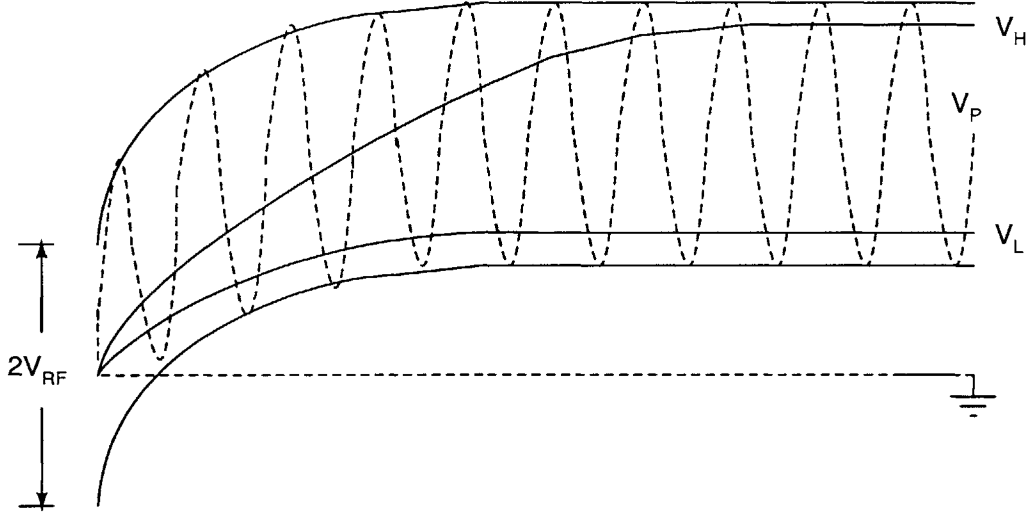


Figure 3-9: Typical transient startup waveforms produced by the four transistor switching cell for sinusoidal inputs.

3.4.3 Threshold Voltage

The biggest problem faced by the circuit shown in Figure 3-3 at low power levels is the threshold (turn on) voltage V_T of the transistors. In subthreshold or weak inversion operation, i.e., when $V_{GS} < V_T$, the current I_{DS} through a MOSFET depends exponentially on V_{GS} and V_T. For a NMOS device

$$I_{DS} = I_S \exp\left(-\frac{V_T}{\phi_T}\right) \exp\left(\frac{\kappa V_{GS}}{\phi_T}\right) \frac{W}{L}\left[1 - \exp\left(\frac{V_{DS}}{\phi_T}\right)\right] \tag{3.2}$$

where I_S is a constant, $\kappa < 1$ is the subthreshold slope parameter, ϕ_T is the thermal voltage and $\frac{W}{L}$ is the aspect ratio of the device. The corresponding equation for PMOS transistors is similar. Thus, for a given RF amplitude (fixed maximum value of V_{GS}), the current through the switch decreases exponentially with V_T. In other words, for a given input amplitude, the output resistance of the rectifier increases exponentially with V_T.

In above threshold or strong inversion operation, i.e., when $V_{GS} > V_T$, two regimes may be distinguished. When $V_{DS} < (V_{GS} - V_T)$ (the triode or linear regime), I_{DS} depends linearly on $(V_{GS} - V_T)$. For a long channel NMOS device, the classic expression

for I_{DS} is

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \tag{3.3}$$

where μ_n is the electron mobility and C_{ox} is the oxide capacitance per unit area. When $V_{DS} > (V_{GS} - V_T)$, the transistor is said to be saturated, and the current depends quadratically (for a long channel transistor) on $(V_{GS} - V_T)$ and is given by

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \tag{3.4}$$

where λ is a fitting parameter. (3.3) and (3.4) are similar for PMOS transistors. All else being equal, therefore, one expects the rectifier performance to improve as V_T is decreased. This should be true for all regions of transistor operation. The output resistance and I_{DS} should decrease, the output voltage should increase, and the efficiency should go up.

Optimum Threshold Voltage

Figure 3-10 shows the simulated effects of adjusting the threshold voltages of the NMOS and PMOS transistors ($V_{TN,0}$ and $V_{TP,0}$, respectively) on the output voltage V_L of a five stage rectifier. The load current $I_L = 2$ μA and the input RF amplitude was $V_{in} = 1$ V at 900 MHz. One sees that, as expected, decreasing the threshold voltages up to a point decreases the output resistance of the rectifier and increases V_L. Beyond this, however, further reductions in threshold voltage do not cause V_L to go up. In fact, V_L starts to decrease slightly. Thus there is an optimum pair of threshold voltages $(V_{TN,opt}, V_{TP,opt})$ which maximizes V_L. In this case, $V_{TN,opt} \approx 0.05$ V and $V_{TP,opt} \approx -0.3$ V. The presence of this optimum is easily explained. When the threshold voltages are low, the predominant loss mechanism in the rectifier is the on resistance R_{on} of the switches. For a given V_{in}, decreasing the threshold voltage rapidly decreases the on resistance to the point where $R_{on} I_L \ll V_L$. Further decreases in R_{on} have almost no effect on V_L. The dominant loss mechanism then becomes reverse conduction caused by the switches not switching off completely during the off

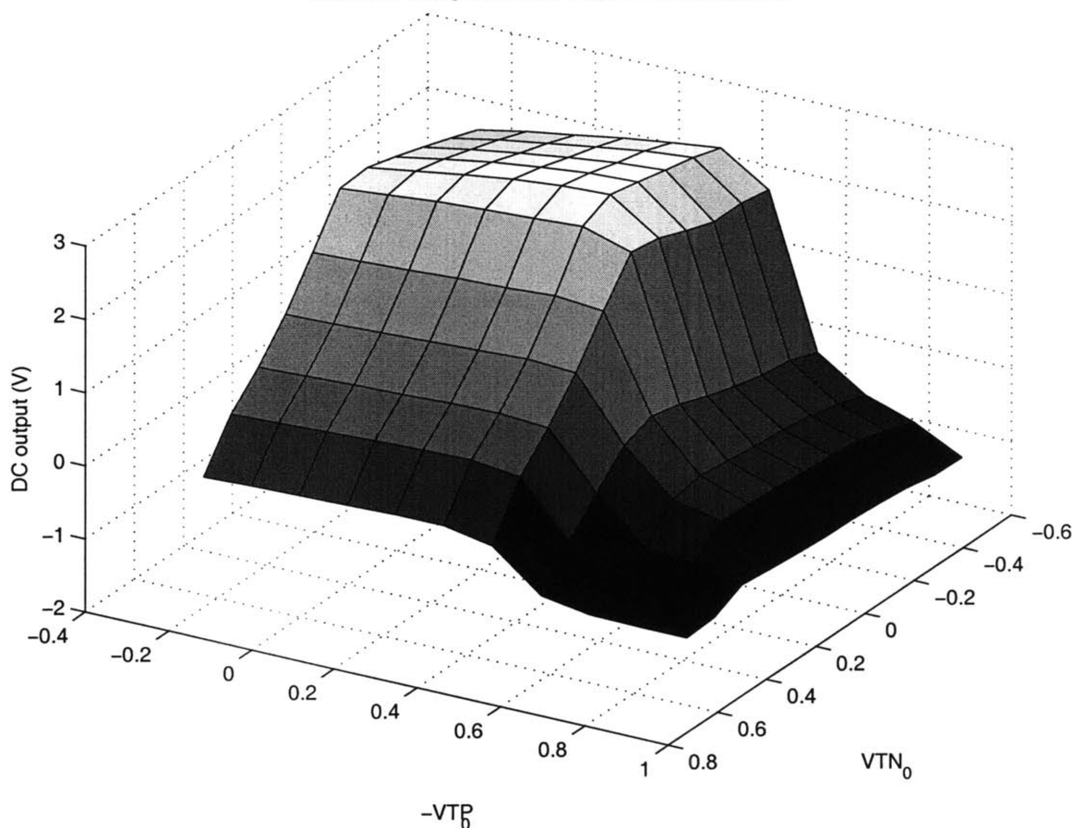


Figure 3-10: Effect of changing the PMOS and NMOS threshold voltages (at zero V_{BS}) VTP_0 and VTN_0 on the output DC voltage of a five stage rectifier (simulated). The input RF amplitude was 1 V at 900 MHz and the load current was 2 μA.

cycle. This effect worsens as the threshold voltage is decreased, causing the slight negative trend in V_L beyond $(V_{TN,opt}, V_{TP,opt})$.

The absolute values of $(V_{TN,opt}, V_{TP,opt})$ were found to vary significantly with V_{in}, I_L and N, the number of stages in the rectifier. The strategy used in this case was to optimize the threshold voltages for the lowest V_{in} which gives the required V_L at the specified load current, and rely on the broadness of the optimum in Figure 3-10 for acceptable performance higher V_{in}. Experiments indicate that this strategy is reasonable, mainly because increasing V_{in} is extremely effective at decreasing R_{on} (the 'dead zone' effect). However, I have not yet indicated *how* I intend to change transistor threshold voltages. This is discussed in Section 3.5.

3.4.4 Body Effects

The circuit shown in Figure 3-3 did not show the bulk terminals, i.e., ignored the fact that the MOSFET is a four terminal device. The real situation for a n-well CMOS process is shown in Figure 3-11. The bulk terminals of both NMOS transistors are at the substrate (ground) potential, while the PMOS bulk (well) terminals have been tied to V_H. From a system viewpoint, the main effect of the bulk is to increase the threshold voltage of the NMOS transistors as V_L increases, as follows

$$V_{TN} = V_{TN,0} + \gamma \left(\sqrt{2\phi_F + V_{SB}} - \sqrt{2\phi_F} \right) \tag{3.5}$$

where $V_{TN,0}$ is the NMOS threshold voltage when $V_{SB} = 0$, $2\phi_F$ is the surface potential of the transistor in strong inversion, γ is the body effect parameter, and V_{SB} is the source-bulk voltage. γ is given by

$$\gamma = \frac{\sqrt{2\epsilon_{Si} q N_A}}{C_{ox}} \tag{3.6}$$

where N_A is the doping density of the bulk, and ϵ_{Si} is the dielectric constant of silicon. Typical values of γ range from 0.5 - 0.75 $\mathrm{V}^{-1/2}$. In this case, $V_{SB,i}$, the value of V_{SB} for the NMOS transistors in the i-th rectifier stage, is equal to V_L at the input of that stage. $V_{SB,i}$ is proportional to $(i-1)$, assuming all the $(i-1)$ previous stages contribute the same amount to V_L. Thus the V_{TN} increase $V_{TN} - V_{TN,0}$ is $\propto \sqrt{i}$. This increase in V_{TN} reduces the efficiency and voltage gain $(V_H - V_L)$ obtained from each stage when many cells are cascaded in series.

The effect of body bias is shown in Figure 3-12, which compares the simulated output voltages V_L obtained from rectifiers with different values of N, the number of stages, for two cases. The first case, shown with solid lines, is when the bulk terminals of NMOS transistors are grounded (as in Figure 3-11). This means that $V_{BS} > 0$ for all NMOS transistors except those in the first stage. This body bias effect causes NMOS threshold voltages to increase as more stages are added to the rectifier. The second case, shown with dashed lines, is when all NMOS bulk terminals are tied to the corresponding source terminal, ensuring $V_{SB} = 0$ and eliminating body

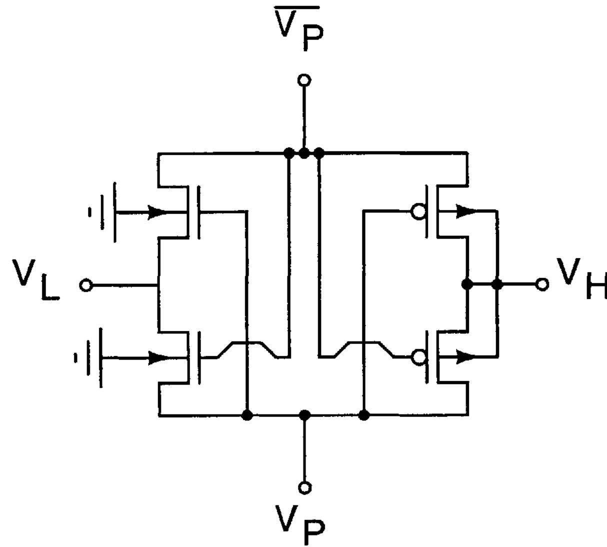


Figure 3-11: The four transistor cell using MOS devices in a n-well CMOS process. NMOS bulk terminals are tied to the chip substrate, which is taken to be the ground terminal.

bias effects. The simulation frequency was 900 MHz and different load currents I_L were considered. The input RF amplitude was $V_{in} = 1$ V. The figure shows that for low output voltages, both cases produce the same output voltage (i.e., body effects are negligible). However, for large output voltages, body bias causes large decreases in rectifier efficiency by increasing V_{TN} (and thus the output resistance of the rectifier). Thus the output voltage V_L of the first case (when body effects are present) becomes smaller than that of the second case. The presence of an optimum number of stages which maximizes V_L when body effects are considered is however harder to understand. One expects V_L to saturate but not actually decrease as N, the number of stages, increases.

The cause of this behavior can be reasoned out. Clearly the voltage gain per stage decreases down the rectifier chain due to the body effect. By itself, this would just cause V_L to saturate as N was increased. V_L can *decrease* with increasing N only if adding another stage at the end of the rectifier chain decreases the voltage gains of the stages before it by an amount large enough to exceed the gain added by the last stage. Simulations support this theory. Succeeding stages load stages before them,

and the voltage gain per stage decreases as the number of stages increases. This causes the first case (which is what one actually gets in practice) to have an value of N (number of stages) which maximizes V_L. In Figure 3-12, this optimal number of stages is $N = 5$. Figure 3-12 also shows that the no load ($I_L = 0$) and loaded curves move linearly further apart as N increases. This means that the output resistance of the rectifier increases approximately linearly with N, as expected. In other words, the resistance per stage is constant. In Figure 3-12, this is $R_S \approx 200$ kΩ.

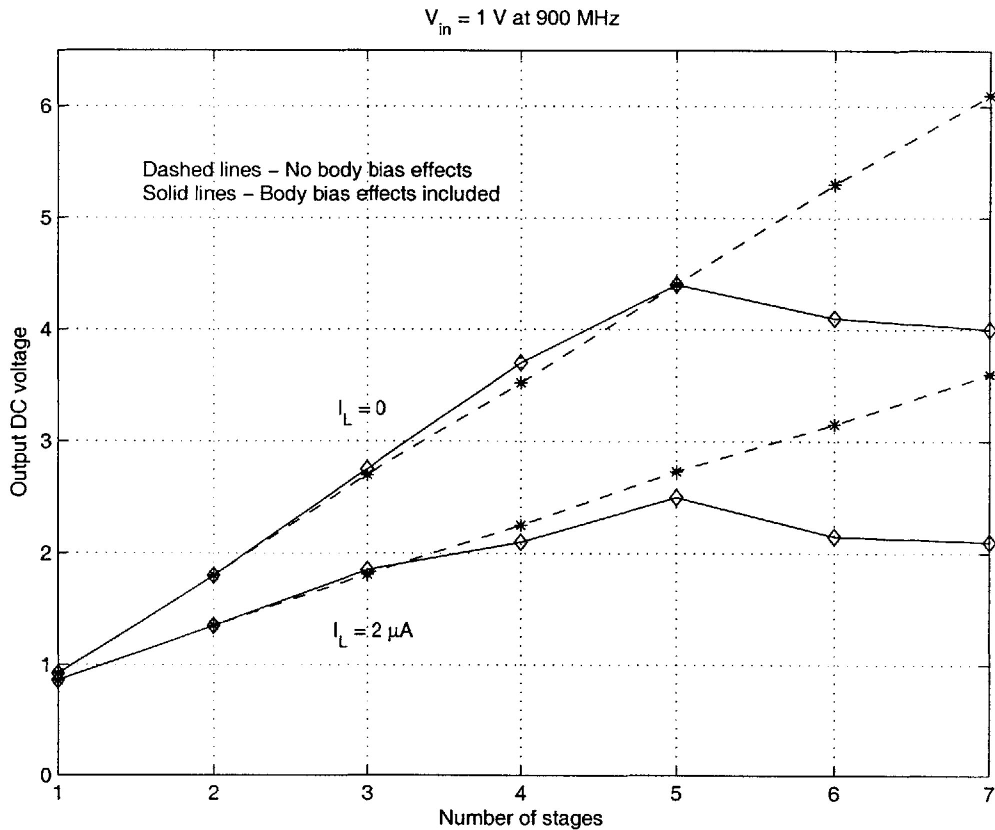


Figure 3-12: Body bias effect on rectifier performance for different load currents I_L. Solid lines show the output voltage when NMOS bulks are grounded ($V_{SB} \geq 0$). Dashed lines show the output when NMOS bulks are tied to sources ($V_{SB} = 0$, no body effects).

Figure 3-13 shows simulated rectification curves of rectifiers with different numbers of stages at $I_L = 2$ μA. These curves show the rectifier function $F(V_{in}, I_L)$ (Section 2.3.4) for a fixed value of I_L. The effects of body bias have been included

in these plots. One notices that for low values of V_{in}, V_L for a fixed V_{in} *decreases* as the number of stages increases (though the effect is not large). This behavior makes sense in the context of the theory developed in the previous paragraph. For low V_{in}, the voltage gain per stage is so small that the decrease in gain per stage when more stages are added exceeds the gain of the added stages themselves. The situation is reversed for larger input voltages since the voltage gain per stage is then large enough for this effect to be overshadowed.

Figure 3-13 can be used to find the minimum value of V_{in} required to obtain the required V_L at a given load current. This value of V_{in} can then be used to find the power up threshold, using the theory developed in Section 2.3.4. For example, from Figure 3-13, one sees that $V_{to} \approx 0.92$ V for a three stage switching rectifier to supply $V_L = 1.0$ V at $I_L = 2$ μA. One can now use the theoretical power up threshold curves shown in Figure 2-10 to evaluate P_{th}. For $C_s = 1.0$ pF, the power up threshold is found to be $P_{th} = 65$ μW ($P_A = 32.5$ μW). This is very close to the value of $P_A = 35$ μW that was achieved experimentally in Section 2.5.2. Note that the theoretical curves in Figure 2-10 assume that the Bode-Fano limit has been reached. Thus the P_{th} measured experimentally will always be somewhat higher than the theoretical value.

3.5 Floating Gate Structures

In this section, I describe how to adjust the threshold voltage of transistors using electron tunneling on and off capacitively coupled (floating) nodes, and describe rectifier structures which use this technique in order to improve their performance at low power levels.

3.5.1 Fowler-Nordheim Tunneling

Fowler-Nordheim (F-N) tunneling is a phenomenon where current flows through thin potential barriers, in this case layers of insulating material [38], usually silicon dioxide ('oxide'). It is a quantum mechanical phenomenon: electron wave functions decay

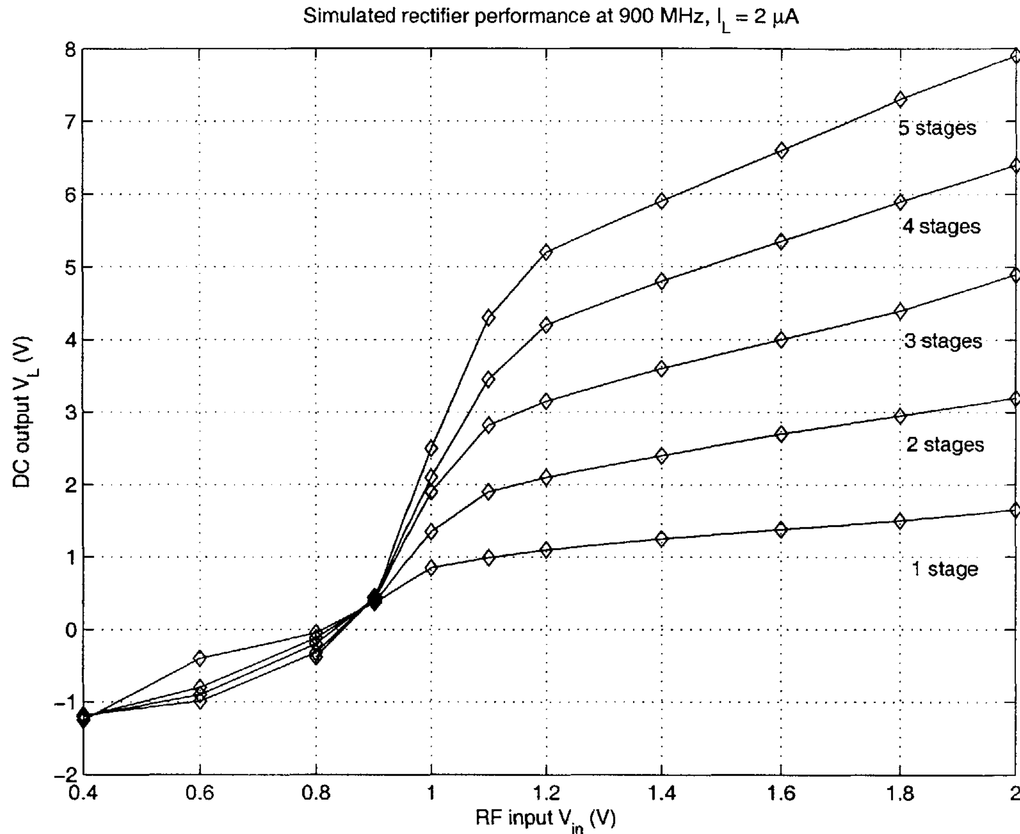


Figure 3-13: Simulated rectification curves (output DC voltage V_L versus RF input voltage V_{in}) for rectifiers with different values of N (the number of stages). $I_L = 2$ μA.

exponentially inside the insulator, but because the insulating layer is thin, retain a non-zero amplitude on the other side. Thus there is a finite non-zero probability for an electron to tunnel through the insulator. This constitutes the tunneling current. The tunneling current is exponentially sensitive to the thickness of the oxide layer (the mathematics is described in Appendix B. F-N tunneling has been extensively used in erasable memories, for trimming analog circuits [7] and adaptively controlling their characteristics.

The band diagram for the standard metal-oxide-semiconductor tunneling structure is shown in Figure 3-14 for different bias conditions. The potential barrier for electron flow between the oxide and metal conduction bands is denoted by ϕ_M and between the

oxide and semiconductor conduction bands by ϕ_S. The potential barrier for holes to flow between the semiconductor and oxide valence bands is denoted by ϕ_{Sh}. Similar band diagrams can be obtained for semiconductor-oxide-semiconductor structures like poly-poly capacitors. In that case one would have symmetry and ϕ_S would appear on both sides of the oxide. Consider Figure 3-14(b), which corresponds to negative bias on the metal electrode w.r.t. the semiconductor. Electron emission occurs by F-N tunneling through the triangular potential barrier of height ϕ_M from the metal into the oxide conduction band. This constitutes the tunneling current. Similarly, electron emission occurs from the semiconductor's conduction band through the barrier ϕ_S in Figure 3-14(c), which corresponds to positive bias on the metal w.r.t. the semiconductor. The tunneling current is now in the opposite direction. In theory, hole tunneling could also occur through ϕ_{Sh}. However, for silicon-silicon dioxide interfaces, $\phi_S < \phi_{Sh}$ ($\phi_S \approx 3.1$ eV, $\phi_{Sh} \approx 3.8$ ev), so hole tunneling is negligible compared to the electron tunneling. This explains why hot hole flow through oxides is not usually seen.

One's main interest in F-N tunneling is a mechanism for controlling the threshold voltage of MOS transistors. For this, one can use transistors with gate terminals which are floating, i.e., are only capacitively coupled (no DC path) to the outside world. These devices are commonly called floating gate transistors. Consider the canonical floating gate transistor structures shown in Figure 3-15. The voltage at the floating gate node V_{fg} controls the operation of the device. Using charge conservation, it is given by

$$V_{FG} = \frac{\sum_{i=1}^{K} C_i V_i + Q_{fg}}{\sum_{i=1}^{K} C_i} \tag{3.7}$$

where the transistors have K inputs each, V_i is the voltage and C_i is the coupling capacitor of the i-th input, and Q_{fg} is the charge on the floating gate when all the inputs are grounded. The floating gate transistor thus acts like a weighted summer (aggregator) of signals. To change the transistor threshold voltage permanently, one needs to change Q_{fg}. Consider the floating gate structures shown in Figure 3-16. The

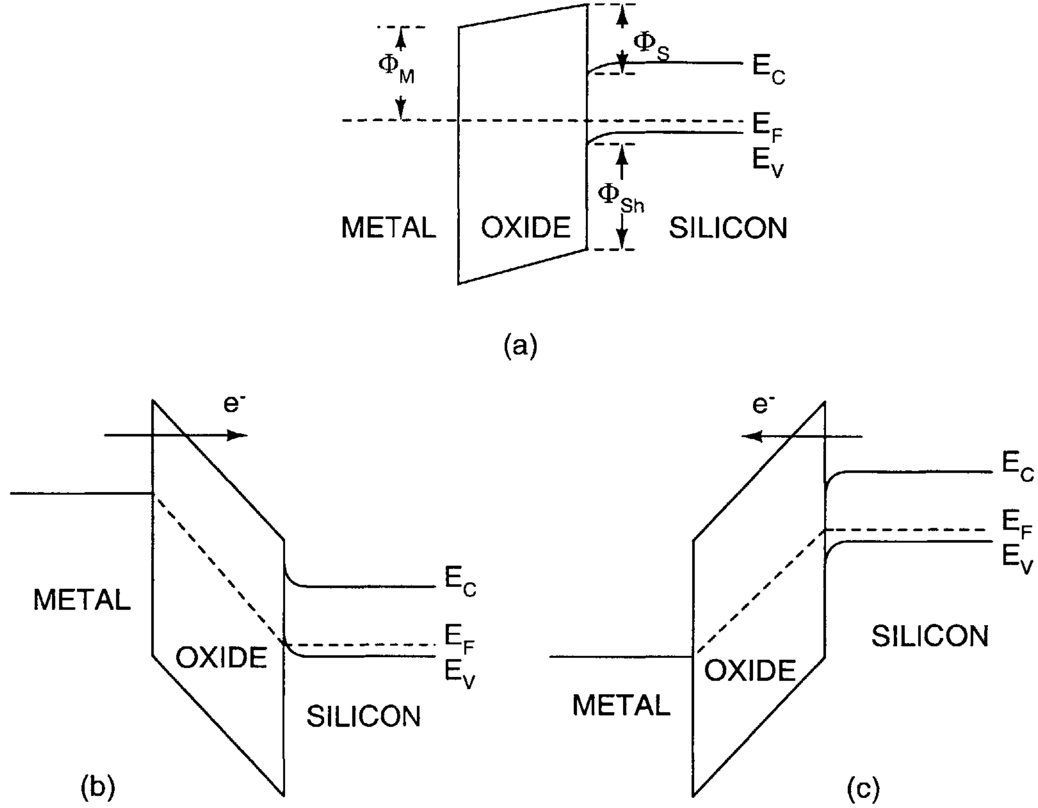


Figure 3-14: Energy band diagrams for F-N tunneling in metal-oxide-semiconductor structures for (a) zero bias, (b) negative bias and (c) positive bias on the metal electrode.

RF input signal V_{in} is fed through C_P (the 'pump' capacitor). C_{par} represents the parasitic capacitance to ground at the floating gate node, and C_{tun} is used to tunnel charge on or off the floating gate. To add charge and increase Q_{fg}, a large positive voltage is placed at V_{NMOS}. Similarly, a large negative voltage at V_{NMOS} is used to remove charge and decrease Q_{fg}. This is called the *programming* phase. During normal operation, V_{NMOS} and V_{PMOS} are both grounded. The effective threshold voltage of the transistor is given by

$$V_{T,eff} = V_T \left(1 + \frac{C_{par} + C_{tun}}{C_P}\right) - \frac{Q_{FG}}{C_P} \tag{3.8}$$

To decrease the threshold voltage, apart from changing Q_{fg} appropriately (increase it for NMOS, decrease it for PMOS) one must also keep the ratio $\frac{C_{par}+C_{tun}}{C_P}$ as small as

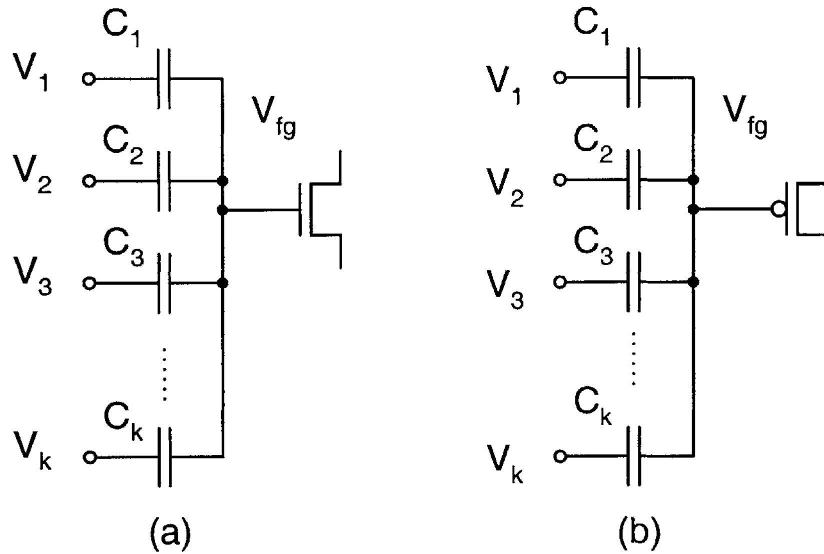


Figure 3-15: Canonical floating gate transistor structures. (a) NMOS and (b) PMOS.

possible. This means that one must have $C_P \gg (C_{par} + C_{tun})$. The tunneling junction can be implemented using any thin dielectric (usually oxide) layer between metal or semiconductor electrodes. The commonest method is to tunnel through gate oxide by using an accumulation mode MOS capacitor. However, this has the disadvantage of not being bidirectional [13]. To overcome this problem, poly1-poly2 oxide is used for tunneling (i.e., C_{tun} is a small poly1-poly2 capacitor). The poly1-poly2 oxide is usually about 2-3 times thicker than the gate oxide and is of lower quality (leading to larger tunneling voltages and more variability in tunneling currents than with gate oxide), but is truly bidirectional. This allows one to freely adjust $V_{TN,0}$ and $V_{TP,0}$.

The tunneling current is only weakly dependent on temperature and increases (exponentially) by a factor of about 10 from 0 to 400 K. This has been explained to be due to electron barrier height lowering at elevated temperatures [26]. Since the leakage from floating gates is very small and does not noticeably worsen at high temperatures, RFID tags using floating gate rectifiers can be safely used across the global supply chain, even with large variations in environmental temperature.

A final point to consider is that most of the tunneling occurs at the edges of the capacitor plates, especially the corners, since, for a given applied voltage, the electric

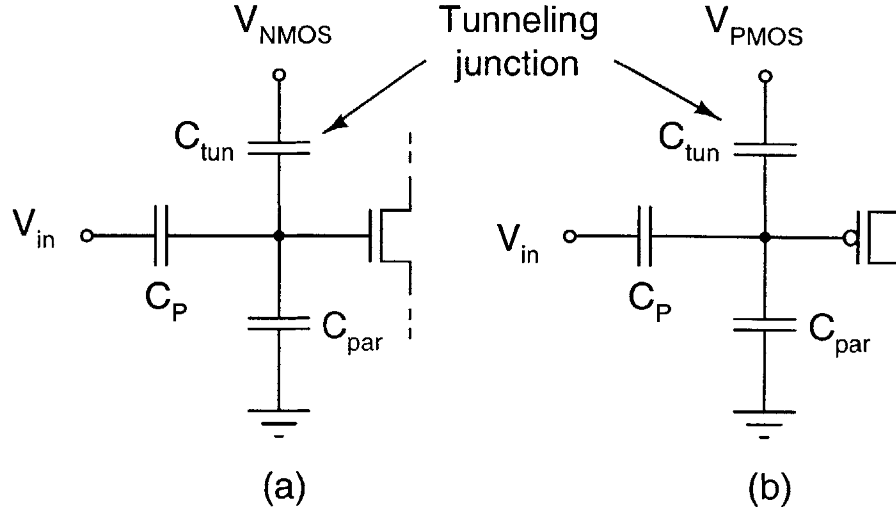


Figure 3-16: Floating gate transistor structures suitable for use in rectifiers to shift transistor threshold voltages. (a) NMOS and (b) PMOS.

field is highest there [28]. As a result, the tunneling current scales with the perimeter (and not the area) of the tunneling junction.

3.5.2 Rectifier Structures

Floating gate structures can be used to improve the performance of rectifiers by reducing transistor threshold voltages. The floating gate version of the four transistor switching cell of Figure 3-3 is shown in Figure 3-17. Each transistor in Figure 3-3 has simply been replaced by its floating gate equivalent from Figure 3-16. V_{NMOS} and V_{PMOS} are programming voltages which are grounded during normal operation.

Figure 3-19 shows simulated rectification curves for rectifiers using different numbers of floating gate switching cell (Figure 3-17) stages. Both NMOS and PMOS threshold voltages were reduced by 0.55 V (in my process, $V_{TN,0} \approx 0.64$ V and $V_{TP,0} \approx -0.91$ V). It was found previously that the optimal threshold voltage varies with V_{in}, I_L and the number of stages N in the rectifier. The threshold voltage used here represents a compromise value that seems to work well within the parameter ranges of interest ($1 < V_L < 2$ V, 1 μA $< I_L < 2$ μA, $1 < N < 5$). The RF frequency was 900 MHz, $I_L = 2$ μA and the pump capacitors were $C_P = 300$ fF. Comparing

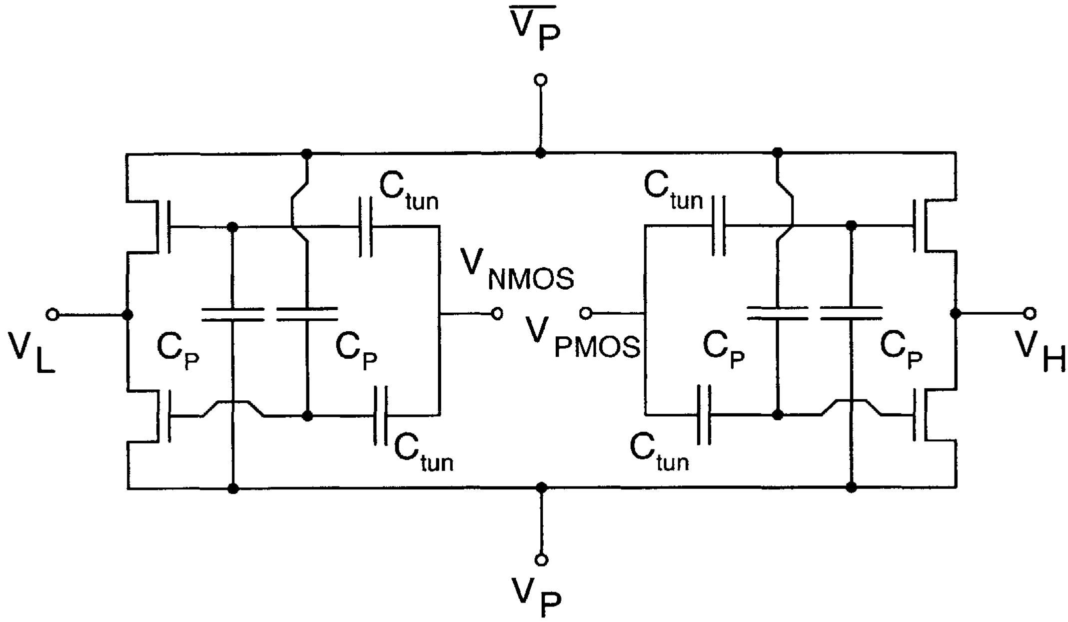


Figure 3-17: The four transistor cell using floating gate MOS transistors. Body (bulk) terminals are not shown

Figure 3-19 with Figure 3-13, which shows the performance of the same rectifiers with unmodified threshold voltages, one sees that the minimum V_{in} needed for a required value of V_L has been significantly reduced. For $V_L = 1.5$ V, for example, V_{in} required has been reduced to 0.7 V (from 0.95 V) for a five stage rectifier. At high values of V_{in}, V_L decreases below that obtained from normal transistors. This is again because reverse conduction, which increases as the threshold voltage decreases, is the dominant loss mechanism for large input voltages. However, this region is not really of interest: assuming that the required minimum V_L is obtained at some value of V_{in}, V_L is certain to go up further as V_{in} increases.

Figure 3-18 shows the measured output voltage of a five stage floating gate rectifier as a function of the NMOS and PMOS threshold voltages V_{TP} and V_{NP}. The RF frequency was 635 MHz and $I_L = 2$ μA. The input amplitude V_{in} is difficult to measure, but was probably in the range 1 - 1.5 V. Absolute threshold voltages cannot be found from Figure 3-18, because one does not know the actual DC voltage on the floating gate. One can, however, measure voltage *changes*. The axes on the plot

thus show relative threshold voltage changes with respect to an arbitrary origin. This origin was set by programming the rectifier to a suitable point. Data was then taken by varying V_{TP} and V_{TN} about the origin. One sees that the experimental data in Figure 3-18 looks broadly similar to the simulated plot of Figure 3-10.

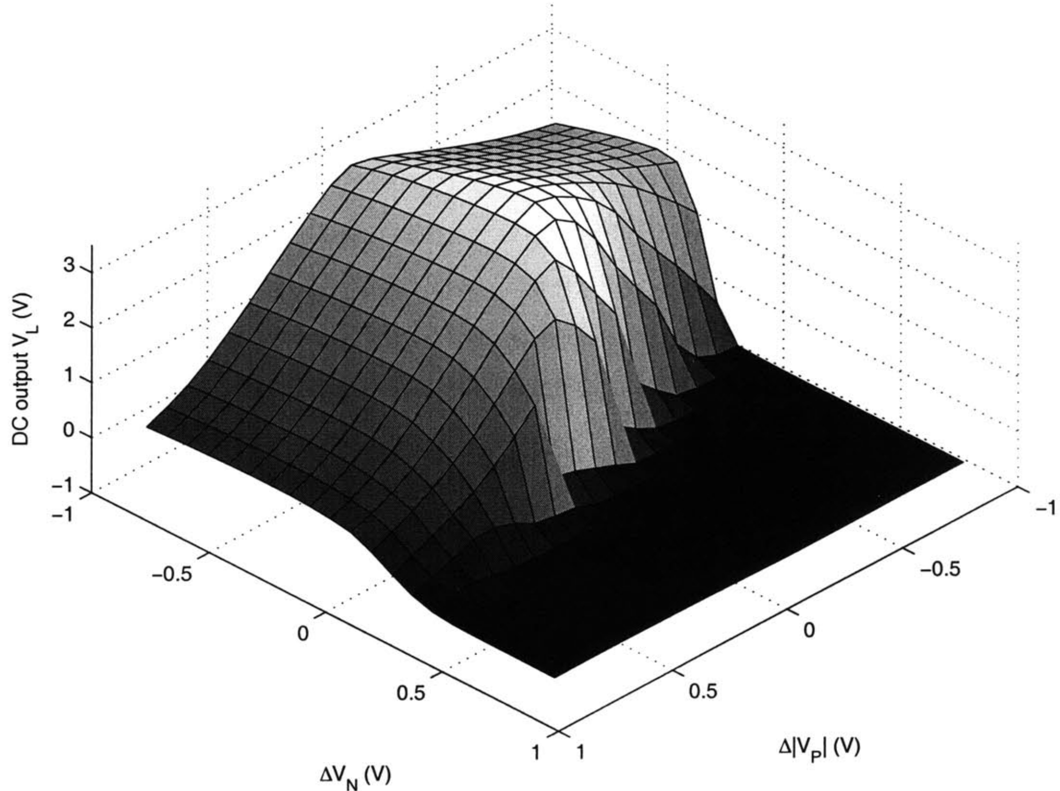


Figure 3-18: Experimental variation of output voltage V_L with change in threshold voltages ΔV_{TN} and $\Delta |V_{TP}|$ for a five stage floating gate rectifier.

The major problem with floating gate rectifiers is increased capacitive overhead. All the extra capacitors (four per switching cell) consume area and increase the input capacitance C_s through their bottom plate parasitics. With $C_P = 300$ fF and an LCC package, a five stage floating gate rectifier in this process can be designed to have $C_s \approx 0.9$ pF. From Figure 2-10, the power up threshold ($V_L = 1.5$ V and $I_L = 2$ μA) in this case should be about 32 μW ($P_A = 16$ μW). This is comparable to the

best published figure of $P_A = 16.7\ \mu$W [22], except that such low power levels have been obtained over a much wider bandwidth (30 MHz in our design versus 2 MHz in the published design). One may of course question whether this performance gain is worth all the added circuit complexity. I believe that the answer depends on the application. Where power up range is at an absolute premium, this seems to be an useful technique. In addition, further gains in performance (which may not be too robust) are expected if one continues to tweak threshold voltages about the optimum point.

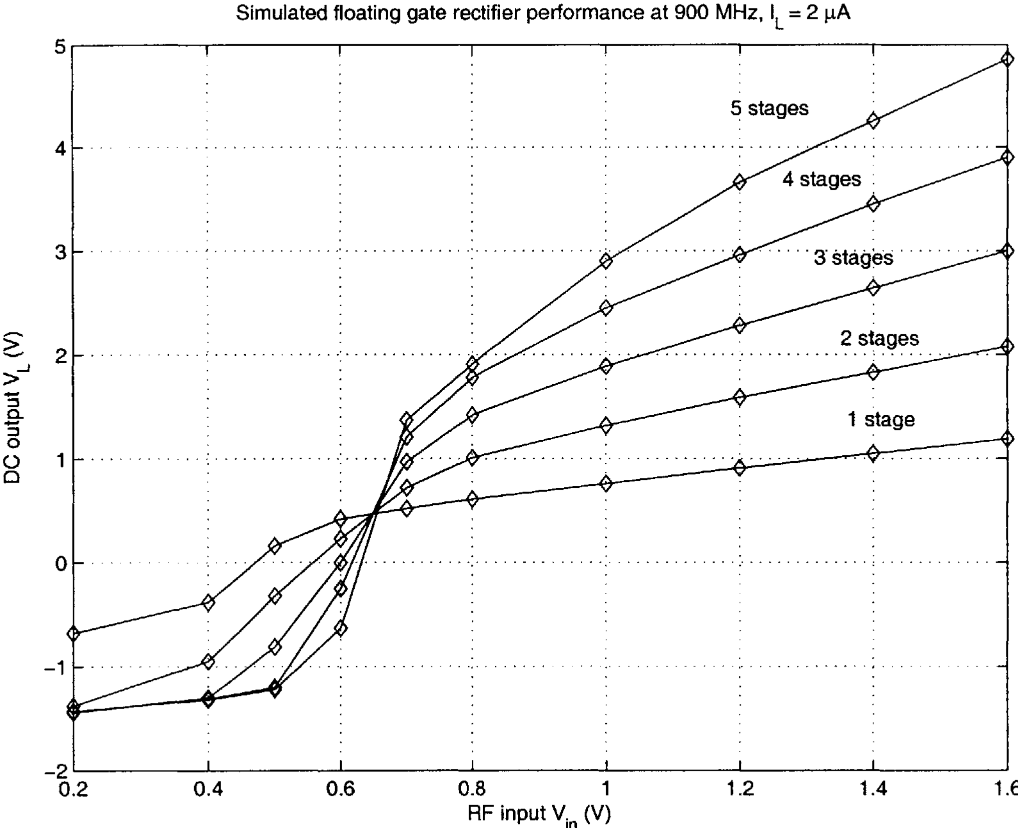


Figure 3-19: Simulated rectification curves (output DC voltage V_L versus RF input voltage V_{in}) for floating gate rectifiers with different values of N (the number of stages). $I_L = 2\ \mu$A, $C_P = 0.3$ pF.

In conclusion, there is one practical issue with using floating gate devices which should be mentioned. This is the initial floating gate voltage. The lack of a DC

path to ground means that these floating nodes pick up charge during various wafer processing steps, which translates into a random initial voltage. This causes the threshold voltage of all NMOS and PMOS transistors to be random, meaning that each chip has to be individually programmed before being used (this is why true bidirectional control of both threshold voltages was needed). My solution was to use a standard hill climbing algorithm to maximize the output voltage of the rectifier at specified input RF voltage and load current (V_{in} and I_L). I implemented a downhill simplex method in two dimensions (the NMOS and PMOS threshold voltages), and maximized the DC output voltage V_L (the objective function) for every chip being tested. This worked, but was rather time consuming and may prove prohibitively expensive to implement in volume production. To avoid this, one can use ultra-violet radiation to remove all charge from the floating node first. This removes the random initial Q_{fg}. After this, applying a fixed voltage at V_{NMOS} and V_{PMOS} for a fixed time results in known, controllable threshold voltages for all the chips.

3.6 Performance Optimization

Once the rectifier topology is fixed, one has to optimize it. The optimization process, like with most engineering systems, is s series of trade-offs. This section describes system optimization strategies. I take into account both layout and circuit design issues.

3.6.1 Layout Issues

To maximize the input quality factor of the rectifier, one wants to minimize the real (resistive) part of the input impedance seen by the antenna, and maximize the imaginary (capacitive or inductive) part. At most frequencies of interest, the input impedance will look capacitive, and one should therefore try to make this input capacitance as small as possible. This section describes layout techniques to achieve this.

Series-Parallel Conversions

In what follows, one shall often need to consider reactive elements with loss, i.e., finite Q factor. Either a series or parallel representation can be used for such an element (see Figure 3-20). It is easy to show that the two representations are interconvertible at a given frequency [25]. The conversion relies on the fact that both representations should give one the same value of Q (by an energy conservation argument). The result is

$$\begin{aligned} X_P &= X_S\left(\frac{Q^2+1}{Q^2}\right) \\ R_P &= R_S\left(Q^2+1\right) \end{aligned} \tag{3.9}$$

In the high Q case ($Q \gg 1$), (3.9) reduces to $X_P \approx X_S$ and $R_P \approx Q^2 R_S$.

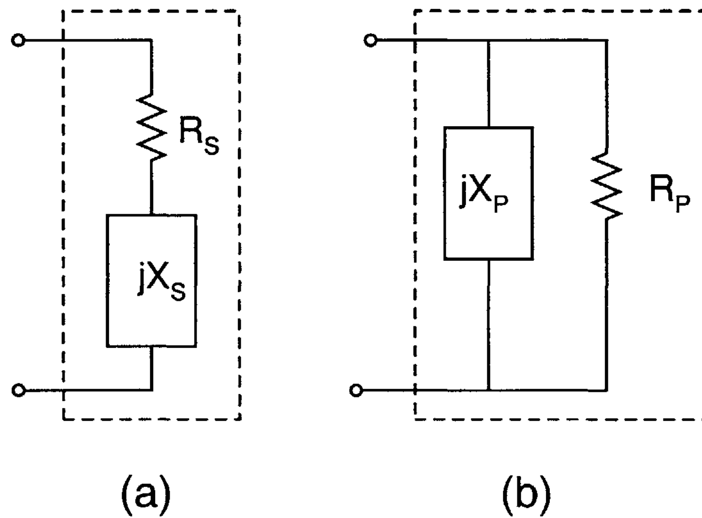


Figure 3-20: Alternative representations of a reactive element with finite Q. (a) Series and (b) Parallel.

A large part of the impedance seen by the antenna at the rectifier input is the result of circuit parasitics. Hence proper layout is essential in order to maximize performance. I shall next consider different sources of layout related parasitics, evaluate their effects and strive to minimize them.

Skin Effect

A potential source of interconnect loss is the skin effect. At high frequencies, the effective resistance of a conductor increases. This is because the current density J exponentially attenuates with depth x in the conductor. In fact, one has

$$J(x) = J_0 e^{-\frac{x}{\delta}} \tag{3.10}$$

where δ is known as the *skin depth* and J_0 is the current density at the surface of the conductor. The skin depth is given by the following expression

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \tag{3.11}$$

where ω is the frequency, and μ and σ are the permeability and conductivity of the conductor, respectively. Since the skin depth is proportional to $\frac{1}{\sqrt{\omega}}$, current flow becomes increasingly confined to the surface of the conductor as the frequency increases. In other words, the cross sectional area available for the RF current to flow effectively decreases as the frequency goes up. This is manifested as an increase in resistance.

To determine whether skin effect is of concern to us, I evaluate the skin depth in interconnect metalization (aluminum) at the nominal operating frequency of 900 MHz. One has $\omega = 2\pi \times 9 \times 10^8$ rad/sec and $\sigma = 3.5 \times 10^7$ Ω−m. This gives $\delta = 2.8$ μm. A typical metal layer in out process has a thickness of ~ 1 μm, which is about a third of the skin depth. Thus the current flow in the conductor will be almost uniform and the DC resistance value can be used for all resistance calculations.

Transistor Gate Resistance

Since transistor gates are made of polysilicon, their resistance can prove to be significant if the process is non salicided. Salicided processes use metal silicides to reduce the resistance of the gate material (typically from the 25 $\Omega/\square$ typical of polysilicon to 1 - 2 $\Omega/\square$).

It is assumed that the sheet resistance of the gate material is ρ and the transistor

has a aspect ratio $\frac{W}{L}$. The minimum gate extension from the active region before a metal contact can be placed (see Figure 3-21) is Δ and the device has N gate fingers. Contacts are placed at one end of the gate and connected together using an equipotential (such as a wide metal line). The resistance of each finger, R_{finger}, is, by square counting

$$R_{finger} = \rho \frac{\frac{W}{N} + 2\Delta}{L} \tag{3.12}$$

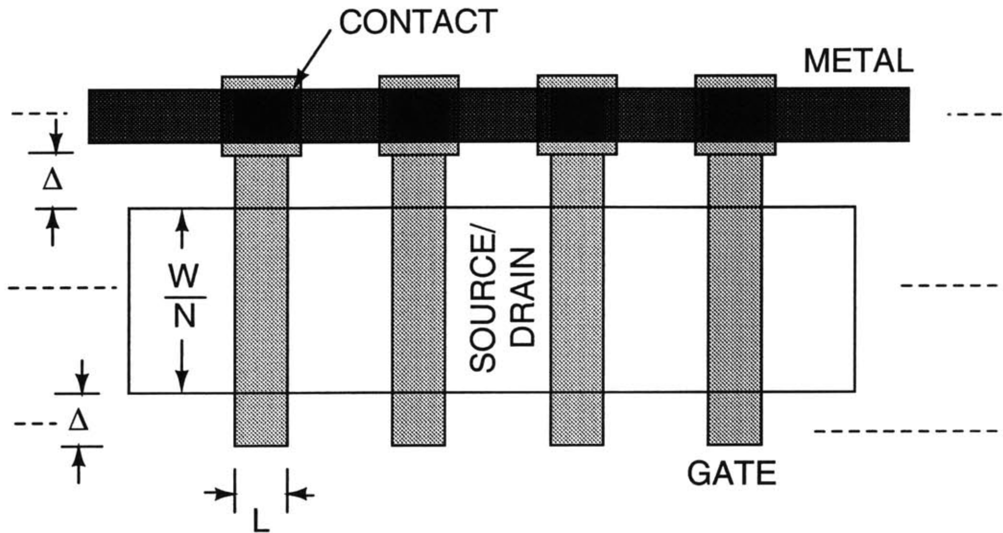


Figure 3-21: Transistor geometry with multiple gate fingers. Only four fingers are shown; in general there can be N fingers.

The net gate resistance R_G is just the resistance of N fingers in parallel, so

$$R_G = \rho \left(\frac{W}{N^2 L} + \frac{2\Delta}{NL} \right) \tag{3.13}$$

One sees that the only way to reduce R_G for a fixed device size is to increase the number of fingers. The resistance goes down approximately as $\frac{1}{N^2}$. If possible, one should also put contacts at both ends of the gate. If the two contacts are shorted together, R_G is reduced by an additional factor of 4 over (3.13). Since my primary process is non salicided, it is critical that both these tricks are used to reduce R_G as much as possible.

Interconnect Resistance

Long lengths of metal interconnect can degrade performance significantly when attached to the input nodes. The worst offenders in this respect are the main RF feed lines, which feed the RF signal into all the switching cell stages in parallel. Generically, the interconnect behaves like a lossy transmission line with frequency dependent R, L and C [34]. For simplicity, I limit myself here to the narrowband case, where one can consider L, R and C to be constant. At UHF, L is usually negligible and the line behaves as an RC transmission line.

The simplest possible model uses a single lumped LRC section (see Figure 3-22). A typical value of the self inductance associated with on-chip interconnect is $L_0 \approx 0.5$ nH/mm. At frequencies of interest, as has been shown, the skin effect is negligible, so the sheet resistance is just the DC sheet resistance R_{dc}. The capacitance C can be calculated from the area and fringe capacitances (to the substrate) of the layer in question. Thus this model for a line of width w and length l has

$$\begin{aligned} L &= L_0 l \\ R &= R_{dc}\frac{l}{w} \\ C &= C_A wl + 2C_F(w+l) \end{aligned} \tag{3.14}$$

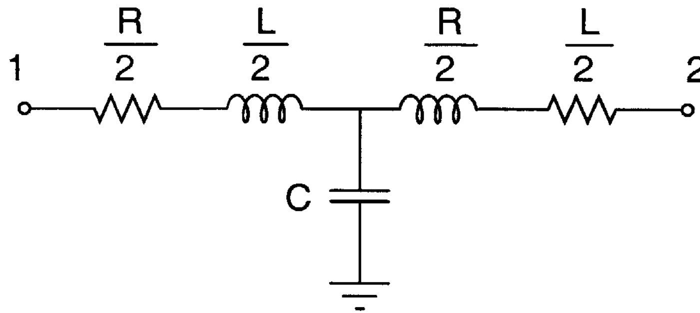


Figure 3-22: Single L-R-C section used to model interconnect between terminals 1 and 2.

where C_A and C_F are the area and fringe capacitance constants of the interconnect layer (see Section 3.6.1).

For more accuracy, I now model the feed line as a RC transmission line of length

l (see Figure 3-23) (the effects of inductance are still ignored). One wants to find its input impedance, assuming it is open circuited at the end and that there are no other loads on it. This last assumption is inaccurate since in reality current is being drawn from the line at intervals by the rectifier stages. However, since this reduces the resistive part of the input impedance, this estimate provides a worst case bound on the resistance seen by the RF input due to the feed line. This is useful for design.

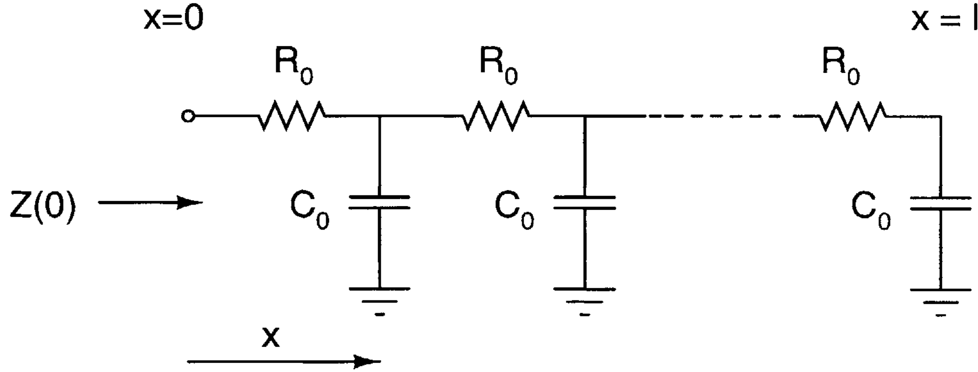


Figure 3-23: RC transmission line model of an interconnect of length l open circuited at one end.

I start by defining the per unit length resistance and capacitance R_0 and C_0 as $R_0 = \frac{\rho}{w}$ Ω/m and $C_0 = (C_A w + 2C_F)$ F/m. By writing down KCL and KVL equations, one gets

$$I(x) - I(x + dx) = sC_0 V(x)dx \Rightarrow \frac{dI}{dx} = -sC_0 V(x) \tag{3.15}$$

where s is the complex frequency variable, and

$$V(x) - V(x + dx) = R_0 I(x)dx \Rightarrow \frac{dV}{dx} = -R_0 I(x) \tag{3.16}$$

By noting that the impedance $Z(x) = \frac{V(x)}{I(x)}$, one gets

$$\frac{dZ}{dx} = \frac{d}{dx}\left(\frac{V}{I}\right) = \frac{I\frac{dV}{dx} - V\frac{dI}{dx}}{I^2} \tag{3.17}$$

After substituting for $\frac{dV}{dx}$ and $\frac{dI}{dx}$ from (3.15) and (3.16), one gets a first order ODE

for $Z(x)$

$$\frac{dZ}{dx} = -R_0 + sC_0Z^2 \tag{3.18}$$

This may be easily integrated from 0 to l to give the input impedance $Z(0)$. Remembering that $Z(l) \to \infty$, I get

$$Z(0) = \sqrt{\frac{R_0}{sC_0}} \coth\left(l\sqrt{sR_0C_0}\right) \tag{3.19}$$

As a check on (3.19), one may evaluate the impedance when $R_0 = 0$. Since $\lim_{x\to 0} x\coth(x) = 1$, Z_0 becomes

$$Z(0) = \frac{1}{sC_0l} \tag{3.20}$$

This is just the capacitive reactance of the line, as expected. One can also simplify (3.19) at low frequencies when $l\sqrt{sR_0C_0} \ll 1$. Using the series expansion for $\coth(z)$

$$\coth(z) = \frac{1}{z} + \frac{z}{3} - \frac{z^3}{45} + \frac{2z^5}{945} + ... \tag{3.21}$$

for $z \ll 1$, one gets

$$\coth\left(l\sqrt{sR_0C_0}\right) \approx \frac{1}{l\sqrt{sR_0C_0}} + \frac{l\sqrt{sR_0C_0}}{1} \tag{3.22}$$

Substituting (3.22) in (3.19) gives

$$Z(0) \approx \frac{1}{sC_0l} + \frac{R_0l}{3} \tag{3.23}$$

If one substitutes $s = \jmath\omega$, the first term in (3.22) (which is the same as (3.20)) is seen to represent the capacitive component of $Z(0)$. The second term is the resistive component. The approximate expression of (3.23) is found to be quite accurate for typical on-chip interconnects at frequencies up to 1 GHz or more . This is because the R_0C_0 product of metal interconnects is quite low, so the approximation (3.22) is usually a good one. One can therefore define $\mathrm{Re}\{Z(0)\} \approx \frac{R_0l}{3}$ to be the effective input

resistance of an interconnect line for a wide range of frequencies and wire geometries with little error. One sees that $\mathrm{Re}\{Z(0)\}$ is one third of the DC resistance $R_0 l$ of the wire. This value shall be used for subsequent calculations.

Well Resistance

Since the n-well in a standard CMOS process is lightly doped (typically, $N_D \approx 10^{16}$ cm^{-3}), it has substantial sheet resistance. In my 0.5 μm process, for example, well sheet resistance is $R_{sh,W} = 820\ \Omega/\Box$. However, the effective resistance can be reduced by using many contact fingers, which divide the well into n unit rectangles of length L_U and width W_U (see Figure 3-24). The contact fingers are made of metal and have negligible resistance, forming a equipotential surface. Current flow is then perpendicular to this surface. It is easy to show that the series resistance R_W produced by this geometry is then approximately

$$R_W \approx \frac{1}{12n} R_{sh,W} \frac{W_U}{L_U} \tag{3.24}$$

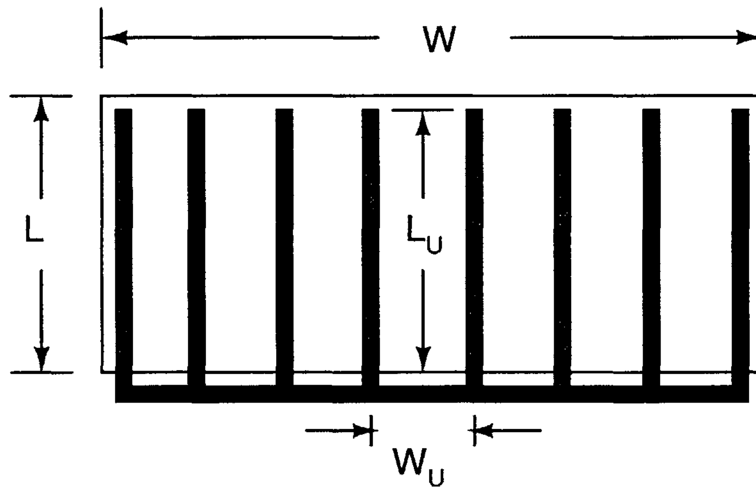


Figure 3-24: Contact finger geometry for reducing well resistance. The well has dimensions $W \times L$ and is divided by the fingers into n rectangles.

The factor of 12 in (3.24) appears because each unit rectangle has metal contacts (near equipotentials) along two sides [24]. In addition, one notes that in terms of the

total well length and width W and L, $L_U = L$ and $W_U \approx \frac{W}{n}$, thus giving

$$R_W \approx \frac{1}{12n^2} R_{sh,W} \frac{W}{L} \tag{3.25}$$

Since the number of fingers is $n+1$, one sees that the resistance goes down quadratically with the number of fingers. For practical layouts, the number of fingers cannot be increased indefinitely because the area overhead causes the substrate capacitance to increase.

Let the case when the total area of the well, $A = WL$, is fixed be considered next. In this case, one has

$$R_W \approx \frac{1}{12n^2} R_{sh,W} \frac{W^2}{A} \tag{3.26}$$

According to (3.26), one should minimize W, i.e., make the well rectangle as long and thin as possible. Practically, however, the contact fingers must be separated by a minimum distance Δ to allow transistors, Schottky diode contacts, etc. to be placed between them (i.e., the minimum value of L_U is Δ). In that case, the maximum value of n (ignoring the area occupied by the fingers themselves) is $n_{max} = \frac{W}{\Delta}$, thus giving the following minimum possible value of R_W using fingering

$$R_{W,min} \approx \frac{1}{12} R_{sh} \frac{\Delta^2}{A} \tag{3.27}$$

Finally, one notes that the same fingering technique can be used to reduce the resistance of any layout layer, not just the well. For example, the Q of poly-poly capacitors can be increased by reducing the plate resistance [24].

Capacitor Non-idealities

A generic circuit model of an integrated capacitor can be constructed as shown in Figure 3-25. Here C_0 is the ideal capacitance that one wants between terminals 1 and 2. L_p is the parasitic series inductance and R_p is the series resistance. Common sources of R_p are the finite sheet resistance of the capacitor plate material (usually

metal or polysilicon), contacts and vias. These elements are common to any physical capacitor, integrated or discrete. The remaining elements are specific to integrated capacitors, and comprise substrate resistances (R_{S1} and R_{S2}) and capacitances (C_{S1} and C_{S2}). These elements arise because of the proximity of the chip substrate, which acts like a ground plane.

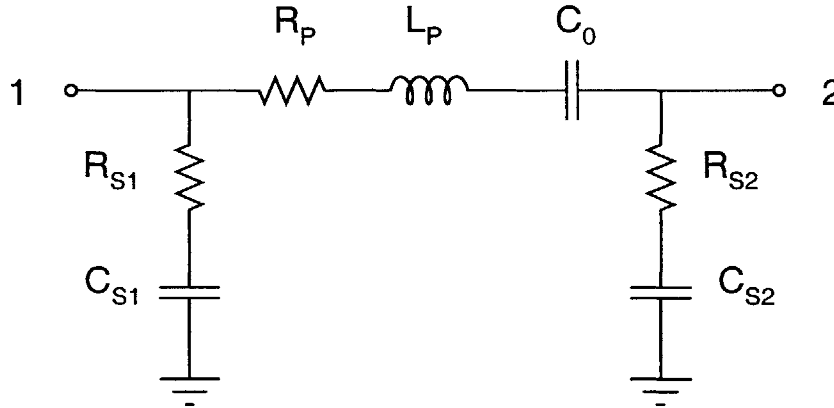


Figure 3-25: Generic model of an integrated circuit capacitor.

In general, the integrated capacitor is an asymmetric device, i.e., $R_{S1} \neq R_{S2}$ and $C_{S1} \neq C_{S2}$. The most common capacitor type, the parallel plate structure, is strongly asymmetric since the top plate is shielded from the substrate by the bottom plate. Thus, if 1 is taken to be the top plate terminal, $R_{S1} \approx 0$ and $C_{S1} \approx 0$. Capacitor structures which utilize lateral flux are more symmetric. To minimize C_{S1} and C_{S2}, one wants to minimize capacitor size, i.e., maximize the capacitance density (capacitance per unit area).

In their comprehensive review of integrated capacitors, Aparicio and Hajimiri [1] study various lateral and vertical flux structures. Their essential conclusion is that major improvements in capacitance density over simple parallel plate capacitors are possible as lateral dimensions shrink with every process generation. A semi-empirical upper bound on capacitance density as a function of minimum lateral dimension is also derived. Unfortunately, my primary process (0.5 μm CMOS) has a minimum lateral dimension (defined as the minimum width of a metal line, or the minimum allowable spacing between two metal lines) of 0.9 μm, which is too large for any improvement

over parallel plate structures to be possible. This is confirmed by my own calculations, and is doubly unfortunate since for comparable capacitance densities, C_{S1} and C_{S2} are typically lower for lateral flux capacitors.

I therefore limit my attention to parallel plate capacitors in this thesis. For parallel plate capacitors, the model of Figure 3-25 can thus be simplified to that of Figure 3-27, where I have relabeled C_{S2} as ϵC_0 to emphasize that C_{S2}, also being a parallel plate capacitance, is proportional to C_0 with a process dependent *bottom plate coefficient* ϵ, and R_{S2} has been relabeled R_S for simplicity. The bottom plate coefficient is approximately given by

$$\epsilon = \frac{x_{top}}{x_{sub}} \tag{3.28}$$

where x_{top} and x_{sub} are the vertical distances of the bottom plate from the top plate and the substrate, respectively (see Figure 3-26). Thus, ignoring fringing fields,

$$\begin{aligned} C_0 &= \frac{\epsilon_0 \epsilon_r A}{x_{top}} \\ \epsilon C_0 &= \frac{\epsilon_0 \epsilon_r A}{x_{top}} \end{aligned} \tag{3.29}$$

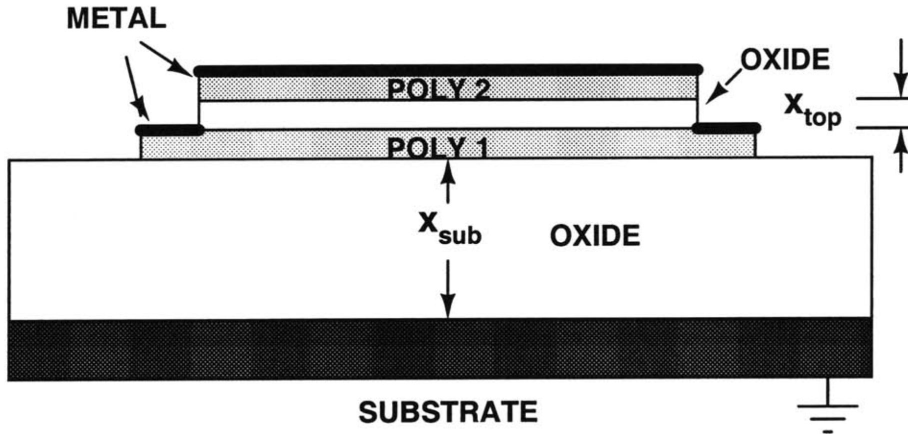


Figure 3-26: Cross sectional geometry of integrated parallel plate capacitors (a poly1-poly2 capacitor is shown). Drawing is not to scale.

where A is the area of the plates, ϵ_0 is the permittivity of free space and ϵ_r is the dielectric constant of the inter-plate dielectric. Also, at frequencies and capacitor

sizes of interest, L_P is usually negligible and has been removed (i.e., one is well below the self resonant frequency of the capacitor). This leaves only two finite Q capacitors in the model: the designed capacitance between terminals 1 and 2, and the unwanted bottom plate parasitic from terminal 2 (the bottom plate) to chip substrate (ground). I shall make extensive use of this simplified model.

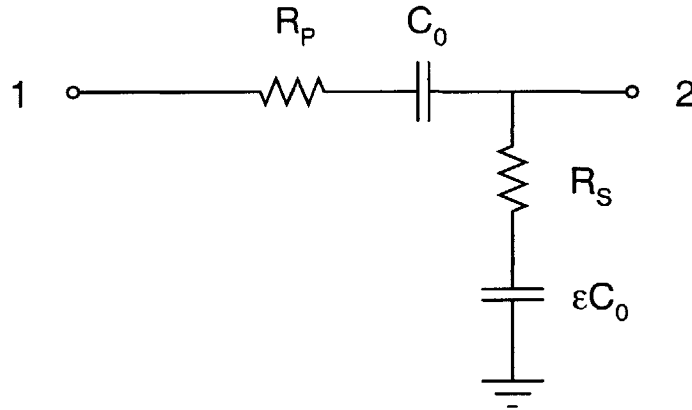


Figure 3-27: Simplified model of an integrated circuit parallel plate capacitor. Terminal 2 is the bottom plate.

Referring to Figure 3-27, one wants as small a value of ϵ as possible. In my process, the highest capacitance density (i.e., smallest value of x_{top}) and smallest value of ϵ is obtained with polysilicon (Poly1 -Poly2) capacitors. A typical value for the capacitance density is 888 aF/μm^2, with $\epsilon \approx 0.15$. Unfortunately the high sheet resistance of polysilicon as compared to metal results in high values of R_P and lowers the capacitor Q. Typical figures are 26 $\Omega/\Box$ for polysilicon and 0.09 $\Omega/\Box$ for metal layers.

Strategies for lowering R_P and improve the Q of parallel plate capacitors rely on using low resistance layers (typically interconnect metallization) as equipotentials. Such layers are used to break up the parallel plates into smaller sections, thereby reducing current flow lengths in the higher resistance plate material (and thus R_P). Here again, the bottom plate is more of a concern. This is because while the top plate can be completely covered with metal, the bottom plate cannot be. Another point to note is that any resistance lowering strategy increases the area of the bottom

plate (due to wiring and contact overheads), effectively increasing ϵ. For a given capacitance value, there exists an optimum bottom plate equipotential density.

While fingering strategies have been tried [24], they entail substantial area overheads. A simple and effective alternative is to simply place an equipotential (a metal line) along the outer edge of the bottom plate. I now analyze this mathematically. The model geometry is shown in Figure 3-28. It consists of a circular parallel plate capacitor of radius R having a lossless top plate and a bottom plate with sheet resistance ρ. The plates are separated by an ideal dielectric with dielectric constant ϵ_r and thickness d. Fringing fields will be neglected. The edge of the bottom plate is an equipotential V_0, and the capacitance density is

$$C_a = \frac{\epsilon_0 \epsilon_r}{d} \tag{3.30}$$

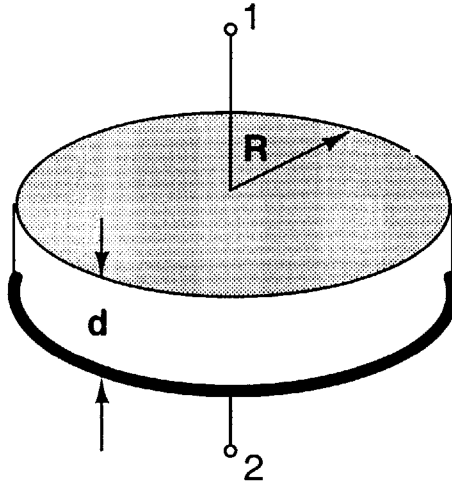


Figure 3-28: View of parallel plate capacitor geometry. The top plate is assumed to be lossless, and the bottom plate has a sheet resistance ρ.

At frequencies much lower than the self-resonant frequency, the structure can be modeled as a distributed RC network. Because of the radial symmetry, I use polar coordinates and start by considering a differential sector subtending an angle $d\theta$ at the center (see Figure 3-29). One then has

$$r(r) = \frac{\rho dr}{r d\theta} \tag{3.31}$$

and

$$r(r) = C_a r d\theta dr \tag{3.32}$$

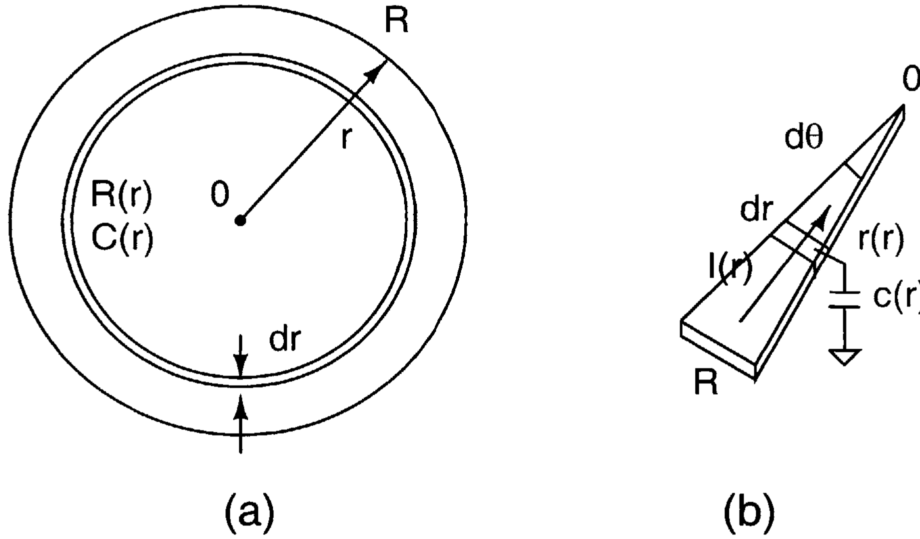


Figure 3-29: Illustrating the geometry of the circular parallel plate capacitor problem. (a) Polar coordinates and (b) a differential sector element.

where $r(r)$ and $c(r)$ are the resistance and capacitance, respectively, of a a differential unit of length dr and width $rd\theta$, and current flow is radial (perpendicular to the equipotentials). I now use radial symmetry and integrate over θ from 0 to 2π to get $R(r)$ and $C(r)$, the resistance and capacitance of an annulus of radius r and thickness dr. Since resistors add up in parallel, one gets

$$R(r) = \frac{1}{\int_0^{2\pi} \frac{r d\theta}{\rho dr}} = \frac{\rho dr}{2\pi r} \tag{3.33}$$

Similarly, capacitors also add up in parallel, so

$$C(r) = \int_0^{2\pi} 2\pi C_a r dr d\theta = 2\pi C_a r dr \tag{3.34}$$

By using $R(r)$ and $C(r)$, I have reduced the problem to a one dimensional tapered RC transmission line, as shown in Figure 3-30. One's goal here is to find the input impedance Z_{in} of this structure (as seen from the equipotential edge at $r = R$) and compare it with that of an ideal capacitor. Writing down Kirchoff's Current Law (KCL) at any node

$$I(r + dr) - I(r) = sC(r)V(r) \Rightarrow \frac{dI}{dr} = s2\pi r C_a V(r) \tag{3.35}$$

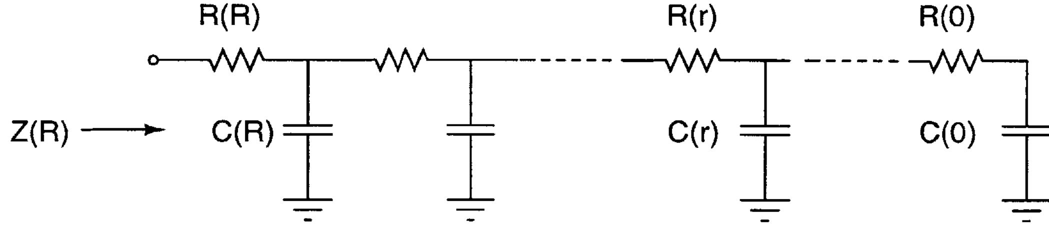


Figure 3-30: Tapered 1D RC transmission line representation of the parallel plate capacitor problem. $R(0) \to \infty$ and $C(0) \to 0$.

In (3.35), s is the complex frequency variable, and the domain of interest is $0 < r < R$. Similarly, by writing down Kirchoff's Voltage Law (KVL), one gets

$$V(r + dr) - V(r) = R(r)I(r) \Rightarrow \frac{dV}{dr} = \frac{\rho}{2\pi r} I(r) \tag{3.36}$$

By differentiating (3.35) and (3.36) w.r.t. r and substituting, one can decouple the current and voltage equations to get

$$\begin{aligned} \frac{d^2I}{dr^2} - \frac{1}{r}\frac{dI}{dr} - sC_a\rho I &= 0 \\ \frac{d^2V}{dr^2} + \frac{1}{r}\frac{dV}{dr} - sC_a\rho V &= 0 \end{aligned} \tag{3.37}$$

However, what one is really interested in is the impedance $Z(r) = \frac{V(r)}{I(r)}$ (or equivalently, the admittance $Y(r) = \frac{I(r)}{V(r)}$) of the structure. This can be done as follows

$$\frac{dZ}{dr} = \frac{d}{dr}\left(\frac{V}{I}\right) = \frac{I\frac{dV}{dr} - V\frac{dI}{dr}}{I^2} \tag{3.38}$$

After substituting for known quantities in (3.38), one gets a nonlinear, first order

ODE for the impedance $Z(r)$ of the structure, as follows

$$\frac{dZ}{dr} + 2\pi r s C_a Z^2 - \frac{\rho}{2\pi r} = 0 \tag{3.39}$$

As a sanity check, one can see if one gets the right answer from (3.39) if one lets $\rho \rightarrow 0$, i.e., by considering the ideal parallel plate capacitor with lossless top and bottom plates. In that case, (3.39) becomes

$$\frac{dZ}{dr} + 2\pi r s C_a Z^2 = 0 \tag{3.40}$$

This can be easily integrated from 0 to r to give

$$Z(r) = \frac{1}{\pi r^2 s C_a} \tag{3.41}$$

Remembering that for a capacitor C, $Z = \frac{1}{sC}$, this is the correct answer for a parallel plate capacitor with plate area πr^2. Thus emboldened, one finds the corresponding equation for $Y(r)$, which is

$$\frac{dY}{dr} + \frac{\rho}{2\pi r} Y^2 - 2\pi r s C_a = 0 \tag{3.42}$$

Since one knows that $Y(0) = 0$, this equation can be solved to find the admittance of the capacitor structure. Unfortunately, this equation cannot be solved analytically. Therefore numerical techniques were used to find the solution. Once $Y(r)$, is found, the quality factor is simply given by

$$Q = \frac{\text{Im}\{Y(r)\}}{\text{Re}\{Y(r)\}} \tag{3.43}$$

Figure 3-31 shows the capacitor Q for various values of ρ. The actual value of ρ for polysilicon in my process is 26.2 $\Omega/\square$. One sees that, for this value of ρ, the Q's are very good and remain above 50 even for caps as big as 3 pF. The reason for this is evident from Figure 3-32, which shows the resistance of the structure for different values of ρ. One sees that for a wide range of values of ρ, the resistance settles rapidly to a value of $\approx \frac{\rho}{25}$. This low value explains the high Q factors in Figure 3-31. The

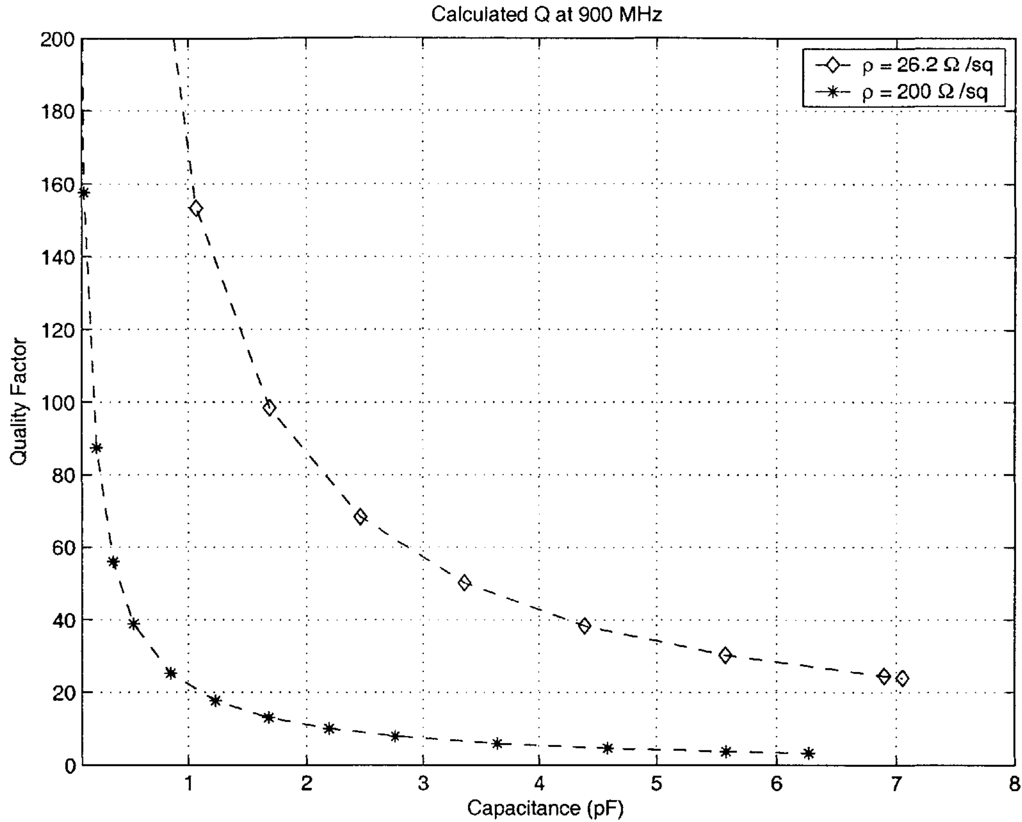


Figure 3-31: Calculated Q of capacitors at 900 MHz for different values of ρ. C_a=900 aF/μm^2, corresponding to poly1-poly2 capacitors in AMI's 0.5 μm process.

resistance value is constant for large spreads in ω and C_a. Because the resistance values are so low, the real Q of such capacitors will likely be set by other parasitic resistances, such as that of interconnect, vias and contacts, and not by the resistance of the bottom plate. Nevertheless, one sees that properly designed poly capacitors of values up to at least 1-2 pF should provide excellent Q's at 900 MHz operating frequency.

Finally, another reason to make integrated capacitors as close to circular as possible is to minimize the bottom plate capacitance. Using circular plates minimizes the perimeter for a given area, which helps in two ways. Firstly, the parasitic fringe capacitance of the bottom plate to the substrate is proportional to the perimeter of the bottom plate and is minimized. In addition, the bottom plate always has to

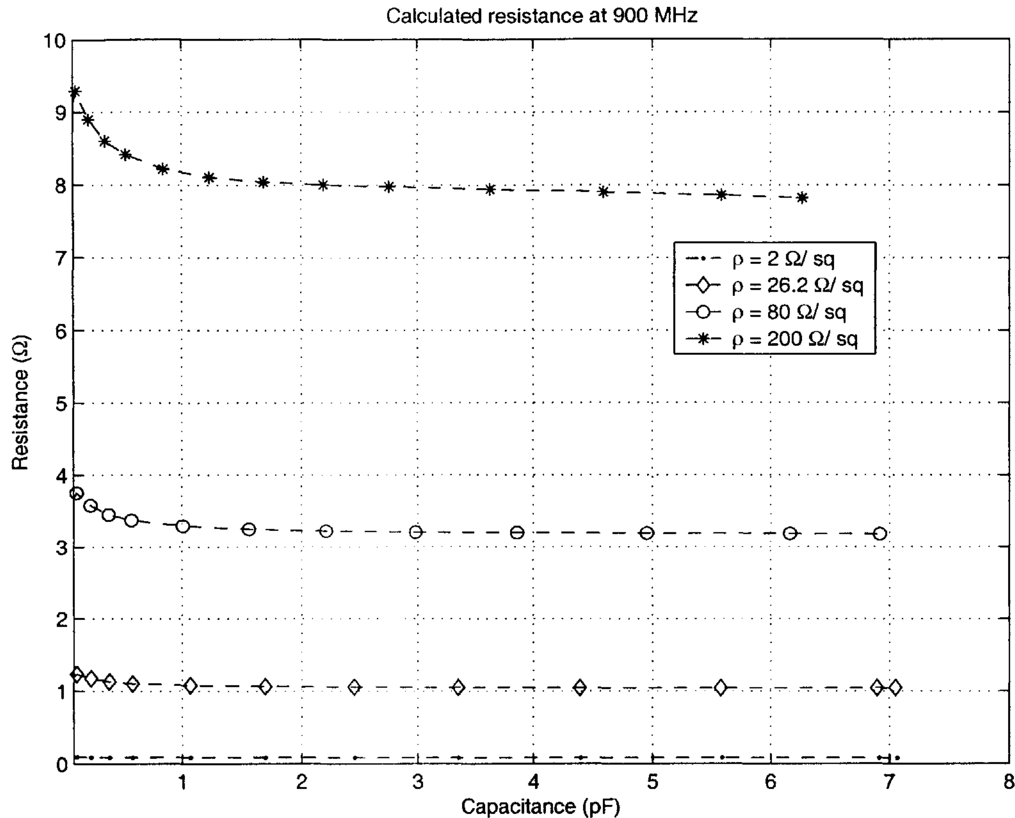


Figure 3-32: Calculated resistance of capacitors at 900 MHz for different values of ρ. C_a=900 aF/μm^2, corresponding to poly1-poly2 capacitors in AMI's 0.5 μm process.

be bigger than the top plate to allow for contacts to be placed at its edges. This overhead area is proportional to the perimeter of the plates. The area of the bottom plate (and hence its parasitic capacitance) is thus minimized if the perimeter is made as small as possible. Both these effects mean that circular capacitor geometries end up minimizing the value of ϵ (the bottom plate capacitance coefficient). In integrated circuit geometries (which typically only allow 90° and 45° polygons), one ends up using regular octagons instead of circles.

Substrate Capacitances

Any overlap or proximity between integrated circuit layers at different potentials generates a capacitance. However, these layers are largely planar and the largest overlaps

are usually with the chip substrate, which is grounded. Hence as long as inductances are negligible (mostly true at UHF), circuit layers which are not at AC ground essentially act like the top plates of parallel plate capacitors, with the bottom plate being the substrate, thus producing parasitic capacitances to ground. The dielectric in between is usually silicon dioxide. The resultant electric field distributions, however, are usually far from simple. The small dimensions (and consequent large perimeter to area ratios) mean that fringing fields are significant and can contribute a major part of the total capacitance.

To avoid having to deal with complex field solutions, approximate expressions are often used to estimate parasitic capacitances. The most pervasive (and certainly the easiest to use) approximation breaks up the capacitance C_{par} to ground into 'area' and 'fringe' contributions, as follows

$$\begin{aligned} C_{par} &= C_{area} + C_{fringe} \\ &= AC_A + PC_F \end{aligned} \tag{3.44}$$

where A is the area of the layer, P its perimeter and C_A and C_F are the area and fringe capacitance constants. Practically, the area and fringe contributions are often of similar magnitude.

Transistor and Diode Capacitances

The parasitic capacitances of active circuit elements - diodes and transistors - are among the major contributors to the total input parasitic capacitance of the rectifier. In this thesis I use the standard capacitance models of MOSFETs and Schottky diodes.

A high frequency model of a diode is shown in Figure 3-33. There are two contributors to C_D: the junction capacitance C_j and the diffusion capacitance C_d. C_j arises due to the charge dipole of the depletion region near the metallurgical junction, while C_d is due to minority carrier charge storage. Schottky diodes are majority carrier devices. As a result, $C_d = 0$, i.e., $C_D = C_j$ for Schottky diodes.

The junction capacitance C_j is nonlinear (voltage dependent). For small signals,

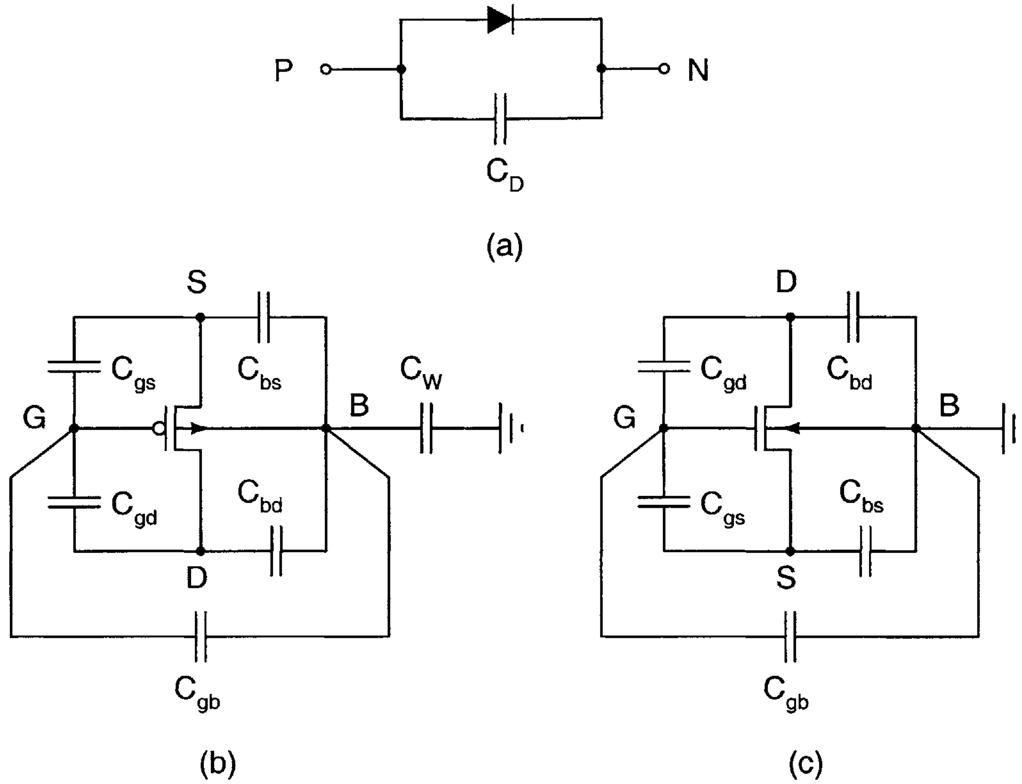


Figure 3-33: AC (charge storage) models for (a) diodes, (b) PMOS transistors and (c) NMOS transistors, showing the various parasitic capacitors. PMOS transistors are assumed to be fabricated inside n-wells.

it may be defined as

$$C_j(V) = \left|\frac{dQ_j}{dV}\right| \tag{3.45}$$

where Q_j is the charge on each side of the junction, and V is the applied voltage across the junction. One generally uses the zero bias value of C_j for design, i.e., assume $C_D \approx C_j(0)$. This approximation is quite crude for RF amplitudes greater than ≈ 100 mV, but proves adequate for hand calculations. For integrated Schottky diodes in my process, one finds

$$C_D = NC_{cnt} \tag{3.46}$$

In (3.46), N is the number of metal-semiconductor contacts constituting the

diode, and C_{cnt} represents the zero bias junction capacitance of a single contact (see AppendixA.1.4). Calculated values of C_{cnt} in my 0.5 μm CMOS process are 0.207 fF/μm^2 for metal / n-well contacts and 0.072 fF/μm^2 for metal / p-substrate contacts.

In most rectifier circuits which use diodes, one diode terminal is connected to a node at RF ground (the DC output voltage), while the other is connected to the input RF signal. The C_D of each diode in the circuit thus appears to the RF as a capacitance to ground and contributes to the total input capacitance of the rectifier.

Calculating transistor parasitic capacitances is slightly more complicated because they arise from different causes and there are more of them to be considered (see Figure 3-33). With four terminals, the MOSFET has ${}^4C_2 = 6$ possible parasitic capacitances. Of these, $C_{ds} = 0$ since the source and drain don't couple directly unless the channel length is extremely small. This leaves five parasitic capacitances to be determined: C_{gs}, C_{gd}, C_{bs}, C_{bd} and C_{gb}.

Like the diode junction capacitance, MOSFET parasitic capacitances are generally nonlinear. The subthreshold MOSFET capacitance model shall be used for calculations, since one is primarily concerned with low input RF amplitudes. To simplify matters, the nonlinear capacitances have also been linearized (the small signal approximation). With these approximations, for subthreshold operation, one gets

$$\begin{aligned} C_{gs} &= WC_{gso} \\ C_{gd} &= WC_{gdo} \\ C_{bs} &= (AS \times C_j + PS \times C_{jsw}) \\ C_{bd} &= (AD \times C_j + PD \times C_{jsw}) \\ C_{gb} &= 2WC_{gbo} + C_{ox}(1-\kappa)WL \end{aligned} \tag{3.47}$$

where C_{gso}, C_{gdo}, C_j, C_{jsw} and C_{gbo} are process dependent constants, W is the width of the device and L is its length. AS, AD, PS and PD are the areas and perimeters of the source and drain junctions, κ is the subthreshold slope parameter and C_{ox} is the oxide capacitance per unit area. $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$, where ϵ_{ox} is the dielectric constant of the gate oxide and t_{ox} is its thickness. In addition, because MOSFET sources and drains are symmetric, $C_{gso} = C_{gdo}$. If device layout is also symmetric (which is frequently

the case), $PS = PD$ and $AS = AD$. In that case, $C_{gs} = Cgd$ and $C_{bs} = C_{bd}$.

In addition to the capacitors in (3.47), transistors fabricated in wells (for example, PMOS devices in a n-well process) have a capacitance C_W between the bulk / well terminal and chip substrate (taken to be ground). This is given by

$$C_W = (AW \times C_{wj} + PW \times C_{wsw}) \tag{3.48}$$

where C_{wj} and C_{wsw} are process dependent constants, and AW and PW are respectively the area and perimeter of the well.

Any of the capacitances in (3.47) and (3.48) which has one end connected to the input RF signal and the other end at a node which is RF ground will appear as a capacitance to ground at the RF input. Since parasitic capacitances scale with device size according to (3.46) and (3.47), to minimize the input capacitance, one has to reduce the number and size of devices connected to the RF input terminals as much as possible. This is true for both transistor and diode based rectifiers.

Bonding Pad Capacitance

Bonding pads are used to connect the chip with the package. The overglass layer that protects the rest of the die has a cutout over the bonding pad, which allows a metal bond wire (in conventional bonding) or a solder bump (in flip chip bonding) to be attached to the top metal layer of the chip. The pad thus has to include top level metal, but to prevent liftoff problems during bond wire attach, most processes require all available metal layers to be present and connected together with vias. Minimum allowable pad dimensions in this process are 60 μm $\times$ 60 μm. This large metal area essentially presents a lumped parallel plate capacitance to the chip substrate. Designers have very little control over the value of this parasitic. One can, however, maximize its Q by reducing the series resistance the RF current sees before it flows to chip ground. To ensure this, each bonding pad was surrounded with a wide ring of substrate contacts tied to ground.

My best estimate of the differential input capacitance generated by the bonding

pads is 140 fF. This value is set by the area and fringe capacitance of the lowest metal layer (metal 1, in this case) used in the bonding pads, as well as the fringe capacitances of the other metal layers.

Parasitic Budget

After playing all possible layout tricks, one finally comes up with a parasitic budget for a given layout. This allows one to predict the input impedance of the rectifier circuit at zero bias (i.e., when the active circuits are off). All the parasitic capacitances to ground are summed up and divided by two to give C_s, the total differential input capacitance seen by the RF signal. The parasitic resistances are similarly added to give the series resistance R_s. The resultant quality factor of the input is $Q = \frac{1}{\omega C_s R_s}$. After fabrication, actual measurements of the input impedance $Z_{in} = R_s - \frac{j}{\omega C_s}$ are made to verify the calculated values of C_s and R_s. Parasitic series inductances (from interconnects and bond wires) have been ignored so far. This is a decent approximation at UHF. These inductances cause slight increases in the measured value of C_s over that predicted theoretically by contributing some positive reactance to Z_{in}. In practice, it was found that even fairly rough estimates of C_s matched measurements well, but that R_s, being much smaller in comparison, was more difficult to estimate accurately.

3.6.2 Circuit Optimization

In this section, I play games with circuit parameters which are similar to the ones that were used to optimize layout. These two aspects of system optimization are of course not really independent, but life can usually be simplified by carrying out circuit optimizations first, and then optimizing the layout of the resultant circuit.

Figure of Merit

One needs a performance metric to optimize the rectifier. From (2.21), which defines upper bounds on V_{in}, the input voltage amplitude as a function of various system

parameters, a suitable metric is $V_{to}\sqrt{C_s}$. Here V_{to} is the minimum value of the input RF amplitude V_{in} for which the rectifier can supply its required load, i.e., produce a DC voltage V_L at a load current I_L, and C_s is the input capacitance of the rectifier. The designer only has control over V_{to} and C_s. All the other parameters in (2.21) are fixed for any given application, and r, the distance from the transmitter, has to be maximized. Thus, one may write

$$r_{max} \propto \frac{1}{V_{to}\sqrt{C_s}} \tag{3.49}$$

where r_{max} is the maximum power up range of the system. In addition, we know that $P_{th} \propto \frac{1}{r_{max}^2}$ for a given transmitter power under free space propagation conditions. Here P_{th} is the power up threshold (minimum available power needed for the tag to power up). Thus we get

$$P_{th} \propto V_{to}^2 C_s \tag{3.50}$$

One can therefore define a quantity called the rectifier Figure of Merit (FOM), as follows

$$FOM = V_{to}\sqrt{C_s} \tag{3.51}$$

The lower the value of the FOM, the better. From (3.49) and (3.50), one gets

$$r_{max} \propto \frac{1}{FOM} \tag{3.52}$$

and

$$P_{th} \propto FOM^2 \tag{3.53}$$

The FOM of a particular rectifier design can be estimated as follows. V_{to} can be found from rectification curves at the required I_L (theoretical or experimental), such as those shown in Figures 3-2, 3-13 and 3-19 by finding the value of V_{in} which corresponds to the required V_L. This amounts to solving $V_L = F(V_{to}, I_L)$ for fixed V_L

and I_L. For example, for the floating gate rectifier curves of Figures 3-19, $V_{to} = 2.0$ V, 1.13 V, 0.84 V, 0.75 V and 0.72 V for $N = 1, 2, 3, 4$ and 5, respectively. C_s can be found experimentally by measuring the input impedance of the chip at the operating frequency.

In the following sections, I try to minimize the FOM by optimizing various rectifier design parameters.

Pump Capacitor Sizing

The size of the pump capacitors C_P in Figure 3-6 has a direct effect on circuit performance. Consider the single ended capacitance model shown in Figure 3-34. The bonding pad and package capacitances are independent of the size of the rectifier circuit (i.e., the number of stages), and lumped together into C_{pak}. C_{par} is the total parasitic capacitance to ground at the output of C_P which is not dependent on the size of C_P, and consists mainly of transistor parasitics. ϵ is the bottom plate coefficient of the capacitor C_P, and N is the number of pump capacitors connected to the (single ended) RF input. This is equal to the number of stages (in this section, denoted by N_S) if floating gates are not used. When floating gates are used, $N \approx 3N_S$, since every floating gate cell needs three pump capacitors connected to the RF input, all of which are supposed to be of the same size, for simplicity (see Figure 3-17). Figure 3-34 shows two cases: the RF input is connected to either the top or bottom plate of C_P.

Consider the top plate case first. The RF amplitude at the transistor gates V_G is the actual input that the rectifier receives. V_G is lower than V_{in}, the theoretical input amplitude, because of the capacitive voltage divisions shown in Figure 3-34. In this case

$$V_G = \frac{C_P}{C_P(1+\epsilon) + C_{par}} V_{in} \tag{3.54}$$

The total input capacitance seen by the RF input is given by

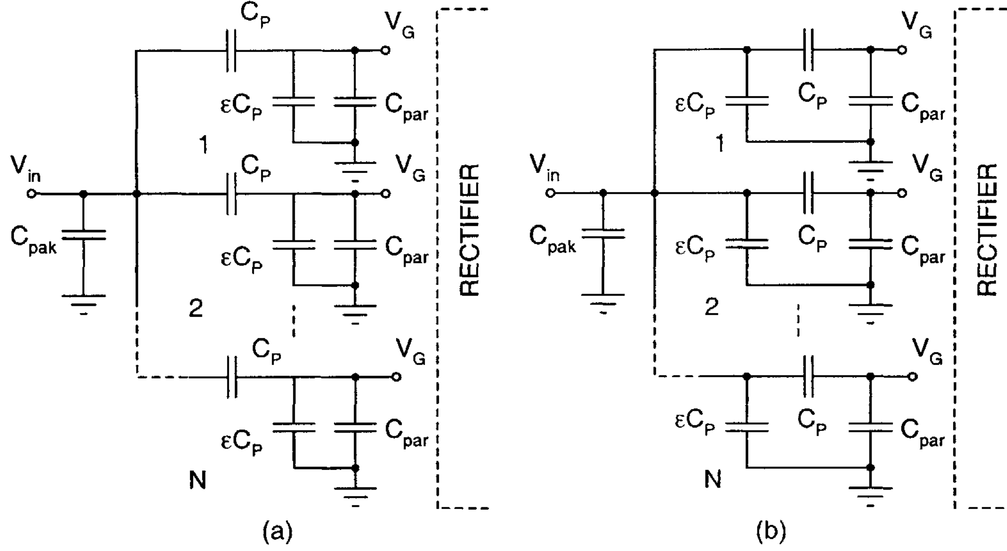


Figure 3-34: Circuit model used to find optimum pump capacitance C_P when the RF signal V_{in} is fed to the (a) top and (b) bottom plate of C_P.

$$C_s = C_{pak} + N\frac{C_P\left(C_{par} + \epsilon C_P\right)}{C_{par} + C_P\left(1+\epsilon\right)} \tag{3.55}$$

Similarly, for the bottom plate case, one gets

$$V_G = \frac{C_P}{C_P + C_{par}}V_{in} \tag{3.56}$$

and

$$C_s = C_{pak} + N\epsilon C_P + N\frac{C_P C_{par}}{C_{par} + C_P} \tag{3.57}$$

Also, for a given available power P_A, one knows from (2.21) that $V_{in} \propto \frac{1}{\sqrt{C_s}}$ (i.e., the FOM is fixed). One can intuitively see from Figure 3-34 and these equations why there should be an optimum value of C_P which will maximize V_G, and thus rectifier efficiency, for given P_A. As C_P increases, the capacitance division ratio with C_{par} increases, increasing V_G; however, C_S also increases (because $\epsilon \neq 0$), decreasing V_{in} (and eventually V_G. Figure 3-35 shows calculated values of V_G as a function of C_P for different values of N. Assumed parameter values were C_{pak} = 300 fF, ϵ = 0.15

and $C_{par} = 25$ fF. The values of V_G have been normalized. One sees the existence of a clear optimum (maximum) for both top and bottom feed cases. The bottom feed works slightly better for all values of N. As expected, performance worsens as N and C_s increases. For a given value of N, one can use this plot to find the optimum value of C_P. For most cases, this lies in the range 100-250 fF.

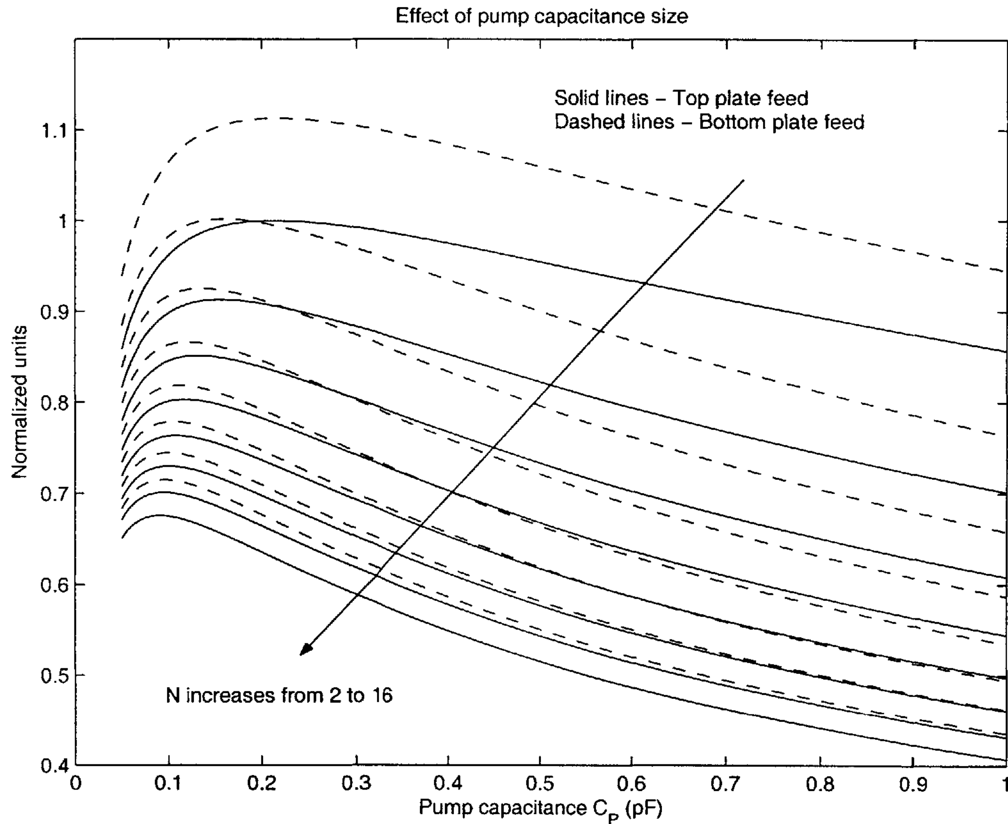


Figure 3-35: Calculated effect of pump capacitor C_P sizing on rectifier input voltage V_G. Values have been normalized.

Finally, it should be noted that this analysis is also valid for other types of charge pumps, such as diode based voltage multipliers. One just has to define appropriate values for N and C_{par}.

Schottky Diode Voltage Multiplier

Once the value of the pump capacitor C_P has been fixed, the main variables in the diode based voltage multiplier shown in Figure 3-1 are the sizes of the diodes and the number of stages N (I am going back to calling the number of stages N, for convenience). The total input capacitance C_s of the rectifier is modeled as

$$C_s = NC_{stage} + C_{pack} \tag{3.58}$$

where C_{stage} is the input capacitance contributed per stage, and C_{pack} represents all parasitic capacitances that do not scale with the number of stages (mainly package and bonding pad capacitance). C_{stage} consists of device and layout parasitics (as discussed in Section 3.6.1), and can be estimated from simulations and layout extraction techniques. The zero bias capacitance model of a single voltage doubler stage (assuming bottom plate feed) is shown in Figure 3-36, where ϵ is the bottom plate coefficient, C_{D1} and C_{D2} are the diode junction capacitances, C_{n1} and C_{p2} are layout parasitics, R_{D1} and R_{D2} are the diode series resistances, R_{CAP} is the series resistance of the pump capacitor C_P, and R_W is the well resistance. For these Schottky diodes $C_n \gg C_p$ because C_p is only stray wiring parasitics, but C_n includes the substrate capacitance of the n-well which forms the negative terminal. Assuming C_P and C_L are much larger than the other capacitances,

$$C_{stage} \approx \epsilon C_P + C_{D1} + C_{D2} + C_{n1} + C_{p2} \tag{3.59}$$

One can also define a series resistance R_{stage} for the stage. Normally the diode series resistances dominate the loss, so

$$R_{stage} \approx R_{D1} || R_{D2} \tag{3.60}$$

Since C_{D1} and C_{D2} are proportional to the junction area, they increase as the diodes get bigger (i.e., they are proportional to I_S). However, the series resistances R_{D1} and R_{D2} of the diodes decrease linearly with junction area, so the diode Q

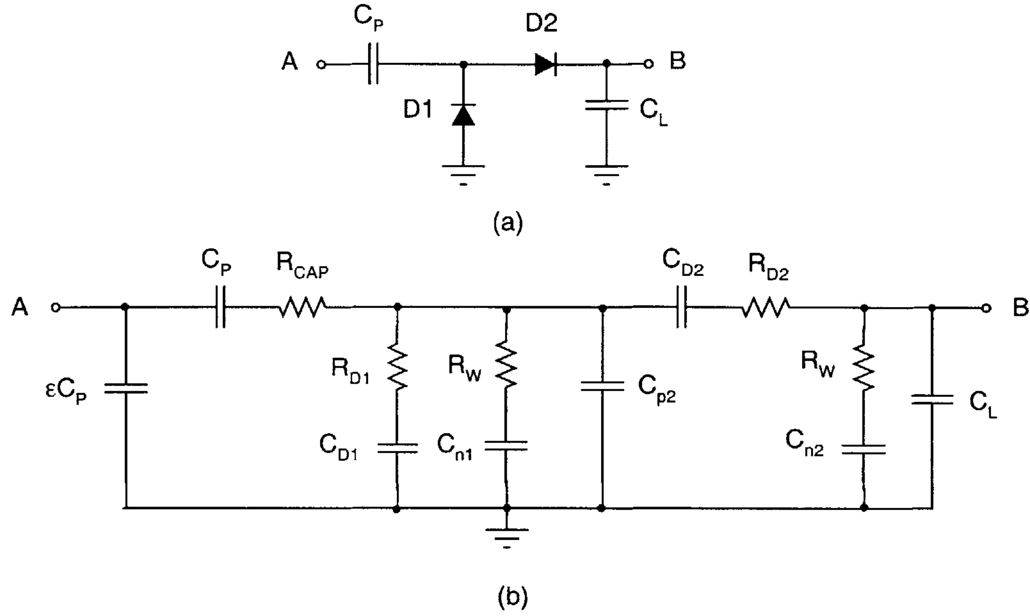


Figure 3-36: (a) A single voltage doubler stage and (b) its model at zero input amplitude.

remains roughly constant. In addition, the voltage drop per stage is $(V_{th1} + V_{th2})$, and the forward voltage drops V_{th1} and V_{th2} of the diodes decrease logarithmically with I_S. This weak dependence of V_{th} on diode size enables one to simplify the optimization problem.

The optimization strategy I follow is to decrease the diode sizes until $(C_{D1} + C_{D2}) \approx (C_{n1} + C_{p2} + \epsilon C_P)$, i.e., the parasitic capacitances become significant. The Q of the stage is still set primarily by R_{D1} and R_{D2}, and the voltage drop $(V_{th1} + V_{th2})$ barely changes. An additional subtlety is that C_n is also roughly proportional to the diode size (since the size of the well increases as the diode becomes larger). To account for this, the size of the clamping diode D1 is decreased till $C_{n1} \approx C_{p2}$. The size of the peak detecting diode D2 is then adjusted to maintain $(C_{D1} + C_{D2}) \approx (C_{n1} + C_{p2} + \epsilon C_P)$. This gives the final design of the stage. This simple optimization strategy usually produces designs that match well with those obtained from a formal function minimization routine. A typical design using Schottky diodes in my process at 900 MHz is $D1 = 10$ contacts, $D2 = 80$ contacts and $C_P = 250$ fF (each contact is 0.6 μm $\times$

0.6 μm).

Next one has to optimize the number of stages N. To first order (as long as $NC_{stage} \gg C_{pack}$), the input Q is constant with N and equal to the Q of each stage. This is because the Q is unaffected when N reactive elements with the same Q are placed in parallel. One optimizes for N by finding the FOM of the rectifier as a function of N. To find V_{to} for different values of N, one uses Figure 3-2. One can already see why there should be an optimum: V_{to} decreases with N, but $\sqrt{C_s}$ increases, so their product should go through a minimum. This is confirmed by Figure 3-37, which shows the calculated figure of merit for different values of C_{pack}. Each stage was optimized as discussed previously, $V_L = 1.5$ V and $I_L = 2$ μA. The graphs have been normalized. Referring to (2.23) and using $G_r = 1.5$ dBi, $G_t = 6$ dBi, P_t=1 W and $B = 30$ MHz, the best performance, i.e., $FOM = 1$, occurs for four stages and corresponds to $r_{max} = 14.1$ meters, i.e., an available power threshold of $P_{th} = 9.7$ μW. This is for $C_{pack} = 0$. With $C_{pack} = 0.4$ pF, P_{th} increases to about 17 μW ($r_{max} = 10.7$ meters, six stages, $FOM \approx 1.32$). This is close to the best published result (16.7 μW, [22]) but over a much wider bandwidth.

Switching Cell Rectifier

Optimization of the switching cell based rectifier is very similar to that of the diode based rectifier. The calculation of C_{stage} and R_{stage} differs in the details, but roughly the same scaling relationships hold. Bigger transistors have bigger parasitic capacitances and smaller series resistances. The size of a transistor is the aspect ratio $\frac{W}{L}$. L is kept fixed at the minimum allowed value in the process and vary W to change the device size. The main contributor to series resistance is the gate resistance R_G; as W increases, more fingers can be used, so R_G decreases. The voltage drop per stage scales as $\log\left(\frac{L}{W}\right)$ for subthreshold operation, and as $\sqrt{\frac{L}{W}}$ above threshold, and thus is still a weak function of device size.

After carrying out the optimization process, one ends up with each device sized at $\frac{W}{L} = 10$ and $C_P = 250$ fF. Figure 3-38 shows the resultant FOM of the floating diode rectifiers (rectification curves in Figure 3-19) for different values of N and C_{pack}. This

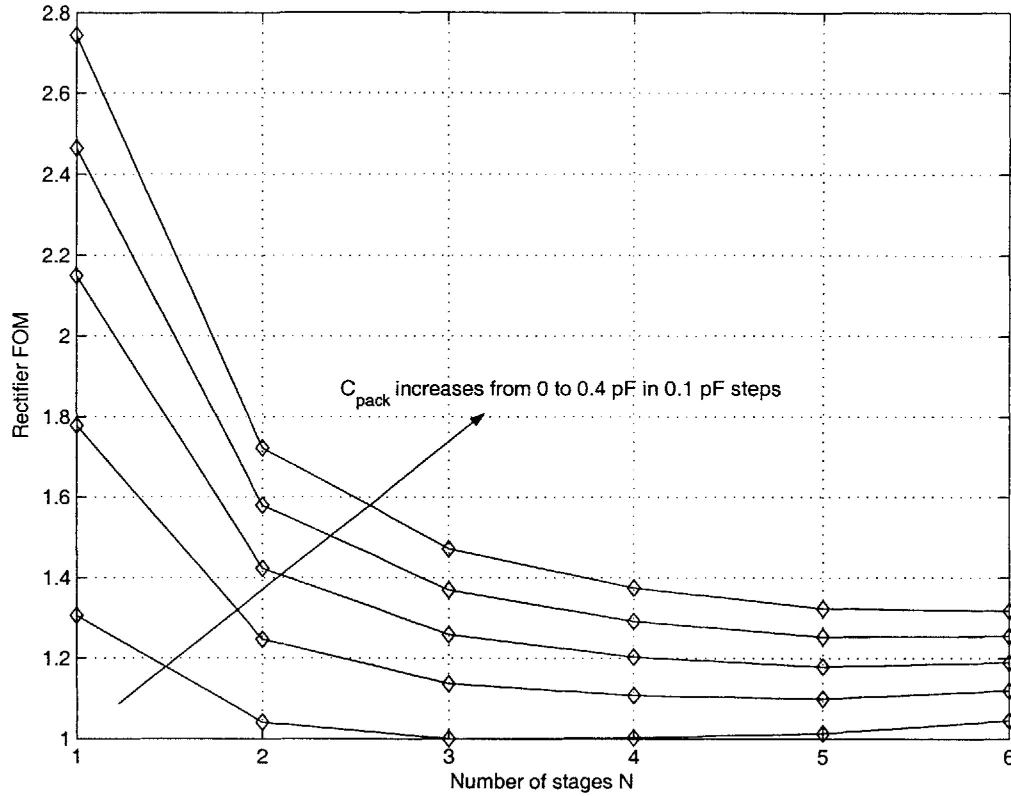


Figure 3-37: Rectifier figure of merit for Schottky diode based voltage multipliers with different package capacitances C_{pack} for $C_P = 0.25$ pF, $V_L = 1.5$ V and $I_L = 2$ μA.

graph has been scaled with respect to the corresponding one for Schottky diode based voltage multipliers (Figure 3-37) so that the performance of the two rectifier types can be compared. One sees that the performances are very similar (similar FOM values) for higher values of C_{pack}. For low C_{pack} values, the floating gate rectifier performs somewhat better. The best value occurs for $C_{pack} = 0$ and a three stage rectifier. This corresponds to $r_{max} = 15.1$ m and $P_{th} = 8.6$ μW. Further incremental improvements are possible by tweaking the threshold voltage values, but will probably not be robust.

The performance degradation which occurs if optimization is not carried out is illustrated in Figure 3-39. This shows the FOM for a non-optimized switching cell rectifier. Floating gates were not used, $C_P = 2$ pF and $\frac{W}{L} = 30$. The large values of

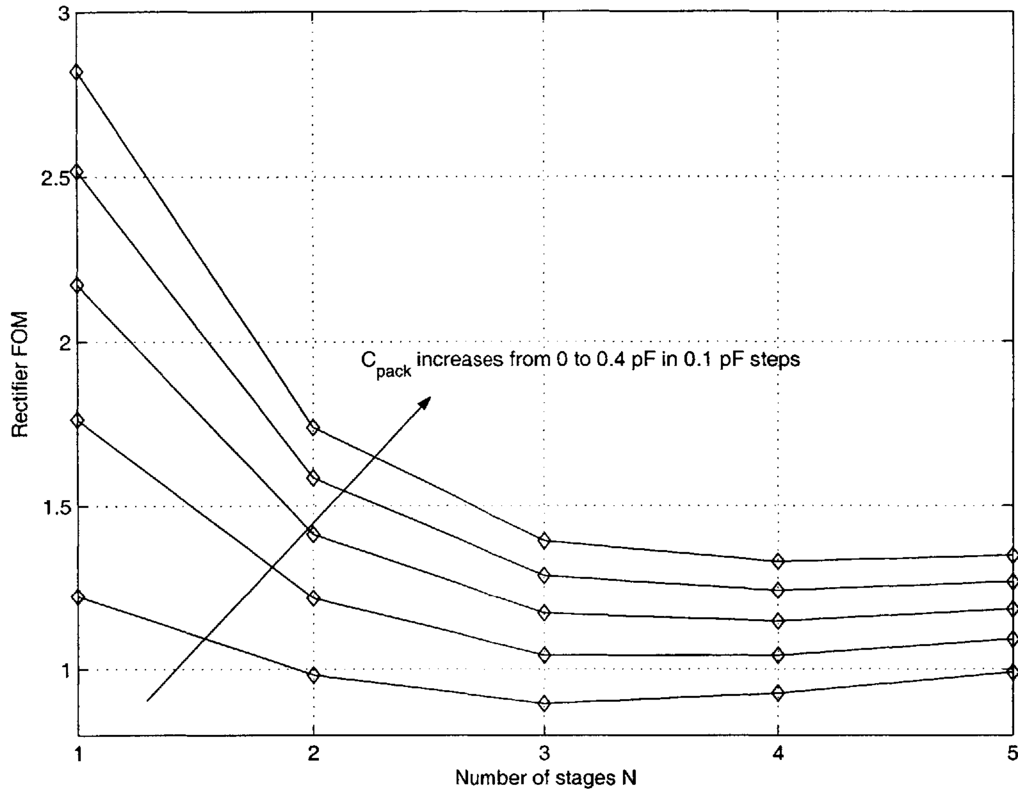


Figure 3-38: Figure of merit for floating gate based switching rectifiers with different package capacitances C_{pack} for $C_P = 0.30$ pF, $V_L = 1.5$ V and $I_L = 2$ μA. Plot normalized w.r.t. Figure 3-37.

C_P and $\frac{W}{L}$ increase C_S and degrade performance. The FOM for a three stage rectifier with $C_P = 0.4$ pF is about 1.9, corresponding to $r_{max} = 7.4$ m and $P_{th} = 35.5$ μW. This is the same design that was tested experimentally in Section 2.5.2, and the observed power up threshold of $P_{th} \approx 40$ μW ($V_L = 1.5$ V, $I_L = 2$ μA, $B = 30$ MHz) matches very well with the theoretical value.

Effect of Load Current

A load current of $I_L = 2$ μA has been assumed throughout the optimization process. This value is not written in stone; it's simply an estimate of the current consumption of most commercial UHF tags. In this final section, I study the effects of changing I_L on V_{to} (and consequently, the FOM and power up threshold P_{th}). Figure 3-40 shows

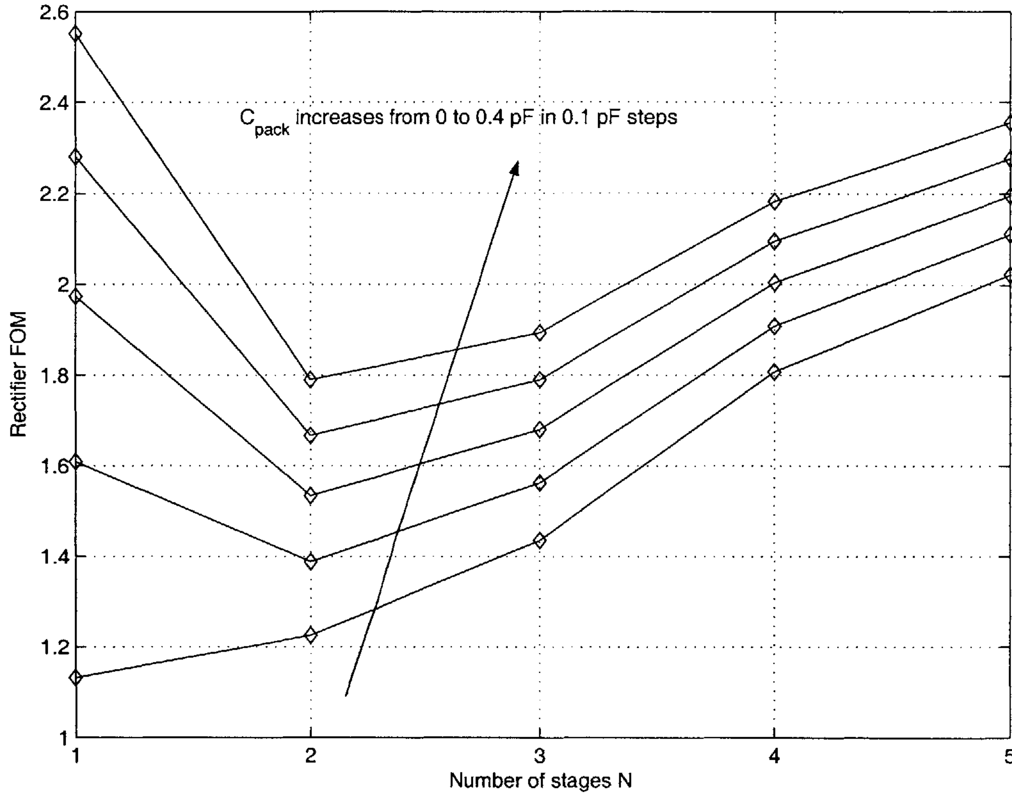


Figure 3-39: Figure of merit for switching rectifiers with different package capacitances C_{pack} for C_P = 2 pF, V_L = 1.5 V and I_L = 2 μA. Plot normalized w.r.t. Figure 3-37.

the simulated rectification curves of a three stage floating gate rectifier for various values of I_L. The RF frequency was 900 MHz and $C_P = 0.3$ pF.

One sees that the shapes and relative positions of the curves in Figure 3-40 match the general theoretical ones shown in Figure 2-7. The value of V_{to} for $V_L = 1.5$ V increases from 0.72 V for no load ($I_L = 0$) to 1.08 V for $I_L = 10$ μA. Since the input capacitance C_s is constant, from (3.51) one sees that $FOM \propto V_{to}$. This means that the rectifier FOM increases as the load current increases. In other words, the power up threshold P_{th} increases and the maximum power up distance r_{max} decreases as the load current I_L increases. In this particular case, however, one notices that going from the no load to the 10 μA load case only increases the power up threshold by a factor of $\left(\frac{1.08}{0.72}\right)^2 = 2.25$. Thus, simply decreasing I_L only gives limited wins in P_{th} (a

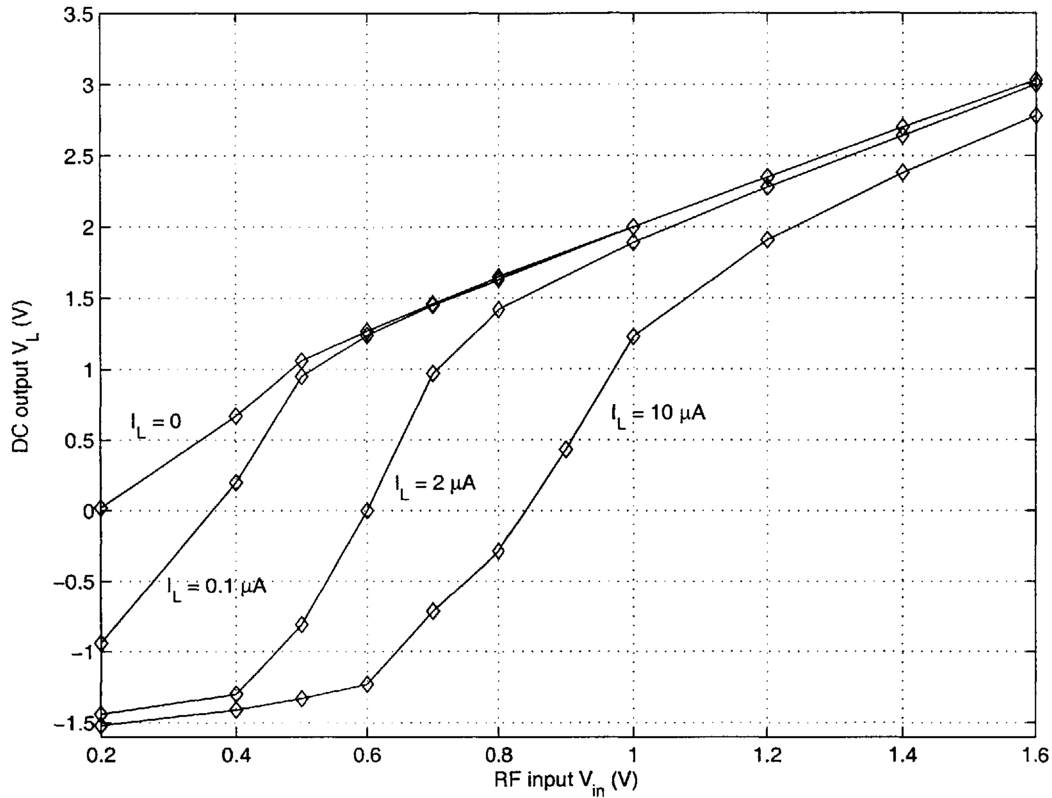


Figure 3-40: Simulated rectification curves of a three stage floating gate rectifier for various load currents I_L.

factor of 2, in this case). To further improve performance, one needs to reduce the required load voltage V_L, possibly by applying low voltage, low power digital design techniques [9].

3.6.3 Input Impedance Measurement

Measuring the input impedance Z_{in} of the chip accurately proved to be surprisingly difficult. The main reason is the high Q, which makes measuring the value of the small real part of Z_{in} challenging. The other problem faced was de-embedding the chip from the test fixture used for measurements. Cables, connectors, feed lines and IC sockets are examples of test fixtures. Any measurement setup can be regarded as two networks in cascade: the text fixture, and the device under test (DUT) (see

Figure 3-41). The characteristics of the DUT are unknown and are to be found. Direct measurements can only be made of the test fixture-DUT combination and thus include the (unwanted) characteristics of the test fixture. If one chooses a cascadable set of network parameters, like ABCD parameters (S-parameters are not cascadable, but can be converted into ABCD parameters), the measured network parameters can be written as a matrix equation

$$[A]_M = [A]_F [A]_D \tag{3.61}$$

where $[A]_M$, $[A]_F$ and $[A]_D$ are measured, test fixture and DUT network parameters, respectively. The purpose of de-embedding is to remove the effects of the test fixture network from the measurements, so that $[A]_D$, the characteristics of the DUT alone, can be found. De-embedding procedures can be based on models or measurements. Model based de-embedding involves the creation of a mathematical model of the test fixture. Analytical techniques or electromagnetic simulations can be used to create this model. Once created, the model can be used to estimate $[A]_F$, the network parameters of the text fixture. This estimate of $[A]_F$ is then used to invert the equation (3.61) to find $[A]_D$.

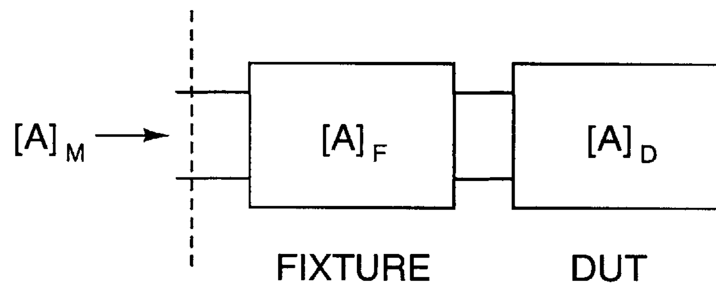


Figure 3-41: Typical measurement scenario. The device under test can only be accessed through a test fixture.

The end result of measurement based de-embedding is identical - an estimate for $[A]_F$ which can be used to invert (3.61). The estimation procedure is, however, different. In this procedure, the DUT in Figure 3-41 is replaced with known networks (opens and shorts are commonly used), and the resultant values of $[A]_M$ are mea-

sured experimentally. Solving a set of circuit equations then allows one to find $[A]_F$. Further details of the de-embedding procedure can be found in [36]. Both model and measurement based de-embedding are used in practice. The measurement based technique is usually found to be more accurate, but technical difficulties limited us to using model based de-embedding in this thesis. The model was created by using a planar electromagnetic field solver (ADS Momentum) to find the network parameters of a simplified physical representation of the test fixture.

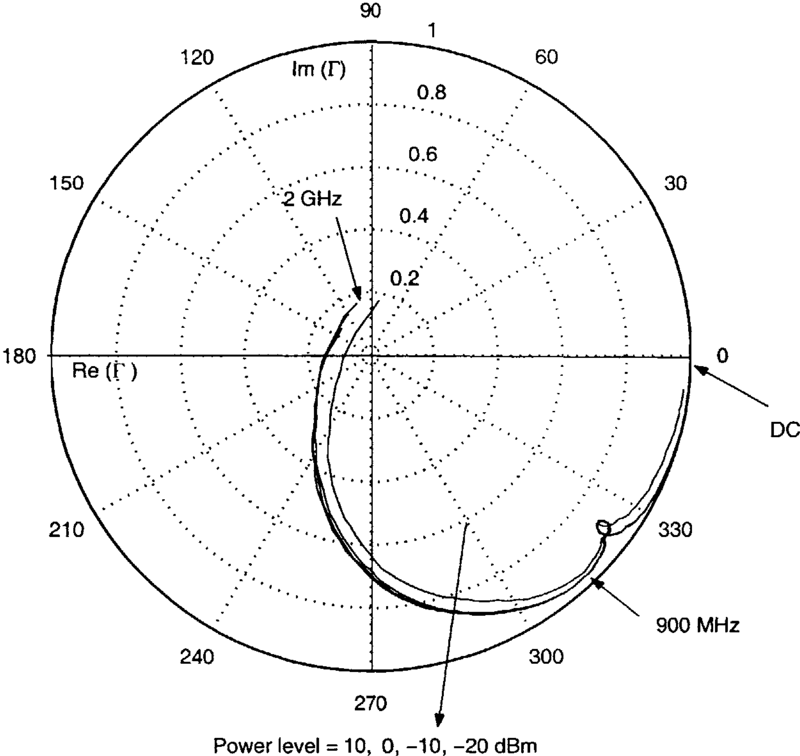


Figure 3-42: Measured input reflection coefficient Γ_{in} of a three stage rectifier in a LCC package after model based de-embedding for different input power levels.

Figure 3-42 shows the measured input reflection coefficient Γ_{in} (on a polar plot) of the three stage rectifier tested in Section 2.5.2 after model-based de-embedding for different input power levels. The rectifier was inside a 40 pin LCC package. The de-embedding model was generated from planar electromagnetic simulations of the test fixture. The reference impedance is 50 Ω. One sees that, as expected, the imaginary

part of Γ_{in} is negative up to about 1.6 GHz, i.e., the rectifier input looks capacitive. The real part of Z_{in} increases at high power levels because the effective resistance of the rectifier decreases, loading the input reactive circuit (see Figure 2-11). At low power levels, the input impedance at 900 MHz is $6 - j125\ \Omega$, corresponding to an input capacitance $C_s = 1.4$ pF and an input Q of 21. This value of Q must be treated as approximate because it is difficult to measure the small real part of Z_{in} accurately.

Experimentally, it was found that model based de-embedding underestimated the series inductance of the text fixture, leading to an inflated estimate for C_s, the input capacitance of the chip. The value of C_s was hence estimated as follows. First the resonant frequency of the tag was measured as described in Section 2.5.2. By using simulated values of the antenna impedance at this frequency, one can then estimate the input impedance of the chip. The resultant estimates were $C_s \approx 1.0$ pF and input $Q \approx 30$ for resonant frequencies around 900 MHz. It would probably be more accurate to use measurement based de-embedding instead. I hope to investigate this technique soon.

Chapter 4

Antenna Design

4.1 Introduction

Most of the area of the power extraction system is occupied by the antenna. Recent attempts have been made to integrate antennas on chip (silicon substrates) [27]. However, fundamental physical constraints on minimum antenna size do not allow one to follow this approach at UHF. For this thesis, I confined myself to designing planar antennas on printed circuit board substrates. A real tag has to be built on a flexible substrate. Deformations of this substrate can lead to variations in antenna performance which have not been explored in this thesis, but are of interest for future work.

Table 4.1 shows relevant physical parameters for the printed circuit boards which were used to fabricate all the antennas described in this chapter. FR-4 was chosen as the board material mainly because of its low cost and easy availability. A maximum of two metal layers (one on either side of the FR-4 dielectric) were used.

4.2 Transmission Lines

In this section, I gather together, for convenience, some design data on transmission lines which have proven useful for making printed circuit antennas and interfacing them to other circuits, but are not widely known or used.

Table 4.1: Printed circuit board parameters

Parameter	Symbol	Value
Material	-	FR-4
Dielectric constant	ϵ_r	4.2 $\pm$ 10 %
Dielectric loss tangent	$\tan\delta$	0.01
Board thickness	h	31 mil
Metallization	-	Copper
Metal conductivity	σ	5.8 $\times 10^7$ ℧/m
Metal thickness	t	1.3 mil

4.2.1 Coplanar Strips Line

Coplanar strips (CPS) line is a planar balanced transmission line geometry which is useful for feeding printed dipole antennas. It consists of two planar conductors on a dielectric substrate separated from each other by a distance $2a$ (see Figure 4-1).

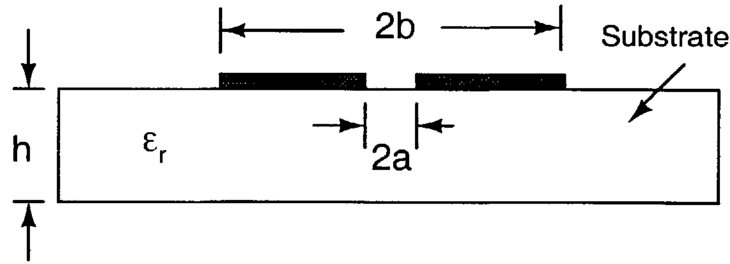


Figure 4-1: Cross section of coplanar strips (CPS) transmission line geometry. The dielectric substrate is infinitely wide and has a thickness h (which may also be infinite).

The analysis of CPS lines, especially when the substrate has finite thickness, is somewhat involved. The analytical expressions may be found in [6] and are graphed in Figure 4-2. For design purposes, one notes that the characteristic impedance Z_0 increases monotonically with the ratio $\frac{a}{b}$. The lowest practical value of $2a$ is set by fabrication tolerances, thus limiting the lowest obtainable Z_0 for reasonable values of b. With normal printed circuit board fabrication processes, this minimum value is about 70 Ω. This necessitates the use of an impedance matching network when feeding the antenna from a 50 Ω source (unless the feed line can be kept much shorter compared to a wavelength, in which case its characteristic impedance doesn't matter).

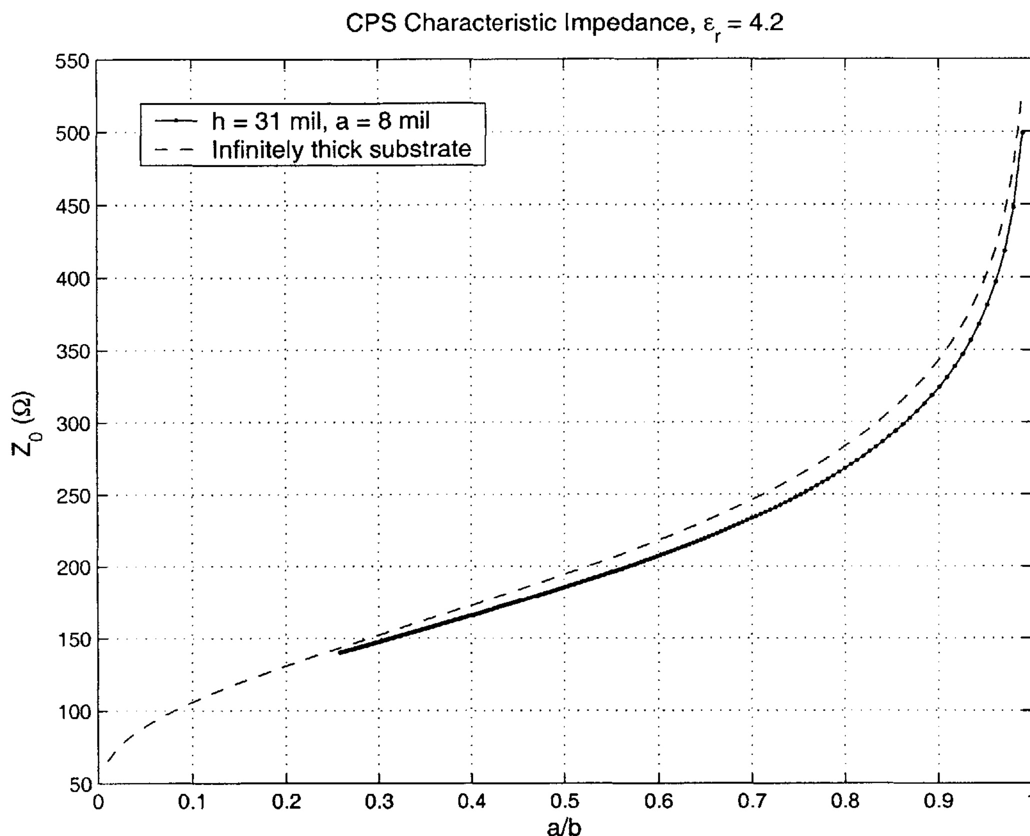


Figure 4-2: Calculated characteristic impedance Z_0 of coplanar strips transmission line.

4.2.2 Parallel Strips Line

Parallel strips line is a balanced planar transmission line geometry (see Figure 4-3) which is useful for feeding double sided printed dipole antennas with edge launch connectors. It consists of two planar conductors of width w separated by a dielectric substrate of thickness d. The analysis of this line is simplified by the observation that the electric field distribution is unchanged when an infinite sized perfect electric conductor (i.e., a ground plane) is inserted into the substrate parallel to the strips [6]. When inserted midway between the strips, the geometry is converted to two microstrip lines back to back. The characteristic impedance $Z_{0,parallel-strips}$ and effective dielectric constant $\epsilon_{re,parallel-strips}$ are thus simply related to that for microstrip line with substrate thickness $h = \frac{d}{2}$ by

$$
\begin{aligned}
Z_{0,parallel-strips} &= 2Z_{0,microstrip}\left(h = \tfrac{d}{2}\right) \\
\epsilon_{re,parallel-strips} &= \epsilon_{re,microstrip}\left(h = \tfrac{d}{2}\right)
\end{aligned}
\qquad (4.1)
$$

The characteristic impedance of the microstrip line with $h = \frac{d}{2}$ is half of $Z_{0,parallel-strips}$ because each microstrip line carries the same current but only half the voltage of the parallel strips line. Alternatively, one can explain this as a differential to single ended impedance conversion (see Figure 4-4). Since widely known analytical formulas exist for the characteristic impedance of microstrip lines [6], that of parallel strips line can now be easily found.

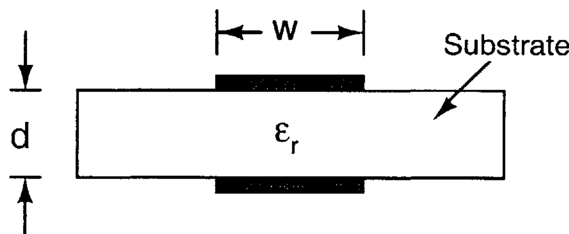


Figure 4-3: Cross section of parallel strips transmission line geometry. The dielectric substrate is infinitely wide and has a thickness d.

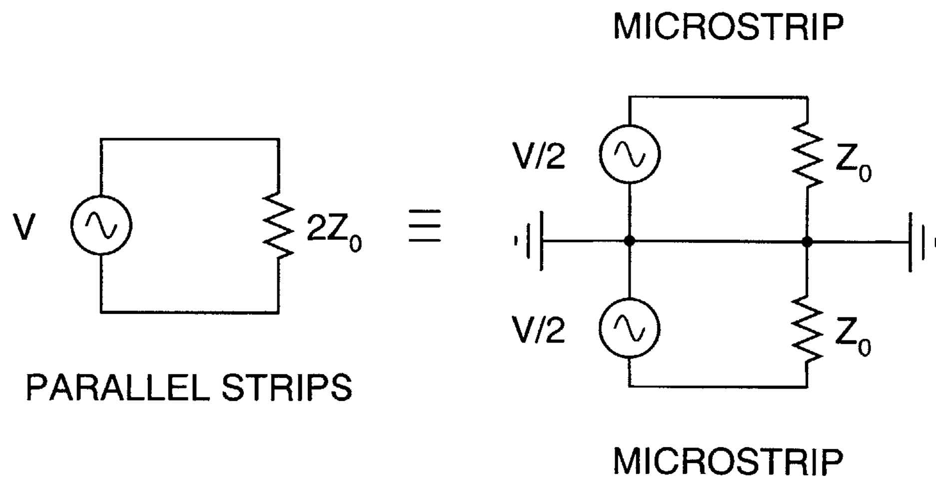


Figure 4-4: Differential (parallel strips line) to single ended (back to back microstrip lines) impedance conversion.

4.3 Transmit Antennas

Transmit antennas are needed for testing the power extraction system/tag. Size requirements on them are not stringent, but they must have high relatively gain (up to 6 dBi is allowed by the RFID protocol). Also, the impedance matching bandwidth of the transmit antenna must be as large as possible in order to impart robustness. This is especially important given that the proximity of readers to humans is likely to cause random shifts in the transmit antenna's resonance frequency. The design challenge here is thus to achieve good gain and enough impedance matching bandwidth when the antenna is connected to a signal source with a 50 Ω output impedance while keeping the physical size small enough to permit easy testing and provide enough portability. Practically, this means a size of about 12" square or less.

4.3.1 Typical Transmit Antennas

Typical transmit (reader) antennas used for RFID systems include Yagi antennas, bowtie antennas and dipole arrays. I shall not describe these further since the main focus here is just to design a working transmit antenna for testing my power extraction system.

4.3.2 Final Design

Among the most popular dipole antennas are Yagis [20]. They can provide relatively high gains (up to $\approx$16 dBi is common), are compact and give good bandwidths when properly designed. I designed a planar, compact, broadband Yagi antenna as the transmit antenna for testing the power extraction system.

The basic idea behind a Yagi antenna is to use constructive and destructive interference of waves in order to increase the gain of a simple half wave dipole. Refer to Figure 4-5, which shows the layout of my Yagi antenna. The central dipole is the only actively fed element and is nominally half a wavelength ($\frac{\lambda}{2}$) long at the center frequency of operation, i.e., $l_l \approx \frac{\lambda}{2}$. This dipole is electromagnetically coupled to several other dipole elements which are parasitic, i.e., not directly fed. The parasitic

elements in front of the active element are called directors, while those behind it are called reflectors.

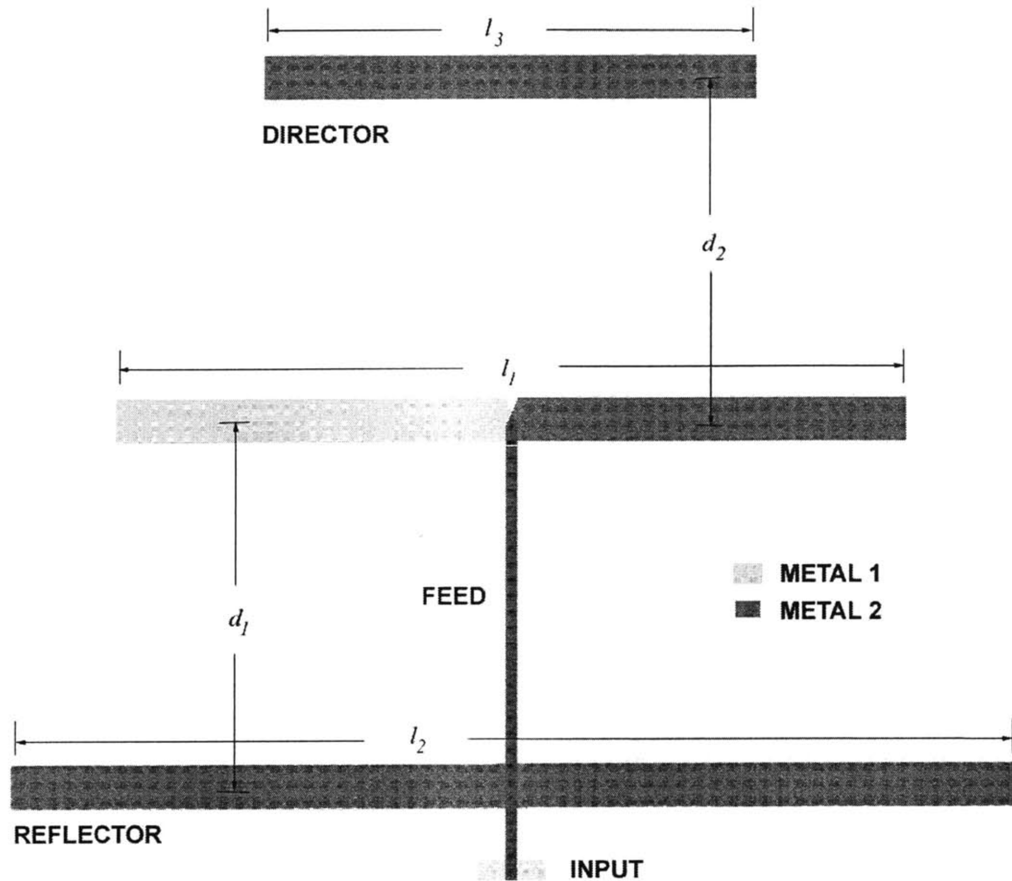


Figure 4-5: Structure of the planar Yagi antenna with parallel strips line feed. Metal 1 and Metal 2 are on opposite sides of a thin dielectric layer.

In my design (Figure 4-5), I have the simplest case of a single director and a single reflector. The director is made somewhat shorter than $\frac{\lambda}{2}$ so that it appears capacitive, while the reflector is made longer than $\frac{\lambda}{2}$ to look inductive, i.e., $l_3 < \frac{\lambda}{2}$ and $l_2 > \frac{\lambda}{2}$. One wants the radiated wave in the forward direction to arrive back at the active element after being reflected and re-radiated by the director in phase with the original wave. The distance d_2 between the active element and the director and the length l_3 of the director is adjusted so that this constructive interference increases the energy radiated forward (upward in Figure 4-5). Similarly, the reflector length l_2

and distance d_1 is adjusted to ensure destructive interference of the backward wave. This increases the antenna gain by making the radiation pattern asymmetric: more energy is radiated forwards than backwards. The gain can be further increased by adding more directors, but incremental improvements rapidly diminish when more than two or three are used.

The antenna in Figure 4-5 has a size of about 6.9" $\times$ 5.5". The feed line is a parallel strips line. This can be easily designed to have 50 Ω impedance with typical printed circuit board fabrication tolerances, simplifying impedance matching to the 50 Ω coaxial cable used to feed the antenna. The active dipole is fabricated on both sides of the printed circuit board. The electric fields at the central feed point are then primarily vertical (across the dielectric), allowing direct interfacing to the parallel strips feed line. The overlap area of the two dipole arms around the feed point is adjusted to vary the input reactance of the antenna appropriately (the overlap area acts like a lumped parallel plate capacitance). The director and reflector are single planar metal lines. They don't need to be split in the middle to be fed since they are only fed parasitically by the radiated wave from the central dipole.

The antenna was simulated using ADS Momentum, a planar (2.5D) method of moments based electromagnetic field solver. It was fabricated by a commercial printed circuit vendor. Relevant physical parameters are shown in Table 4.1. Figure 4-6 compares the experimental reflection coefficient magnitude $|S_{11}|$ of the antenna as seen from a 50 Ω source with simulation. The experiment was not carried out in anechoic conditions, but efforts were made to approximate free space conditions as much as possible. The figure shows that the experimental and simulated responses, especially the reference frequencies, match very well. The -10 dB bandwidth is 200 MHz about a center frequency of 900 MHz. The origin of the spurious resonances in the experimental response is unknown. Possible causes are parasitic L's and C's added by the edge mount coaxial connector that was used to feed the antenna, and reflections off floors and walls (these could couple the resonant antenna to itself with a time delay, splitting the resonant frequency). Measurements have to be taken in an anechoic chamber to be sure about this.

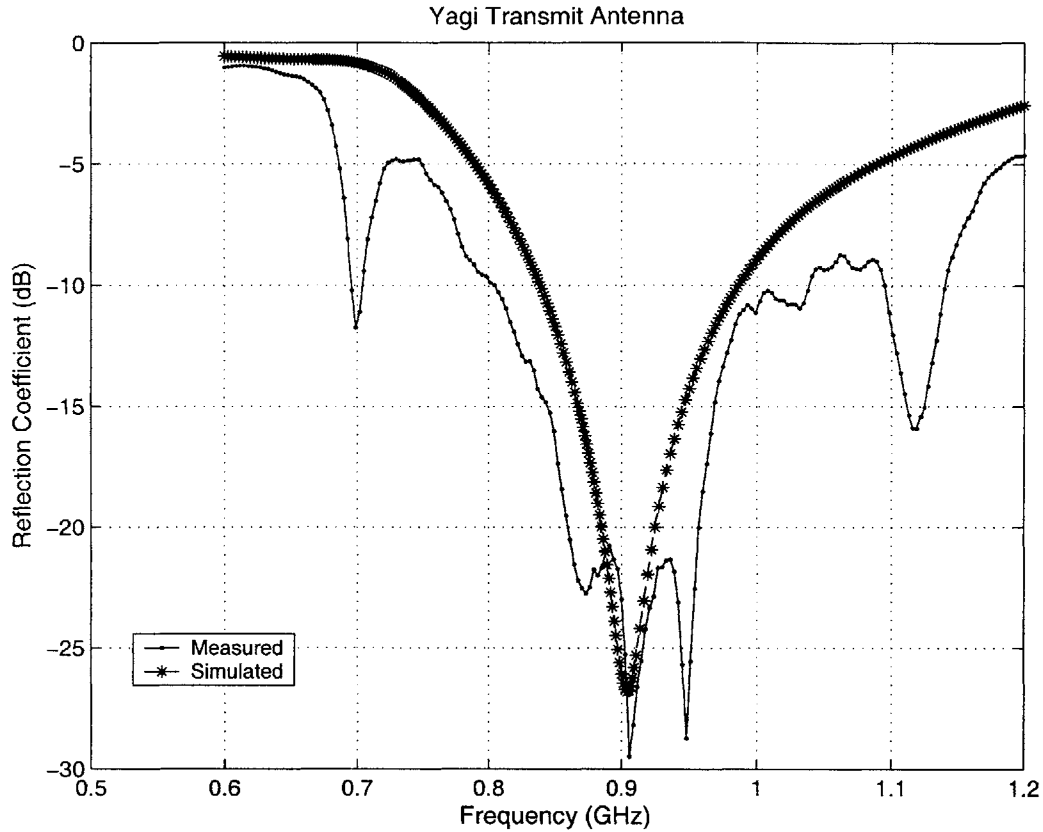


Figure 4-6: Measured and simulated reflection coefficients (in dB) of the Yagi antenna into a 50 Ω load.

The gain G of the antenna was measured experimentally by using two identical antennas, placing them a known distance r apart and measuring the transmission loss T (in dB) between them using a network analyzer. I assume that r is large enough for the two antennas to be in each other's far fields, and also that free space propagation occurs. Then

$$T = 2G + 20\log\left(\frac{\lambda}{4\pi r}\right) \tag{4.2}$$

where G is in dBi (dB gain over that produced by an isotropic radiator). Experimentally, this procedure gives an average value of G from 900 - 930 MHz of 4.4 dBi. This is very close to the simulated value of G, which is 4.6 dBi over the same frequency range.

4.4 Receive Antennas

Receive antennas and associated matching networks must be designed to ensure maximum power transfer to the rectifier. The design procedure is somewhat complicated by the fact that the input impedance of the rectifier is dependent on the RF input level, i.e., is nonlinear. My solution to this problem is to match to the input impedance at the lowest power level which ensures circuit power up, and assume that any mismatch at higher power levels is compensated for by the increased power available. Small antennas (characteristic physical dimension $\ll \lambda$, the wavelength in the medium) are primarily reactive. The antenna impedance can only become purely real under certain conditions. This is known as a resonance and requires that the characteristic dimension be comparable to λ.

4.4.1 Typical Receive Antennas

Short linear dipoles appear capacitive. This is because their radiation resistances are low; the antenna essentially looks like a short length of open circuited transmission line. They cannot be directly impedance matched to capacitive loads. Resonances happen when the dipole length l becomes comparable to λ, the electromagnetic wavelength in the medium, for example at $l = \frac{\lambda}{2}$ and $l = \lambda$.

Folded Dipoles

Folded dipoles are formed by adding a second radiating element in parallel with a normal linear dipole and shorting the ends of the two elements together. Short folded dipoles look inductive, i.e., like a short length of short circuited transmission line. Another way to see this is to notice that in essence a folded dipole is a planar loop antenna where one dimension has been reduced almost to zero. Since small loops look like inductors, one expects the input impedance of a folded dipole to be inductive. This property makes small folded dipoles easy to impedance match to capacitive loads (forming the L-match shown in Figure 2-6).

Loaded Dipoles

The input impedance Z_{in} of a lossless transmission line with characteristic impedance Z_0 and termination (load) impedance Z_L is given by

$$Z_{in} = Z_0 \frac{Z_L + \jmath Z_0 \tan(\beta l)}{Z_0 + \jmath Z_L \tan(\beta l)} \tag{4.3}$$

where β is the propagation constant and l is the length of the line. For a short (compared to the wavelength) length of short-circuited line ($Z_L = 0$, $\beta l \ll 1$), this expression reduces to

$$Z_{in} = \jmath Z_0 \beta l \tag{4.4}$$

For a dispersionless, TEM line like coaxial cable and stripline, one has the simple dispersion relation $\beta^2 = \omega^2 \mu \epsilon$, where ω is the frequency of operation and μ and ϵ are the permeability and permittivity of the line, respectively (these can also be interpreted as the inductance and capacitance of the line per unit length). This relation is also approximately true for quasi-TEM lines such microstrip line. In such cases, one thus gets $\beta = \frac{\omega}{c}$, where c is the propagation velocity along the line. Substituting this in (4.4), one gets

$$Z_{in} = \jmath\omega \left\{ \frac{Z_0 l}{c} \right\} = \jmath\omega L \tag{4.5}$$

Thus the short length of short circuited transmission line behaves like an inductor of inductance L given by

$$L = \frac{Z_0 l}{c} \tag{4.6}$$

A similar argument, applied to short lengths of open circuited transmission line, shows that they behave as capacitors of capacitance C given by

$$C = \frac{1}{Z_0 c} \tag{4.7}$$

Short lengths of open and short circuited transmission line thus behave as lumped inductors and capacitors. They can be used in matching networks. They can also be used to load antennas in order to reduce their physical size while maintaining the same operating frequency. For example, placing open circuited stubs (which look like lumped capacitors) at the ends of a linear dipole results in a capacitively loaded dipole which is shorter than a regular dipole for the same resonant frequency, but has a higher Q (lower bandwidth). This falls in line with what one expects - smaller antennas have higher Q's.

Loops

Small loops are the electromagnetic duals of small dipoles. Most of their near field energy is magnetic and thus their input impedance appears inductive. The problem with using small loops is that their radiation resistance is very low. This reduces their radiation efficiency, since loss sources begin to dominate the total resistance of the loop. The radiation resistance can be increased by N^2 by using N turns in parallel; however this is not practical for planar antennas. Resonances happen when p, the perimeter of the loop becomes comparable to a wavelength, for example at $p = \frac{\lambda}{2}$ and $p = \lambda$.

4.4.2 Final Design

In this section, I describe my final tag antenna design. As has been shown in Section 2.3.2, the impedance matching bandwidth of the chip to the antenna can be increased by increasing the order of the matching network. In other words, the ideal brick wall bandpass characteristic (see Figure 2-3) can be better approximated by higher order matching networks. A powerful method for synthesizing such networks for relatively narrow band (fractional bandwidth <15 %) applications is to used coupled resonator topologies [29].

Coupled Resonator Matching Networks

A general coupled resonator impedance matching network is shown in Figure 4-7. The idea is to use resonant circuits of a particular type - either series or parallel - and couple them using impedance inverters (gyrators) known as K and J inverters. There is no basic difference between them, but K-inverters are more suitable for coupling series resonant circuits, and J-inverters for parallel resonant ones. The impedance transformation between the two ports is defined by

$$\begin{aligned} Z_2 &= \frac{K^2}{Z_1} \\ Y_2 &= \frac{J^2}{Y_1} \end{aligned} \tag{4.8}$$

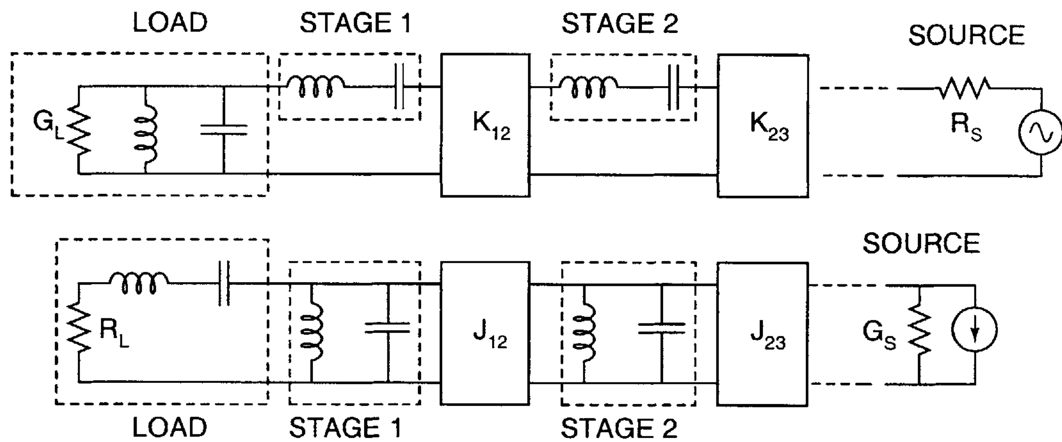


Figure 4-7: Coupled resonator impedance matching networks. The two circuits are duals of each other.

for K and J inverters, respectively. Thus impedance inverters transform parallel resonances to series resonances and vice versa. They can be distributed, like quarter wavelength transmission lines, or lumped. Several common lumped impedance inverters are shown in Figure 4-8. For example, the T-type inductive K-inverter has $K = \omega L$, and transforms a load Z_1 connected at terminal 1 to an impedance Z_2 looking into terminal 2, where

$$Z_2 = \frac{\omega^2 L^2}{Z_1} \tag{4.9}$$

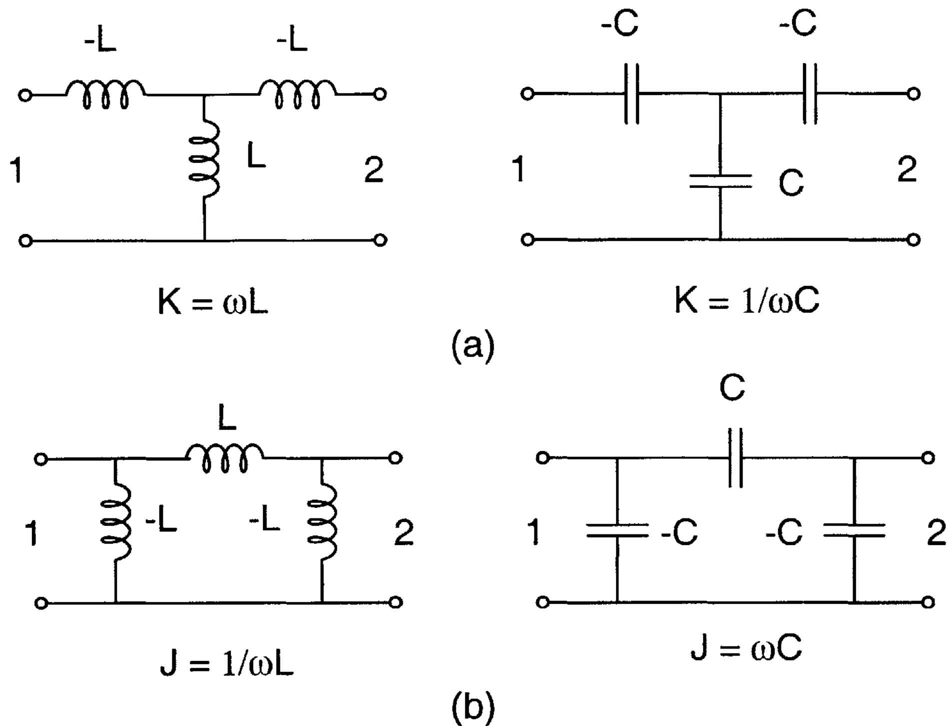


Figure 4-8: Examples of lumped impedance inverters (a) K-type and (b) J-type.

Negative L and C values in Figure 4-8 can be absorbed into the resonant circuits connected at both terminals. Coupling several resonators in this way splits their resonant frequencies and produces the bandpass response one is looking for. Equiripple, maximally flat and other responses can be generated by adjusting the coupling coefficients K_{ij} and J_{ij} appropriately.

Antenna Design

Applying the ideas from the previous section, I now try to create a broadband antenna for the tag. The chip impedance at frequencies of interest looks capacitive. It is thus represented as C_{in} in parallel with a resistance R_P. To create the first resonator, C_{in} is resonated out with a parallel inductor L_1 (this is easier than using a series inductor). This creates the first (parallel) resonant circuit. The antenna itself can now be made to act as a second resonant circuit. One can either use a series resonant antenna and couple it directly to the first resonator, or use a parallel resonant antenna and couple

it to the first resonator using a J-type impedance inverter.

Examples of series resonant antennas include half wavelength dipoles and full wavelength loops. Examples of parallel resonant antennas include full wavelength dipoles and half wavelength loops. I chose to go with a half wavelength loop because it is simple and area efficient. The reason the half-wavelength loop looks like a parallel resonator is that any current injected into one feed point comes back to the second feed point 180° out of phase and cancels it. The impedance seen between the two feed points thus looks like an open circuit, i.e., a parallel resonator. The finite Q of the resonator is because of radiation resistance (and other losses, like conduction losses). From a circuit viewpoint, the antenna thus looks like C_2 in parallel with L_2 and a resistance R_2. One now couples the two resonators - the chip and the antenna - using a J-type impedance inverter to obtain a broadband impedance match. The resultant equivalent circuit is shown in Figure 4-9.

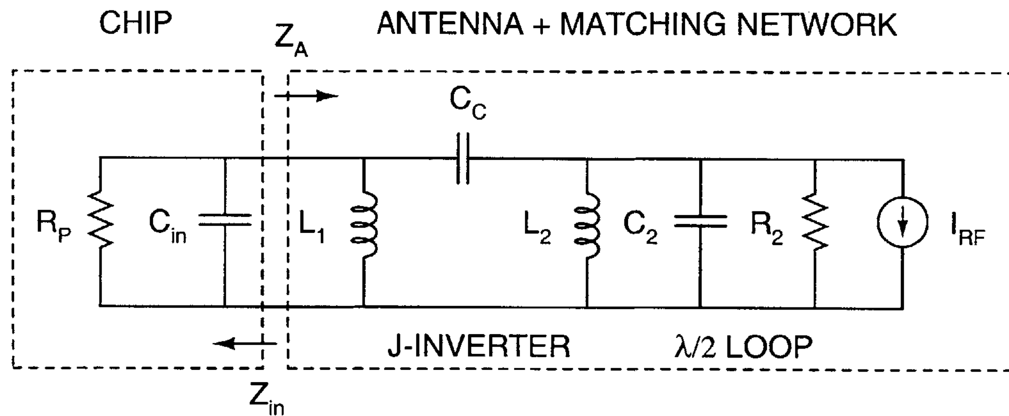


Figure 4-9: Equivalent circuit of the rectifier (chip), antenna and impedance matching network.

Figure 4-10 shows the physical structure of the tag antenna. The antenna is planar and uses a single metal layer on a printed circuit board. The small loop creates the inductance L_1 which resonates out the chip capacitance. The large loops have a perimeter equal to half a wavelength ($\frac{\lambda}{2}$) at the resonant frequency, and produce the parallel resonant circuit consisting of L_2, C_2 and R_2. The coupling capacitor C_C is produced by the transmission lines connecting the inner loop with the larger ones.

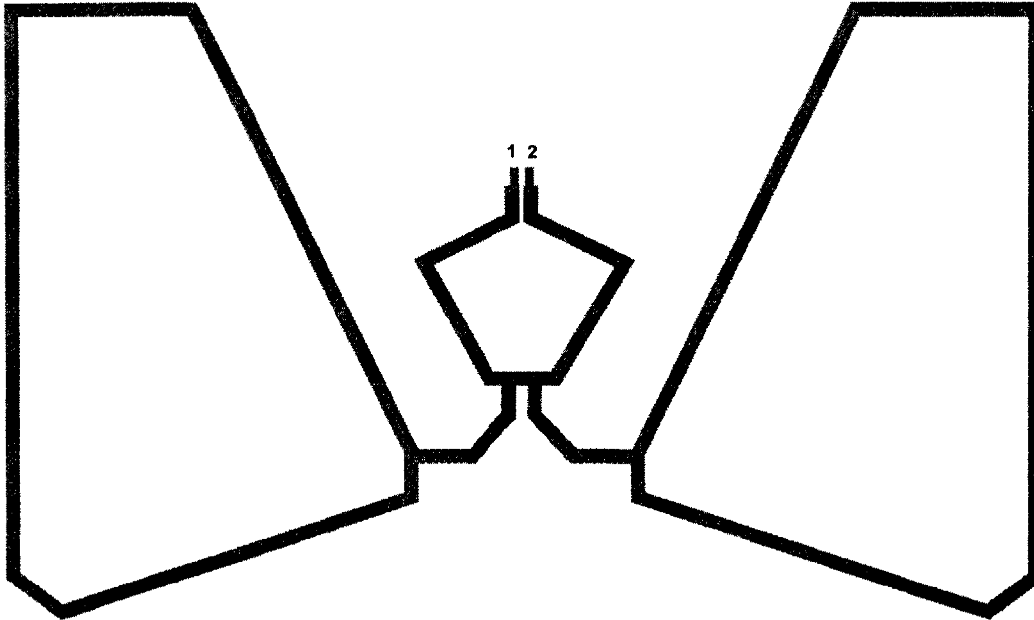


Figure 4-10: Physical structure of the final tag antenna. The rectifier chip is fed differentially from ports 1 and 2.

The short lengths of line look capacitive because the resonant antenna terminating it looks like an open circuit. Two resonant loops and two coupling lines are used for fully differential operation. The line of symmetry that splits the antenna into identical parts is a virtual ground. Thus the entire antenna can be regarded as the combination of two single ended antennas back to back. This is similar to the parallel strips line - microstrip line conversion technique shown in Figure 4-4. The antenna feeds the rectifier chip differentially (between terminals 1 and 2) though a short length of coplanar strips line. The total size of the antenna is about 2.8" × 1.7", and the simulated gain is 1.4 dBi at 950 MHz.

The structure shown in Figure 4-10 was simulated using ADS Momentum, a planar electromagnetic solver. Loop sizes and coupling line lengths were adjusted to match Z_A to Z_{in} around 900 MHz. Figure 4-11 shows the simulated input impedance $Z_A = R_A + jX_A$ that was obtained when the antenna was designed to match to a chip impedance Z_{in} consisting of $R_{in} = 7\ \Omega$ in series with a capacitance $C_{in} = 1$ pF

($R_{in} = \frac{R_P}{Q^2+1}$, where Q represents the quality factor of the load). This gives $Z_{in} = 7 - j177\ \Omega$ at 900 MHz, which corresponds to the measured input impedance of the three stage rectifier described in Section 3.6.2 when encapsulated in a Leadless Chip Carrier (LCC) package. The R_A and X_A curves in Figure 4-11 can be viewed as the result of damping out the antiresonance (parallel resonance) of the $\frac{\lambda}{2}$ loop by coupling it with another resonant circuit (the chip and L_1 in Figure 4-9).

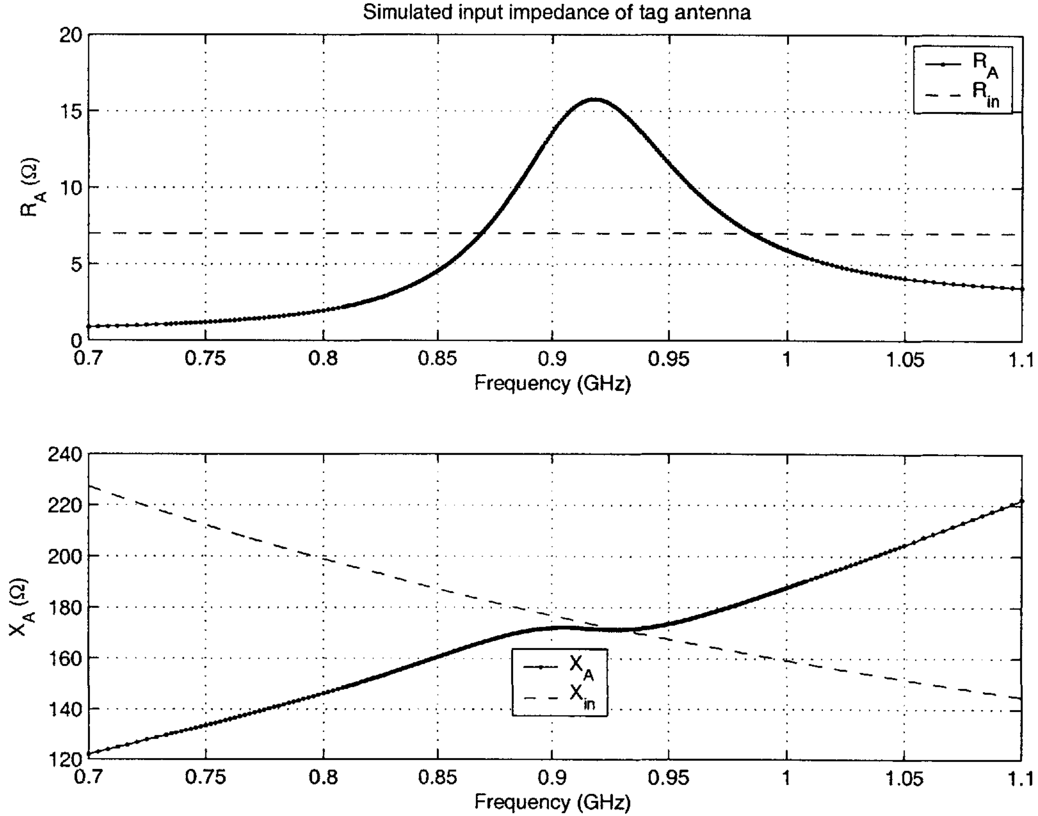


Figure 4-11: Simulated input impedance $R_A + jX_A$ of the tag antenna with matching network, matched to $R_{in} - jX_{in}$, the input impedance of the chip.

The reflection coefficient magnitude $|S_{11}|$ seen by the antenna looking into the load (Z_{in} of the rectifier chip) is shown in Figure 4-12 for three different values of R_{in}. $S_{11} = -10$ dB corresponds to 90 % of available power being absorbed by the rectifier. The second order behavior of the matching network can be seen from the figure. Investigating the behavior of the antenna with different values of load resis-

tance is important because R_{in} is very difficult to measure accurately for high Q loads like Z_{in} (which in this case has a Q of about 25 at 900 MHz). As expected, the resonant frequency in Figure 4-12 remains relatively unchanged as R_{in} varies. However, impedance matching improves (i.e., $|S_{11}|$ remains low over a larger bandwidth) as R_{in} increases and the quality factor of the load decreases. This is what one expects based on the Bode-Fano criterion. However, impedance matching only improves up to a point as R_{in} increases. The reflection coefficient begins to increase (i.e., more power starts getting reflected) as R_{in} is increased above the peak value of R_A, which is about 15 Ω. The impedance matching network was not designed to operate in this regime. Thus one sees that the antenna will perform well over a bandwidth of 50 MHz or more when R_{in} is in the range 7-15 Ω. The actual value of R_{in} in this case is probably about 6 - 7 Ω.

4.5 Antenna Tuning Networks

There is one basic chicken-and-egg problem that bedevils any active, power consuming schemes for improving the performance of a self powered system. Granted that the system is not good enough and needs to be improved in some way, what provides the power for the auxiliary system that improves it? An auxiliary to the auxiliary system? Clearly there is an infinite regression involved. There does not appear to be an easy solution to this problem.

In this section I briefly describe some work on improving the performance of a power extraction system by automatically tuning the resonant frequency of its antenna to changing environmental conditions. Most of this work is still at a very preliminary stage and exists only as simulations. The very real chicken-and-egg problem of how to power up the tuning circuitry in the absence of a rectified power supply has also not been solved. Solar power, in the form of CMOS solar cells, was considered, but experiments show that in standard CMOS, solar cells are subject to severe practical problems (see Appendix A.2).

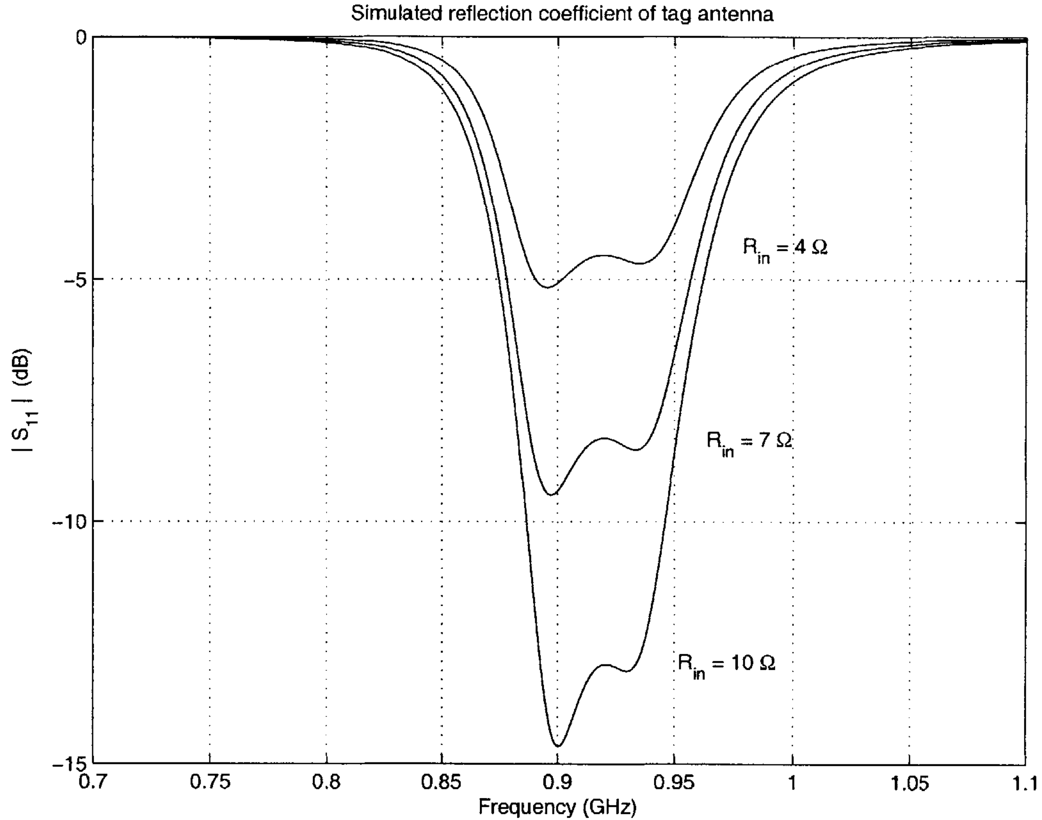


Figure 4-12: Simulated reflection coefficient magnitude $|S_{11}|$ of the tag antenna when connected to the chip input impedance for $R_{in} = 4$, 7 and 10 Ω.

4.5.1 Algorithm

The output voltage V_L of the tag as a function of frequency roughly tracks the resonance curve of the antenna (for example, Fig 2-17), since the performance of the rectifier itself is largely frequency independent. The objective of designing an antenna tuning network is to adaptively control the antenna resonant frequency so that the maximum value of $V_L(\omega)$ always occurs at the operating frequency ω_0, thus maximizing the performance of the power extraction system. This amounts to a function maximization problem, where V_L is the objective function to be maximized. Many function maximization algorithms exist [17]. To minimize the power consumption of the adaptation circuitry, one wants to implement as simple an algorithm as possible.

A particularly simple discrete time hill climbing algorithm is given by the control

law

$$\Delta\omega_{res,n+1} = \text{sgn}\left(\frac{dV_L}{dt}\right) \oplus \Delta\omega_{res,n} \tag{4.10}$$

where $\Delta\omega_{res,n+1}$ denotes the change in the antenna resonant frequency ω_{res} at the $n+1$-th timestep, sgn denotes the signum function and $\oplus$ is the logical exclusive-or (XOR) operarion. This control law makes the same change to the resonant frequency as made at the last time step if the output voltage V_L is increasing. Otherwise it reverses the previous change. This constitutes a hill climbing strategy that should eventually find the value of ω_{res} that maximizes V_L. The situation is complicated by the presence of local maxima in the antenna resonance curve.

Several enhancements can be made to this algorithm. The step size $\Delta\omega_{res,n+1}$ can be made adaptive, for example proportional to $\frac{dV_L}{dt}$. This decreases the step size as one approaches the maximum and increases both the stability and convergence time of the algorithm. The algorithm also tends to get stuck in local maxima. To avoid this, a certain amount of random frequency jitter can be introduced into the control law. Of course, this only works if the jitter amplitude is significantly larger than the width of the local maximum. Such large values of jitter may not be desirable for other reasons, such as stability and noise considerations.

Circuit implementations of this algorithm are being developed. ω_{res} can be controlled with a MOS varactor C_V, which unfortunately adds to the total input capacitance C_s and degrades rectifier performance. There is a tradeoff between the amount of performance degradation and the frequency tuning range, both of which go up as C_V increases. The rest of the circuitry can be implemented using low power analog and digital design techniques. The next section describes how one basic element, a reference circuit, can be built.

4.5.2 Current Reference

A current or voltage reference is useful for any analog circuit. The reference circuit produces a fixed current I_{ref} or voltage V_{ref} which can then be scaled appropriately

to DC bias the rest of the system. I_{ref} and V_{ref} should have low noise and be stable to changes in power supply voltage and temperature. Large decoupling capacitors are usually needed between V_{ref} and the power supply rails V_{DD} and ground to reduce high frequency noise. In addition, current and voltage references are equivalent, since a single transistor can be used to convert between I_{ref} and V_{ref}. A good reference for a low power adaptation circuit such as an antenna controller should consume very little power. Since the power consumption of the reference $P_{ref} \propto I_{ref} V_{DD}$, this means a low value of I_{ref}.

A well known constant-g_m CMOS current reference circuit is shown in Figure 4-13. Ignore the startup circuit shown in gray (transistors $M5$ and $M6$ and capacitor C_S) for now. $M1$ and $M2$ have the same aspect ratio, while the aspect ratio of $M4$ is m times larger than that of $M3$. C_P and C_N are used to decouple the reference voltages V_P and V_N which are fed to the rest of the system. In subthreshold operation, the current through each branch of the circuit is given by

$$I_{ref} = \frac{\phi_T}{\kappa R} \ln(m) \tag{4.11}$$

where $\phi_T = \frac{kT}{q}$ is the thermal voltage, κ is the NMOS subthreshold slope parameter and $m > 1$ is the ratio of the aspect ratios of $M4$ and $M3$, i.e., $\left(\frac{W}{L}\right)_4 = m\left(\frac{W}{L}\right)_3$. Ideally, I_{ref} is independent of V_{DD}, i.e., the circuit has infinite output resistance. To get low values of I_{ref}, one can increase R or decrease m. Since $\phi_T \propto T$, the absolute temperature, I_{ref} is also proportional to T. This makes I_{ref} PTAT (Proportional To Absolute Temperature).

The main advantages of this circuit are simplicity, power supply independence (self-biased) and the fact that no external components are required. The circuit also has several disadvantages, which are summarized below.

1. To get small values of I_{ref}, large values of R are required. Large integrated resistors occupy chip area and have absolute values which match poorly, increasing the variability in I_{ref} across chips and wafers.

2. The circuit has two possible operating states. The stable state has I_{ref} given

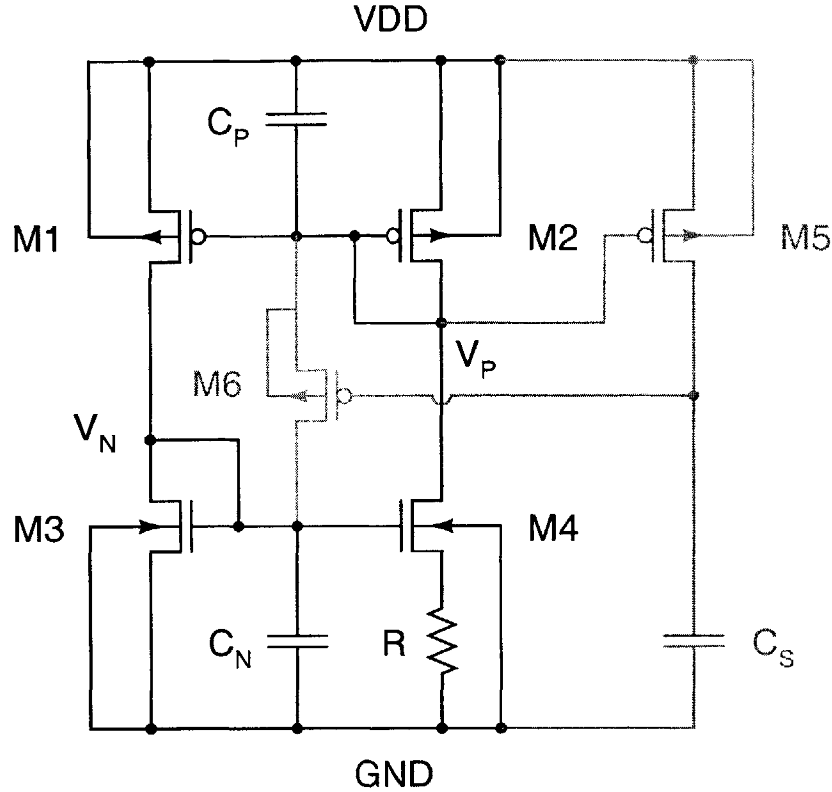


Figure 4-13: CMOS current reference. The startup circuit is shown in gray.

by (4.11). The other state is $I_{ref} = 0$, which is metastable. The presence of this state means that a startup circuit is often needed to speed transitions out of it.

3. When V_{DD} is switched on, the circuit starts is in the $I_{ref} = 0$ state. Leakage currents cause V_P to fall and V_N to rise as C_P and C_N charge. This causes I_{ref} to rise exponentially with time and eventually saturate to (4.11). This leads to a finite startup time T_{start} for the circuit.

4. T_{start} increases dramatically at low power supply voltages and large values of R, because the leakage currents through the transistors decrease. Increases in C_P and C_N also increase T_{start}.

5. (4.11) assumes that all transistors are saturated. This sets $V_{DD,min}$, the minimum usable value of V_{DD}.

6. The finite output resistance of the transistors causes I_{ref} to increase slowly with V_{DD} for $V_{DD} > V_{DD,min}$. Cascodes can be used to increase the output resistance, but only at the cost of higher $V_{DD,min}$.

7. I_{ref} is not constant with temperature. In fact, it is PTAT.

I want a low voltage, supply independent low I_{ref} reference for adaptation circuitry which starts up quickly (so that the tag read rate is not reduced). To reduce T_{start}, I designed a novel startup circuit. Refer to Figure 4-13. When V_{DD} is switched on, all capacitors in the circuit are discharged, so $V_P = V_{DD}$, $V_N = 0$ and the gate of $M6$ is at ground. This means $V_{SG} = V_{DD}$ for $M6$. A large current thus flows through $M6$, reducing V_P and increasing V_N and 'jump starting' the current reference. As V_P decreases, current starts to flow through $M5$, charging up C_S and decreasing the V_{SG} of $M6$. Eventually, the gate of $M6$ reaches V_{DD} and $M6$ turns off. Since $M5$ does not carry any static current, the startup circuit dissipates no power once the reference is on. It is also supply voltage independent, i.e., works at all values of V_{DD}. Both these properties contrast favorably with conventional startup circuits, which consume static power and are designed to work for particular values of V_{DD}. The startup time can be decreased further if required by increasing the aspect ratios of $M5$ and $M6$.

Figure 4-14 shows measured values (at room temperature) of I_{ref} versus V_{DD} for the circuit in Figure 4-13 with $m = 1.5$ and $R \approx 8$ MΩ for five different chips on the same wafer. The nominal value of I_{ref} one expects from (4.11) is 1.9 nA. Experimental current levels are somewhat lower (see Figure 4-14), probably because R was higher than expected. The fairly wide spread in absolute current values (about ± 30 %) is due to variability in R. $V_{DD,min} \approx 0.9$ V, and is set by $V_{SG,p} + V_{DS,sat}$, where $V_{SG,p}$ is the source-to-gate voltage of the PMOS transistors (about 0.7 V at this current level), and $V_{DS,sat} \approx 0.1$ V is the saturation voltage of the NMOS transistors.

Figure 4-15 shows the measured startup time of the current reference with $C_P = C_N = C_S = 10$ pF as a function of the power supply voltage V_{DD}. T_{start} is exponential in V_{DD}. The inverse of the slope is approximately 160 mV/decade decrease in T_{start}. This corresponds to $\frac{2\phi_T}{\kappa}$ mV/decade, with $\phi_T = 25.9$ mV and $\kappa \approx 0.74$. It can be

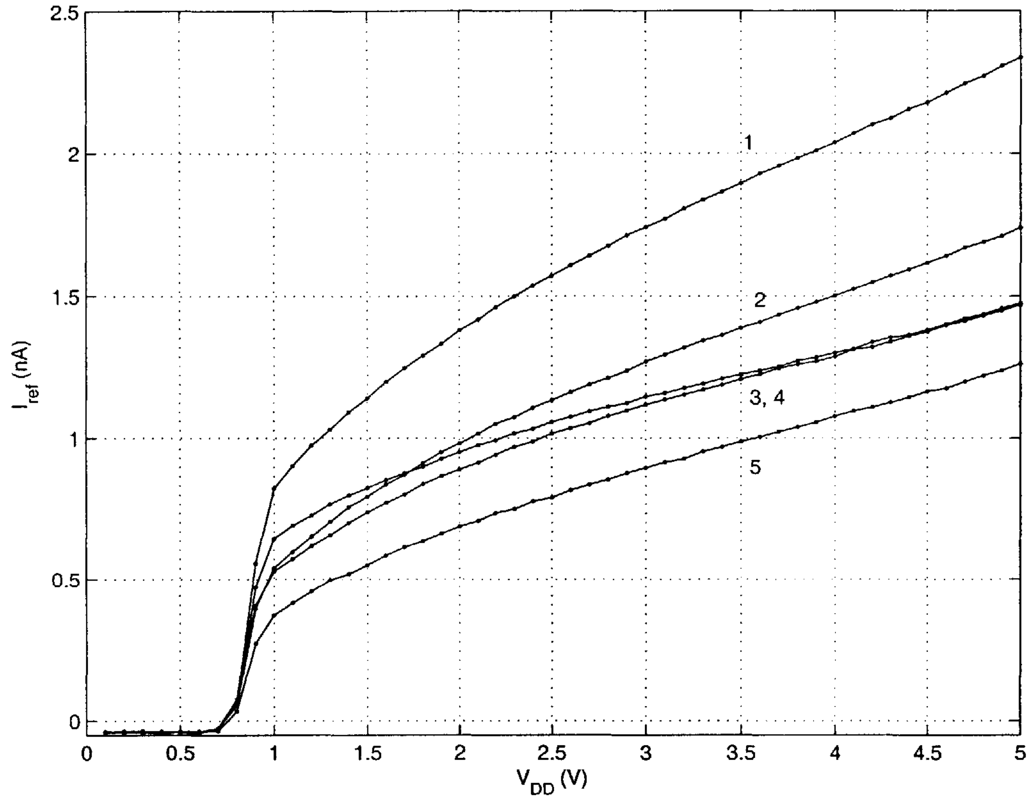


Figure 4-14: Measured output current I_{ref} of the current reference versus power supply voltage V_{DD}. Results from five different chips are shown.

shown that $T_{start} \propto R_{DS,6}$, where $R_{DS,6}$ is the series resistances of $M6$. $R_{DS,6}$ is proportional to $\exp\left(-\frac{\kappa V_{SG,p}}{\phi_T}\right)$. The observed dependence of T_{start} on V_{DD} is easy to explain if one remembers that $V_{SG,p} = V_{DD} - V_P$. This means that $R_{DS,6}$ decreases exponentially with V_{DD}, leading to the exponential decrease of T_{start} with V_{DD}. In addition, measurements of I_{ref} as a function of temperature confirm PTAT behavior.

Measurements of such low values of current can only be made with a highly accurate multimeter. Available multimeters are limited to a maximum sampling frequency of about 100 Hz and a timing resolution of 10 ms in this current range. Thus I was not able to take T_{start} data for $V_{DD} > 1.35$ V because it decreased to values too low to be measured (less than 10 ms). However, on extrapolating the measurements made at lower values of V_{DD} to $V_{DD} = 1.5$ V, one sees that the startup time should be about 2 ms. This is quite a few orders of magnitude faster than without a startup circuit,

and allows this reference to be used for adaptation circuitry even at low power supply voltages.

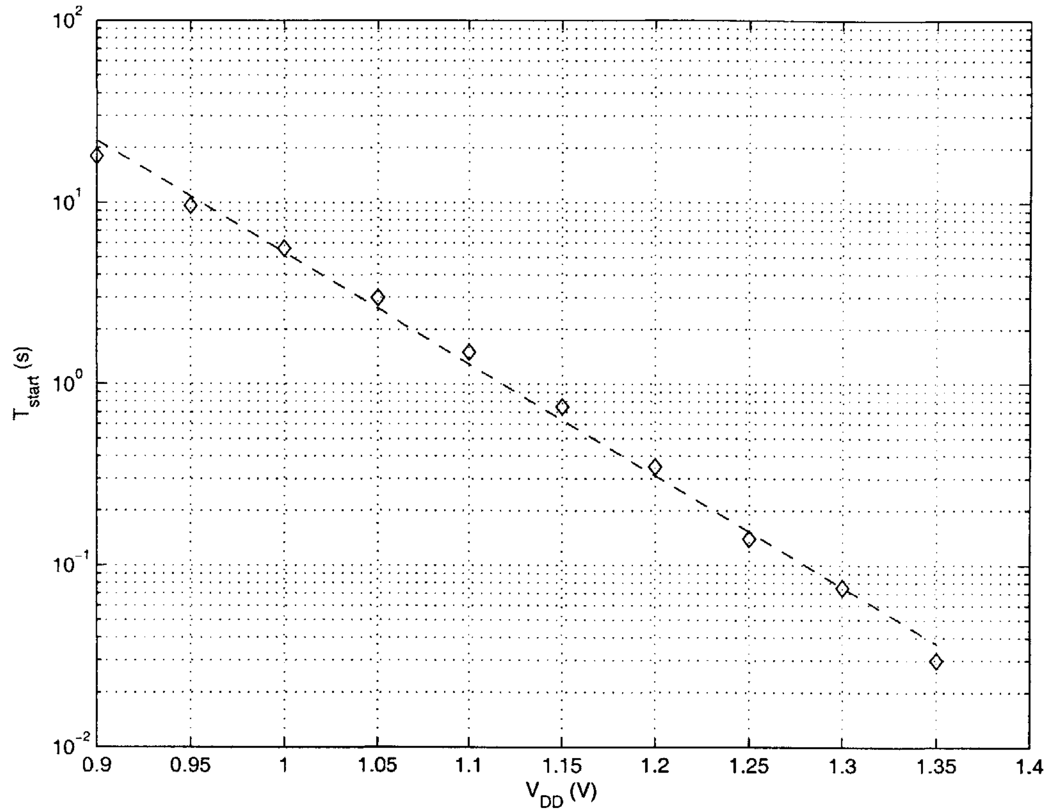


Figure 4-15: Measured startup time T_{start} of the current reference versus power supply voltage V_{DD}.

Chapter 5

Summary and Conclusions

5.1 Summary

In this section, I summarize the key contributions of this thesis.

- I have evaluated the fundamental upper bounds on received input voltage amplitude for a reactive load which forms part of a radio link.

- I have used these bounds to theoretically evaluate maximum power up range and efficiency for generic power extraction systems operating over a given bandwidth.

- Various rectifier structures have been devised for power extraction, including threshold voltage programmable floating gate rectifiers.

- The effects of device threshold voltage (dead zone) on rectifier performance have been simulated, explained theoretically and compared with experiment.

- Layout optimization has been treated quantitatively. The effects of several commonly used circuit layout 'tricks' have been evaluated mathematically.

- An innovative compact, broadband antenna has been developed for power extraction applications. Its performance has been verified experimentally.

- Innovative Schottky diode structures have been created in a standard CMOS process. These diodes have been used to create rectifiers.

- The ideas described above were combined in order to create a working RF power extraction system which was measured to have an available power up threshold of 35 μW and a bandwidth of 30 MHz. The read range of this system was 7.5 meters when the transmitted power was 4 W (EIRP).

5.2 Future Work

This thesis has dealt with the extraction of power from RF signals at low power levels. However, this is but one side of the coin. The usage of the extracted power is equally important. In effect, they form two parts of one system and the design of one cannot proceed independently of the other. Throughout this thesis, I have modeled the load, i.e., the circuitry being powered by the extracted power, as a current source with a certain minimum voltage requirement. However, both the value of this current source and its voltage requirements are adjustable through proper circuit design. I have started to address some of these issues, but much work remains to be done. Some of my immediate and future plans in this respect are summarized below.

- Several of the key optimizations described in this thesis were not implemented on the tag which was tested, since it was an early version. I hope to improve the power up threshold by at least a factor of 2 by using all the optimization strategies developed so far in a new power extraction system.

- The long term reliability of floating gate rectifiers remains to be evaluated. There is some evidence for very slow dynamics inside my floating gate structures which needs more investigation.

- I want to carry out experiments inside an anechoic chamber in order to more accurately characterize antenna gains, radiation patterns and tag power up ranges.

- I want to implement the whole power extraction system on a flexible substrate in order to investigate the effects of physical deformations on system performance.
- My work on antenna tuning networks is still very incomplete and needs to be extended further.
- A theoretical explanation for the I-V curves of the Schottky diodes developed as part of this project is still incomplete.
- I have not yet implemented backscatter modulation on the tag. Some preliminary studies indicate that using MOS varactors and PSK might be most efficient from a power collection viewpoint.
- Perhaps most importantly, I want to design the rest of a functional RFID tag using novel, power efficient circuit design techniques. The knowledge thus gained should also be useful for other low power system applications.

Appendix A

Non-standard CMOS Devices

This appendix describes some of my work aimed at creating non-standard (not foundry supported) devices in a standard CMOS process. The disadvantages of creating one's own device from scratch are obvious. There are no available simulation models, performance is usually suboptimal since the fabrication process was not designed for the device, and device matching may be poor. I investigate some of these aspects in the following sections.

A.1 Schottky Diodes

In this section, I describe my attempt to create metal-semiconductor barrier diodes (Schottky diodes) in AMI's 0.5 μm CMOS process. Access to this process was obtained through the MOSIS fabrication service.

The electron energy band diagrams of rectifying metal - semiconductor junctions at equilibrium (no applied bias) are shown in Figures A-1 (n-type semiconductor) and A-2) (p-type semiconductor). In the figures, E_0 is the vacuum energy, E_F is the Fermi Level, W_M and W_S are the work functions of the metal and semiconductor, respectively, χ is the electron affinity of the semiconductor, E_C and E_V are the semiconductor conduction and valence band energies, ϕ_{bi} is the built up potential at the junction and V_{SBn} and V_{SBp} are the Schottky barrier potentials of the junction. These are characteristic of a particular metal-semiconductor pair, but are independent of

the doping levels in the semiconductor [11].

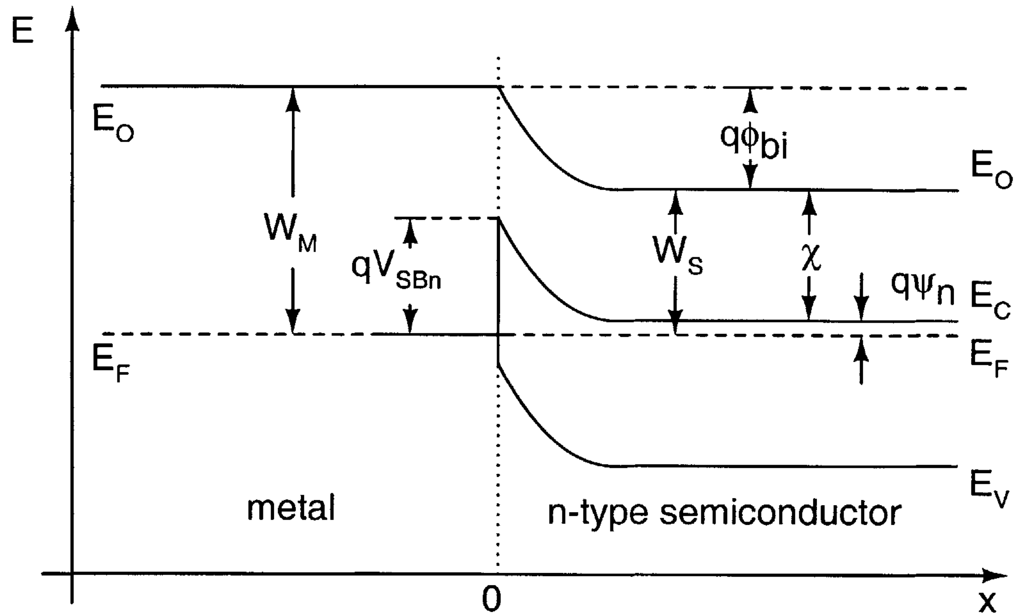


Figure A-1: Energy band diagram of a metal / n-type semiconductor junction at equilibium.

The polarity of Schottky diodes formed by rectifying metal-semiconductor junctions is easy to remember. The semiconductor doping type determines the diode terminal polarity. N-type semiconductor is negative, and p-type semiconductor positive. The polarity of the second (metal) terminal is set accordingly.

In most integrated circuit applications, one wants metal-semiconductor contacts that have Ohmic and not rectifying I-V characteristics. The most effective way to get Ohmic contacts is to heavily dope (degenerate) the semiconductor near the contact interface so that the Fermi level enters the conduction or valence bands. The depletion region in the semiconductor then becomes thin enough to allow electrons to tunnel bidirectionally through the V_{SB} potential barrier (see Appendix B for an account of the mathematics involved). To produce Schottky diodes in a process which does not support them, one can remove the heavily doped region below an Ohmic contact. I describe how to do this in the next section.

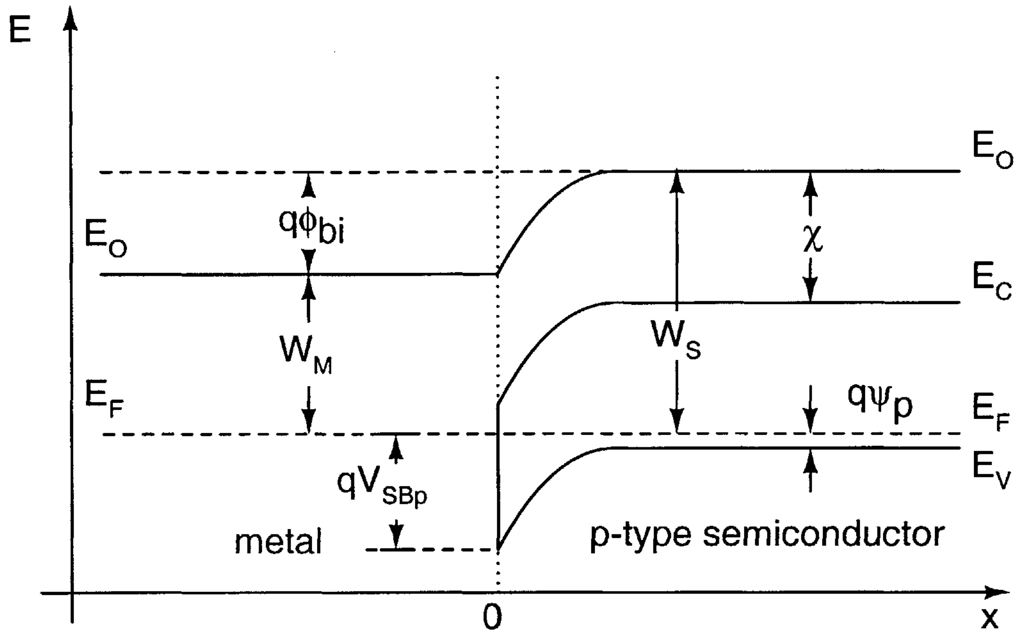


Figure A-2: Energy band diagram of a metal/ p-type semiconductor junction at equilibrium.

A.1.1 Type 1: metal to n-well

Contact vias connect the lowest metal layer (metal 1) with lower layers like substrate, wells and polysilicon. They are usually made of a metal-silicon alloy to reduce series resistance and create a good interface with the contacted layer. One wants to create floating (this means that both diode terminals should be controllable) Schottky diodes using contact vias. Probably the only way to do this in a standard single-well CMOS process which does not support Schottky diodes is to use well contacts. This is because the doping levels typically present in wells, $\approx 10^{16}$ cm^{-3}, is low enough to allow rectifying (Schottky) contacts to be formed with metal layers. Normally, of course, one does not want rectifying contacts, but Ohmic ones. To get low resistance Ohmic contacts, the area under each contact is therefore heavily doped ($\approx 10^{20}$ cm^{-3}). To mask layer that controls this doping is typically called SELECT. To get Schottky diodes, therefore, one places contact vias in regions of the well without the SELECT mask layer. This is usually a violation of the process design rules.

See Figure A-3 for a cross sectional view of the n-well case. Contacts over a heavily doped n^+ layer are Ohmic, the others are rectifying. Each contact size is fixed by the process rules and cannot be changed. To get diodes with large junction areas, many contacts are connected in parallel. The diode is fingered to reduce the path length for current flow between the diode terminals. This helps to reduce the overall series resistance of the diode. The layout of a diode with 10 contacts is shown in Figure A-4. The well diffusion happens first. The ACTIVE layer prevents growth of thick field oxide. SELECT layers are placed over the ACTIVE layer in order to heavily dope it (creating a n^+ region for N-SELECT and a p^+ region for P-SELECT). Any ACTIVE area not covered by N-SELECT remains at the background doping concentration of the well, and this is where Schottky contacts are created.

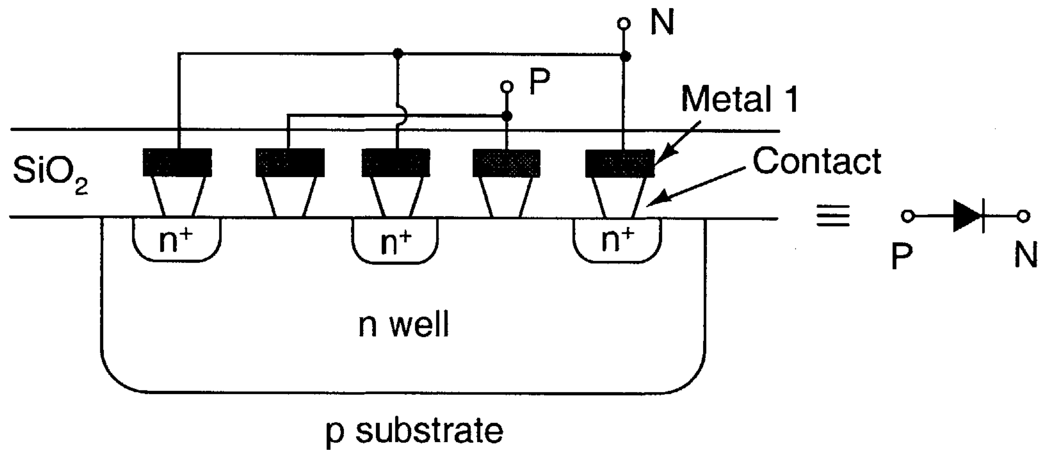


Figure A-3: Cross section of metal / n-well Schottky diode. The device is fingered in order to reduce series resistance.

Figure A-5 shows measured DC I-V curves for these devices. The curves are asymmetric, i.e., one sees definite rectifying behavior. The forward current ($V > 0$) follows a straight line on a logarithmic plot (except at high current levels, when the series resistance starts to dominate the current), as expected from the exponential I-V characteristic of a diode. The reverse current ($V < 0$) is not constant with reverse bias as in an ideal diode, but increases slowly with increasing reverse bias. Nevertheless, one can justifiably call these devices Schottky diodes. The main contributor to enhanced reverse current is electron tunneling across the Schottky barrier V_{SBn}.

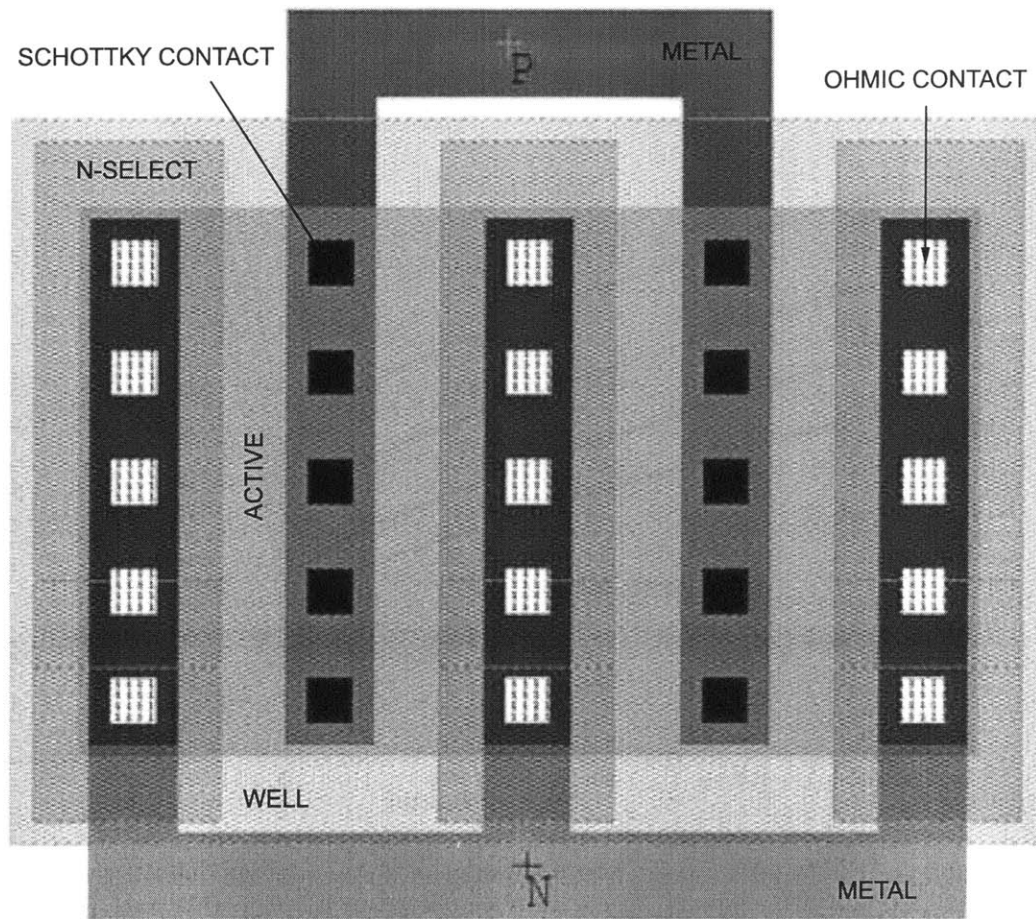


Figure A-4: Layout of a metal / n-well Schottky diode with 10 contacts.

A simple circuit model for these diodes is shown in Figure A-6(a). D represents the Schottky diode and R_S is its series resistance. D_W represents the n-well / p-substrate junction diode, and R_W is the series resistance of the well.

To qualitatively verify the model of Figure A-6(a), I left the well (terminal 2) floating, and varied V_P, the voltage on the metal (terminal 1). One expects that when $V_P < 0$, the well voltage will gradually become negative, turning D_W on and causing current to flow. Otherwise, one only expects the small reverse saturation current of D_W to flow. The results of this experiment are shown in Figure A-7 and are as expected. For $V_P > 0$, I hit the 10 pA noise floor of the measuring instrument, while for $V_P < 0$, current rises approximately exponentially with voltage. However,

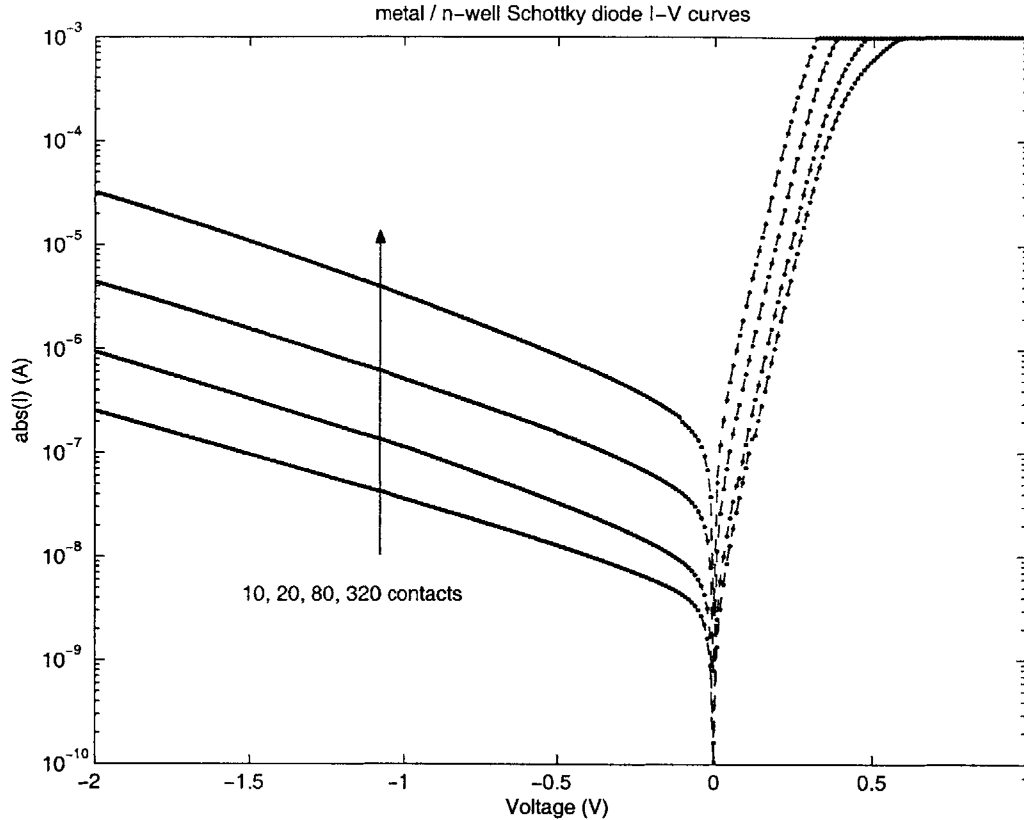


Figure A-5: Measured DC I-V characteristics of metal / n-well Schottky diodes with 10, 20, 80 and 320 contacts. Wells were connected to the chip substrate (ground).

the rate of rise is very slow. The non-ideality factor $n \approx 18$. This means that the well voltage V_W is only weakly dependent on $V_P < 0$. This is because the Schottky diode D is reverse biased and acting like a large resistor, creating a large potential division between V_P and V_W). The situation is complicated by the fact that the value of this potential divider will be voltage dependent, since the reverse current of D is not constant with bias.

To avoid this effect, all subsequent data on metal / n-well diodes was taken with the well fixed at ground or at a positive voltage w.r.t. ground. This ensures that the current through the reverse biased well / substrate diode D_W is negligible. Experimentally, it was found that the well voltage made no measurable difference to the I-V characteristics (i.e., only the voltage *difference* between the metal and the well

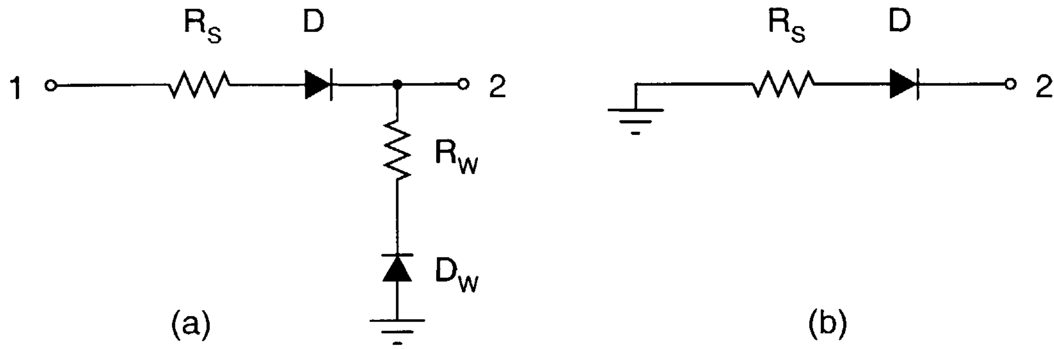


Figure A-6: Simple circuit models for integrated circuit diodes. (a) metal / n-well type and (b) metal / p-substrate type.

controls the current), as long as it was positive. This confirms one's expectations.

Finally, the effect of the minimum layout distance d between Schottky and Ohmic contacts on the I-V characteristics, especially on the reverse leakage current, was investigated. One possible cause for reverse leakage is physical proximity between the two contact types. If this is a factor, one expects reverse currents to decrease (and series resistance to increase) as the contacts are moved further apart. Experimental results are shown in Figure A-8 for diodes with 320 contacts each, but different values of d. One finds that the reverse current does decrease with increasing d. However, the effect is very small, and R_S does not change. In most cases, it should be safe to neglect it. This suggests that the current flow paths between Ohmic and Schottky contacts (which have lengths $\approx d$) have series resistances which are much smaller than that of the Schottky contacts themselves, i.e., the bulk resistance is much smaller than the contact resistance. In addition, these results are strong evidence for the reverse leakage current being dependent only on properties of the Schottky contacts. One concludes that, over the range of d shown, diode characteristics are essentially independent of d. To save layout space and minimize substrate capacitance, one nevertheless uses as small a value of d as possible.

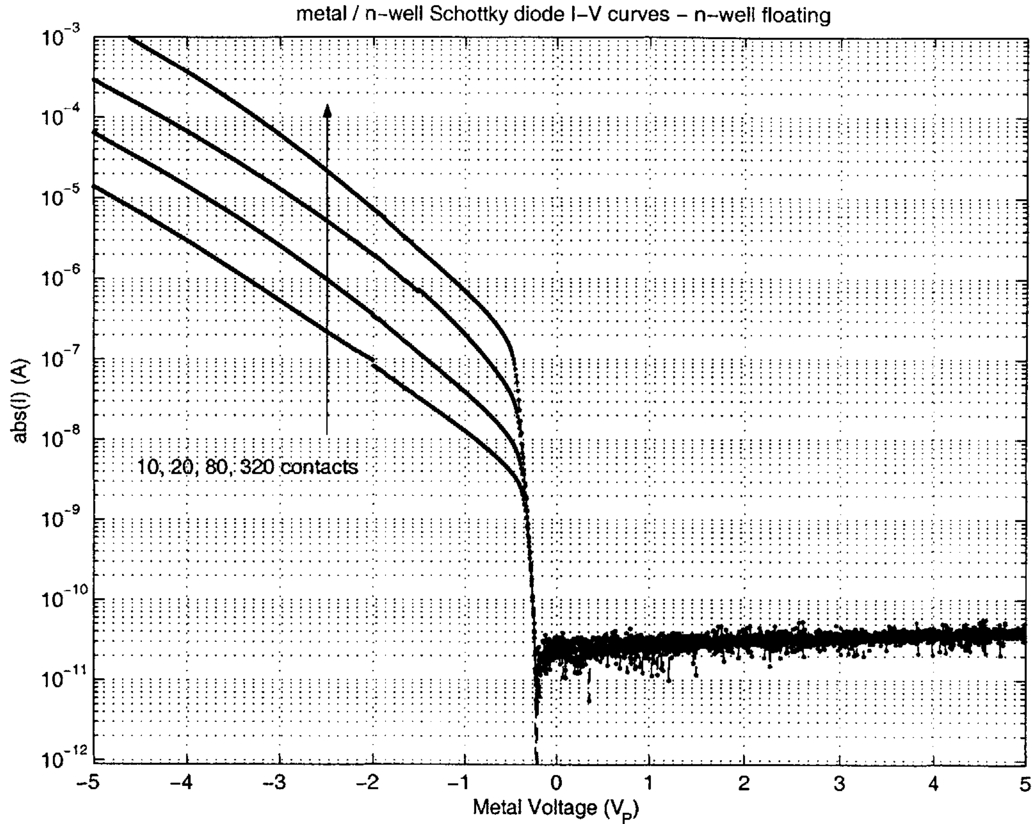


Figure A-7: Measured DC I-V characteristics of metal / n-well Schottky diodes with 10, 20, 80 and 320 contacts. Wells were unconnected (floating).

A.1.2 Type 2: metal to epitaxial p-substrate

An alternative way to create Schottky diodes is to use substrate contacts. This results in diodes which cannot be floated, since one terminal is always fixed at the substrate potential (usually taken to be ground). In addition, the substrate is usually heavily doped to ensure low resistance (precluding Schottky diode formation). However, in many modern CMOS processes, the heavily doped substrate is covered by a relatively lightly doped ($\approx 10^{15}$ cm^{-3}), thin ($\approx 3\mu$m thick) epitaxial layer having the same dopant type as the substrate. All active circuits are built in this epitaxial layer, which improves device isolation, controls transistor threshold voltages and also allows me to create Schottky diodes by using substrate contacts.

The cross section of a Schottky diode using substrate contacts (doped p-type) is

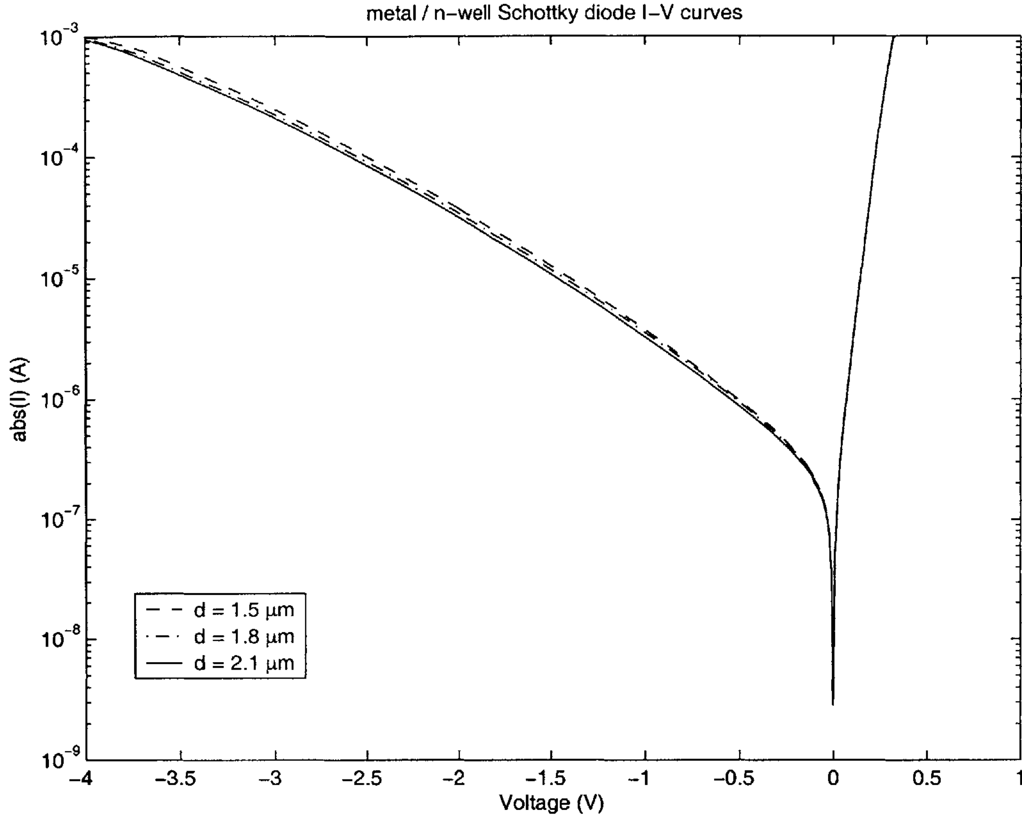


Figure A-8: Measured DC I-V characteristics of metal / n-well Schottky diodes. Each diode has 320 contacts but different values of d, the minimum spacing between Schottky and Ohmic well contacts.

shown in Figure A-9. As before, Ohmic contacts are formed by heavily doping (p^+) the area under the contact via. Schottky contacts are formed by not allowing this doping to take place (by leaving out the P-SELECT layer on top of an ACTIVE layer placed on the substrate). Layout is similar to the n-well case, with multiple fingers used to reduce the series resistance of the diode. Figure A-10 shows measured I-V characteristics of these devices. They are similar to those obtained from the metal/n-well structure (Figure A-5), again verifying the formation of Schottky diodes. However, for the same junction area, these devices have higher series resistance and lower reverse currents.

A simple circuit model for these diodes is shown in Figure A-6(b). D represents the Schottky diode and R_S is its series resistance. The grounded substrate is the

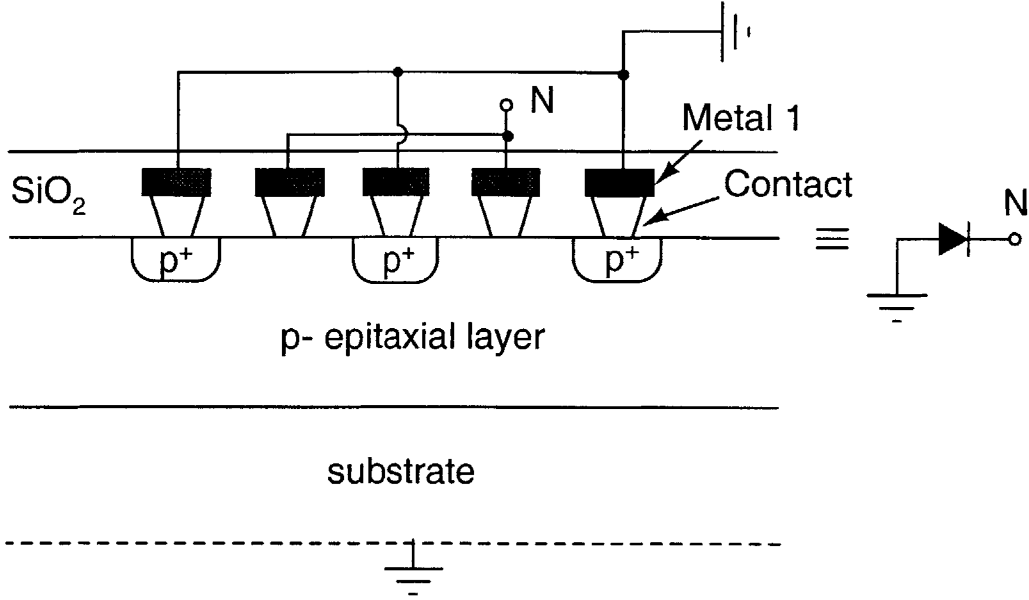


Figure A-9: Cross section of metal / p-epitaxial layer Schottky diode. The device is fingered in order to reduce series resistance.

positive terminal of the diode.

A.1.3 DC Modeling

In this section, I describe a simple model for predicting the DC behavior of Schottky diodes constructed in my process. All data was taken at room temperature (295 K). The current I through a Schottky diode when a voltage V is applied across it is given by the standard diode current equation

$$I = I_S \left[\exp\left(\frac{V}{n\phi_T}\right) - 1 \right] \tag{A.1}$$

where I_S is the reverse saturation current, ϕ_T is the thermal voltage and n is known as the non-ideality factor of the diode. The saturation current I_S of a Schottky diode is given by the thermionic emission equation [37]

$$I_S = NA_C A_{RC} K_{RC} T^2 \exp\left(\frac{V_{SB}}{\phi_T}\right) \tag{A.2}$$

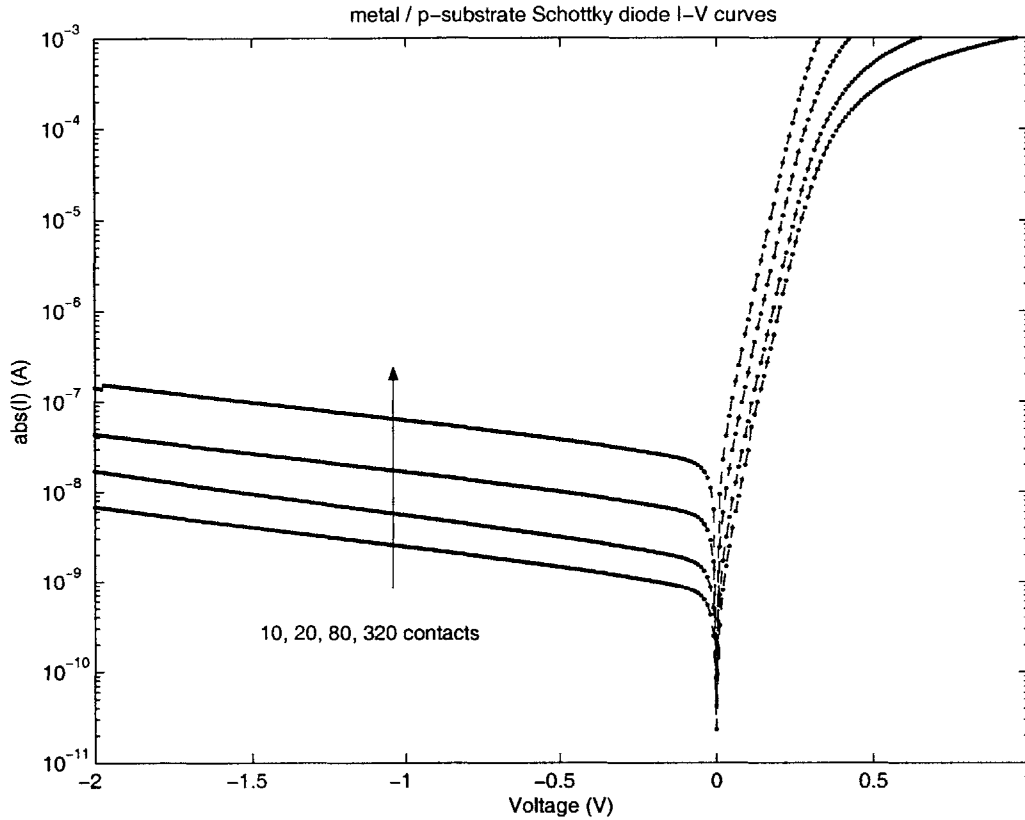


Figure A-10: Measured DC I-V characteristics of metal / p-substrate Schottky diodes with 10, 20, 80 and 320 contacts.

where A_C is the area of each rectifying metal-semiconductor contact, N is the number of such contacts, $A_{RC} = 1.2 \times 10^6$ $\mathrm{Am^{-2}K^{-2}}$ is known as Richardson's constant, K_{RC} is known as the Richardson's constant non-ideality parameter, T is the operating temperature, V_{SB} is the Schottky barrier potential and $\phi_T = \frac{kT}{q}$ is the thermal voltage. K_{RC} is characteristic of the semiconductor material and the type of carrier [11]. Typical values of K_{RC} are ≈ 0.66 for metal / n-type silicon junctions, and ≈ 2.15 for metal / p-type silicon junctions. Experimentally, one can measure I_S for given values of N and A_C. Thus one can experimentally find the value of V_{SB}. One finds that V_{SB} does not correspond to that for aluminum on silicon, but is closer to that for titanium. This may be due to the presence of a titanium 'barrier layer'. Such layers (normally composed of a Ti/TiN alloy) are deposited on the contact surfaces

to prevent silicon and aluminum from diffusing into each other and causing unwanted short circuits.

Any practical diode has associated with it a parasitic series resistance R_S. For Schottky diodes, R_S has three main components: the Schottky contact resistance, the bulk resistance and the Ohmic contact resistance [14]. The Schottky contact resistance arises because of the finite conductivity of the semiconducting layer below the rectifying junction. The bulk resistance arises because of current flow between the rectifying and Ohmic contacts through the semiconducting substrate. The Ohmic resistance is the resistance of the Ohmic contact itself and is usually negligible. It has been shown that for these diodes the bulk resistance is much smaller than the Schottky contact resistance. Thus, for a Schottky diode with N rectifying contacts in this process, $R_S \approx \frac{R_{SC}}{N}$, where R_{SC} is the series resistance of each Schottky contact.

One can now extract these parameters for both diode types by fitting the I-V data shown in Figures A-5 and A-10 to (A.1). The results are shown in Table A.1.3 and Figure A-11. The deviation of the non-ideality factor n from 1 for the n-well diodes implies that thermionic emission is not the sole contributor to current flow through the device. I obtain a good fit to experimental data for forward bias (except at high current levels, when the series resistance becomes significant), but the reverse currents of the real diodes are not constant and are thus not well modeled by the constant I_S of (A.1). Similar results are obtained for metal/ p-substrate diodes and are shown in Figure A-12.

Table A.1: Calculated Schottky diode parameters

Parameter	**Type 1 (metal/n-well)**	**Type 2(metal/p-substrate)**
I_S (per unit area)	1.75×10^3 A/m^2	2.60×10^3 A/m^2
I_S (per contact)	4.38×10^{-10} A	6.50×10^{-11} A
V_{SB}	0.48 V	0.50 V
n	1.3	1.04
R_S(per contact)	2.50 kΩ	6.25 kΩ
$C_j(0)$	0.58 fF/μm^2	0.20 fF/μm^2
$C_j(0)$(per contact)	0.21 fF	0.072 fF

To obtain a better estimate of the reverse leakage current, I add another diode

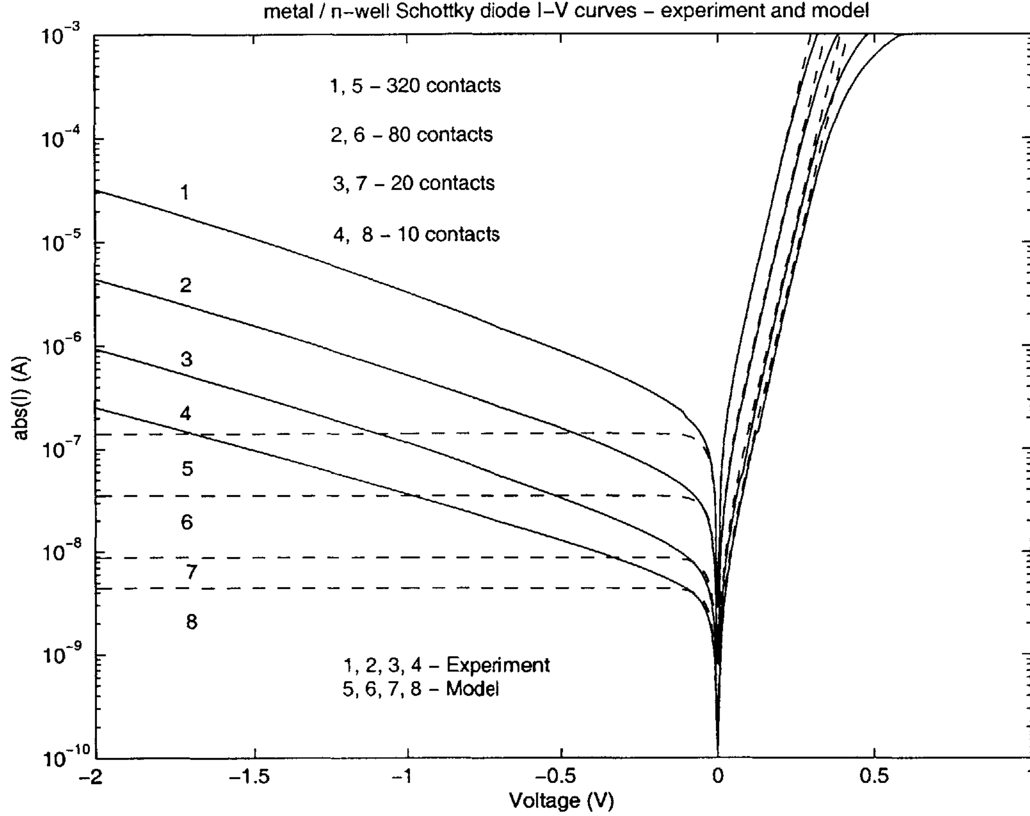


Figure A-11: Modeling DC I-V characteristics of metal / n-well Schottky diodes using a single diode (ignoring series resistance).

D_2 to the model. This is prompted by the fact that the reverse current looks almost exponential, i.e., approximately linear on a logarithmic plot. It can therefore be modeled by a diode. The original model consisted only of diode D_1, which now models the forward current. Its parameters are given in Table A.1.3. As shown in Figure A-13, D_2 is added in parallel and with opposite polarity to D_1. Assuming the voltage across the diodes is V, the currents I_{D1} and I_{D2} through D_1 and D_2, respectively, are given by

$$\begin{aligned} I_{D1} &= I_S\left(\exp\left(\frac{V}{n\phi_T}\right) - 1\right) \\ I_{D2} &= -\alpha(N) I_S\left(\exp\left(\frac{-V}{n_2\phi_T}\right) - 1\right) \end{aligned} \tag{A.3}$$

where the parameters for D_1 are given by Table A.1.3, as before, α_N is a parameter

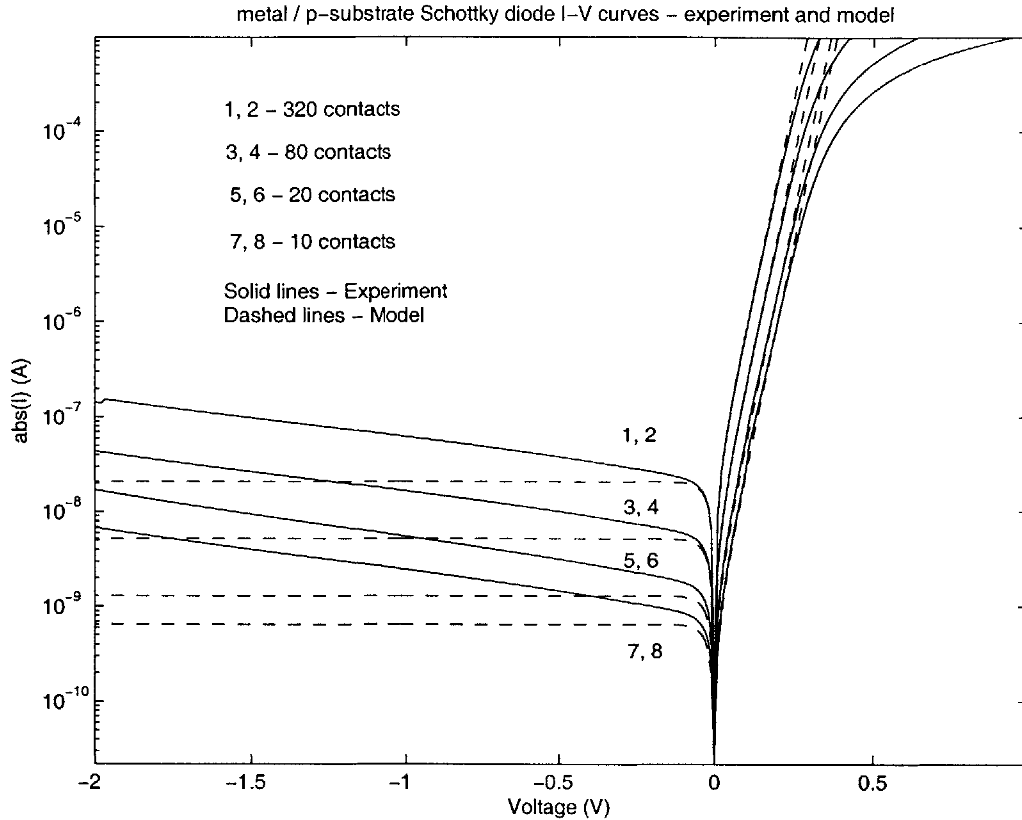


Figure A-12: Modeling DC I-V characteristics of metal / p-substrate Schottky diodes using a single diode (ignoring series resistance).

which depends on the size of the diode (N is the number of rectifying contacts) and n_2 is the non-ideality parameter of D_2. The total current predicted by the model is then $I = I_{D1} + I_{D2}$. This model predicts both forward and reverse bias currents fairly accurately, as shown in Figure A-14. One finds that typically $\alpha(1) \approx 1$, and that α scales approximately as $\sqrt{N}$. Thus bigger diodes have higher reverse leakage current densities than smaller ones. A typical value of n_2 is 18.

I can carry through the same exercise for metal / p-substrate diodes, and get similar results (see Figure A-15). For these diodes, $\alpha \approx 1$ is approximately constant with diode size, and $n_2 = 30$. Physically, the main reason for increased reverse current (as compared to the constant I_S predicted by the diode equation) and its dependence on reverse bias is electron tunneling across the Schottky barrier V_{SB} [41]. As the

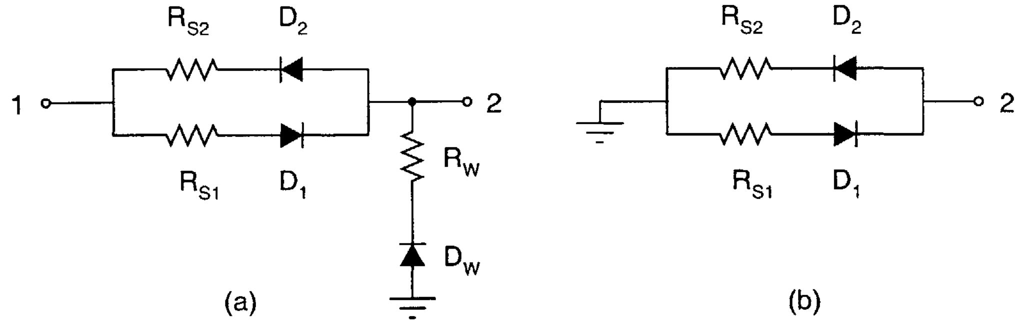


Figure A-13: Improved circuit models for integrated circuit Schottky diodes. (a) metal / n-well type and (b) metal / p-substrate type.

reverse bias V goes up, the electric field ε_{max} at the interface increases as

$$\varepsilon_{max}(V) = \varepsilon_{max}(0)\sqrt{1 - \frac{V}{\phi_{bi}}} \tag{A.4}$$

Thus the increase in ε_{max} with bias V is proportional to $\sqrt{V}$. Since the tunneling current density J_{tun} is exponentially sensitive to ε_{max} (Appendix B), this leads to the nearly exponential reverse currents seen on the experimental I-V plots. The absolute value of J_{tun} depends strongly on physical conditions at the interface, such as surface states and variations in doping levels, as well as on the barrier height V_{SB}.

A.1.4 RF Performance

Schottky diodes have been widely used as rectifiers for many years. As a result, their modeling at RF frequencies is well developed [14]. Approximate (ignoring the second diode in Figure A-13) zero bias RF models for integrated Schottky diodes in my process are shown in Figure A-16.

As discussed in Section 3.6.1, a Schottky diode has no diffusion capacitance because it is a majority carrier device. C_D is thus entirely due to the junction capacitance C_j of the metal-semiconductor junction, which may be analytically estimated as follows. The built in potential ϕ_{bi} at a metal-semiconductor junction is given by

$$q\phi_{bi} = qV_{SBn} - q\psi_n \tag{A.5}$$

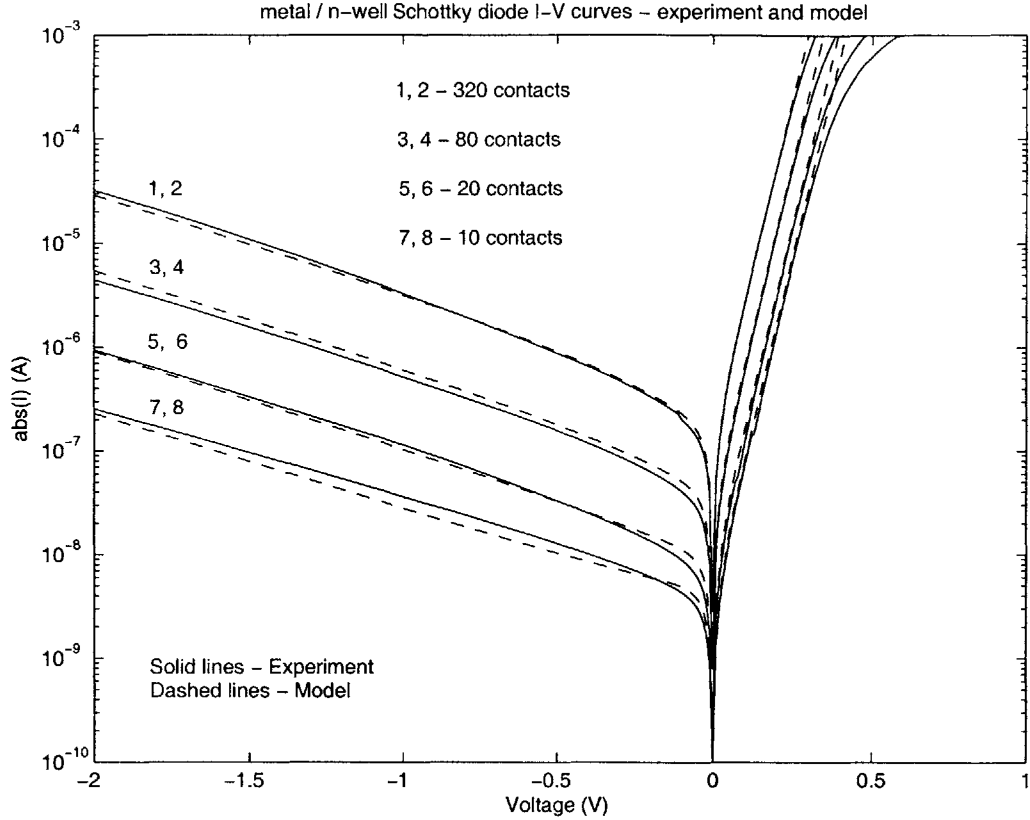


Figure A-14: Modeling DC I-V characteristics of metal / n-well Schottky diodes using two back to back diodes (ignoring series resistance).

for a n-type semiconductor, and

$$q\phi_{bi} = qV_{SBn} - q\psi_p \tag{A.6}$$

for a p-type semiconductor, where V_{SBn} and V_{SBp} are the respective Schottky barrier potentials. An approximation for the Schottky barrier potential for n-type semiconductors is $V_{SBn} = W_M - \chi$, i.e., V_{SBn} is approximately the difference in work functions between the metal and the semiconductor. In practice, this formula can lead to large errors and measured values of V_{SBn} (and V_{SBp}) are used instead. Also, ψ_n and ψ_p are defined as

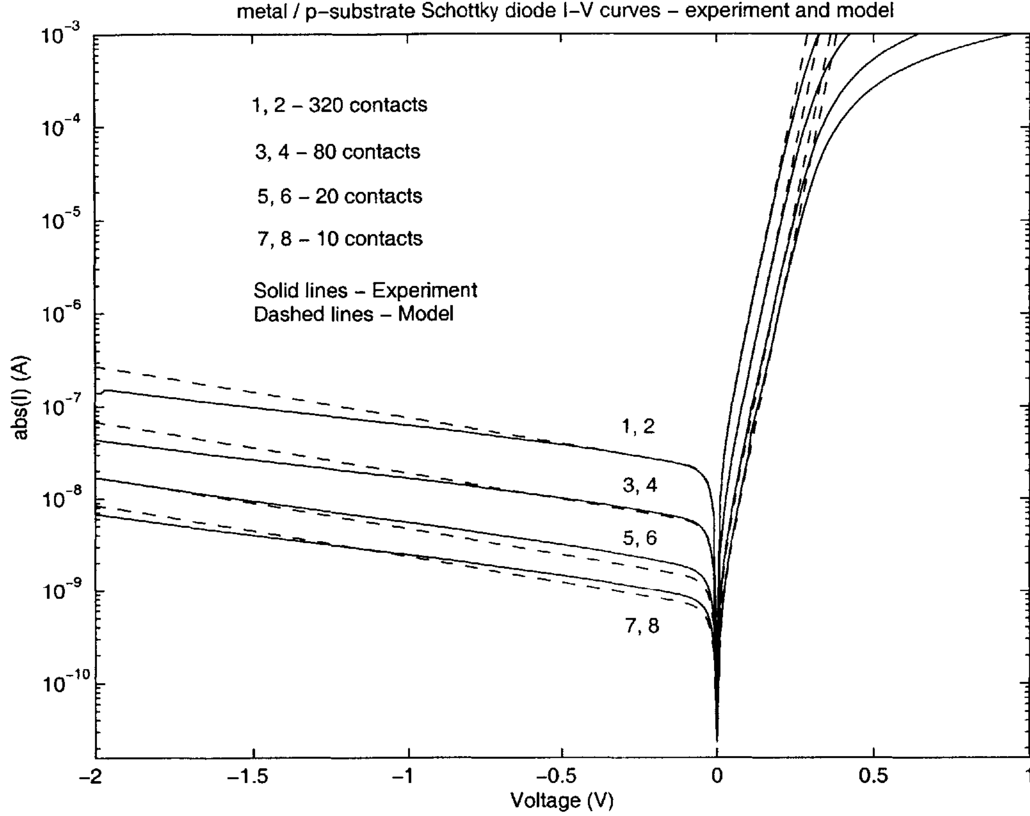


Figure A-15: Modeling DC I-V characteristics of metal / p-substrate Schottky diodes using two back to back diodes (ignoring series resistance).

$$\begin{aligned} q\psi_n &= E_C - E_F \\ q\psi_p &= E_F - E_V \end{aligned} \qquad \text{(A.7)}$$

Here E_C, E_V and E_F are respectively the energies of the conduction band, valence band and Fermi energy in the semiconductor at equilibrium (see Figures A-1 and A-2). If the dopant densities for n and p-type semiconductors, respectively are N_D and N_A, ψ_n and ψ_p are given by

$$\begin{aligned} q\psi_n &= kT \ln\left(\frac{N_C}{N_D}\right) \\ q\psi_p &= kT \ln\left(\frac{N_V}{N_A}\right) \end{aligned} \qquad \text{(A.8)}$$

where N_C and N_V are the effective densities of states in the conduction and valence

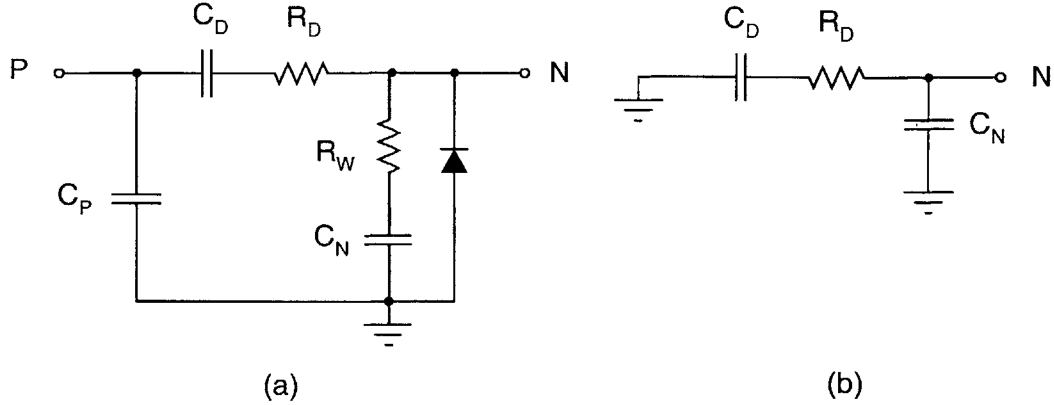


Figure A-16: Approximate zero bias RF models of integrated Schottky diodes. (a) metal / n-well type and (b) metal / p-epitaxial type.

bands, respectively, of the semiconductor. For silicon at room temperature, $N_C = 2.86 \times 10^{19}$ cm^{-3} and $N_V = 3.10 \times 10^{19}$ cm^{-3} [11]. One can use these formulas and measured values of V_{SBn} and V_{SBp} to calculate ϕ_{bi} for both Schottky diode types in my process. I get $\phi_{bi} = 0.27$ V for metal / n-well diodes and $\phi_{bi} = 0.23$ V for metal / p-substrate Schottky diodes. Finally, by using the abrupt junction approximation, one obtains the per-unit-area junction capacitance $C_j(V)$ for the n-well diodes to be

$$C_j(V) = \sqrt{\frac{\epsilon_{Si} q N_D}{2(\phi_{bi} - V)}} \tag{A.9}$$

Here V is the applied voltage across the diode, and N_D is the dopant density in the n-well, and ϵ_{Si} is the dielectric constant of silicon. By replacing N_D in (A.9) with N_A, the dopant density in the substrate epitaxial layer, and using the appropriate ϕ_{bi}, one can get $C_j(V)$ for the p-substrate diodes. For most RF designs, one can usually approximate the junction capacitance by its zero bias ($V = 0$) value, so the result is

$$C_D \approx N A_C C_j(0) \tag{A.10}$$

where N is the number of rectifying contacts in the diode, and A_C is the area of each contact ($A_C = 0.25$ μm^2 in my process). Calculated values of $C_j(0)$ are shown in Table A.1.3. Note that these values (like most of the other parameters in the table)

are likely to be fairly rough estimates, useful mainly for hand calculations.

For the metal / n-well diode, additional parameters of interest are the well resistance R_W and the well-substrate capacitance C_N, which are given by (3.27) and (3.48) respectively. The diode represents the n-well / p-substrate junction diode. Finally, C_P denotes any stray parasitics from the metal terminal to the substrate, such as interconnect capacitance. In most cases $C_N \gg C_P$. Similarly, for the metal / p-substrate diode, C_N is just stray wiring parasitics to ground.

To see whether the single diode model defined in the previous section gives one enough accuracy for circuit simulations, I compare predictions made by both models (with series resistance and junction capacitance added) with experimental results. The circuit used for comparison was the simple diode doubler stage shown in Figure 3-36(a), with $C_{IN} = 1$ pF. Figure A-17 shows the results when both diodes 1 and 2 were of the metal / n-well Schottky type, with 80 contacts each. The input RF frequency was 2 MHz. One sees that the single diode model is extremely inaccurate at high output voltages, because it does a very bad job of modeling the reverse current at large reverse bias voltages. The two diode model is much better and quite accurate for both load currents.

Figure A-18 compares the same models versus experiment when diode 1 in Figure 3-36(a) is implemented with a metal / p-substrate type Schottky diode having 80 contacts (diode 2 is the same as before). The results are very similar. The two diode model predicts the output voltage quite accurately, while the one diode model is very inaccurate for large output voltages. For given RF amplitude, the output voltage is higher for this circuit, with the difference increasing with increasing RF amplitude. This is primarily because of the lower reverse current of the metal / p-substrate diode as compared to the metal / n-well diode.

One sees that for most circuit simulations, the two diode model (with the addition of series resistance and junction capacitance) is necessary for getting acceptable accuracy. When the reverse voltage across diodes in the circuit is limited to low values, however, the single diode model may suffice.

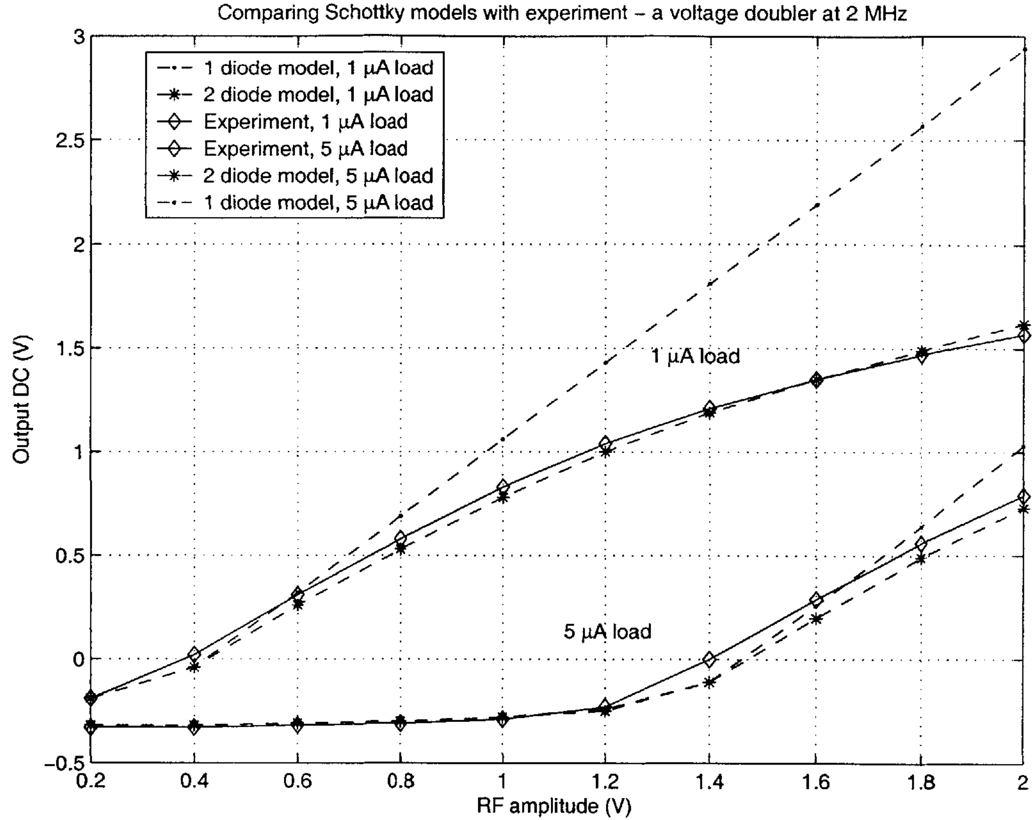


Figure A-17: Experimental rectification curve (at 2 MHz) of a voltage doubler implemented with Schottky diodes compared with that predicted by simulation. Both diodes were of the metal / n-well type.

A.1.5 Matching across dies and wafers

In this section, I investigate the matching of these new devices across different chips (dies) and process runs (wafers). Figures A-19 and A-20 show measured I-V curves for various device geometries (number of contacts) from seven separate chips on the same wafer. The curves were measured using the same voltages. One sees that the forward current is well controlled, i.e., matches well across chips, but the reverse current is more variable. A measure of matching in this case is $S = \frac{std(I)}{\bar{I}}$, where $std(I)$ is the standard deviation of the currents obtained from any device geometry, and $\bar{I}$ is the mean current. Worst case values of S for these devices ranges from 10% to 20%. This level of performance is broadly similar to that obtained with foundry

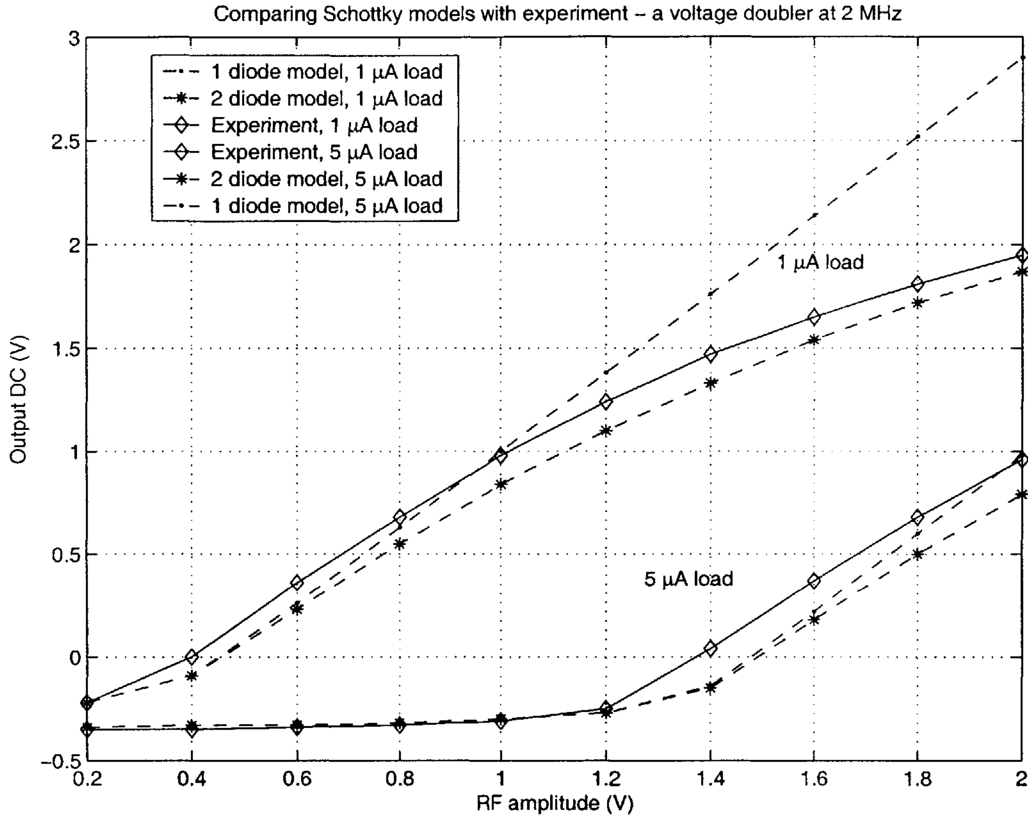


Figure A-18: Experimental rectification curve (at 2 MHz) of a voltage doubler implemented with Schottky diodes compared with that predicted by simulation. The freewheeling diode was of the metal / p-substrate type.

supported devices like MOSFETs, capacitors and resistors.

Finally, Figure A-21 shows the matching performance across two different wafers. The currents being compared are wafer averages obtained by averaging across different chips on the same wafer. Again, the forward current is relatively well controlled, but the reverse current variations are much worse than between wafers. The reverse current variation seems to be of the order of 50% in this case, but data from more wafers is needed for an accurate assessment.

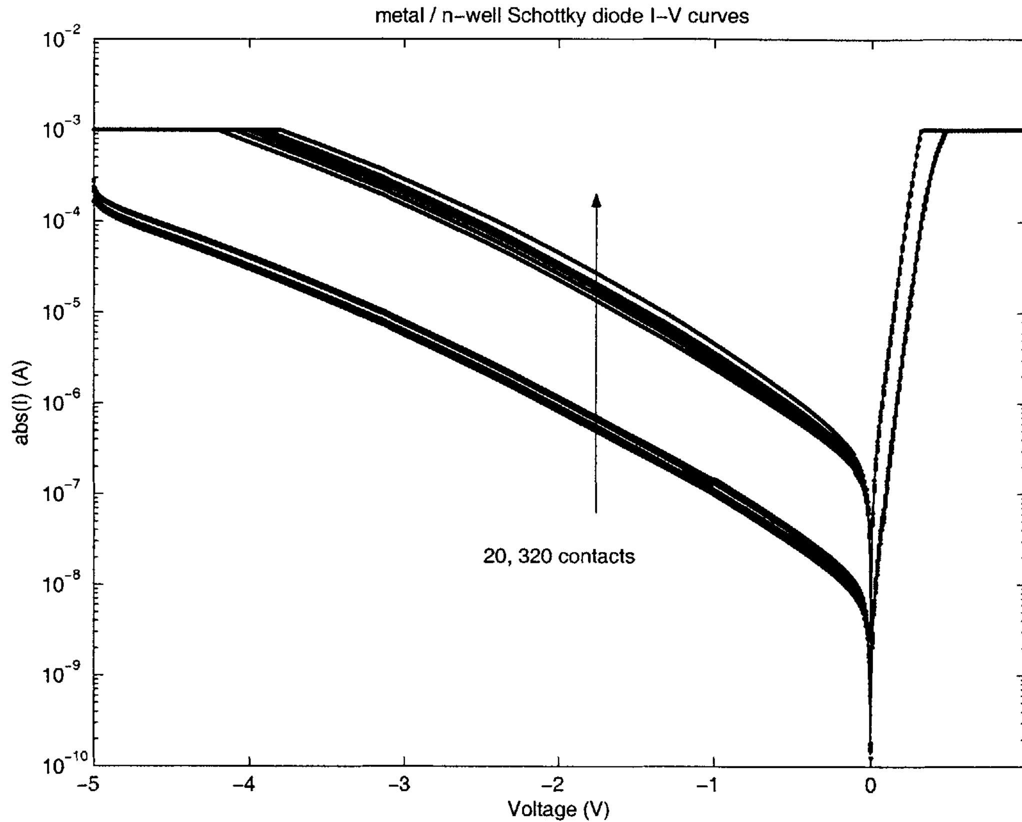


Figure A-19: Measured DC I-V characteristics of metal / n-well Schottky diodes with 20 and 320 contacts for 7 different dies on the same wafer. Data was current limited to 1 mA.

A.2 Solar Cells

In this section, I describe my attempts to create solar cells in a standard CMOS process. The attempts were not entirely successful. I did learn that creating new integrated devices is difficult without control over the chip fabrication process.

A.2.1 Properties of Solar Cells

Typical solar cells are just chains of junction diodes exposed to light. Light is absorbed by the semiconductor material, generating electron hole pairs. Majority carriers disappear into the background, but the minority carriers now diffuse. Those minority carriers generated within a diffusion length of the depletion region at the junction

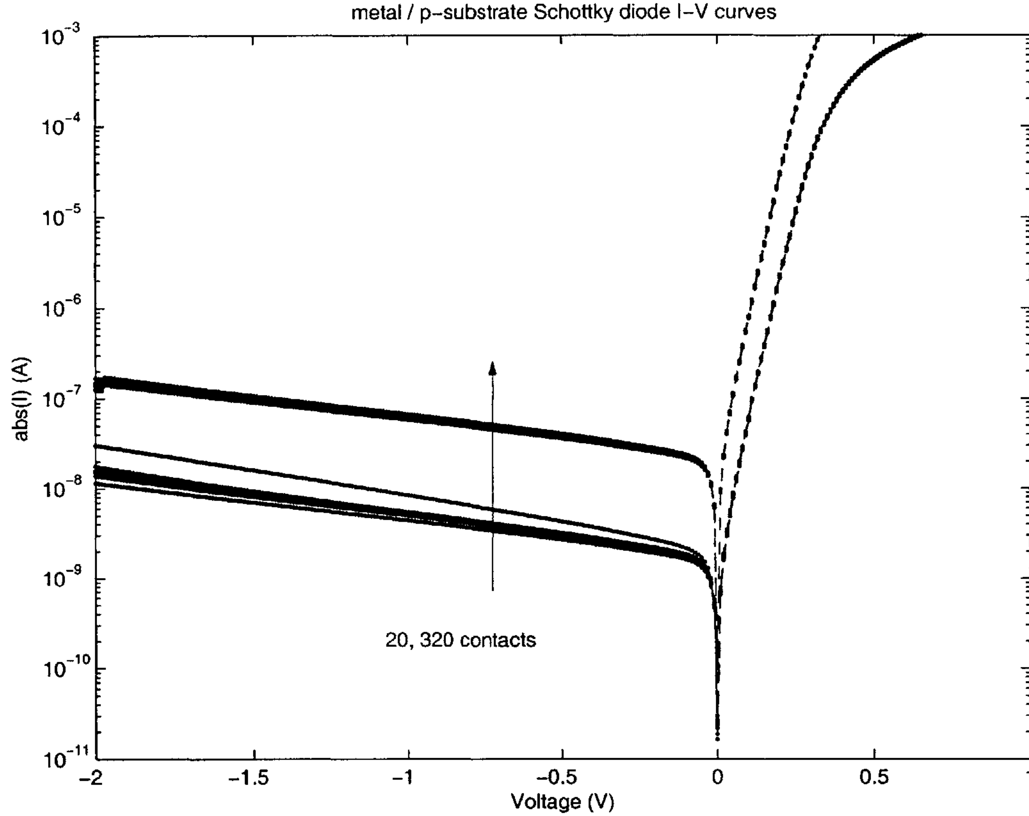


Figure A-20: Measured DC I-V characteristics of metal / p-substrate Schottky diodes with 20 and 320 contacts for 7 different dies on the same wafer.

eventually reach it, after which they are swept across the junction by the built in electric field. This causes the generation of a photocurrent proportional to the light level and shifts the diode I-V characteristic downward (see Figure A-23). Now for low forward bias voltages, the diode current is negative (operation in the fourth quadrant of the I-V plane). In this region the diode acts as a battery, i.e., a source of energy. A solar cell is created by stacking many of these batteries in series (Figure A-22).

The output voltage can be calculated as follows. The optical current I_{op} is given by

$$I_{op} = qAg_{op}\left(L_p + L_n + W\right) \tag{A.11}$$

where A is the cross sectional area of the p-n junction, g_{op} is the optical carrier

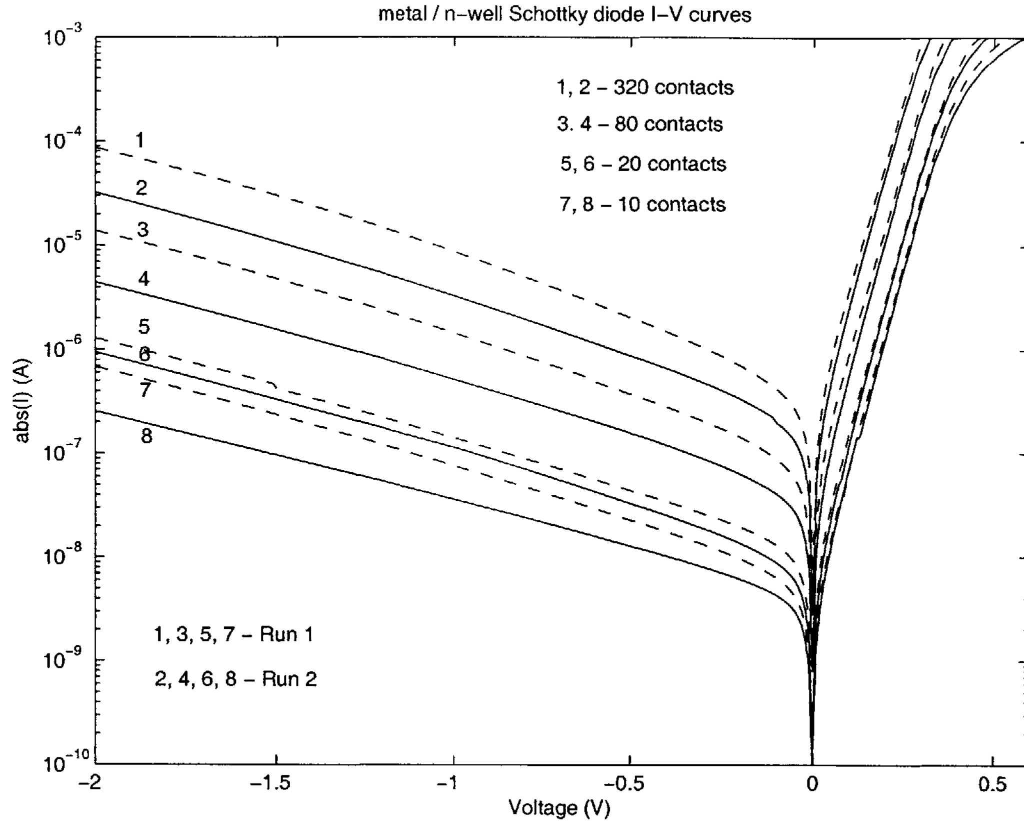


Figure A-21: Measured DC I-V characteristics of metal / n-well Schottky diodes with 10, 20, 80 and 320 contacts for two different process runs (wafers)

generation rate (electron-hole pairs produced per unit volume), L_p and L_n are carrier diffusion lengths in the p-type and n-type semiconductor, respectively, and W is the width of the depletion region at the junction.

The total diode current consists of drift, diffusion and optical currents and is given by

$$\begin{aligned} I &= I_{drift} + I_{diff} + I_{op} \\ &= I_{th}\left(e^{\frac{qV}{kT}} - 1\right) - I_{op} \end{aligned} \tag{A.12}$$

where V is the voltage across the diode, I_{th} is the reverse saturation (thermal) current of the diode in the dark ($I_{op} = 0$), q is the electronic charge, k is Boltzmann's constant and T is the temperature. One sees that the light increases the reverse current of the

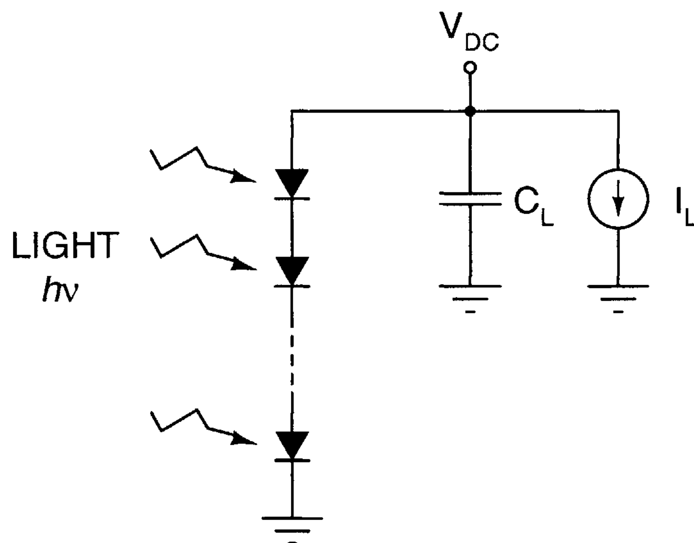


Figure A-22: Schematic of a generic solar cell. I_L is the load current and V_{DC} is the output voltage.

diode. The short current I_{sc} of the solar cell is the current supplied by the cell at zero bias ($V = 0$). From (A.12), $I_{sc} = I_{op}$.

The open circuit voltage V_{oc} is the voltage developed by the solar cell at zero load current ($I = 0$). Solving (A.12) for V at $I = 0$, one gets

$$V_{oc} = \frac{kT}{q} \ln \left(\frac{I_{op}}{I_{th}} + 1 \right) \tag{A.13}$$

When $g_{op} >> g_{th}$, where g_{th} is the thermal carrier generation rate, (A.13) simplifies to

$$V_{oc} \approx \frac{kT}{q} \ln \left(\frac{g_{op}}{g_{th}} \right) \tag{A.14}$$

For n diodes connected in series to form a solar cell, the output no load voltage $V_{out} = nV_{oc}$. Since g_{op} is proportional to the incident light intensity L , a plot of V_{out} versus $\ln(L)$ should give a straight line with a slope $\frac{nkT}{q}$. One observes this in Figure A-24, which shows the measured open circuit voltage from a CMOS solar cell as a function of incident illuminance (in lux) from a spectrally white source. Illuminance and intensity are interconvertible using the fact that 1 lux corresponds

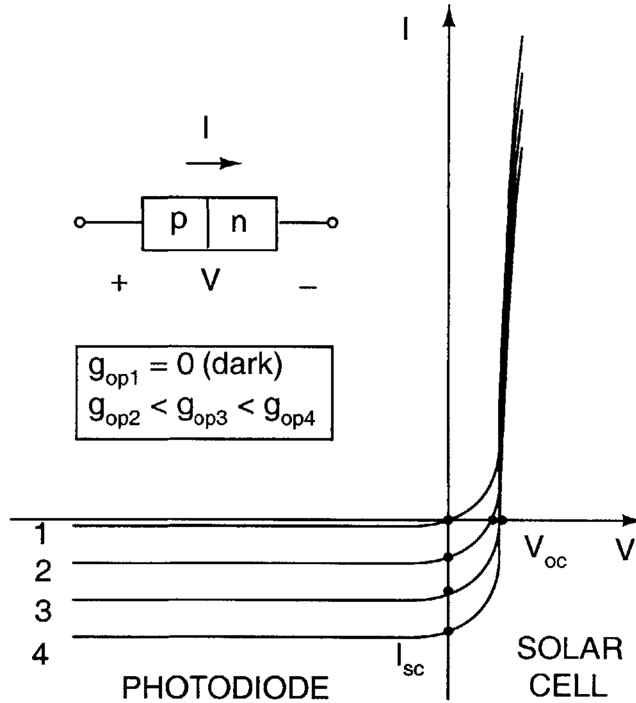


Figure A-23: Theoretical I-V curves of a p-n junction exposed to light.

to an intensity of 1.46 mW/m^2 at 555 nm (green light). The solar cell consisted of eight p$^+$-active to n-well junction diodes connected in series. For reasons discussed in the next section, each diode was on a separate chip and all n-wells were zero biased w.r.t the chip substrates. The best fit slope in Figure is 0.462 V/decade, corresponding to $\frac{kT}{q} = 25.1$ mV. This means that $T = 291.4$ K (65 °F), which matches the actual room temperature very closely.

At high light intensities, however, g_{th} increases because of the increased minority carrier concentrations in the device [37]. Thus (A.14) no longer holds true. Practically, V_{oc} is limited to the built in junction voltage V_0, i.e., $V_{oc} \to V_0$ as $g_{op} \to \infty$.

A.2.2 CMOS Solar Cells

There is a limited selection of diode types which can be used to create solar cells (or photodiodes) in standard planar IC fabrication processes. These are shown in Figure A-25 for a normal vanilla CMOS process (an epitaxial wafer is assumed).

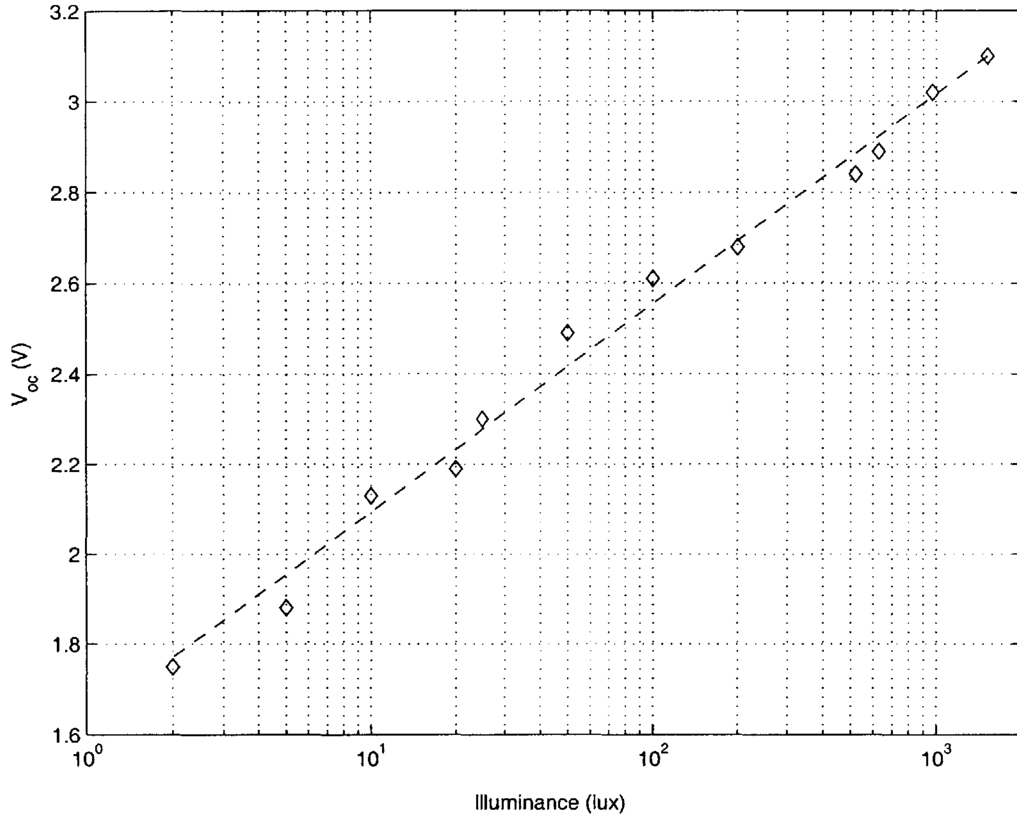


Figure A-24: Measured open circuit voltage obtained from a eight diode CMOS solar cell as a function of the incident illuminance.

Relevant physical parameters are summarized in Table A.2.2. Actual doping levels vary from process to process; those shown must be treated as order of magnitude estimates only.

Assume that light of wavelength λ is incident on the surface of a silicon wafer. The radiation intensity (power/unit area) decreases with depth x inside the wafer, as the incident photons excite atoms in the lattice to higher energy levels and are absorbed. The intensity drop-off with x is exponential, as follows

$$I(x) = I_0 \exp\left(-\frac{x}{L(\lambda)}\right) \tag{A.15}$$

where I_0 is the intensity on the surface ($x = 0$), and $L(\lambda)$ has units of length and is known as the photon absorbtion length of the material (in this case, silicon). $L(\lambda)$ is

Table A.2: Typical CMOS process parameters

Parameter	Symbol	Value
NMOS source/drain (active)) doping	n^+	10^{20} cm^{-3}
PMOS source/drain (active)) doping	p^+	10^{20} cm^{-3}
Well doping	n	10^{16} cm^{-3}
Substrate doping	p^-	10^{15} cm^{-3}
Active junction depth	X_A	0.15 μm
Well junction depth	X_W	2-3 μm

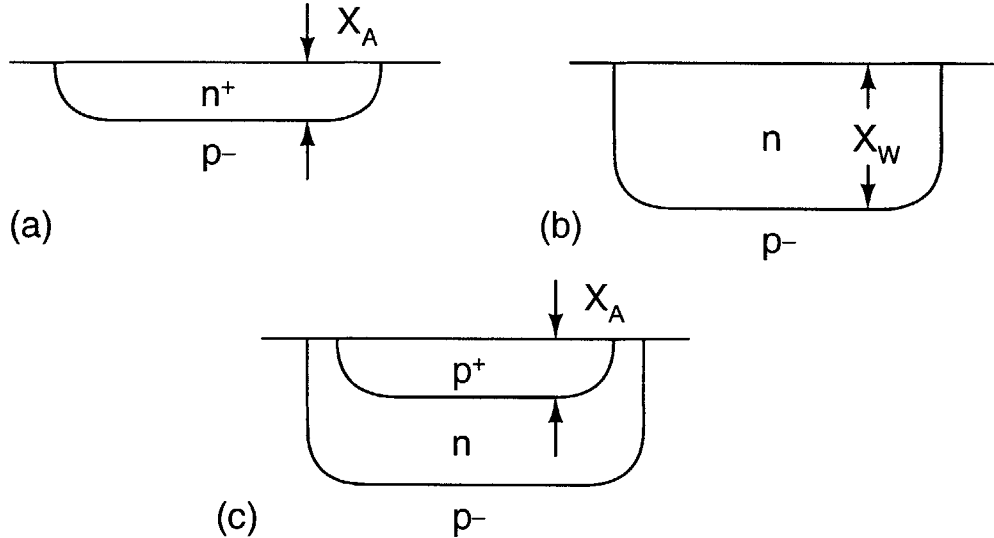


Figure A-25: Possible phototransducing structures in a vanilla CMOS process (a) n^+ active to p^- substrate, (b) n-well to p^- substrate, (c) p^+ active - n-well - p^- substrate.

the length over the intensity $I(x)$ falls to $\frac{1}{e}$ of the incident level I_0. The absorption length is a strong function of λ. Figure A-26 shows $L(\lambda)$ for silicon at room temperature (data adapted from [10]). The absorption length decreases monotonically with λ. The sharp rise in the slope of the curve in Figure A-26 around 1.1 μm corresponds to a photon energy equal to the band gap of silicon. In other words, silicon looks transparent to light with $\lambda > 1.1$ μm. In other words, these photons have insufficient energy to excite electrons from the valence to the conduction band. More importantly, one sees that even over the narrow human optical range, $L(\lambda)$ varies by an order of magnitude, from 0.4 μm at 450 nm (blue) to 3 μm at 650 nm (red).

To fabricate a solar cell on a standard CMOS process, one needs floating diodes

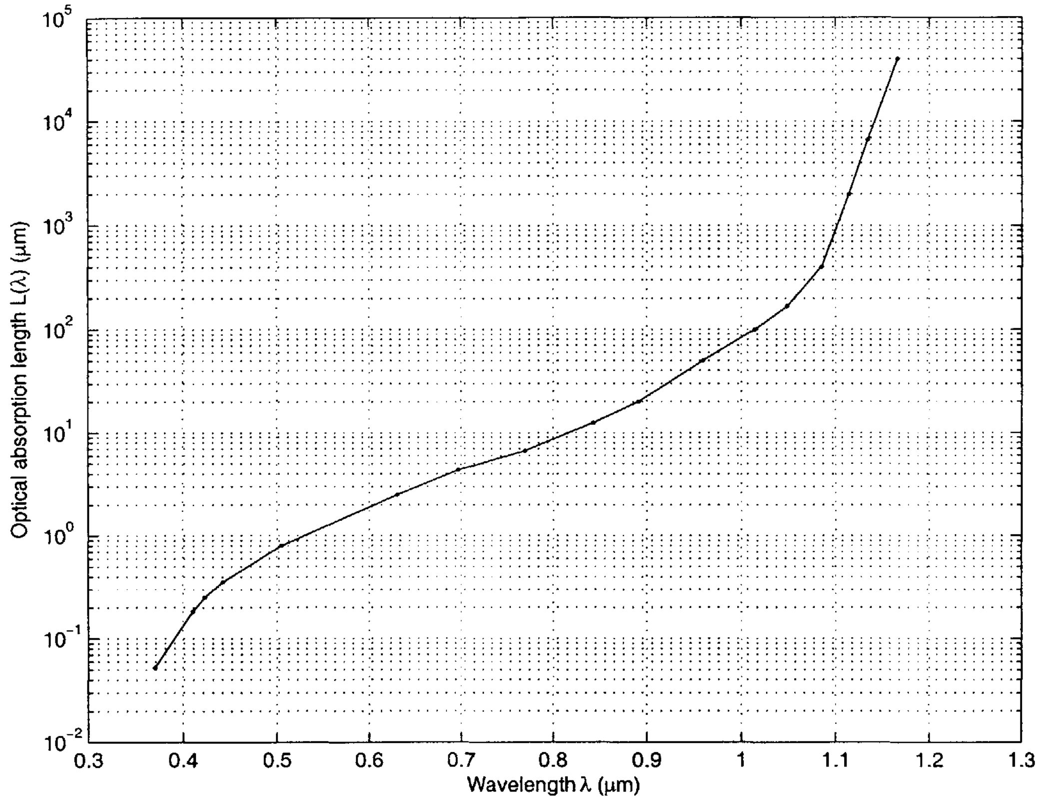


Figure A-26: Optical absorption length $L(\lambda)$ for silicon at room temperature (300 K).

(the substrate potential is taken to be ground). From Figure A-25, the only possibility is the p^+ active - n-well junction (Figure A-25(c)). Herein lies the problem. There is no way to create a floating diode without involving an extra junction. The structure in Figure A-25(c) has two junctions and actually behaves like a vertical pnp phototransistor, with the active as the emitter, the well as the base and the substrate as the (grounded) collector. The resultant equivalent circuit when N of these structures are connected in series on the same substrate to create a CMOS solar cell is shown in Figure A-27.

From Figure A-27, one sees that the effect of the second unwanted junction (n-well to substrate) is to add a current I_{Ci} to ground at each node $(2 < i < N)$. This is the collector current of the i-th transistor; it reduces V_{DC} for a given light intensity. For the cell to work well, one should therefore have $I_{Ci} \ll I_{Ei}$. Also, since

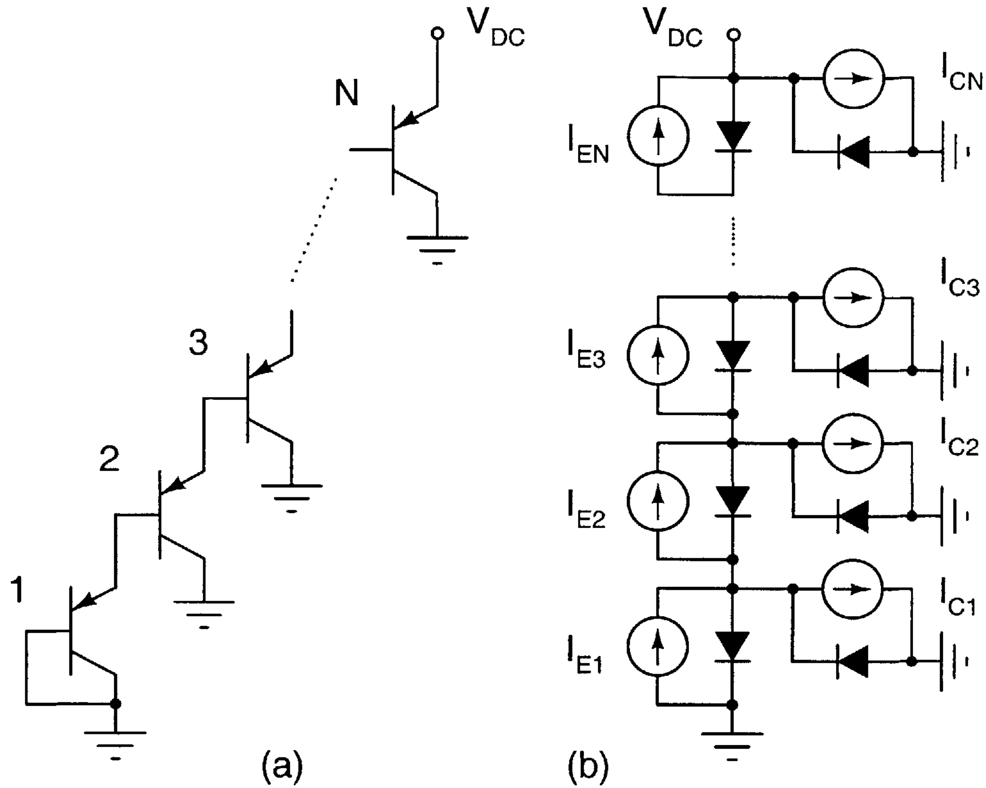


Figure A-27: Equivalent circuit of a CMOS solar cell (a) using transistors and (b) using diodes and photocurrents.

$$I_{Ci} = \frac{I_{Ei}}{\left(1 + \frac{1}{\beta_i}\right)} \tag{A.16}$$

where β_i is the current gain of the i-th transistor, this implies that one wants $\beta_i \ll 1$. The quantum efficiency $Q(\lambda)$ of a photo-transducer is defined as the ratio of the number of charge carriers collected to the number of incident photons at wavelength λ. $Q(\lambda) < 1$ for photodiodes, but can be > 1 for phototransistors. The quantum efficiency $Q(\lambda)$ of most phototransducers available in CMOS and BiCMOS processes has been thoroughly studied [12]. Unfortunately for us, these studies show high $Q((\lambda)$ and β for the phototransistor in Figure A-25(c). Over the visible range, β for this structure was found to range from 100 - 150. The high value of β means that this structure will not work well as a solar cell. Simulations and experiment confirm this.

Measurements were made with diode chains of different lengths (in my 0.5 μm CMOS process) and red, green, blue and white light sources. For given illuminance, V_{oc} was found to increase as the wavelength decreased (moved from red to blue). However, the value of V_{oc} remained small (<0.5 V) even for an intense blue light source.

These observations may be explained qualitatively as follows. From Table A.2.2, one sees that the active junction depth $X_A \approx 0.15$ μm. This is lower than $L(\lambda)$ for silicon over the entire visible range, even at the blue end. In contrast, $X_W \approx 3$ μm means that most photons are absorbed before reaching the well junction even at the red end of the spectrum. The main light absorption region (at visible wavelengths) is thus between the two junctions shown in Figure A-25(c). These junctions then compete to attract minority carriers as they diffuse. The deep well junction wins out over the whole visible range, leading to high values of β (i.e., most of the photocurrent is shunted to ground). The shallow active junction gets an increasing share of carriers as λ and $L(\lambda)$ decrease, which explains why V_{oc} was found to be highest for blue light.

These experimental findings are somewhat depressing. Clearly it is impossible to create working CMOS solar cells at visible wavelengths without control over the fabrication process. There is an alternative, which is to fabricate each diode in Figure A-22 on a separate chip, and hook them up externally (possibly within the same package). In this case, one can connect well and substrate together, i.e., short collectors and bases for all transistors in Figure A-27. This converts them into true diodes. Figure A-28 shows experimental I-V relationships (load curves) of a solar cell formed by cascading eight diodes in this way for different illuminance levels (normal office lighting levels are about 750 lux). Each diode had an area of 3 $\times$ 250 μm $\times$ 250 μm. The source was spectrally white. The lack of good calibration structures makes it difficult to be precise about the actual on-chip illumination level, but the best available estimate of the peak efficiency (electrical power delivered / power in incident radiation) is $\eta_{max} \approx 5$ %. This is not bad considering my complete lack of control over the fabrication process. Even with specialized processes designed for solar cells, it is very difficult to get $\eta_{max} > 10$ %.

To create a solar cell using multiple chips, one can also simply use a single junction

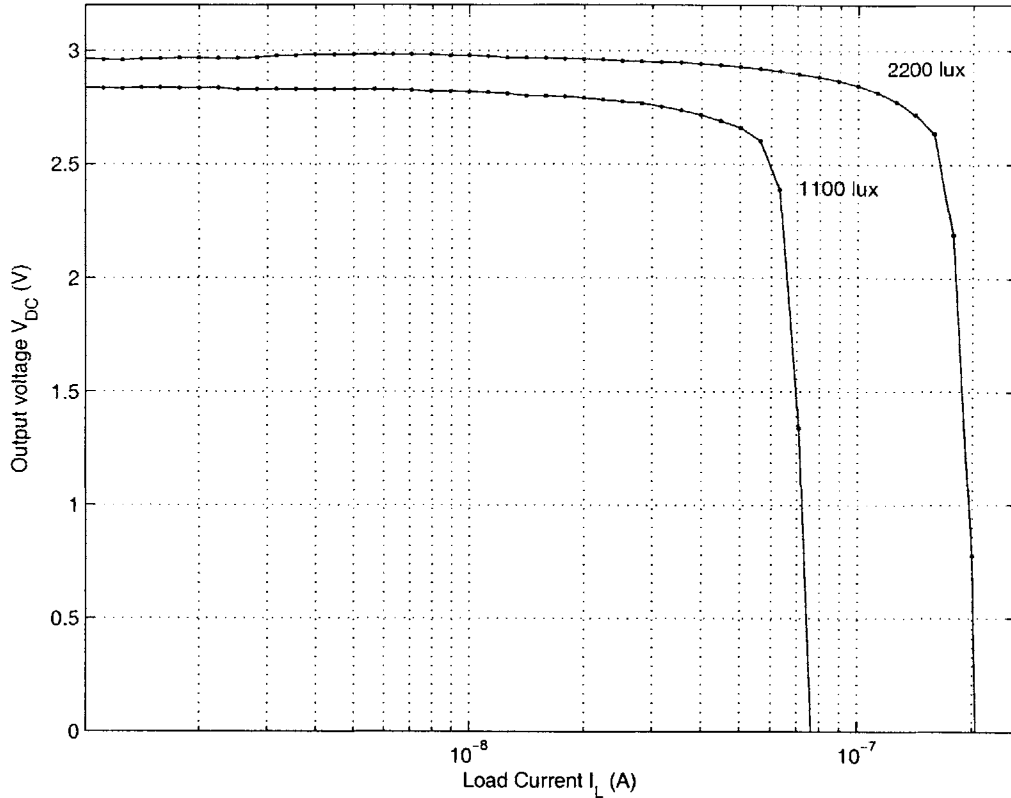


Figure A-28: Measured load curves of a eight diode CMOS solar cell for different illuminance levels. Each diode was on a separate chip. Absolute illuminance values should be treated as approximate.

structure like Figure A-25(a) or Figure A-25(b), and use the substrate as a diode terminal directly. The circuit to be powered up by the solar cell can be placed on the same chip as the lowest (grounded) diode. Incidentally, Silicon-On-Insulator (SOI) processes allow substrate blocks on the same wafer to be electrically isolated from each other. Using a CMOS SOI process thus allows one to create, on one chip, solar cells that behave similarly to the cascade of multiple standard CMOS chips which were proposed in the previous paragraph. Indeed, SOI CMOS solar cells have been reported in the literature [4].

Appendix B

The Mathematics of F-N Tunneling

B.1 Introduction

In this appendix, I provide a brief account of how to use the mathematics of Fowler-Nordheim (F-N) tunneling in practical circuit applications. Tunneling is a quantum mechanical phenomenon.

B.2 Tunneling Current

Electron wave functions decay exponentially inside a potential barrier which is higher than their energy, but are not zero. This leads to a finite probability that the potential barrier can be penetrated. For tunneling through thin oxide films with metal or semiconductor electrodes on both sides, the potential barrier ϕ_0 is between the oxide (dielectric) conduction band and the conduction band of the electron emitting electrode (either ϕ_S or ϕ_M in Figure 3-14). The classical equation for F-N tunneling current density J_{tun} is given by [38]

$$J_{tun} = J_0 E_{ox}^2 \exp\left(-\frac{\beta}{E_{ox}}\right) \tag{B.1}$$

where E_{ox} is the electric field within the oxide dielectric. J_0 and β are constant for a particular choice of electrodes and dielectric and are given by

$$J_0 = \frac{q^3}{8\pi h \phi_0} \frac{m_0}{m_{ox}} \tag{B.2}$$

and

$$\beta = \frac{8\pi}{3} \frac{\sqrt{2m_{ox}}}{qh} \phi_0^{3/2} \tag{B.3}$$

where q is the electronic charge, m_0 and m_{ox} are the electron masses in free space and the oxide, respectively, h is Planck's constant and the barrier height ϕ_0 is expressed in eV. A typical value of $m_{ox} = 0.5m_0$. Substituting known physical constants, one gets $J_0 = 1.5 \times 10^{-6} \frac{m_0}{m_{ox}} \frac{1}{\phi_0}$ A/V^2 and $\beta = 6.8 \times 10^9 \sqrt{\frac{m_0}{m_{ox}}} \phi_0^{3/2}$ V/m.

(B.1) was derived under the assumption that the electrons in the emitting electrode can be described by a free Fermi gas at zero temperature. Several corrections can be made to get J_{tun} more accurately. The more important ones include the effects of finite (non-zero) temperature, variations of m_{ox} within the oxide, quantization of electron potential levels in the semiconductor and lowering of ϕ_0 due to image forces and finite temperature [26]. The resultant ϕ_0 varies from 2.9 - 3.2 eV for a silicon-silicon dioxide interface. The net effect of all these correction factors is to increase J_{tun} obtained in practice at room temperature by a factor of 2-6 over that predicted by (B.1) [30, 3]. Thus (B.1) should be treated more as an order of magnitude calculation than as an accurate prediction of J_{tun}. The best way to characterize F-N tunneling is experimentally. A logarithmic plot of $\frac{J_{tun}}{E_{ox}^2}$ versus $\frac{1}{E_{ox}}$ should give a straight line with a slope $-\beta$. This is known as a Fowler-Nordheim plot.

The actual tunneling current is given by $I_{tun} = A_{tun} J_{tun}$, where A_{tun} is the effective area of the tunneling structure. This is different from the physical area of the electrode-dielectric junctions, because E_{ox} is not uniform across the structure. In particular, E_{ox} (and thus J_{tun}) is higher at sharp edges and corners. Most of the tunneling current flows through these regions of the structure. I_{tun} thus scales with the perimeter (not the area) of the tunneling structure.

(B.3) assumes that the applied voltage $V_{ox} = t_{ox} E_{ox} > \phi_0$, which corresponds to F-N tunneling (where t_{ox} is the oxide thickness). This is usually the case for circuit

applications. In circuits, one usually knows the voltage V_{ox} across the dielectric. In terms of V_{ox}, (B.1) becomes

$$J_{tun} = J_0 \left(\frac{V_{ox}}{t_{ox}}\right)^2 \exp\left(-\frac{\beta t_{ox}}{V_{ox}}\right) \tag{B.4}$$

The equation above can be used for order of magnitude calculations of the tunneling current in a particular structure, but experimental characterization and/or an adaptive programming strategy is still needed, primarily because of the exponential sensitivity of J_{tun} to V_{ox}. The approximate poly-poly oxide thickness in the AMI 0.5 μm process is $t_{ox} = 400$ Å, giving $\beta t_{ox} \approx$1050 V.

B.3 Tunneling Into Floating Nodes

Since one's main interest in F-N tunneling is as a method for lowering the threshold voltage of transistors, one expects to be tunneling on and off floating nodes (transistor gates). Using floating gate transistors allows one to modify the threshold voltage permanently. It is of interest, therefore, to evaluate the tunneling current in this case as a function of time. The situation is complicated by the fact that the voltage on the floating node V_{fg} changes as tunneling occurs. Assuming that the tunneling voltage V_{tun} is fixed, this changes the electric field E_{ox} across the oxide dielectric. Since the change in V_{fg} has the same sign as V_{tun}, E_{ox} is reduced as tunneling continues. This negative feedback self limits V_{fg} to V_{tun}. The tunneling current density J_{tun} is given by B.1, with J_0 and β being constants. Consider the case when $V_{tun} > V_{fg}$, so that electrons tunnel out from the floating gate, causing V_{fg} to increase. Assuming E_{ox} has a constant value inside the oxide, it is given by

$$E_{ox} = \frac{V_{tun} - V_{fg}}{t_{ox}} \tag{B.5}$$

The floating gate voltage V_{fg} is given by

$$V_{fg}(t) = \frac{Q_0}{C_{fg}} + \frac{A_{tun}}{C_{fg}} \int_0^t J_{tun}(t)dt \tag{B.6}$$

where Q_0 is the initial charge and C_{fg} the total capacitance of the floating node, and A_{tun} is the effective area of the tunneling junction. Differentiating w.r.t. time and substituting for J_{tun}, one gets a differential equation for $V_{fg}(t)$

$$\frac{dV_{fg}}{dt} = \frac{A_{tun}}{C_{fg}}\frac{J_0}{t_{ox}^2}\left(V_{tun} - V_{fg}\right)^2 \exp\left(-\frac{\beta}{V_{tun} - V_{fg}}\right) \tag{B.7}$$

(B.7) can be solved to give $v_{fg}(t)$, as follows

$$V_{fg}(t) = V_{tun} - \frac{\beta}{\ln\left(\beta\gamma t + \kappa\right)} \tag{B.8}$$

where $\gamma = \frac{A_{tun}J_0}{C_{fg}t_{ox}^2}$ and $\kappa = \exp\left(\frac{\beta}{V_{tun}-V_{fg0}}\right)$ are constants, and $V_{fg0} = \frac{Q_0}{C_{fg}}$ is the initial voltage on the floating gate node. (B.8) shows that, as expected, V_{fg} saturates to V_{tun} as $t \to \infty$. The time scale is set by $\frac{1}{\beta\gamma}$.

Bibliography

[1] R. Aparicio and A. Hajimiri. Capacity Limits and Matching Properties of Integrated Capacitors. *IEEE Journal of Solid State Circuits*, 37(3):384–393, March 2002.

[2] J. U. M. Araiza. Wireless Transmission of Power for Sensors in Context Aware Systems. Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, June 2002.

[3] R. A. Ashton. Gate Oxide Thickness Measurement using Fowler-Nordheim Tunneling. *Proc. International Conference on Microelectronic Test Structures*, 1:57–60, March 1991.

[4] C. L. Bellew, S. Hollar, and K. S. J. Pister. An SOI Process for Fabrication of Solar Cells, Transistors and Electrostatic Actuators. *Proc. International Conference on Transducers, Solid-State Sensors, Actuators and Microsystems*, 2:1075–1079, June 2003.

[5] S. R. Best. On the Performance of the Koch Fractal and Other Bent Wire Monopoles as Electrically Small Antennas. In *Proc. IEEE Antennas and Propagation Society International Symposium*, volume 4, pages 534–537, June 2002.

[6] P. Bhartia, I. Bahl, R. Garg, and A. Ittipiboon. *Microstrip Antenna Design Handbook*. Artech House, Boston, First edition, 2001.

[7] L. R. Carley. Trimming Analog Circuits Using Floating-gate Analog MOS Memory. *IEEE Journal of Solid State Circuits*, 24(6):1569–1575, December 1989.

[8] H. Carlin and B. Yarman. The Double Matching Problem: Analytic and Real Frequency Solutions. *IEEE Journal on CIrcuits and Systems*, 30(1):15–28, January 1983.

[9] A. P. Chandrakasan and R. W. Brodersen. Minimizing Power Consumption in Digital CMOS Circuits. *Proceedings of the IEEE*, 83(4):498–523, April 1995.

[10] W. C. Dash and R. Newman. Intrinsic Optical Absorption in Single-Crystal Germanium and Silicon at 77°K and 300°K. *Physical Review*, 99(4):1151–1155, August 1955.

[11] J. A. del Alamo. Integrated Microelectronic Devices: Physics and Modeling. Lecture Notes for 6.720, Integrated Microelectronic Devices, September 2002.

[12] T. Delbrück. *Investigations of Analog VLSI Visual Transduction and Motion Processing.* PhD thesis, California Institute of Technology, Computation and Neural Systems Program, January 1993.

[13] C. Diorio, D. Hsu, and M. Figueroa. Adaptive Cmos: From Biological Inspiration to Systems-on-a-Chip. *Proceedings of the IEEE*, 90(3):345–357, March 2002.

[14] D. B. Estreich. A Simulation Model for Schottky Diodes in GaAs Integrated Circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2(2):106–111, April 1983.

[15] R. M. Fano. *Theoretical Limitations on the Broadband Matching of Arbitrary Impedances.* DSc dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering, May 1947.

[16] Klaus Finkenzeller. *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification.* John Wiley, Chicester, Sussex, Second edition, 2003.

[17] N. Gershenfeld. *The Nature of Mathematical Modeling.* Cambridge University Press, Cambridge, UK, First edition, 1998.

[18] J. A. Hagerty, F. B. Helmbrecht, W. H. McCalpin, R. Zane, and Z. B. Popović. Recycling Ambient Microwave Energy with Broad-Band Rectenna Arrays. *IEEE Transactions on Microwave Theory and Techniques*, 52(3):1014–1024, March 2004.

[19] J. Joe and M. Y. W. Chia. Voltage, Efficiency Calculation and Measurement of Low Power Rectenna Circuit. *Proc. IEEE Antennas and Propagation Society International Symposium*, 4:1854–1857, June 1998.

[20] R. C. Johnson and H. Jasik (Editors). *Antenna Engineering Handbook.* McGraw-Hill Professional, New York, Third edition, 1992.

[21] David F. Tuttle Jr. *Network Synthesis*, volume 1, Chapter 12, pages 710–722. John Wiley, New York, First edition, 1958.

[22] U. Karthaus and M. Fischer. Fully Integrated Passive UHF RFID Transponder IC with 16.7-μW Minimum RF Input Power. *IEEE Journal of Solid State Circuits*, 38(10):1602–1608, October 2003.

[23] J. G. Kassakian, M. F. Schlecht, and G. C. Verghese. *Principles of Power Electronics.* Addison-Wesley, New York, First edition, 1991.

[24] S.-G. Lee, J.-T. Lee, and J.-K. Choi. High-Q Poly-to-Poly Capacitor for RF Integrated Circuits. *Electronics Letters*, 37(1):25–26, January 2001.

[25] Thomas H. Lee. *The Design of CMOS Radio Frequency Integrated Circuits.* Cambridge University Press, Cambridge, UK, 2nd edition, 2004.

[26] M. Lenzlinger and E. H. Snow. Fowler-Nordheim Tunneling Into Thermally Grown SiO_2. *Journal of Applied Physics*, 40(1):278–283, January 1969.

[27] J.-J. Lin, L. Gao, A. Sugavanam, X. Guo, R. Li, J. E. Brewer, and Kenneth K. O. Integrated Antennas on Silicon Substrates for Communication over Free Space. *IEEE Electron Device Letters*, 25(4):164–166, April 2004.

[28] S. Marchetti and T. Rozzi. Electric Field Singularities at Sharp Edges of Planar Conductors. *IEEE Transactions on Antennas and Propagation*, 39(9):1312–1320, September 1991.

[29] G. Matthaei, E. M. T. Jones, and L. Young. *Microwave Filters, Impedance-Matching Networks and Coupling Structures.* Artech House, Boston, First edition, 1980.

[30] C. A. Mead. *Scaling of MOS Technology to Submicrometer Feature Sizes*, chapter 14, pages 385–404. Analog VLSI: Circuits and Principles. MIT Press, Cambridge, Massachusetts, First edition, November 2002.

[31] S. Meninger. A Low Power Controller for a MEMS Based Energy Converter. Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, June 1999.

[32] T. K. Sarkar and D. L. Sengupta. An Appreciation of J.C. Bose's Pioneering Work in Millimeter Waves. *IEEE Antennas and Propagation Magazine*, 39(5):55–62, October 1997.

[33] T. A. Scharfeld. An Analysis of the Fundamental Constraints on Low Cost Passive Radio-Frequency Identification System Design. Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, August 2001.

[34] S.-P. Sim, S. Krishnan, D. M. Petranovic, N. D. Arora, K. Lee, and C. Y. Yang. A Unified RLC Model for High-speed On-chip Interconnects. *IEEE Transactions on Electron Devices*, 50(6):1501–1510, June 2003.

[35] G. Smith. Efficiency of Electrically Small Antennas Combined with Matching Networks. *IEEE Transactions on Antennas and Propagation*, 25(3):369–373, May 1977.

[36] Agilent Staff. De-embedding and Embedding S-Parameter Networks using a Vector Network Analyzer. Application Note 1364-1, Agilent Technologies, March 2001.

[37] B. G. Streetman and S. Banerjee. *Solid State Electronic Devices.* Prentice Hall, Upper Saddle River, NJ, fifth edition, 1999.

[38] Z. A. Weinberg. On Tunneling in Metal-Oxide-Silicon Structures. *Journal of Applied Physics*, 53(7):5052–5056, July 1982.

[39] H. A. Wheeler. Small Antennas. *IEEE Transactions on Antennas and Propagation*, 23(4):462–469, July 1975.

[40] T. G. Wilson. The Evolution of Power Electronics. *IEEE Transactions on Power Electronics*, 15(3):439–446, May 2000.

[41] L. S. Yu, L. Jia, D. Quio, S. S. Lau, J. Li, J. Y. LIn, and H. X. Jiang. The Origins of Leaky Characteristics of Schottky Diodes on p-GaN. *IEEE Transactions on Electron Devices*, 50(2):292–296, February 2003.