Design of mm-Sized Wirelessly Powered Sensors

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

of Doctor of Philosophy in Electrical Engineering

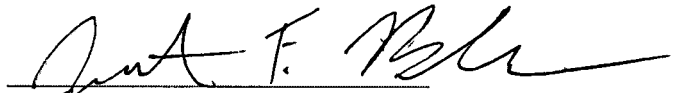
by

Jonathan Frederick Bolus

August 2014

APPROVAL SHEET

The dissertation is submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering



Jonathan F. Bolus (Author)

This dissertation has been read and approved by the examining Committee:

Dr. Travis N. Blalock (Advisor)

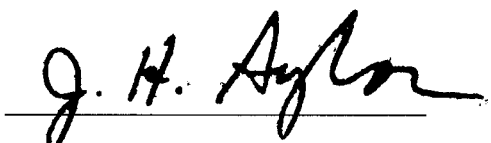
Dr. Benton H. Calhoun (Committee Chair)

Dr. Alfred C. Weaver (Committee Member)

Dr. Ronald D. Williams (Committee Member)

Dr. N. Scott Barker (Committee Member)

Accepted for the School of Engineering and Applied Science:



Dr. James H. Aylor
Dean, School of Engineering and Applied Science

August 2014

Abstract

This work presents several circuits that can be used to improve the range and decrease the size of millimeter-sized wirelessly powered sensors. It is shown that this can be accomplished by decreasing the power consumption of the sensor, and several circuits common to wireless sensors are presented that consume very low amounts of power, in the 100 nW range. A wireless power transmission system that can generate power in the range of 1 μW is also presented.

A circuit that allows for sensor identification is presented, which generates a unique ID number based on variation inherent to the CMOS manufacturing process. The ID generator consumes just 39 fJ/bit from a 0.4 V supply. A method of ensuring reliable identification in the presence of unreliable bits is also presented, and a system that can identify 1000 unique chips, using 31 bits per chip and with an error rate of less than 10^{-6} is presented.

A combined demodulator and clock generator circuit is presented. Both operations are based on a low-power delay line, and it is shown that accuracy problems related to low-power operation can be mitigated by using an adaptive approach in which the delay is tuned to a transmitted reference signal. The same calibrated delay line can be used to perform demodulation and generate a 100 kHz on-chip clock, consuming 220 and 190 nW, respectively, from an 0.4 V supply. An improved version is also presented, with simulated power consumption of 130 nW during demodulation and

70 nW during clock generation.

A wireless power transmission system is also presented for a wireless sensor is also presented. Using a loop antenna formed from a bond wire, with a loop area of 12.3 mm^2, the system is able to transmit 2 μW over a range of 20 cm, or a minimum of 0.8 μW over a 800 cm^2 area, with a minimum output voltage of 0.4 V DC. The system uses a low-cost patch antenna for transmission, and could be used to cover an arbitrarily large area by tiling transmitting antennas.

Acknowledgements

In completing my doctorate I relied, in no small part, on the support of many colleagues, family and friends. Without their support the work presented here would not have been possible, and I am very grateful for their efforts. A few of these people in particular I would like to distinguish here.

First, I would like to thank my advisor, Travis Blalock, for his guidance and support. He allowed me the opportunity to work on a wide variety of interesting and challenging projects during my time at UVa, and was an invaluable help during all stages of my research. He always displayed great confidence in my abilities, and was always receptive and encouraging of my ideas.

I would also like to thank the my doctoral committee, Ben Calhoun, Alf Weaver, Ron Williams and Scott Barker, for all their efforts and helpful feedback. In particular I would like to thank my committee chair, Ben Calhoun. He provided me with space on several chips to implement my designs, offered substantial constructive feedback on a great deal of my writing, and was always highly motivating.

Many of my fellow students are also deserving of my thanks. Stuart Wooters was always available to help with any difficulties I was having, and our conversations about research proved very useful to my work. He also constructed the pad ring for the demodulator and wireless power transmission chips used in Chapters 3 and 4. Yanqing Zhang helped me with the synthesis of some of the digital logic used for

the demodulator. Kyle Alexander Craig constructed the pad ring for the random identification chip described in Chapter 2. I would also like to thank Steve Jocke, Andrew Jurik, and Peter Beshay for their assistance.

Many members of my family are deserving of my thanks. In particular, I would like to thank my grandfather, Fred Wikner, who has always maintained a great interest in my education, and been a strong motivating force in the pursuit of my doctorate.

Lastly, but most importantly, I would like to thank my parents John and Andrea, and my brother Christian, for their unwavering love and support. They have always encouraged my interest in technical things, and have provided me with everything I needed to pursue my education. Their confidence in me has been invaluable. I cannot imagine a more supportive family, and for this I will always be grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Small, wireless sensors have the potential to allow for the development of many medical and biological applications that have previously been unrealizable. Such sensors are currently a popular area of academic research, and a wide variety of applications have been proposed, including cardiac [1] [2] [3] [4] [5], neural [6] [7] [8], ocular [9] [10], blood-pressure [11], blood glucose [12], orthodontic [13], and temperature [14] [15] monitors, which would be of use in both research and clinical environments. Additionally, this kind of instrumentation is not limited to use in humans, but has been proposed for monitoring laboratory animals [16] [11] [17], commercial livestock [18] [19], household pets [20], and wildlife [21] [22] [23]. A universal feature of these applications is the desire for minimal sensor size. For implantable sensors this is critical, as it facilitates insertion and reduces invasiveness, and it is a primary feature of external sensors as well, as it minimizes patient discomfort. For many applications there are hard limits to sensor size: insertion into small cavities, such as arteries or eyes, or any type of attachment to small animals or insects, requires some minimum size for safe, non-interfering operation. For these reasons, the study of sensor miniaturization is of significant academic and commercial interest.

1.1 Wireless Sensor Architecture

A block diagram of a generic, wireless sensor is shown in Fig. 1.1. This figure shows elements that are common to a wide variety of wireless sensors, but specific sensors may vary significantly from this architecture depending on the particular application. To achieve the maximum possible miniaturization, the architecture integrates all components that can be practically integrated. External transducers, which convert environmental properties into electrical signals, are often not integrable. These include leads for bioelectric signals such as electrocardiogram (ECG), electromyogram (EMG), and electroencephalogram (EEG) signals. Many other transducers are integrable, such as temperature sensors, since a small IC will be in thermal equilibrium with its environment. The energy source, or energy harvesting transducer, and the antenna used for communication, are often not integrated.

The analog front-end (AFE) is responsible for converting the transducer signals into digital form for processing and transmission. This frequently includes a low-power, low-noise instrumentation amplifier and an analog-to-digital (A/D) converter. Sensors that collect data from multiple transducers may have multiple parallel paths, with multiple amplifiers and A/Ds, or may use multiplexing to collect data from multiple transducers using only a single acquisition path by rapidly switching between transducers. Again, the particular application may vary from this architecture. For example, a temperature sensor can employ a temperature controlled oscillator and frequency counter, which can be fully integrated and requires no amplifier or A/D.

Acquired signals are usually digitized prior to transmission, which allows for a variety of digital processing techniques to be performed, including filtering, feature detection, and compression. Feature detection includes any method by which key features of the acquired signal are transmitted rather than the complete acquired

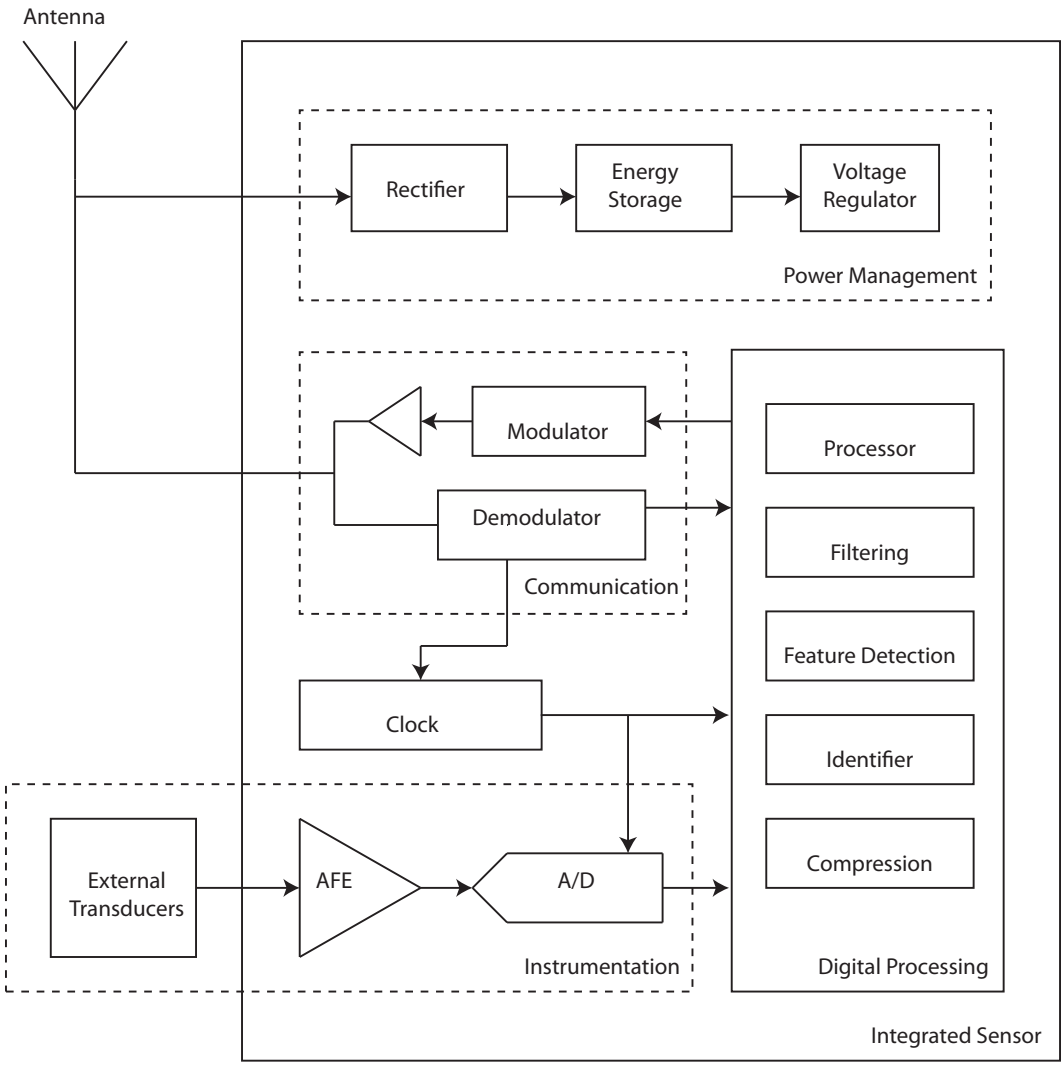


Figure 1.1: Block diagram of generic, wirelessly powered sensor.

waveform, usually in the interest of reducing the data rate. For example, a wireless ECG monitor could derive heart rate, or detect instances of cardiac arrhythmia, rather than transmitting the complete ECG signal. This often has the effect of reducing the total power consumption, because wireless data transmission is frequently the dominant consumer of power in a wireless sensor, and the energy expended to perform feature detection and compression can frequently be made less than the energy saved by reducing the amount of data to be transmitted [24] [25].

Communication with a remote sensor reader is frequently bidirectional. In the simplest applications, the reader can transmit a simple inquiry signal, to which all external sensors will respond. This signal does not need to be a modulated wave, but could simply be the initiation of a high-frequency radio wave that energizes wirelessly powered sensors. In a system where multiple sensors are within range of the reader, communication is often made simpler if individual sensors can be addressed and interrogated by the external reader. This requires at least the transmission of some unique identification data, which the sensors can use to determine if they are being addressed. More advanced sensors can receive a wide variety of data from the external reader, which specifies things such as what type of signal is to be acquired, the type of processing and filtering to be performed, and what identifying features or other data is to be transmitted back to the external reader. Of course, for the acquired data to be of any use, the wireless sensor must be able to transmit back to the external reader. For these reasons, most wireless sensors include both a demodulator for receiving data, and a modulator for transmitting data.

The source of energy or power used to operate the wireless sensor is one of the most variable areas of current wireless sensors, and frequently the most important in determining the size and weight of the complete sensor. There are many possibly sources of energy, which will be discussed in the following section, but in general there is an external transducer that converts a source of non-electrical energy into electrical energy. The output is rarely appropriate to drive the sensor circuity directly, and so frequently requires an AC/DC rectifier and/or a voltage regulator to provide stable DC voltage. In the case of wirelessly powered sensors, the antenna used for communication may also be shared by the power transmission circuitry.

1.2 Overview

Due to the incredible circuit density available in modern CMOS processes, circuit area is not the primary constraint in reducing sensor size. For example, several recent SoC implementations of wireless sensors achieve significant integration of analog instrumentation, A/D conversion, digital filtering, and digital processing with total silicon areas of less than 10 mm^2 [2] [3] [13] [5]. The primary difficulty is in finding a suitably small source of reliable energy to power the sensor, and reducing the energy consumption of sensor components.

This work focuses on resolving some of these difficulties. Specifically, it examines how to transmit power wirelessly to a small sensor, and presents circuits that allow for communication to the sensor, clock generation, and sensor identification. Although the amount of power transmitted is small, this work shows that these circuits can be constructed such that their power consumption is low enough for operation in such a system. An overview of these components is given below.

1.2.1 Wireless Power Transmission

Many sources of energy are potentially available to power wireless sensors, such as thermal, solar, mechanical, chemical, biochemical, and RF energy. Sensors powered by thermal energy are possible by using thin thermoelectric generators (TEGs) [26] [5] to extract energy from the temperature differential between skin and air. This, however, limits placement to near the surface of the skin, which severely restricts the placement of implantable sensors [27], and is limited to use in endothermic animals. Solar energy is harvestable by miniature solar cells, although this is restricted to sensors exposed to ambient light, which excludes implantable devices, with the interesting exception of intra-ocular sensors [10]. Sensors powered by mechanical vibration

are also possible [28], although this is restricted to applications in which motion is guaranteed to occur regularly and with sufficient intensity.

In some cases it may be possible to extract energy directly from biochemical sources inside the subject, such as from the inner ear [29], although the feasibility of this depends heavily on the adjacent physiology. In addition to these, thin-film Li-ion batteries can be used to power wireless sensors [30], and a recent example can store 1 μA-h of energy in an area 1 mm^2 [10]. Batteries, however, are only energy storage elements, and will require periodic recharging unless the total energy consumption of the sensor over its lifetime can be brought below the battery capacity.

Another possible energy source for wireless sensors is RF energy [31]. While a common choice for many wireless sensors [14] [3] [15] [11] [32] [8] [5], it is not without its drawbacks. Most notably, RF antennas tend to be most efficient when their physical dimensions are on the same order as the wavelength of the operating frequency. At dimensions smaller than this, the received power decreases as the antenna dimensions decrease. This would suggest operating at high frequency to decrease the wavelength, which may be possible in some instances, but in general is difficult because of increased attenuation by biological tissue at high frequencies. The result is that the efficiency of RF power for small sensors is generally poor. Additionally, power transfer decreases rapidly as the separation between the RF power source and sensor increases.

Wireless power transmission does, however, have several advantages. Most importantly, and unlike the previously discussed options, the duration and intensity of the RF power is under direct user control, which allows for operation when no other energy sources are available and makes reliable operation easier to guarantee. And although the separation between the RF power source and sensor must be limited, the sensor location is less constrained than under most of the previously discussed

options: as long as the sensor is within the volume of the radiated field, power transfer can occur.

Although the power received by wireless transmission decreases as antenna size decreases, this property is not unique to wireless power transmission. Almost all energy harvesting transducers extract energy at some rate proportional to their volume or surface area [33], which makes low-power operation a universal requirement for small sensors. Similarly, any design innovations that decreases the power consumption of wireless sensors can be used to reduce their size.

This work focuses on a wireless power transmission system intended for monitoring laboratory animals that are free to move within the confines of a cage. The system includes a transmitting patch antenna, and a receiving chip with an external bond-wire antenna. While most of the current research on wirelessly powered sensors is designed to generate standard supply voltages for integrated circuits with devices operating in strong inversion, the work presented here focuses on generating a lower supply voltage for operation in sub-threshold. In doing so, the effective range of the link is increased relative to similarly published work. Furthermore, this work presents measured results showing that a planar area can be effectively covered by the system, allowing movement of the sensors within a plane above the transmitting antenna.

1.2.2 Random Identification System

Multiple wireless sensors are often combined to form wireless sensor networks, in which multiple sensor nodes collect data independently, which is then aggregated in a central location by a remote reader of some form. This number of sensors could range from two, as in the case of ocular sensors [10], to several thousand, as in the case of sensors attached to honey bees [23]. In such networks, when communication happens

wirelessly, the sensor nodes must communicate some unique piece of information so that the remote reader is able to identify from which node each piece of data is originating.

Although this may appear straightforward, it is in fact difficult to program individual sensors with a unique identifier. There are several manufacturing steps that can be added to the CMOS process that create small, one-time programmable memories, such as arrays of electrical- or laser-blown fuses, but these extra manufacturing steps increase the cost of what should be inexpensive devices. Other forms of non-volatile memory, such as EEPROM or Flash, require for programming voltages much larger than any other wireless sensor circuit needs, and larger than what most energy harvesting transducers can generate. These large voltages could be applied for programming at the time of manufacture by wafer probing, but again this increases the complexity of the manufacturing process as well as cost. The ideal identification scheme would provide some unique identifier for each sensor without the need for any extra manufacturing steps or large voltages.

This work proposes that this can be achieved by using the manufacturing variation unique to each chip to generate a random number that can be used for identification of individual sensors. Although this idea has been previously explored [34] [35] [36] [37] [38] [39], the approach presented here allows for each sensor to positively determine whether it is being addressed by a remote reader, which has important energy saving implications. This is usually difficult, because an unavoidable property of such systems is that not all of the bits that form the ID number are necessarily reliable. The approach presented here allows the reliability of individual ID bits to be quickly determined. The "reliability" of previous systems is poorly defined, and this approach presents a much more rigorous analysis of the reliability of the complete system, as well as experimentally determined temperature effects. Additionally, the energy con-

sumed per ID bit of this approach is much lower than any previously published.

1.2.3 Demodulator and Clock Generator

Some form of communication from the remote reader to the sensor is necessary for the operation of most wireless sensor systems. In the simplest case, this could simply be a large pulse of wireless energy, possibly the same one used to energize a wirelessly powered sensor, that would instruct the sensor to begin collecting and transmitting data. However, in many applications, it is necessary to transmit data from the remote reader to the sensor. This could be the identification number of a remote sensor in a network to be interrogated, configuration data specifying the operating mode of the sensor and what data to transmit, or some other data relevant to the application. A demodulator of some kind is required to receive this data.

Similarly, the components mentioned so far, including the A/Ds, digital processing, demodulator and transmitter require one or more clocks or time references for their operation. Standard clock generation with a crystal oscillator typically consumes too much power for a small, wireless sensor. This work observes that because demodulation and clock generation are pervasive components of wireless sensors, and both require an accurate time reference, they can be combined in such a way that a single, low-power time reference can be shared, and their combined power consumption decreased.

The time reference used in this work is a low-power delay line, and an analysis of low-power delay lines is presented based on minimizing the product of the amount of delay generated per unit of energy consumed. This allows for the design of an efficient delay line, however, the amount of delay produced is sensitive to manufacturing variation. To mitigate this, the delay line is made adjustable, and it is shown that

the delay line can be tuned with reference to the transmitted signal using low power. The tuned delay line then be used to as the basis for a a low-power demodulator and an accurate, low-power on-chip clock. The result is a fully integrated, combined demodulator and clock generator that consumes less power than similar published systems.

1.3 Summary

This work consists of circuits for sensor identification, communication and clock generation, as well as a wireless power transmission system. Together, these demonstrate substantial power savings over previously published work, and it is shown that these power savings can be used to extend the range and decrease the size of a wirelessly powered sensor. And, although intended for wirelessly powered medical applications, it is hoped that much of the work here will prove useful in other systems as well. For example, the random identification circuit is useful for any application that needs to uniquely identify integrated circuits, and the demodulator and clock generator would be useful for many systems that require a low-power data reception and an on-chip clock.

Chapter 2

Random Identification System

For wireless systems composed of multiple wireless sensors, some method of uniquely identifying the sensors is required. The small amount of energy available to each sensor makes this difficult, since conventional non-volatile memories based on floating-gate transistors, such as EEPROM and Flash, require relatively large voltages and energies to program [40]. Another option would be to use a system of laser-blown, or otherwise one-time programmable fuses to store a unique identifier. This however, requires additional masks and manufacturing steps, increasing the cost of what should be inexpensive devices.

An ideal identification system would require no additional programming or manufacturing steps, provide a unique identification number that does not vary between power-on cycles, and would require little energy to operate. The circuit presented here performs this function using low energy, just 39 fJ/bit, and without requiring initial programming, by using the natural variation inherent to the CMOS fabrication process as the basis for ID generation. This essentially forms a non-volatile memory: one which stores unique information about the die, in the form of physical variation, and is "programmed" once at the time of fabrication.

An unavoidable complication is that not all the random identification bits produced by the circuit are necessarily reliable: due to electrical noise, some ID bits may vary between successive read operations. This may cause errors in identification. Presented here are several addressing schemes that could be used to compensate for this unreliable behavior. An example implementation is presented that could address 1000 unique ICs, where each IC contains 19 random ID bits, up to 9 of which can be unreliable, with a total error rate of less than 10^{-6}.

2.1 System Overview

The purpose of the random ID system is to uniquely identify individual chips using very little energy. This is accomplished by including on each chip a random ID generator, which is a circuit designed to be sensitive to variations inherent to the CMOS manufacturing process. Although each chip will contain the same circuit, built from the same set of masks, the output of the ID generator will differ between chips. The generated ID is "random" in the sense that it is unknown before manufacture. The random event that determines the value of the ID bits is the manufacture of the IC, which occurs only once, and successive reads of the same ID generator should produce identical IDs.

2.2 Previous Work

Several circuits have been proposed to generate ID numbers for chip identification. One approach is sensing the drain currents of individual FETs and comparing them to some non-varying reference [34]. Other implementations are based on latches and cross coupled inverters, both sides of which are pre-charged to equal voltages and

then released, settling in one of two possible states based on their variation [35] [36]. Neither of these methods allows for the reliability of individual ID bits to be determined without reading from the ID bits a large number of times.

Additionally, several methods have been proposed to identify integrated circuits based on the manufacturing variation present in conventional memories. For example, variations in SRAM bit-cell power-up state [37] can be used to derive a unique ID, although again the reliability of this ID can is difficult to determine. The static-noise margin of SRAM cells can also be used to generate a unique ID [38], although this method requires a 1 Mb SRAM to to generate 128-bit ID. Similarly, identification can be based on variations in retention-rate in DRAM cells [39]. Most low-power sensors, however, are unlikely to have large memories with which these methods could be employed.

All random ID implementations have the property that not all of the random ID bits are necessarily reliable. This problem seems to be unavoidable, because for any identification scheme based on manufacturing variation, there exists the possibility that for a particular bit, the amount of underlying variation is small. If the variation is small enough, then the generated bit will be determined by the effects of electrical noise rather than manufacturing variation. If this is the case, the value of the ID bit will change between power-on cycles or read operations. This is the opposite of the desired behavior, as the the generated ID should be temporally static.

A major disadvantage of previous random ID systems is that they have no fast way of determining the reliability of the random IDs generated. Current literature solves this problem by requiring each sensor to transmit its generated ID number back to the external reader. For a given sensor, the external reader then compares the generated ID number with its recorded ID for that sensor: if the Hamming distance between them is small it concludes it has found a match. The primary disadvantage of this is

that it requires each sensor to transmit its generated ID number back to the external reader. This is unfortunate, since a major energy consumer in most wireless sensors is the transmitter, and any reduction in the amount of transmitted data directly reduces the energy required by the sensor.

This work improves upon previous random ID systems by presenting a method by which sensors can conclude, locally, whether they are being addressed. No transmission of the generated ID is required. This reduces the amount of data that needs to be transmitted, and thus the energy consumption of the sensor. It has been observed that prior knowledge of which bits are unreliable can be used to increase the accuracy of identification [35] [41], and this system is novel in its ability to characterize the value and reliability of each random bit using only two read operations. It does this by examining the amount of manufacturing variation on which each bit is based, and determining if it is large enough to outweigh the effects of electrical noise.

2.2.1 Approach

Although there are many properties of ICs that exhibit random variation, the property selected for this application is the FET threshold voltage, V_T. This was chosen because V_T can vary by a significant fraction of its nominal value; the V_T variation among multiple FETs has a large, uncorrelated component due to random dopant fluctuations; and FET drain current, which can be easily detected by a sense amplifier, is very sensitive to V_T.

A single random bit can be generated by designing two identical FETs, and then comparing their values of V_T after manufacture. If a particular FET has a larger value of V_T then the other, this could represent a logical '1', and if the value of V_T is smaller than the other, this could represent a logical '0'. This has the advantage of

not requiring the precise value of V_T to be determined. Rather the drain currents of both transistors, which depend on V_T, can be easily compared with a sense-amplifier.

If the difference in V_T between the two FETs is small, it may be that the effect of electrical noise is large enough to change the sense-amplifier output between reads. This is problematic because it would lead to changes in the generated ID number. The circuit presented here includes a method to determine which transistor pairs have sufficient difference in V_T such that the effect of electrical noise is insignificant.

2.3 Circuit Implementation

For the identification system presented here, each chip is equipped with an ID generator circuit. Ideally, this ID generator would produce the same, random N-bit ID each time the ID generator is activated. To ensure that the ID generators on different chips are likely to vary from each other, the ID is based on the manufacturing variation inherent to the die, which is likely to be different for every chip.

2.3.1 Random ID Cell

The N-bit ID generator is composed of N ID cells, each of which produce a single random bit. The requirements of such a cell are low read energy and high reliability. The implementation proposed here is shown in Fig. 2.1. Each ID cell contains two identically drawn NMOS devices, who's drain currents are compared to determine the output of the ID cell. If $I_{D1} > I_{D2}$, the value of the ID cell is '1', and if $I_{D1} < I_{D2}$, the value of the ID cell is '0'.

Equivalently, the value of the ID cell is determined by the random variable ΔI_D, where

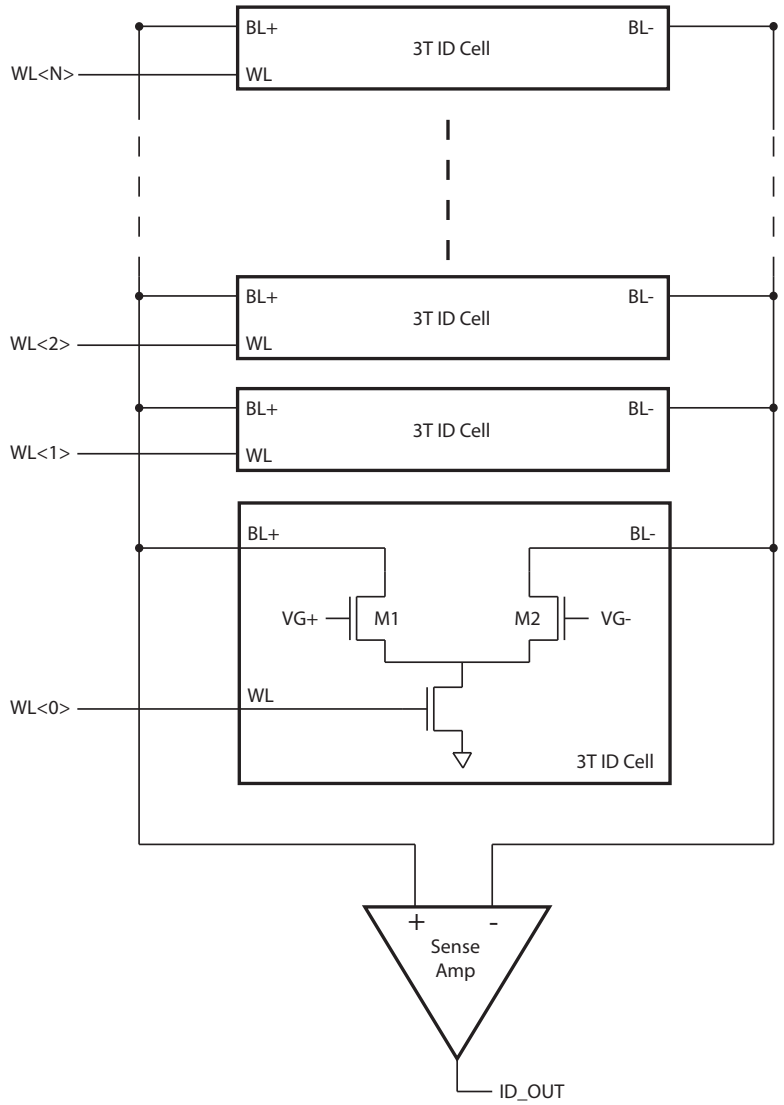


Figure 2.1: N-bit random ID generator. The organization is similar to that of a memory, where multiple ID cells share a common set of bit-lines and a sense amplifier.

$$\Delta I_D = I_{D1} - I_{D2} \tag{2.1}$$

Ideally, the mean of ΔI_D would be very small to ensure a uniform distribution of '0' and '1' bits. This is accomplished by ensuring identical layouts for M1 and M2 and

minimizing the distance between them, to minimize the effect of any process gradients across the die. Also, the variance of ΔI_D should be maximized, as a larger current difference is easier to sense and leads to more reliable operation. This is accomplished by minimizing the size of M1 and M2, since the variance of V_T is inversely proportional to the device area [42].

The results of a Monte Carlo simulation of the expected common-mode current, I_{CM} and differential current, I_{DIFF} of a single 3T cell is shown in Fig. 2.2, where these are defined as:

$$I_{CM} = \frac{I_{D1} + I_{D2}}{2} \tag{2.2}$$

$$\Delta I_{DIFF} = \frac{I_{D1} - I_{D2}}{2} \tag{2.3}$$

For the sense amplifier to function properly, the differential current must be a significant fraction of the common mode current. The average common-mode current is 173 nA, and the standard deviation of the differential current to is 82 nA. This indicates the 3T cell fulfills its goal of providing a large difference in current, which should be easily resolvable by a sense amplifier.

Random ID Cell Layout

The physical layout of the random ID cell is shown in Fig. 2.3, corresponding to the schematic of the ID cell shown in Fig. 2.1. The layout has perfect lateral symmetry, with M1 on the left and M2 on the right, which should ensure that any difference in the electrical properties between devices M1 and M2 is due to random mismatch and not to systematic differences in their layout. Note that device M3 is split into two devices wired in parallel to preserve the symmetry. The distance between M1

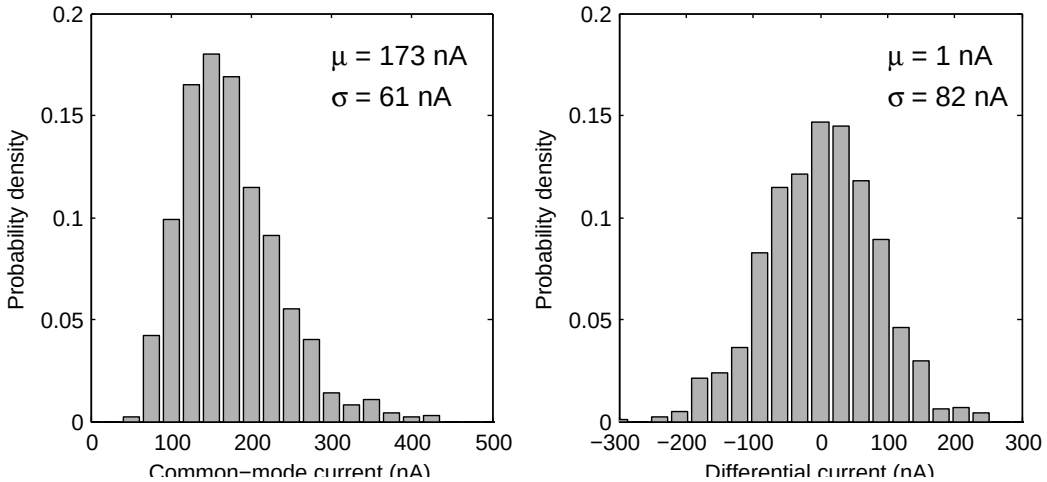


Figure 2.2: Monte Carlo simulation of random ID cell output, showing distribution of common-mode and differential current. The mean, μ, and standard deviation, σ, are indicated. 1000 samples.

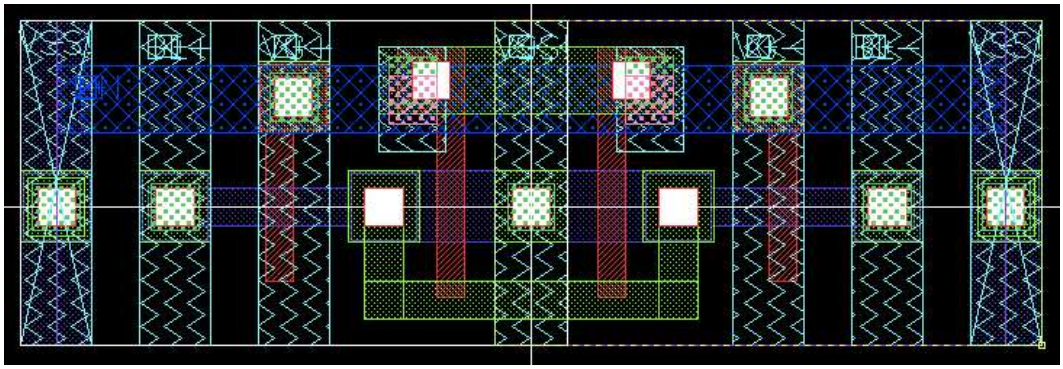


Figure 2.3: Physical layout of random ID cell.

and M2 is made small in order to minimize the effects of any process gradients. The cell is designed to tile horizontally and vertically, where the horizontal tilings share the power, ground, VG+, VG-, BL+ and BL- nets, forming a column. The vertical tiling pitch is 1.36 μm, and the horizontal tiling pitch is 4.00 μm. In principle, other features of the IC near the random IC cell could effect the values of V_T in the ID cell. This could be mitigated by tiling "dummy" cells around the actual row or array of ID cells intended for use. However, as will be shown in Sec. 2.7, no unusual behavior is observed around the edges of the array.

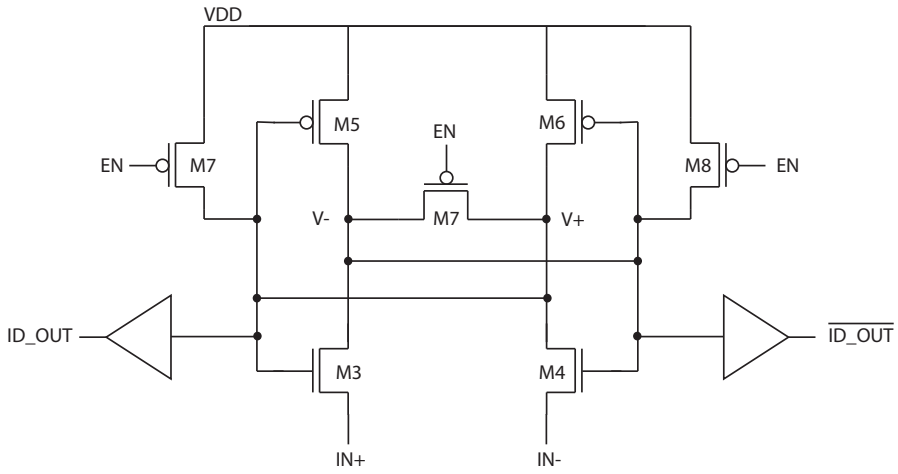


Figure 2.4: Schematic of the latch-based sense amplifier in Fig. 2.1.

2.3.2 ID Generator Periphery and Organization

As shown in Fig. 2.1, several ID cells can be arranged in a column. The organization of the ID cells is similar to a memory. There is no need for random access; ID cells can be read sequentially, which allows for the row enable signals to be generated by a simple shift register rather than a row decoder.

A column of multiple ID cells can share a single sense amplifier (SA). The SA is a latch-type implementation as shown in Fig. 2.4. This topology is preferred since it draws no static power once the latch has settled to its final state. Additionally, sharing the SA across multiple ID cells amortizes any leakage current through the SA.

It is important that the sense amplifier exhibit little variation, since any variation it does exhibit will effect the reading of every ID cell in the column. If the SA has a large input referred offset due to variation, this could bias the output of the column away from the ideal value of $p = 0.5$. By making the devices that compose the SA large relative to the devices in the ID cell, this problem is mitigated. A Monte Carlo

simulation, consisting of 100 runs of the SA with variation was performed, in which an input signal with a common-mode current of 170 nA and a differential current of 4 nA was applied to the SA input. In all cases the SA produces the correct result. Note that the applied differential current of 4 nA is just 5% of the standard deviation of the ID cell current, as shown in Fig. 2.2. This indicates that the expected input referred offset of the SA is much lower than the expected differential current of the ID cell.

A potential complication of arranging multiple cells per column is the effect of the leakage of unselected cells. In a 32-bit column, when one bit is selected, the 31 other cells will draw some leakage current on the column. The same threshold variation that leads to a difference current in selected cells will also lead to a differential leakage current in the unselected cells. This current appears in parallel with the differential current of the selected cell. The effect of this may be to skew the ID cell probabilities of a given column away from the ideal of $p = 0.5$. Or, in other words, the leakage may add a correlated component to what should be 32 random and uncorrelated difference currents. However, a Monte Carlo simulation, shown in Fig. 2.5, indicates that the value of the total leakage current of the unselected cells is much smaller than the expected 3T cell difference current. The standard deviation of the differential leakage current is 13 pA, while the standard deviation of the difference current of a selected cell is 82 nA, over three orders of magnitude larger.

2.3.3 Noise Analysis

In the absence of electrical noise, reading from a single ID cell would be entirely deterministic. However, when the sense amplifier is activated, the total difference between the ID cell currents is the combination of the inherent offset due to device

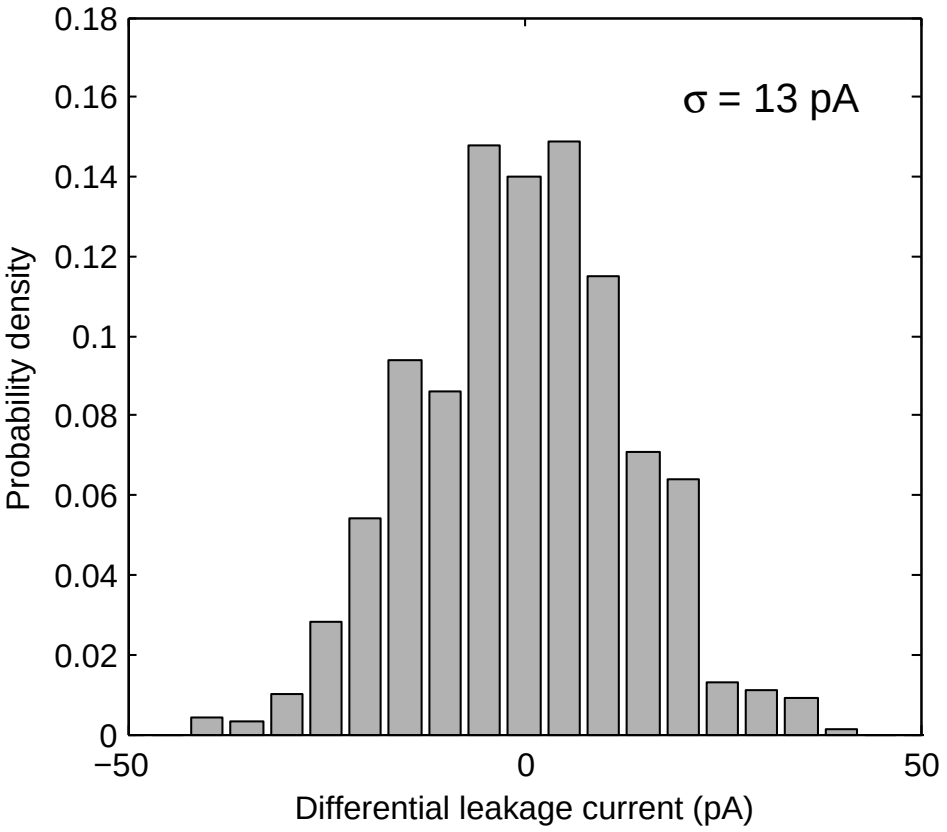


Figure 2.5: Monte Carlo simulation of the combined differential leakage current of 32 ID cells. 1000 samples.

mismatch and any electrical noise, such that

$$\Delta I = \Delta I_m + \Delta I_n \tag{2.4}$$

The internal nodes of the sense amplifier, V_+ and V_-, are pre-charged to VDD prior to sensing, and devices M5 and M6 are off at the start of a read operation. When the sense amplifier is activated, the internal nodes are discharged by the ID cell current. When one of the nodes discharges sufficiently to turn on M5 or M6, the positive feedback is engaged and the amplifier settles to a stable state. A simulation of this is shown in Fig. 2.6, showing the three stages of operation.

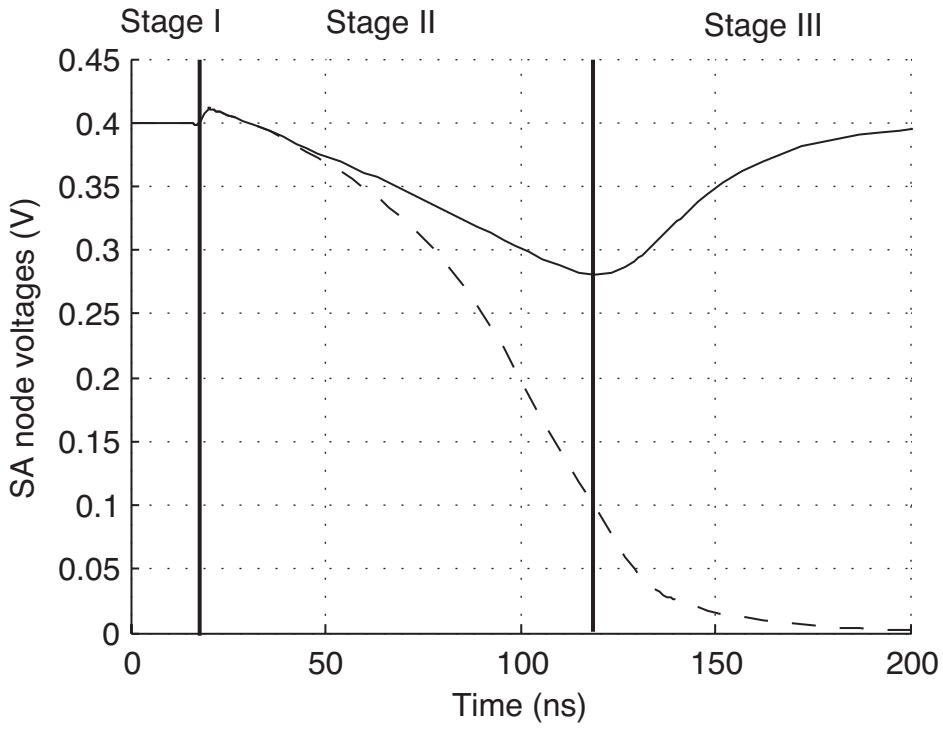


Figure 2.6: Simulated operation of SA. Stage I: Pre-change, Stage II: integration, Stage III: positive feedback.

The arrangement can be modeled as an integrator, where the difference current ΔI is integrated on the parasitic capacitances of nodes 1 and 2, for a time T_i before either M1 or M2 is turned on and the positive feedback is activated.

The total difference voltage ΔV is the sum of the individual difference voltages due to mismatch and noise.

$$\Delta V = \Delta V_m + \Delta V_n \tag{2.5}$$

The relationship between the integrated difference voltage and the input difference current is

$$\Delta V = \frac{1}{C} \int_0^{T_i} \Delta i(t) dt \tag{2.6}$$

This can be modeled as a transfer function of the form

$$h(t) = \frac{1}{C}\text{rect}\left(\frac{t}{T_i}\right) \tag{2.7}$$

with Fourier transform

$$H(jw) = \frac{T_i}{C}\text{sinc}\left(\frac{T_i\omega}{2}\right) \tag{2.8}$$

The mean-squared voltage difference due to thermal noise can then be found

$$\overline{\Delta V_{n,th}^2} = \int_0^\infty H(jw)^2 \frac{\overline{\Delta i_d^2}}{\Delta f} d\omega \tag{2.9}$$

$$\overline{\Delta V_{n,th}^2} = \frac{4kT\gamma g_m T_i}{C^2} \tag{2.10}$$

where

$$\frac{\overline{\Delta i_d^2}}{\Delta f} = 8kT\gamma g_m \tag{2.11}$$

Note that $\overline{\Delta i_d^2}$ is twice the value of the thermal noise power of a single FET [43], since it is the combination of the noise power from devices M1 and M2.

In addition to thermal noise, 1/f noise significantly effects the behavior of the circuit. The 1/f noise current is modeled by:

$$S_{1/f}(f) = \frac{S_0}{f} \tag{2.12}$$

Substituting Eqn. 2.12 into Eqn. 2.9 leads to an integral that does not converge. In fact, there is not complete agreement on the best way to model 1/f noise in integrators [44]. As an approximation, we model the noise voltage by

$$\overline{\Delta V_{n,1/f}^2} = \frac{8S_0T_i^2\ln 2}{C^2} \tag{2.13}$$

similar to the result in [44]. The RMS value of the total noise can then be found.

$$\overline{\Delta V_n^2} = \overline{\Delta V_{n,th}^2} + \overline{\Delta V_{n,1/f}^2} \tag{2.14}$$

$$\sigma_{\Delta V_n} = \frac{1}{C}\sqrt{4kT\gamma g_m T_i + 8S_0T_i^2\ln 2} \tag{2.15}$$

The ΔV_T between devices M1 and M2 creates a difference voltage between the sensing nodes. Because the difference current due to mismatch does not vary with time, it is simply

$$\Delta V_m = \frac{T_i g_m \Delta V_T}{C} \tag{2.16}$$

The effect of the mismatch and noise on the circuit output can be calculated. The probability of ΔV being positive at the end of the integration is equal to the probability that:

$$0 < \Delta V_m + \Delta V_n \tag{2.17}$$

This is given by:

$$p = P\left(\Delta V_n > -\Delta V_m\right) \tag{2.18}$$

$$= \int_{-\Delta V_m}^{\infty} \Phi_{\Delta V_n}(v)dv \tag{2.19}$$

$$= \frac{1}{2}\left[1 + \text{erf}\left(\frac{\Delta V_m}{\sqrt{2}\sigma_{\Delta V_n}}\right)\right] \tag{2.20}$$

where $\Phi_{\Delta V_n}(v)$ is the probability density function of the normally distributed random variable ΔV_n, with variance $\sigma_{\Delta V_n}$.

Finally, this can be rewritten in terms of the equivalent input noise, σ_{V_T}.

$$p = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\Delta V_T}{\sqrt{2}\sigma_{V_T}}\right)\right] \tag{2.21}$$

$$\sigma_{V_T} = \frac{1}{g_m}\sqrt{4kTg_m\frac{1}{T_i} + 8S_0\ln 2} \tag{2.22}$$

The value of σ_{V_T} can then be calculated by extracting the appropriate variables from simulation. Using models supplied by the foundry the following parameters were extracted: $g_m = 3.2\ \mu$A/V, $T_i = 100$ nS, and $S_0 = 3.1 \times 10^{-18}$ A^2 Hz. Additionally, T = 290 K, k = 1.38×10^{-23}, and $\gamma = 2/3$. Using Eqn. 2.22, this gives $\sigma_{V_T} = 1.3$ mV. As will be shown, this is close to the experimentally determined value of 1.5 mV.

It should be noted that although this model appears to agree reasonably well with the experimentally determined value of σ_{V_T}, the model makes several assumptions that are not strictly valid. The most significant of these is modeling the behavior of the sense amplifier as a linear, time-invariant system, when in fact the operation of a sense amplifier is neither linear nor time-invaraint. For this reason, it is suggested that the model only be used as an estimator of σ_{V_T}.

2.3.4 ID Cell Reliability

The primary advantage of characterizing the input referred noise, σ_{V_T}, is that it allows for the reliability of individual ID cells to be determined with a small number of

$\Delta V_T < -V_R$	$\Delta V_T > V_R$	Classification	C	M
True	False	'0'	0	1
False	False	Unreliable	X	0
False	True	'1'	1	1

Table 2.1: Classification of ID cells based on reliability tests, where V_R is the magnitude of the threshold voltage difference required for reliable operation, C is ID code, and M is the ID mask.

measurements, whereas, in most previous random ID implementations, the reliability of ID cells can only be determined by reading from each cell a large number of times. This is accomplished by discriminating between cells based on whether the magnitude of their threshold voltage mismatch, ΔV_T, is greater than the threshold voltage mismatch required for reliable operation, V_R. Given a threshold for reliability, ϵ, and substituting $\Delta V_T = V_R$ and $p = 1 - \epsilon$ into Eqn. 2.21 and solving for V_R gives

$$V_R = \sqrt{2}\sigma_{V_T}\text{erf}^{-1}(1 - 2\epsilon) \tag{2.23}$$

The reliability of an ID cell can then be determined by performing two tests. In the first test, a difference voltage $+V_R$ is applied between the gates of M1 and M2, such that $V_{G1} - V_{G2} = V_R$, and the sense amplifier activated. If the output of the SA is '0', then it must be that $\Delta V_T < -V_R$, and the cell is a reliable '0'. In the second test, a difference voltage of $-V_R$ is applied between the gates of M1 and M2. If the output is '1', it must be that $\Delta V_T > V_R$, and the cell is a reliable '1'. If the cell is neither a reliable '0' nor a reliable '1' then it is unreliable. This classification system is shown in Table 2.1.

2.4 Addressing Procedures

As discussed previously, the output of the ID generator circuit may change between power-on cycles: if the V_T mismatch between the devices within an ID cell is small, noise currents may cause the output bit associated with the cell to vary between power-on cycles in a non-deterministic fashion. This might cause several types of identification failures when the die are addressed by an external system. The focus of this section is to analyze several addressing procedures, which are designed to reliably identify chips even in the presence of unreliable bits.

2.4.1 Overview

All addressing procedures described here share some common elements. Each chip is understood to have a *chip code*, an N-bit number based on the physical structure of the chip, which never changes over the lifetime of the chip. Each chip contains an *ID generator* that, on power-on, outputs an N-bit *generated code*. Ideally, the *generator code* would always be equal to the *chip code*, but in general this is not always the case since the ID generator does not behave reliably. A collection of chips is a *population*, and no two chips in the same population should have the same chip code.

A population of chips is addressed by an *external reader*, which attempts to identify a *target chip* with a particular chip code by broadcasting a *chip address*. Generally, a chip address will be a set of one or more chip codes. Identification can be done either externally or on-chip. in *external identification*, positive identification of a chip is not complete until a chip address is broadcast to the population, one or more chips reply with some information, and the external reader makes a determination about the identity of the chips. Conversely, in *on-chip* identification, a given chip is able to positively determine whether or not is is being addressed without transmitting any

data back to the external reader.

Each addressing procedure can be characterized by several performance metrics:

Code length N The number of bits that make up the chip code.

False-negative rate p_{FN} The probability that a target chip, A, is not positively identified during an addressing operation.

False-positive rate p_{FP} The probability that a second chip, B, is erroneously identified as the target chip, A, during an addressing operation.

Total error rate p_E The combined probability of a false-negative or false-positive error.

$$p_E = (1 - p_{FN})(1 - p_{FP}) \tag{2.24}$$

Maximum population size S The maximum number of chips that can be uniquely addressed.

Yield Y The fraction of manufactured chips that are suitable for use.

The general objective of all addressing procedures is to uniquely address chips in such a way that in the presence of unreliable ID bits, the identity of the chips can still be reliably determined.

To clarify the above discussion, several example addressing procedures are presented below.

2.4.2 External Identification Procedures

In external identification addressing, individual chips are unable to positively determine whether they are being addressed; only the external reader is able to make this

determination after aggregating the responses of multiple chips.

Broadcast Addressing

To identify a target chip with a particular chip code in broadcast addressing, the external reader queries all chips in the population, and all chips reply with their generated code. The reply who's generated code has the smallest Hamming distance from the chip code is determined to come from the target chip. This is the procedure used in most previously published random identification schemes [34] [36]. This has the advantage of being relatively tolerant of multiple ID bit errors, but has the disadvantage of requiring the chip to transmit its generated code back to the external reader.

2.4.3 On-Chip Identification Procedures

The following is a brief overview of three examples of on-chip identification procedures: procedures whereby a given chip can determine whether or not it is being addressed. A detailed analysis of these procedures follows in Section 2.6.

Direct Addressing

In direct addressing, the external reader communicates with individual chips by simply broadcasting the chip code. Using the above terminology, the chip address is equal to the chip code. The chip then compares the received chip code to its generated code: if they are equal the chip concludes that it has been addressed, and if they are unequal the chip concludes it has not been addressed. Although having the virtue of simplicity, note that the false-negative rate may become large for large code lengths, since even a difference of a single bit between the generated code and the

chip code results in a false-negative.

Hamming Addressing

In Hamming addressing, the external reader transmits an N-bit chip code along with a particular Hamming distance D. The chip then determines whether the received chip code varies from its generated code by less than or equal to D bits. If so, the chip concludes that it is being addressed. In this procedure the chip address comprises a block of codes, containing all possible codes within hamming distance D of the chip code. In this way, the false-negative rate is reduced since the system can tolerate up to D ID bit errors. Note that the direct addressing procedure is a special case of Hamming addressing with $D = 0$.

Masked Addressing

In masked addressing, the external reader transmits an N-bit chip code along with an N-bit mask. The chip compares the received chip code to its generated code only for the bits indicated by the mask, and if they are equal concludes that it is being addressed. In this addressing procedure, the chip address contains all codes that differ from the chip code only among the bits not selected by the mask. This approach may be more efficient, in the sense that the false-negative rate can be decreased with a smaller increase in the false-positive rate, than using Hamming addressing, assuming the some bits of the ID generator output are more reliable than others and that these bits can be identified in advance. Note again that direct addressing is a special case of masked addressing with all the mask bits equal to '1'.

2.5 ID Generator Model

Each chip contains an ID generator, which generates an N-bit code when power is applied. The output of the ID generator is modeled as a discrete, N-bit random variable, which we will refer to as the generated code G, where each bit is modeled as a Bernoulli trial with probability mass function:

$$f_X(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases} \tag{2.25}$$

where p is the probability that a given ID cell will evaluate to '1'. Each bit G_i has its own probability p_i, and a particular ID generator is completely described by a sequence of N probabilities $(p_0, p_1, ..., p_{N-1})$. The N-bit chip code, C, can then be defined:

$$C_i = \begin{cases} 1 & p_i > 1/2 \\ 0 & p_i \leq 1/2 \end{cases} \tag{2.26}$$

In other words, the chip code C is the code most likely generated by the ID generator.

2.5.1 ID Cell Model

As discussed previously, the ID cell probability is the result of physical variation between MOSFETs in the cell. These physical variations create an offset current ΔI_m, between the cell FETs. The function of the sense amplifier is to determine the polarity of ΔI_m, but it does not do this reliably due to the presence of input-referred noise ΔI_m. Both ΔI_m and ΔI_n are modeled here as random variables with normal distribution. Note that although both ΔI_m and ΔI_n are random variables,

they correspond to different random events. The random variable ΔI_m is sampled once for each ID cell, at the time of manufacture, whereas the random variable ΔI_n is sampled every time the ID cell is read from.

Ideally, the ID cell probability, p, would depend only on ΔI_m, and could be modeled as a discrete random variable with probability mass function:

$$p = \begin{cases} 1 & \Delta V_T > 0 \\ 0 & \Delta V_T \leq 0 \end{cases} \tag{2.27}$$

And since I_m is normally distributed with zero mean, p could then be described by a discrete probability distribution with probability mass function:

$$f_P(p) = \begin{cases} 1/2 & p = 1 \\ 1/2 & p = 0 \end{cases} \tag{2.28}$$

This is the ideal distribution of p for the ID cells: half the cells always evaluate to '1', the other half always to '0'.

However, when the ID cell is activated, the sense amplifier samples the random variable ΔI_m in the presence of noise, so that values of p are not necessarily '0' or '1', but some value in between. This was described previously in Sec. 2.3.3. However, because variance of ΔV_T is much larger than the variance of the input referred noise, the distribution of p values is most dense around '0' and '1'. In other words, most ID cells will have manufactured offsets larger than any electrical noise. This suggests the following approximation of the ID cell probabilities:

$$p' = \begin{cases} 1-\epsilon & 1-\epsilon \leq p \leq 1 \\ 1/2 & \epsilon < p < 1-\epsilon \\ \epsilon & 0 \leq p \leq \epsilon \end{cases} \tag{2.29}$$

Plainly, we designate two categories of ID cells based or their value of p: cells whose value of p is within some range ϵ of 0 or 1 are said to be *reliable*, and ID cells outside this range *unreliable*. Reliable ID cells of value '0' or '1' are treated as either evaluating to '0' or '1', with probabilities $1-\epsilon$. Unreliable cells are treated as completely random, evaluating to either 0 or 1 with equal probability. This approximation greatly simplifies the following analytic analysis of addressing procedures. Also note that the approximation is extremely conservative, and the error rates calculated in the following sections represent an upper bound on the error rates likely encountered in actual practice.

2.5.2 ID Generator Approximation

The output of the ID generator can be modeled as an N-bit number whose bits can be categorized into three types: reliable bits that always evaluate to 1, reliable bits that always evaluate to 0, and unreliable bits that evaluate to either 0 or 1 with equal likelihood. Every ID generator will have some number of unreliable bits u, where u is a discrete random variable between 0 and N with binomial distribution. If p_U is the likelihood a given ID cell being unreliable, then the probability mass function of u is:

$$f_U(u) = \binom{N}{u} p_U^u (1-p_U)^{N-u} \tag{2.30}$$

This distribution is directly related to the chip yield: chips with low values of u

can be correctly identified with high probability and are suitable for use, while chips with large values of u may be very difficult to identify and are unusable.

The probability and number of bit flips, l, will prove to be important in the analysis that follows. For a given N-bit ID generator with u unreliable bits, the number of bit flips among the reliable and unreliable bits, l_R and l_U, are random variables with probability mass functions:

$$f_{L_R}(l_R) = \binom{N-u}{l_R} \epsilon^{l_R}(1-\epsilon)^{N-u-l_R} \tag{2.31}$$

$$f_{L_U}(l_U) = \binom{u}{l_U} (1/2)^{l_U} \tag{2.32}$$

The probability mass function of the total number of bit flips, $l = l_R + l_U$, is then given by:

$$f_L(l) = \sum_{i=0}^{l} f_{L_R}(l-i) f_{L_U}(i) \tag{2.33}$$

Substituting Eqns. 2.31 and 2.32 into Eqn. 2.33:

$$f_L(l) = \sum_{i=0}^{l} \binom{N-u}{l-i} \binom{u}{i} \epsilon^{l-i}(1-\epsilon)^{N-u-l+i}(1/2)^i \tag{2.34}$$

2.6 Addressing Procedure Performance

Because the ID generator is unreliable, it is necessary to rigorously show that chips can still be identified with a high degree of accuracy. We will show that the rates of false-positives and false-negatives can be reduced to arbitrarily low levels with the appropriate choice of addressing procedure and appropriate preselection of chips.

The price of this reliability will be its impact on minimum ID code length, addressing overhead, chip yield, and maximum population size.

Because the number of unstable bits, u, varies among chips, each chip has its own particular value of p_{FN} and p_{FP}. One possible set of metrics would be the average rates of false-positives and false-negatives. While this may be appropriate for some applications, in many others it is inadequate, since fixing a maximum value of $\overline{p_{FN}}$ and $\overline{p_{FP}}$ allows for the possibility of individual chips that cause errors at a much higher rate than $\overline{p_{FN}}$ or $\overline{p_{FP}}$. For this reason, rather than focusing on average error rates, we will adopt worst-case error rates as our metric. This guarantees that no chip among a given population will have an identification error rate greater than the thresholds p_{FN} and p_{FP}.

2.6.1 Direct Addressing

In direct addressing, the chip address contains only the target chip code. A difference of only a single bit between the N-bit chip code C and the generated code G will result in a false-negative. For this reason, the only way to ensure low rates of false-negatives is to only allow the use of chips with no unstable bits. The yield, Y, is then be found using Eqn. 2.30, with $u = 0$.

$$Y = P(U = 0) = f_U(0) = (1 - p_U)^N \tag{2.35}$$

The false-negative rate, p_{FN} is then equal to the probability that one or more of the target chip's bits deviate from their usual value. Because $u = 0$, this is given by:

$$p_{FN} = 1 - f_{flips,R}(0) \tag{2.36}$$

Substituting Eqn. 2.31 into Eqn. 2.36 with $u = 0$ gives:

$$p_{FN} = 1 - (1 - \epsilon)^N \tag{2.37}$$

If $(1 - \epsilon)^N$ term of Eqn. 2.37 is expanded, it can be seen to contain higher order terms of ϵ from ϵ^1 to ϵ^N. Because ϵ is very small, these higher order terms can be ignored and Eqn. 2.37 can be approximated by

$$p_{FN} \approx N\epsilon \tag{2.38}$$

A virtue of direct addressing is that with no unstable bits, the maximum population, M, is simply equal to the size of the N-bit code space:

$$M = 2^N \tag{2.39}$$

A false-positive will occur if any of the chips that are not the target chip have a generated code, G, equal to the target chip code, C. This is more likely for chips who's chip code has a small Hamming distance from the target chip code. For example, consider chips with a chip code which varies only by a single bit from the target code. If this bit is flipped, and the other bits remain unflipped, a false-positive will occur. This occurs for a given chip with probability

$$p'_{FP,H=1} = \epsilon(1 - \epsilon)^{N-1} \tag{2.40}$$

Because ϵ is normally chosen to be quite small, in actual practice the second term, $(1 - \epsilon)^{N-1}$ is very nearly equal to 1, and so an approximation can be made

$$p'_{FP,H=1} \approx \epsilon \tag{2.41}$$

In other words, we will assume that if the differing bit is flipped, the other bits remain unchanged.

There are N chips who's chip code varies by one bit from the target chip. That is, there are N chips that have chip codes a Hamming distance of 1 from the target chip, which makes the total probability of a false-negative not occurring among the $H = 1$ chips equal to $(1 - \epsilon)^N$. Similarly, for chips codes that vary by two bits from the target chip code, the probability of both of these bits flipping is ϵ^2. Because there are $\binom{N}{2}$ chips with Hamming distance 2 from the target chip, the total probability of a false-negative not occurring among the $H = 2$ chips is equal to $(1 - \epsilon^2)^{\binom{N}{2}}$. The same calculation can be extended to all possible Hamming distances, from 1 to N, giving the total false-positive rate:

$$p_{FP} = 1 - \prod_{h=1}^{N} (1 - \epsilon^h)^{\binom{N}{h}} \tag{2.42}$$

Again, because ϵ is normally chosen to be small, an approximate solution to this equation can be found by ignoring the higher order terms of ϵ^h. In other words, we can consider only the case of single bit flips, since multiple bit flips are far less likely. In this case only the $h = 1$ term is important, and the false-positive rate reduces to

$$p_{FP} \approx 1 - (1 - \epsilon)^N \tag{2.43}$$

This can be further approximated just as in Eqn. 2.37 to

$$p_{FN} \approx N\epsilon \tag{2.44}$$

2.6.2 Hamming Addressing

In Hamming addressing, the chip address contains the target chip code and all codes within some Hamming distance, D, of that code. In this way, positive identification is still possible even in the presence of up to D unstable bits. This may significantly improve the yield, Y, since now any chip with $u \leq D$ is acceptable for use.

$$Y = P(U \leq D) = \sum_{u=0}^{D} f_U(u) \tag{2.45}$$

Substituting Eqn. 2.30 into Eqn. 2.45 gives

$$Y = \sum_{u=0}^{D} \binom{N}{u} p_U^u (1 - p_U)^{N-u} \tag{2.46}$$

A false-negative will occur if the chip has a number of bit flips, l greater than D.

$$p_{FN} = P(L > D) = \sum_{l=D+1}^{N} f_L(l) \tag{2.47}$$

Substituting Eqn. 2.33 into Eqn. 2.47:

$$p_{FN} = \sum_{l=D+1}^{N} \sum_{i=0}^{l} \binom{N-u}{l-i} \binom{u}{i} \epsilon^{l-i} (1-\epsilon)^{N-u-l+i} (1/2)^i \tag{2.48}$$

The worst-case false-negative rate will occur for chips that have $u = D$. Substituting this into Eqn. 2.47 gives:

$$p_{FN} = \sum_{l=D+1}^{N} \sum_{i=0}^{l} \binom{N-D}{l-i} \binom{D}{i} \epsilon^{l-i} (1-\epsilon)^{N-D-l+i} (1/2)^i \tag{2.49}$$

For values of ϵ approaching zero, the terms of the summating containing higher

orders of ϵ^{l-i} become very small. The largest of these terms has $l = D+1$ and $i = D$, and can be used as an approximation of Eqn. 2.49:

$$p_{FN} \approx \frac{N-D}{2^D}\epsilon \tag{2.50}$$

A disadvantage of Hamming addressing is that the maximum population is significantly smaller than that of direct addressing. Consider a target chip A with chip code C_A, and another chip B with chip code C_B, which in the worst case will have D unstable bits. In Hamming addressing, the chip address for chip A will be all codes, C, within Hamming distance D of C_A: all C such that

$$H(C_A, C) \leq D \tag{2.51}$$

If chip B has D unstable bits, then the distance between its chip code C_B and all the address codes must be greater than D.

$$H(C_B, C) > D \tag{2.52}$$

Combining Eqns. 2.51 and 2.52 gives the requirement

$$H(C_A, C_B) > 2D \tag{2.53}$$

Eqn. 2.53 is the reason for the small maximum population of Hamming addressing. Rather than densely packing chip codes within the N-bit code space, chips with D unreliable bits must have codes separated from all other chip codes by at least $2D+1$ bits. The maximum population is equal to the maximum number of N-bit codes with mutual Hamming distance of at least $2D+1$, which is given by

$$M = 2^{\mathrm{Int}\left(\frac{N}{2D+1}\right)} \tag{2.54}$$

where $\mathrm{Int}(x)$ is the whole part of x.

Any chip that produces a generated code within hamming distance D of the target code will generate a false-positive. The chips that are likeliest to do this are the chips with Hamming distance $D+1$ from the target chip. If any of the $D+1$ differing bits flip, this will move the generated code to within D of the target chip, producing an error. For a given, chip the likelihood of this occurring is

$$p'_{FP} = 1 - (1-\epsilon)^{D+1} \tag{2.55}$$

$$p_{FP} \approx (D+1)\epsilon \tag{2.56}$$

In the worst case, there are $\binom{N}{D+1}$ codes within Hamming distance $D+1$ of the target codes. However, this number is usually much larger than the maximum population size, M. Therefore in actual practice, there are a maximum of $M-1$ chips that could cause a false positive. The combined probability of a false positive due to any of these chips is therefore

$$p_{FP} \approx 1 - \big(1 - (D+1)\epsilon\big)\big)^{M-1} \tag{2.57}$$

or, ignoring the higher terms of ϵ, and substituting Eqn. 2.54

$$p_{FP} \approx (2^{\mathrm{Int}\left(\frac{N}{2D+1}\right)} - 1)(D+1)\epsilon \tag{2.58}$$

2.6.3 Masked Addressing

Addressing performance is improved if it is determined in advance which bits of each chip are reliable and which are unreliable. This information can be broadcast, along with the target chip code, to a population of chips, and the target chip can then exclude its unstable bits from comparison. Specifically, the external reader will broadcast two N-bit numbers: the target chip code, C_A, and the target chip mask, M_A, where each bit of M_A is equal to 1 if chip A's corresponding ID generator bit is stable, and 0 if the bit is unstable. If chip A's generated code is G_A, then a positive identification will occur if:

$$C_A \ \& \ M_A = G_A \ \& \ M_A \tag{2.59}$$

Like Hamming addressing, this allows for an improvement in the yield over direct addressing, since any chip with up to D unstable bits is now acceptable for use. The yield, Y, is:

$$Y = \sum_{u=0}^{D} \binom{N}{u} p_U^u (1 - p_U)^{N-u} \tag{2.60}$$

Because the unreliable bits are excluded from comparison, only an error in one of the reliable bits can cause a false-negative. In the worst-case, the likelihood is highest for a chip with a number of unstable bits $u = D$, and is equal to the probability that one or more of the $N - D$ reliable bits are flipped:

$$p_{FN} = 1 - f_{L_R}(0) \tag{2.61}$$

Substituting Eqn. 2.31 into Eqn. 2.61, with $u = D$ gives the worst-case false-

negative rate:

$$p_{FN} = 1 - (1 - \epsilon)^{N-D} \tag{2.62}$$

or approximately

$$p_{FN} = (N - D)\epsilon \tag{2.63}$$

To find the maximum population, consider that broadcasting an N-bit chip code with an N-mask with D unreliable bits is equivalent to broadcasting 2^D unique N-bit codes. Because there are a total of 2^N possible codes, the total number of non-overlapping addresses is therefore

$$M = \frac{2^N}{2^D} \tag{2.64}$$

$$M = 2^{N-D} \tag{2.65}$$

Errors among reliable bits could also cause false-positive identifications. Calculating the combined probability of any false-positive event is difficult, so we will restrict the calculation to single bit errors, since the probability of E bit errors is ϵ^E, which decreases rapidly for $E > 1$. For $E = 1$, only chips with codes within Hamming distance 1 of the target chip code can cause false-positives. In the worst case, for an N bit code with U unreliable bits, there are $(N - U)2^U$ possible chip codes within a Hamming distance of 1. However, this number is usually much larger than the maximum number of uniquely addressable chips, M, given by Eqn. 2.65. In actual practice, therefore, there are a maximum $M - 1$ other chips that could cause false-positive errors. The probability of any of these chips causing a false-positive is

then

$$p_{FP} = 1 - (1 - \epsilon)^{2^{N-U}-1} \tag{2.66}$$

or approximately

$$p_{FP} = (2^{N-U} - 1)\epsilon \tag{2.67}$$

The procedures for chip characterization and chip identification under the masked addressing scheme are given below.

Chip Characterization Procedure

Chip characterization occurs once for each chip, recording C and M, which are necessary to identify the chip in the future.

1. A single chip A, to be characterized, is placed in range of the external reader. The external reader transmits a characterize signal.
2. The chip applies a voltage difference ΔV_R to the ID cells, activates the ID generator, and records the output G_+.
3. The chip applies a voltage difference $-\Delta V_R$ to the ID cells, activates the ID generator, and records the output G_-.
4. The chip transmits G_+ and G_- to the external reader.
5. The external reader computes:

$$C_A = \overline{G_+} \ \& \ G_- \tag{2.68}$$

$$M_A = G_+ \ || \ G_- \tag{2.69}$$

6. The external reader stores C_A and M_A for chip A.

Chip Identification Procedure

Chip identification occurs when the system needs to locate a particular chip from a group of chips. Following identification, the chip can reply with a simple acknowledgement, and any other data to be collected.

1. To identify chip A, the external reader transmits $C_A \ \& \ M_A$ and M_A. All chips within range receive this message.

2. Each chip activates its ID generator, with zero voltage difference applied to the ID cells, and records the output G.

3. Each chip evaluates the statement

$$C_A \ \& \ M_A = G \ \& \ M_A \tag{2.70}$$

4. If the preceding step evaluates to true, the chip concludes it is being addressed.

2.6.4 Comparison

The performance metrics of the direct, Hamming, and masked addressing protocols are shown in Table 2.2. The most critical metric is the yield, Y, since this has a direct impact on manufacturing costs. Assuming $N = 10$ and a fairly low value of $p_U = 0.05$, the fraction of unreliable bits, the yield of the direct addressing scheme is less than 60%. The Hamming and masked addressing protocols have the same expression for their yields, which depend on both N and D, the maximum allowed number of unreliable bits. With appropriate selection of N and D, the yield can be

Table 2.2: Comparison of addressing protocol metrics: yield (Y, maximum population (M), false-negative rate (p_{FP}), and false-positive rate (p_{FP}).

	Direct	Hamming	Masked
Y	$(1-p_U)^N$	$\sum_{u=0}^{D} \binom{N}{u} p_U^u (1-p_U)^{N-u}$	$\sum_{u=0}^{D} \binom{N}{u} p_U^u (1-p_U)^{N-u}$
M	2^N	$2^{\mathrm{Int}\left(\frac{N}{2D+1}\right)}$	2^{N-D}
p_{FN}	$N\epsilon$	$\frac{N-D}{2^D}\epsilon$	$(N-D)\epsilon$
p_{FP}	$N\epsilon$	$(2^{\mathrm{Int}\left(\frac{N}{2D+1}\right)}-1)(D+1)\epsilon$	$(2^{N-U}-1)\epsilon$

made greater than 99.9%. For this reason, the direct addressing protocol is determined to be impractical for real-world use.

The second most important metric is the maximum population size, M. The number of chips that need to be uniquely addressed is likely fixed by the application, and so this metric determines the number of bits N. The amount of energy required by the ID generator is directly proportional to N, so minimizing N is advantageous from an energy perspective. Inspection of the expressions for M shows that masked addressing has unequivocally better performance in this respect than Hamming addressing, for any values of N and D.

The improved performance of the masked addressing protocol compared to the Hamming addressing protocol in terms of population size can be seen graphically in Fig. 2.7. In this case, imagine a single chip with an 4-bit ID of where the first three bits are a reliable '1' and the last bit is unreliable. In other words, the chip may generate either '1110' or '1111' as its ID, and these are marked with X's in plots (a) and (b). To identify this chip with Hamming addressing protocol requires sending a target code, '1111', and a Hamming distance of 1. There are however, 5 codes within Hamming distance 1 of '1111', and these are shaded in gray in plot (a). In the masked

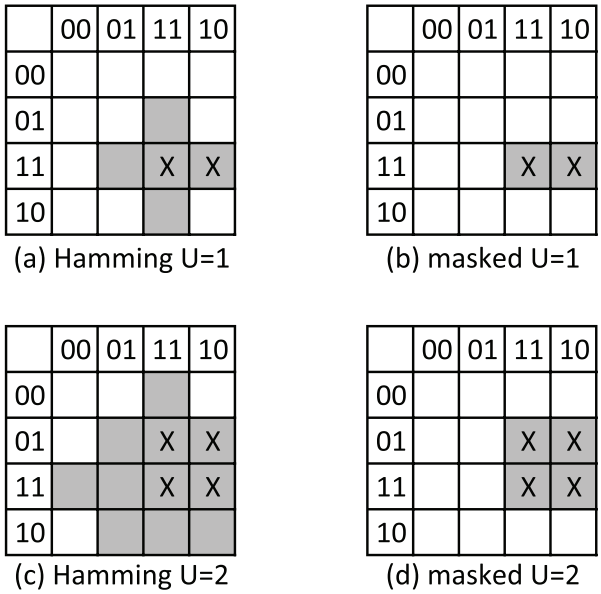


Figure 2.7: Karnaugh maps illustrating relative efficiency of Hamming and masked addressing.

addressing system, a target code '1111' and mask '1110' would be transmitted. This specifies only two possible codes, shaded in gray in plot (b). In other words, the Hamming system is inefficient in the sense that the block of codes specified by the protocol is much larger than the number of possible codes that can be generated. Plots (c) and (d) show the same thing except for a chip with two unreliable bits. Again, the block of codes specified by the Hamming address is much larger than the number of possible codes.

The false-negative and false-positive rates also differ between protocols, but for all protocols they can be made quite small simply by reducing the error threshold, ϵ. For these reasons, masked addressing was chosen as the best performing option.

For the masked addressing system, with $\sigma_{\Delta V_T} = 1.5$ mV, $\epsilon = 10^{-10}$, $p_U = 0.18$, and a required yield of 99.9%, the maximum number of uniquely addressable chips is

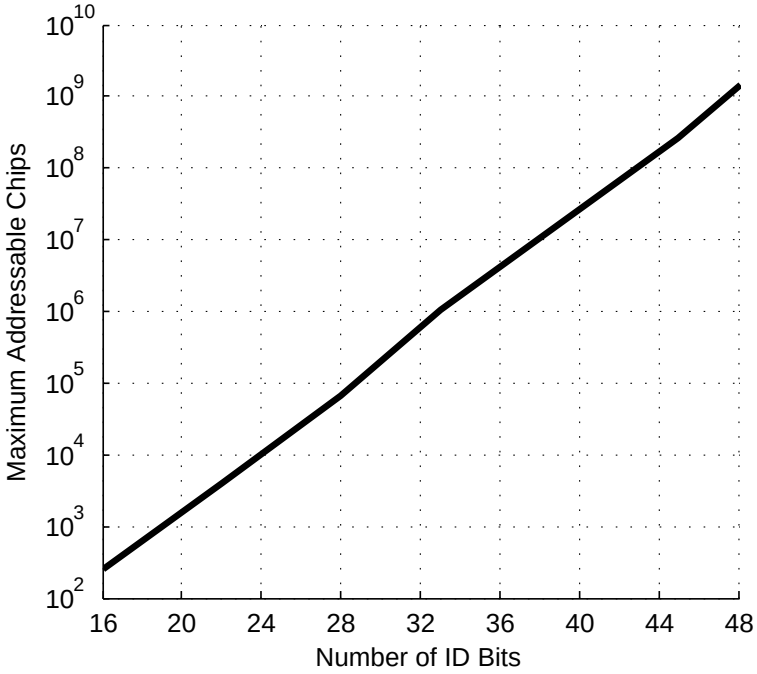


Figure 2.8: Maximum number of addressable chips under the masked addressing system for a given number of random ID bits, N.

shown in Fig. 2.8 (the values of $\sigma_{\Delta V_T}$ and p_U being taken from Sec. 2.7). This shows reasonable performance as the addressing scheme is scaled to a large number of chips.

2.7 Experimental Results

To evaluate the preceding theory, the random ID circuit was fabricated in a standard 130 nm CMOS process. A photograph of the die is shown in Fig. 2.9

The memory-like structure of the circuit allows for many random ID cells to be fabricated on a single die. Random ID cells are organized into columns, with 32 bits per column, and each column sharing a sense amplifier. Each die contains 480 columns, for a total of 15360 random ID cells per chip. A complete 32-bit column, including the SA, occupies an area of 68.5 x 4.3 μm^2. The effective area per bit is 9.2

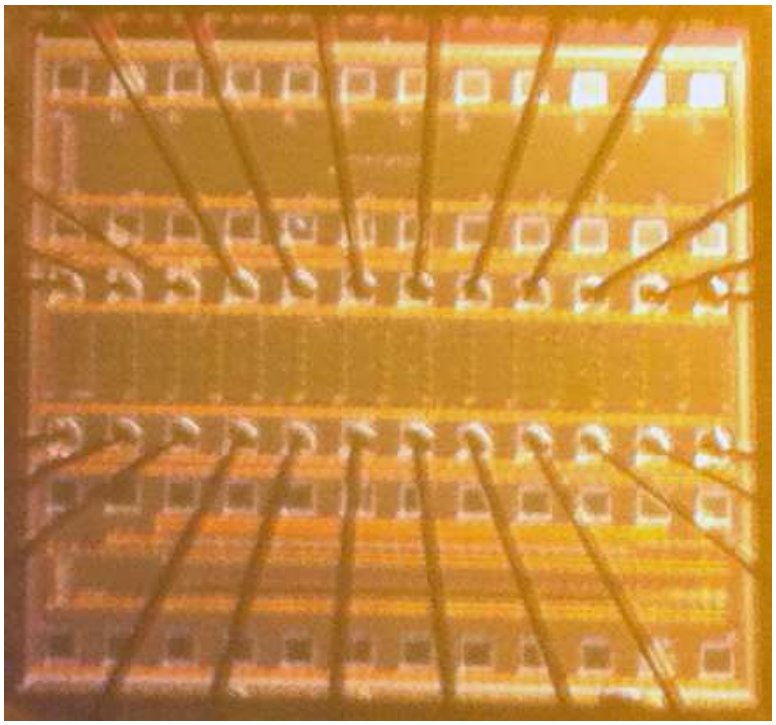


Figure 2.9: Photograph of random ID chip with 15360 ID bits, fabricated in 130 nm CMOS process.

μm^2 (32 bits plus SA divided by total area). For the following experiments, V_{DD} was set to 400 mV.

2.7.1 Method

As discussed in Section 2.3.4, determining the value of the input referred noise, $\sigma_{\Delta V_T}$ allows for the reliability of ID cells to be determined based on their threshold voltage differences, ΔV_T. However, $\sigma_{\Delta V_T}$ represents the combined effects of all noise sources in the ID cell and SA, and cannot be measured directly. Determining $\sigma_{\Delta V_T}$ requires determining the values of p and ΔV_T for a large number of cells, and then fitting Eqn. 2.21 to that data with $\sigma_{\Delta V_T}$ as the free parameter.

This requires two measurements to each cell, the first to determine p and the second to determine ΔV_T. The procedure for determining p is simply to read from

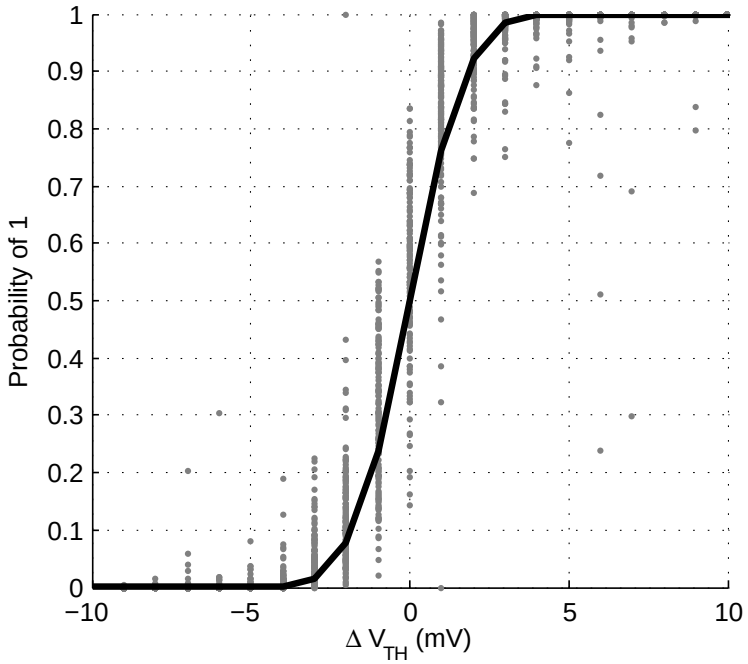


Figure 2.10: Relationship of ID cell probability, p, to threshold voltage mismatch ΔV_T, measured over 15360 bits. Eqn. 2.21 is also plotted with $\sigma_{V_{in}} = 1.5$ mV.

each cell 1000 times and calculate the average value of the output. To determine the value of ΔV_T, a programmable DC voltage source is placed between the gates of M1 and M2, applying a difference voltage ΔV_{DC}. ΔV_{DC} is swept in 1 mV increments, and the output of the ID cell recorded at each increment. To reduce the effect of the input referred noise, the ID cell output is recorded 16 times at each 1 mV increment. The value of ΔV_{DC} that causes the output to flip from '0' to '1' is determined to be that ID cell's value of ΔV_T. The results of these measurements are plotted in Fig. 2.10, along with the best fit of Eqn. 2.21 with $\sigma_{\Delta V_T} = 1.5$ mV.

With $\sigma_{\Delta V_T}$ determined, the value of V_R required to obtain a particular threshold of reliability, ϵ, can be determined. For example, if we choose $\epsilon = 10^{-10}$, we can use 2.23 to determine $V_R = 10$ mV. That is, cells that satisfy the inequality $|\Delta V_T| > 10$ mV will flip from their nominal value during read at a rate less than 10^{-10}, and can

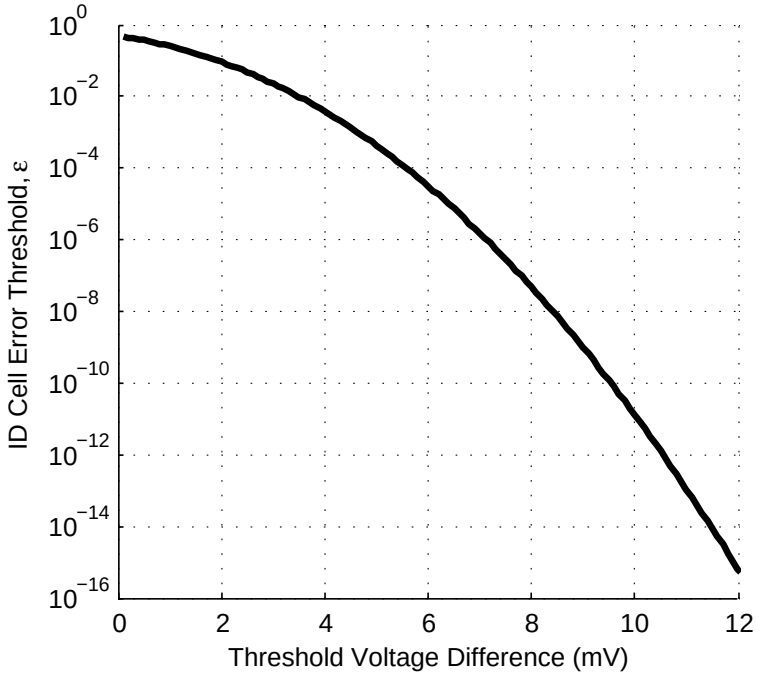


Figure 2.11: Relationship between ID cell error rate, ϵ, and the magnitude of the threshold voltage mismatch in the ID cell, $|\Delta V_T|$.

be said to be reliable. More generally, Fig. 2.11 plots Eqn. 2.23 for $\sigma_{V_T} = 1.5$ mV, which shows the value of value of V_R required for a given threshold of reliability, ϵ.

Of the cells examined, 82% have $\Delta V_T > 10$ mV, ($p_U = 0.18$). Decreasing the threshold of reliability reduces the fraction of bits classified as reliable. This is shown in Fig. 2.12.

The location of unreliable cells for an arbitrarily chosen 32 x 32 array is shown in Fig. 2.13. The unreliable cells are shown in black. The magnitude of ΔV_T is also shown for the same array. This indicates that unreliable bits have random spatial distribution in the array. There is no significant difference between the distribution of ΔV_T around the edges of the array and the array interior, which suggests that edge effects due to nearby features are not significant.

The measured distribution of ΔV_T is shown in Fig. 2.14, along with a Monte

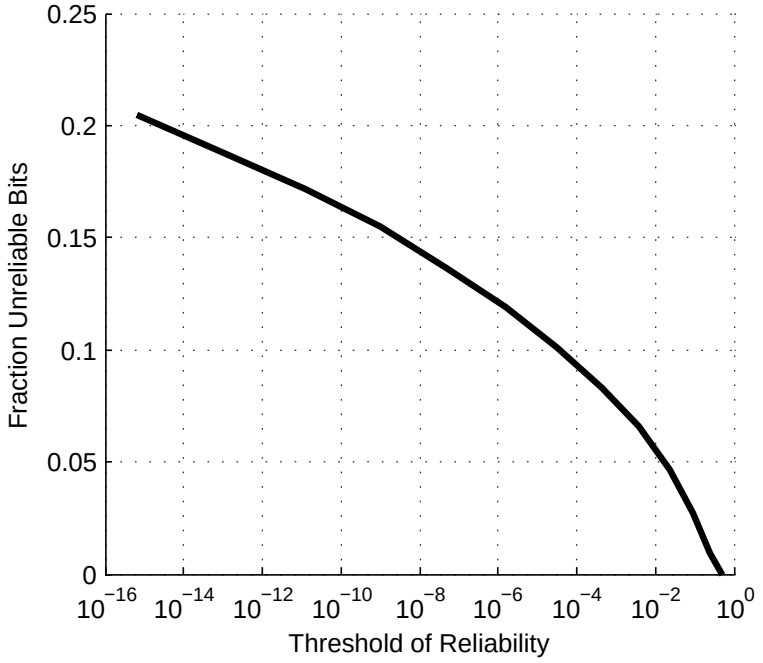


Figure 2.12: Fraction of manufactured bits that can be expected to be unreliable for a given threshold of reliability, ϵ.

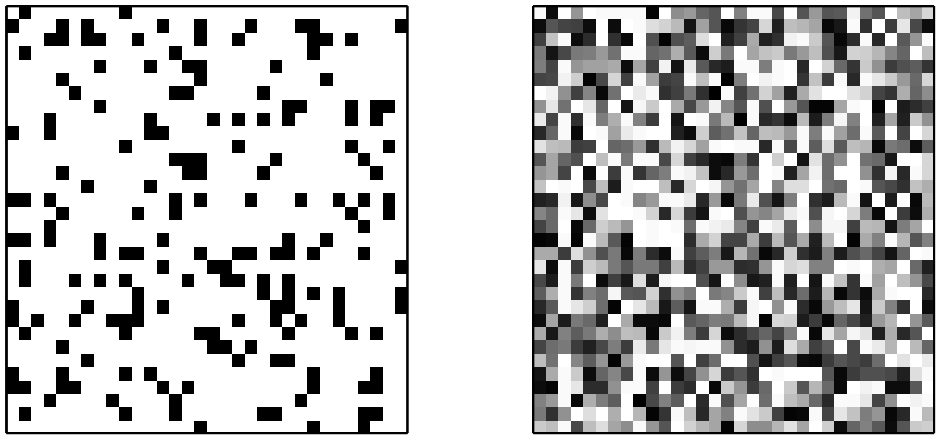


Figure 2.13: (A) Location of unreliable bits ($|\Delta V_T| < 10$ mV), shown in black, in a 32 x 32 array of random ID cells. (B) Magnitude of $|\Delta V_T|$ in the same array, where black is 0 mV and white is 50 mV.

Carlo simulation of the same distribution. This shows the distribution of ΔV_T can be accurately predicted in advance for well modeled processes.

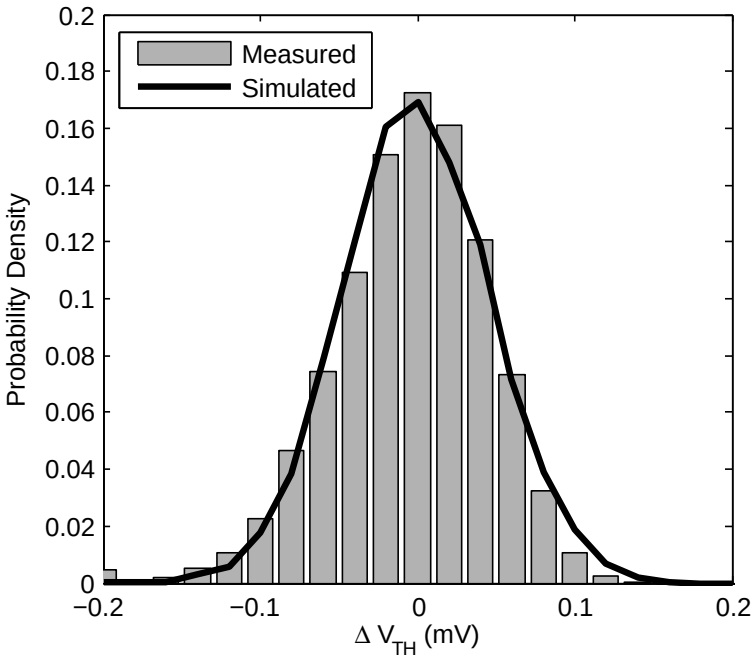


Figure 2.14: Normalized histogram showing measured and simulated distribution of ΔV_T. Measured sample size is 15360 bits, and simulated sample size is 10000 bits.

The value $\sigma_{\Delta V_T} = 1.5$ mV determined is unique to the particular 130 nm process used for manufacturing, and will probably vary in other technologies. It can be seen from Eqn. 2.23 that the value of V_R required for a particular reliability threshold, ϵ, is directly proportional to the value of $\sigma_{\Delta V_T}$. Therefore, if $\sigma_{\Delta V_T}$ is substantially larger in an alternate technology, the fraction of reliable bits will be reduced, unless the distribution of ΔV_T is also wider.

The distribution of '1's among 480 32-bit columns is shown in Fig. 2.15. If the probability of manufacturing an ID cell that evaluates to '1' is independent for every cell in the column, the results should be described by a binomial distribution, with $N = 32$ and $p = 0.5$. However, the measured standard deviation of this distribution is larger than the expected standard deviation of an ideal binomial distribution (4.18 and 2.83, respectively). This suggests that the manufactured ID cell probabilities are

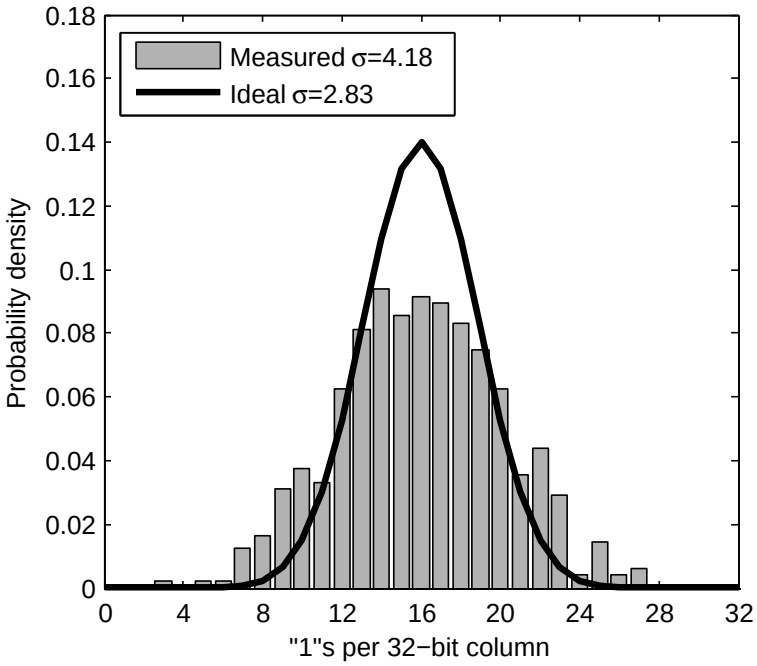


Figure 2.15: Distribution of '1's per 32-bit column.

not totally uncorrelated. A possible reason for this is that each column shares a single sense amplifier, with non-zero input referred offset. Despite this, the deviation from ideal is not excessively large, and should not effect the operation of the random ID system.

The energy consumption was determined by measuring the power consumption of the chip while repeatedly reading from the array, and dividing by the total ID cell read rate. The energy consumption was 39 fJ/bit at a readout rate of 40 kbps.

2.7.2 Temperature Effects

Variations in operating temperature can have a significant effect on the random ID system due to the temperature dependence of MOS transistors. In saturation, the MOS drain current is simply

$$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2 \tag{2.71}$$

Both the carrier mobility, μ, and the threshold voltage, V_T, depend strongly on temperature. Consider first a pair of NMOS devices, for example M1 and M2 in Fig. 2.1. If M1 and M2 are identical, they will have the same drain currents and equal temperature dependence.

$$\Delta I_{D1,2} = I_{D1} - I_{D2} = 0 \tag{2.72}$$

$$\frac{dI_{D1}}{dT} = \frac{dI_{D2}}{dT} \tag{2.73}$$

Therefore the difference in their drain currents will have no temperature dependence:

$$\frac{d\Delta I_D}{dT} = 0 \tag{2.74}$$

Now consider a second pair of devices, M3 and M4, which have unequal carrier mobility and threshold voltage in such a way that their drain currents remain equal.

$$\Delta I_{D3,4} = I_{D3} - I_{D4} = 0 \tag{2.75}$$

$$\mu_3 \neq \mu_4 \tag{2.76}$$

$$V_{T3} \neq V_{T4} \tag{2.77}$$

Because the mobility and threshold voltage are unequal, their temperature dependence is also unequal, and the temperature dependence of the drain currents will be unequal

$$\frac{dI_{D3}}{dT} \neq \frac{dI_{D4}}{dT} \tag{2.78}$$

The magnitude of this difference will increase the larger the difference between the mobility and the threshold voltage. Note also that the polarity of the difference is reversed by swapping the positions of M3 and M4.

Because the sense amplifier in the random ID generator is sensitive only to ΔI_D, the first and second cases just described are initially indistinguishable. This causes the temperature dependence to appear random between the ID cells. The effect of this is that the polarity of ΔI_D in the ID cells may flip as temperature changes, causing bits of the generated ID to flip.

To evaluate the effects of temperature variation, the change in ΔV_T was measured for all ID cells over 30 $^\circ C$ and 50 $^\circ C$ increases in temperature. The observed temperature dependence is in part random, and in part proportional to ΔV_T. Over this range the dependence is roughly linear with temperature, and so can be expressed as

$$\frac{d\Delta V_T}{dT} = K \tag{2.79}$$

where K is a random variable with mean

$$\mu_K = a\Delta V_T \tag{2.80}$$

The temperature coefficient, a, was found to be -6.1×10^-3 $^\circ C$. Because a is negative, the magnitude of ΔV_T tends to decrease as temperature decreases. The

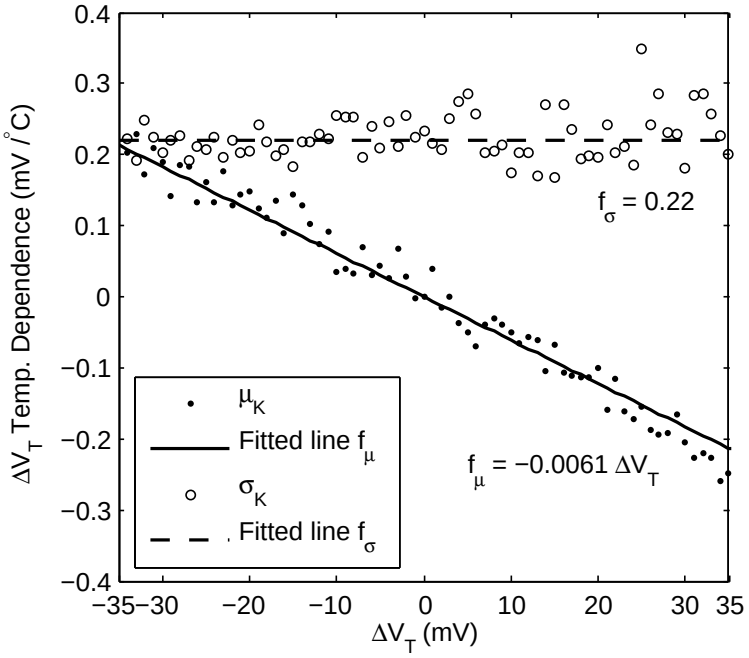


Figure 2.16: Mean and standard deviation of the change in ΔV_T per $1°C$ change it temperature.

standard deviation, σ_K was found to be 0.22 mv / $°C$. A plot of the mean and standard deviation of the the temperature dependence for a range of ΔV_T values is shown in Fig. 2.16.

The practical implication of this is that a larger value of V_R must be selected to ensure reliable operation if the temperature of the random ID generator will change during operation. For positive values of ΔV_T, and if all temperature dependencies are assumed to fall within 6σ, then the worst case shift from ΔV_T to $\Delta V_T'$ due to an increase in temperature ΔT is given by

$$\Delta V_T' - \Delta V_T = (a\Delta V_T - 6\sigma_K)\Delta T \tag{2.81}$$

We can find the value of V_R' required at nominal temperature to ensure that $|\Delta V_T'| \geq V_R$ at increased temperature by substituting $\Delta V_T' = V_R$ and $\Delta V_T = V_R'$ into

Eqn. 2.81.

$$\Delta V_R' = \frac{V_R + 6\sigma\Delta T}{1 + a\Delta T} \tag{2.82}$$

That is, to ensure reliable operation, V_R is first determined from Eqn. 2.23 for some value of ϵ. A larger value, V_R', is then determined from Eqn. 2.82 based on the expected temperature range. Then, at nominal temperature, only ID cells with $|\Delta V_T| \geq V_R'$ are classified as reliable. This ensures that after a temperature increase, ΔT, the new value of the threshold voltage difference, $\Delta V_T'$ satisfies the initial reliability requirement $|\Delta V_T'| \geq V_R$. For example, using the previously determined value of $V_R = 10$ mV, and assuming a possible temperature increase of 10 $°C$, Eqn. 2.82 indicates that $V_R' = 25$ mV. This increases the fraction of unreliable bits to $p_U = 0.41$.

The previously described method to ensure reliable operation over temperature shifts has the advantage of not requiring any extra measurements beyond the two already required by the characterization scheme, and these measurements need not be taken at any particular temperature. It is, however, inefficient due to the constraint imposed by Eqn. 2.81, which assumes the worst case temperature dependence. This is highly unlikely for any given ID cell; a large fraction of cells characterized in this way as unreliable would in fact behave reliably. An obvious alternative would be simply to characterize each cell twice, once at each extreme of the expected operating temperature range, and only select cells that have the same classification at both extremes as reliable. This is more efficient, in the sense that the fraction of unreliably bits will be smaller, but requires a temperature controlled testing environment, which is frequently time consuming and expensive.

2.7.3 Addressing Parameter Selection

The addressing scheme presented here is applicable to a wide variety of applications, and the particular parameters of the addressing scheme depend highly on the specifics of the application. In general, there are three numeric requirements that must be chosen: 1 the maximum number of addressable chips S, (2) the yield Y, and (3) the error rate p_E. After these requirements are specified, the other parameters of the addressing protocol can be determined. As an example, the parameters for a hypothetical use-case are determined here.

Consider a system in which up to 1000 unique chips will be addressed ($M = 1000$). At least 99.9% of fabricated chips should be suitable for use ($Y = 0.999$), and the maximum total error rate should be less than $p_E = 10^{-6}$). The expected temperature shift is $+10°C$. Solving Eqn. 2.65 with $M = 1000$ gives

$$N - D \geq 10 \tag{2.83}$$

Substituting this constraint into Eqn. 2.60 with $Y = 0.999$ gives a minimum value of $N = 31$ with $U = 21$. Finally the total error rate can be found using Eqn. 2.24, $p_E = 1.03 \times 10^{-7}$.

2.7.4 Comparison to Previous Work

Table 2.3 compares this work to previously published random ID generators. The columns indicate the manufacturing technology, the energy consumed to generate one random bit, the silicon area per bit, the percentage of manufactured bits expected to be reliable, and whether identification can happen remotely or on-chip. Unfortunately, many other published methods of generating identification data from manufacturing variation do not include simulations or measurements of the energy consumption [35]

Table 2.3: Comparison of random ID generator publications.

	Tech.	E/bit	Area/bit	% Reliable	Type
00' ISSCC [34]	0.35 μm	8.3 nJ	209 μm^2	-	Remote
08' JSSC [36]	0.13 μm	1.6 pJ	144 μm^2	94%[1]	Remote
This work	0.13 μm	39 fJ	9.2 μm^2	82%	On-chip

[1]Reliability threshold not specified.

[37] [38] [39], so their suitability for use in a low-power, wireless sensor is difficult to evaluate.

Table 2.3 shows that the random ID circuit presented here requires less area and energy per bit than previously published work. The next lowest energy per bit, 1.6 pJ [36], is in fact 41 times higher than value of 39 fJ presented here. The lower supply voltage of 0.4 V used here is significantly lower than the 1 V supply used in [36], and because energy consumption is proportional to the square of the supply voltage, this is likely the most significant cause of the energy savings of this work. Additionally, the use of a shift register rather than a row decoder also reduces the energy consumption compared to [36]. The work presented here could replace the random ID generators in the previously published remote identification systems and consume lower energy. Note as well that although the fraction of reliable bits appears lower, the threshold of reliability for this work is set very high, and the threshold of reliability in previous works not well defined.

Furthermore, this work is the first that allows for the reliability of individual ID cells to be quickly determined. This, as previously discussed, allows identification to occur on-chip, further saving energy by reducing the amount of data that must be transmitted by the sensor. The exact amount of energy saved is difficult to predict in general because it depends on the characteristics of the transmitter chosen for the particular application. As an example, however, 330 fJ/bit is required for trans-

mission in [32], and this is one of the lowest reported values for transmission from wireless sensors. For an ID system with an 32-bit ID, this would save 10.5 pJ over a system that required transmitting the ID back to the external reader. From another perspective, consider a wireless sensor with an 32-bit ID designed to transmit 8-bit temperature measurements. The total amount of data necessary for transmission by the sensor could be reduced by a factor of 5 by adopting a system that allows for on-chip identification.

2.8 Conclusion

The random ID system presented here can be used to remotely identify integrated circuits in a wireless sensor network using very low energy. The low energy comes not only from the particular circuits presented here, which consume only 39 fJ/bit, but from the reduction in data that must be transmitted wirelessly. Other random ID implementations require the complete N-bit generated ID to be transmitted back to the external reader before identification can occur, which consumes a significant amount of energy. It is the ability to easily characterize random ID cells as reliable or unreliable, along with the masked addressing scheme presented here, that allows for identification to occur on-chip. This is also the first presentation of a random ID system that includes a mathematical model of reliability and yield for on-chip identification, both of which are critical metrics to consider when adopting such a system.

Chapter 3

Low-Power Demodulator and Clock Generator

3.1 Introduction

Most small, wireless sensors intended for medical applications require the ability to receive data from an external reader. In the simplest case, this could simply be an inquiry signal to trigger some predetermined behavior in any sensor node that receives it, such as replying with some sensor data. More elaborate data may also be transmitted: identification information to address a particular node within a group, configuration data, or data to select from among multiple sensor interfaces and different options for data encoding and analysis.

Many sensors networks employ pulse-width modulation (PWM) to transmit from a high-power external reader to low-power sensor nodes. A major advantage of PWM is that the circuitry required to demodulate it is relatively simple and low-power, as will be discussed in the following sections. Additionally, in the case of wirelessly powered sensors, data can be transmitted by the high-power carrier used to energize

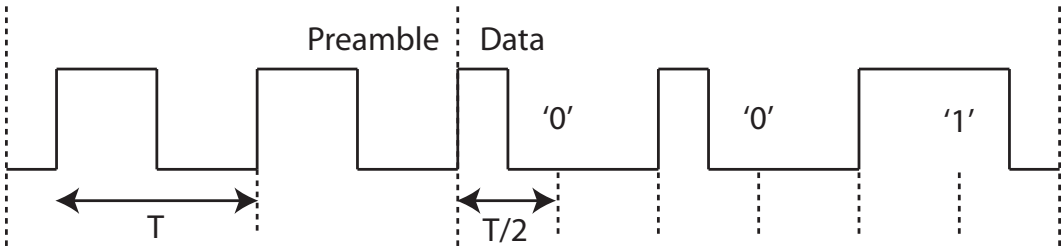


Figure 3.1: Sample of PWM signal. The signal begins with several cycles of unmodulated preamble, followed by the modulated data.

the sensors. Modifying the transmitter to carry PWM data is usually straightforward, since it requires only some way to gate the input of the output power amplifier.

It will be shown that the construction of a PWM demodulator requires a delay element with well controlled timing. This timing element can be reused to construct an on-chip clock source, which is required for any sensor that employs some form of clocked digital logic, analog-to-digital converters, modulator circuits for wireless data transmission.

3.1.1 Pulse Width Modulation

An example of a 100 kHz PWM signal is shown in Fig. 3.1. In fact, this signal is used to further modulate a much higher frequency carrier signal, which is transmitted to the sensor. Recovering the PWM signal is straightforward with a simple envelope detector circuit, which removes the high-frequency carrier, leaving the signal shown in Fig. 3.1.

A short pulse represents a logic '0', and a long pulse represents a logic '1'. Decoding each PWM symbol can be done simply by detecting the value of the signal one half period after each rising edge. For example, Fig. 3.1 shows this applied to the symbols '0', '0' and '1'. In other words, for a PWM symbol with modulation frequency f and period T, the demodulator should latch the value of the PWM signal at a time $T/2$

after each rising edge. This requires a moderately accurate delay element with delay equal $T/2$, which should be low-power and resistant to process variation.

More precisely, the PWM symbols representing '0' and '1', having duration T and starting at time $t = 0$ are described by

$$s_0(t) = \begin{cases} 1 & 0 \leq t < \frac{T}{2}(1-d) \\ 0 & \frac{T}{2}(1-d) \leq t < T \end{cases} \tag{3.1}$$

$$s_1(t) = \begin{cases} 1 & 0 \leq t < \frac{T}{2}(1+d) \\ 0 & \frac{T}{2}(1+d) \leq t < T \end{cases} \tag{3.2}$$

where d is the modulation depth ($0 < d < 1$).

A signal, $m_x(t)$, representing an N-bit message $x = \{x_0, x_1, ..., x_{N-1}\}$ can then be constructed from these symbols.

$$m_x(t) = \begin{cases} s_{x_0}(t) & 0 \leq t < T \\ s_{x_1}(t-T) & T \leq t < 2T \\ ... & \\ s_{x_{N-1}}(t-(N-1)T) & (N-1)T \leq t < NT \end{cases} \tag{3.3}$$

For transmission, this signal is used to further modulate a high-frequency carrier $c(t)$ given by

$$c(t) = A_c \cos(2\pi f_c t) \tag{3.4}$$

where A_c is the peak carrier amplitude and f_c is the carrier frequency.

The modulated wave, $s(t) = m_x(t)c(t)$, is given by

$$s(t) = A_c m_x(t) \cos(2\pi f_c t) \tag{3.5}$$

3.1.2 Demodulation Techniques

Most published examples of wireless sensors include some form of demodulator circuit and clock generator, although these circuits are frequently not combined as they are in the approach presented here. A variety of methods have been proposed to accomplish these tasks. A common approach for demodulating PWM symbols is to integrate each symbol over its duration:

$$y[n] = k \int_{nT}^{(n+1)T} m_x(t) dt \tag{3.6}$$

where k is the time-constant of the integrator. The output signal $y[n]$ is then fed into a comparator, which discriminates between low and high values of $y[n]$, where low values correspond to logic '0' and high values correspond to logic '1'. This approach is attractive because it is relatively insensitive to any noise added to message, $m_x(t)$, since an integrator is a type of low-pass filter. This method employed is employed in [45]. By clocking the integrator with the rising edges of the input signal, $m_x(t)$, this approach appears at first to not require any precise timing element on chip. It does, however, require an integrator with a reasonably well controlled time-constant, k, to ensure that values of $y[n]$ corresponding to logic '0' are below the comparator threshold, and values corresponding to logic '1' are above the comparator threshold. The integrator in [45] is not used to generate an on-chip clock.

The primary advantage of this approach, however, is noise immunity, which is not a priority for most wirelessly powered sensors for two reasons. The first is that, since the transmitted signal is large enough to power the sensor, the magnitude of

the modulated signal is likely already quite large relative to the noise. The second is that the envelope detector, which extracts the modulated envelope from the high-frequency carrier, also functions as a low-pass filter, which acts to reduce any noise present in the received signal. The cost of this approach is the power consumed: constructing an integrator usually requires an op-amp with negative feedback, often consisting of several branches drawing static current.

The work presented in [46] observes that a delay element is useful for both demodulating symbols and for providing an on-chip clock, used in this case for modulating transmitted data. The delay element is used to form an oscillator that can be placed in either astable (oscillating) or monostable (single shot) mode. The monostable mode can be used for demodulation, where the delay element is used to trigger a register that stores the demodulated symbol a set time after the beginning of the symbol. The delay element consists of a three-stage, current limited ring oscillator, with a capacitor at one of the internal nodes to increase the delay. The delay element is, however, not adjustable, and it is observed that due to manufacturing variation, this will cause inaccuracy in the modulation frequency, making demodulation by the external reader more difficult.

Another approach to clock generation for wireless sensors, presented in [47], is to use a current controlled oscillator with digital calibration. In this system, the a delay line is adjusted by summing binary weighted currents onto a capacitor, with digital control of which currents are selected. The circuit, however, is not symmetric: only current sources are available to charge the capacitor, with no current sinks. This makes only the low-to-high transitions of the delay line adjustable. The adjustable delay is followed by a further delay line composed of a series of inverters. The effect of this is that when used as a clock, the frequency of the clock is adjustable, but the pulse width is fixed. In other words, the duty cycle of the clock changes as the

frequency changes.

Phase-locked loops (PLLs) have been used for clock and data recovery for wireless sensors [48] [4]. In [48], the PLL voltage-controlled oscillator (VCO) is composed of a ring oscillator, where the control input adjusts an RC load formed from two MOSFETs at each internal node. In [4], the VCO is a basic relaxation oscillator formed by two comparators, where the control input adjusts the comparator switching voltage. The use of a PLL is advantageous because with proper design, the gain of the VCO (the relationship between the control voltage and the output frequency) need not be well controlled since the application of negative feedback will adjust the VCO frequency to that of the input frequency. This is helpful since the VCO gain is sensitive to manufacturing variation. The principle downside to the use of a PLL is the large number of components required. At a minimum, it requires a phase-detector, charge pump, loop filter and VCO. An implementation with fewer components may achieve lower power consumption.

An approach similar to a delay-locked loop (DLL) is presented in [49]. If a local clock is present on the sensor, with the same frequency as the modulated signal, then the problem is simply to ensure the local clock is in phase with the modulated signal. This can be done quickly, and with low-power, by using adjustable delay line, to delay either the modulated signal or local clock so that both are in phase. However, this approach requires a relatively accurate local clock. If derived from a crystal oscillator, this will increase the sensor size, since the crystal cannot be integrated. And if the local clock is some form of on-chip oscillator, it is likely sensitive to manufacturing variation.

3.1.3 Clock Generation Techniques

In addition to demodulation, clock generation is a task required of nearly all wireless sensor circuits, as a stable clock is required for the operation of clocked logic, A/D converters, and return modulation. Because of this, and the fact that the clock generator is likely active when these other components are active, minimizing the power consumption of the clock generator is critical to achieving minimum sensor power.

Currently, there are several methods under investigation for constructing low-power clocks for small, wireless sensors. However, it will be shown that few of these are suitable for the type of small, wirelessly powered sensors under consideration in this work, as they either increase the number of non-integrable components, or assume a different type of energy harvesting architecture.

The lowest power implementation of timing elements exploit the small gate-leakage current found in all MOSFETs. The gate leakage current is used to charge a capacitor, which triggers a Schmitt trigger when the capacitor voltage crosses some threshold. Because the gate leakage current is so small, these implementations typically achieve very low power consumption, on the order of one to several hundred pW [50] [51]. The small current, however, also means the time constant of the circuit is large, and the period of such timers is typically tens of milliseconds to several seconds. This is particularly useful for energy constrained sensors, in which a battery supplies all of the sensor's power, and the sensor spends long periods of time in a sleep state, waking periodically to perform a measurement or other task. This, however, is not the operating model of a wirelessly powered sensor, in which the sensor sources its power from the RF carrier for some period of time, and then from a local energy store for a short period of time. Furthermore, the long period of these timers makes them

unsuitable for a general purpose clock source: the highest frequency attained is 16 Hz in [51].

The next lowest power clock generator topology is a type of crystal oscillator. Crystal oscillators are used in a wide variety of clock generation applications, due to their excellent accuracy, small jitter, and insensitivity to temperature. They are, however, not normally know for low-power operation. Despite this, several implementations have been proposed that with reduced power consumption [52] [53] [54] [55] [56], the lowest consuming less than 10 nW [55] [56]. Unlike the gate-leakage oscillators, these are able to generate frequencies appropriate for a general purpose clock, usually in the range of several kHz to several MHz. Unfortunately they can not be integrated, which means they will substantially increase both the cost and size of the sensor. As discussed in Chapter 1, minimum size is of paramount importance, and for this reason the use of an external crystal is far from ideal.

Low-power clock generators that operate in the range of 100 kHz have been constructed. Ring oscillators [57], RC oscillators [58] [59] and relaxation oscillators [60] are common. Despite topological differences, these oscillators are all similar in that they generate periodic switching events at a rate controlled by charging some capacitance either through a resistance or with constant current. A central difficulty with all of these approaches is that the sensitivity to manufacturing variation is typically high, and so employing some sort of compensation or tuning is commonly required [58] [57] [60] [59].

3.2 Demodulator Architecture

The approach to demodulation adopted here is based on a variable length delay-line within a delay-locked loop. A low-power delay line is used to sample the value of the

PWM wave one-half period after each rising edge, as shown in Fig. 3.1. However, this delay line is sensitive to manufacturing variation, making the actual manufactured delay difficult to predict. To compensate for this, the length of the delay line is digitally tuned such that the tuned delay becomes equal to one-half period of the pulse-width modulation frequency.

3.2.1 Variable Length Delay Line

From Fig. 3.1, it can be seen that if the length of the delay lines varies substantially from $T/2$, the PWM symbols will not be properly decoded. If the delay is too short, all symbols will appear as '0', and if the delay is too long, all symbols will appear as '1'. The specific accuracy required depends on the modulation depth, d, in Eqns. 3.1 and 3.2. The PWM symbols must be sampled at a time t_S after each rising edge, where

$$(1-d)\frac{T}{2} < t_S < (1+d)\frac{T}{2} \tag{3.7}$$

or

$$t_S = \frac{T}{2} \pm d\frac{T}{2} \tag{3.8}$$

Sampling at t_S can be accomplished by applying the PWM signal to a delay line with delay equal to t_S, and using the output of the delay line to clock a register that samples the PWM signal. If the delay line is constructed from inverters, the manufactured accuracy of the delay line depends mainly on the threshold variation of the devices that form the inverters. In particular, global process variation will cause highly correlated variations in the device thresholds that will directly effect in

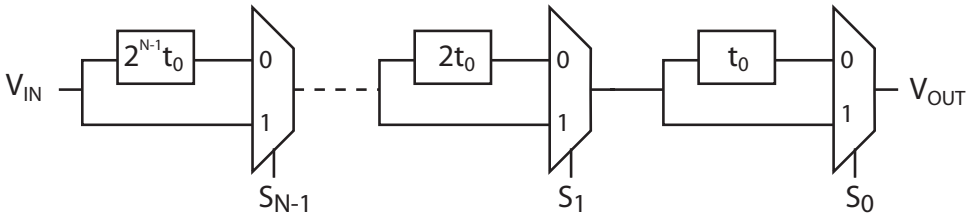


Figure 3.2: Structure of N-bit, variable length delay line.

the manufactured delay. To illustrate this, a delay line with 5 μs nominal delay was simulated, and found to vary from 3.8 μs to 7.8 μs across process corners.

To ensure reliable operation of the demodulator, a delay line with low inherent accuracy could be employed if the delay was made tunable. In the case of a chain of inverters, this can be accomplished by adding or removing inverters from the chain. A diagram of this type of variable length delay line (VLDL) is shown in Fig. 3.2, in which elements with binary weighted delays can be added or subtracted from the chain with a digital control signal.

In practice, some delay of fixed length is usually added in series with the adjustable structure shown in 3.2. The total delay is then

$$t_D = t_F + (2^N - 1)t_0 \tag{3.9}$$

where t_F is the delay of the fixed element, t_0 is the delay of the unit element, and N is the control signal. The individual delay elements are composed of chains of slow inverters, shown in Fig. 3.3. Devices with large lengths are added in series to a regular inverter structure, which has the effect of reducing the charging current and therefore increasing the switching time. The control signal is generated by a feedback loop, as described in the following section.

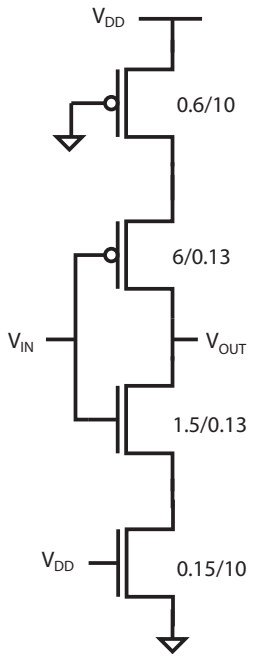


Figure 3.3: Slow inverter, used as basic delay element in the variable length delay line of Fig. 3.2. Device dimensions are given in microns.

3.2.2 Delay Locking

In order to mitigate the effects of manufacturing variation, the negative feedback is applied to the VLDL to form a delay-locked loop. A block diagram of this arrangement, as well as the accompanying clock generator, is shown in Fig. 3.4.

The principal idea behind this arrangement is to tune the inaccurate but adjustable delay line to an accurate time reference broadcast by the external transmitter. This time reference is contained in the preamble shown in Fig. 3.1, which consists of an unmodulated square wave, with 50% duty cycle and period T, and precedes the modulated data.

In locking mode, the preamble is applied to the input labeled ENV_IN, from the previous stage (an envelope detector). A delayed version of the preamble is generated by two inverting delay lines, each with delay T_D, creating total delay $2T_D$. The phase detector then compares the rising edges of the original and delayed preambles. If the

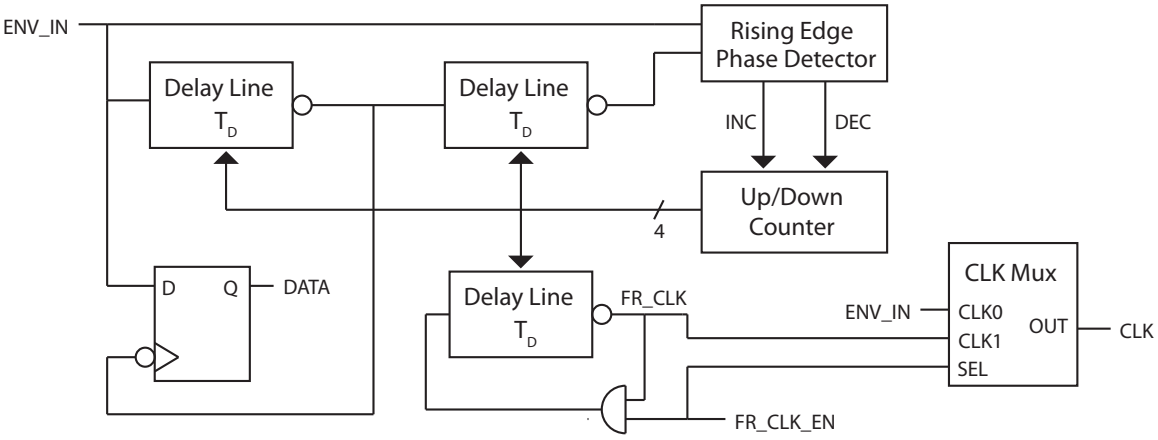


Figure 3.4: Block diagram of the demodulator and clock generator. ENV_IN is output from the envelope detector, FR_CLK_EN enables the free-running on-chip clock.

rising edges of the delayed preamble arrive before the original edges, then the delay is too short, and the phase detector increments a counter. If the rising edges of the delayed preamble arrive after the original edges, then the delay is too long, and the counter is decremented. The value of the counter is then used as the control signal for the two VLDLs.

By initializing the 4-bit counter with value '1000', the delay will be locked such that $T = 2T_D$ within a maximum of 8 cycles of the signal preamble. The phase detecter can then be disabled and demodulation can occur. When the modulated data appears at ENV_IN, the value of the waveform will be latched by a register at time $T_D = T/2$, just as shown in Fig. 3.1. For experimental purposes, the transition from locking to demodulation was controlled by an external signal, but this signal could easily be implemented on-chip by counting the appropriate number of preamble cycles.

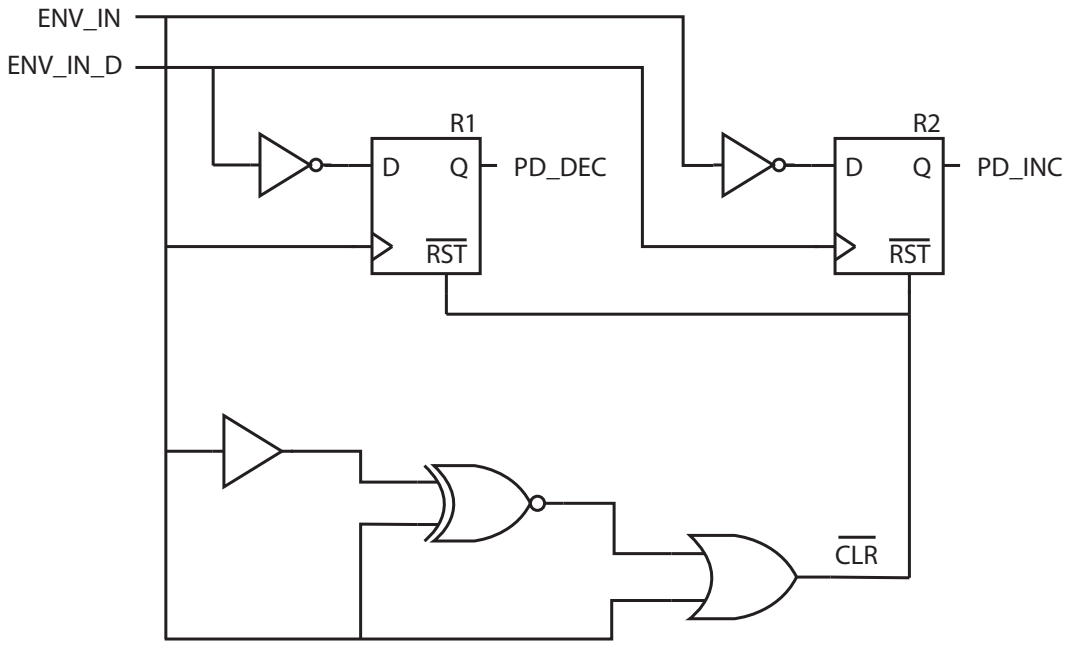


Figure 3.5: Schematic of positive edge triggered phase detector.

Rising Edge Phase Detector

A schematic of the rising edge phase detector is shown in Fig. 3.5. If the rising edge of ENV_IN precedes the rising edge of the delayed signal ENV_IN_D, register R1 will latch high, and the output PD_DEC will go high. This causes the counter to decrement, decreasing the delay of the VLDLs. Simultaneously, register R2 latches low and the output PD_INC is unchanged. Similarly, if the rising edge of ENV_IN lags the rising edge of ENV_IN_D, register R2 will latch high, and the output PD_INC will go high, causing the delay of the VLDLs to increase. The other gates shown in Fig. 3.5 generate a reset pulse after the falling edge of ENV_IN, clearing the registers. The buffer and XNOR gate generate low pulses after each edge in ENV_IN, and the OR gate selects only the pulses triggered by falling edges.

Note that because the VLDL delay is only adjustable in discrete increments, the delay of the VLDLs will never exactly match the period of the preamble. This causes

the VLDL control signal to never settle to a stable value, but oscillate between the values that are just shorter and just longer than the preamble period. This has no practical effect on the delay if the delay of the smallest increment is sufficiently short, although it may consume a small amount of extra switching energy.

3.2.3 Clock Generation

Once the VLDL has been tuned to the delay found in the signal preamble, it can be used to generate an on-chip clock. This eliminates the need for an external clock source, such as a quartz crystal, which would increase the size and cost of the sensor. There is little cost involved in this approach, since it simply reuses the already calibrated delay line as the timing element.

The clock generator is shown in Fig. 3.4. A copy of the variable length delay line is used to form a ring oscillator, using the same control input generated by the delay locked loop for the demodulator delay lines. This clock is referred to as the free-running clock. Prior to its activation, the ENV_IN signal is used as the clock for any digital logic that requires it. The proposed sequence of operation is:

1. Sensor is energized from wireless carrier or other power source.
2. External transmitter broadcasts carrier modulated with preamble.
3. Sensor locks delay elements to preamble.
4. Sensor receives modulated data.
5. Sensor activates on-chip, free-running clock.
6. Sensor switches from ENV_IN clock to free-running clock.

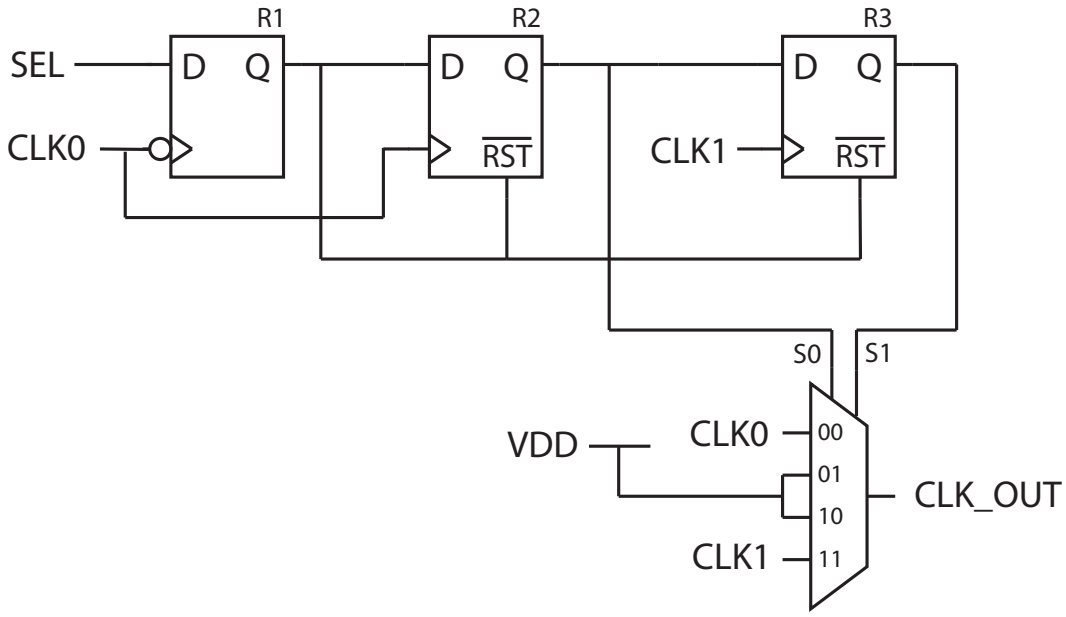


Figure 3.6: Schematic of glitch-free clock multiplexer.

Clock Multiplexer

The purpose of the clock multiplexer is to switch between the clock signal initially provided by the external transmitter, and the chip's internal clock once it it has been tuned and activated. Because there is nothing keeping the external and internal clocks in phase, switching between them at an arbitrary time with a regular multiplexer may cause very short pulses at the output. This glitching may upset the state of the chip's digital logic, since very short clock pulses may cause errors in digital registers.

This problem is mitigated by (1) waiting for the external clock (CLK0) to go high, (2) setting the output clock high, (3) waiting for the internal clock (CLK1) to go high, and (4) switching the output to CLK1. This is accomplished using the circuit shown in Fig. 3.6.

Initially R1 stores a '0', which holds R2 and R3 reset, and the multiplexer selects CLK0 for the output. A high signal on SEL directs the circuit to switch from CLK0 to CLK1. First, R1 latches high on the falling edge of CLK0. Now when CLK0 goes

Table 3.1: VLDL delay in μs. The FF and SS corners represent the shortest and longest possible manufactured delays, respectively.

Control Bits	FF	TYP	SS	Measured
0000	2.0	2.8	3.8	3.7
1000	3.8	5.2	7.2	7.0
1111	5.3	7.3	10.2	9.9

high, R2 latches high and the multiplexer holds the output high. Finally, when CLK1 goes high, R3 latches high and the multiplexer selects CLK1 for the output.

3.2.4 Experimental Results

The demodulator and clock generator shown in Fig. 3.4 was fabricated in a standard 0.13 μm CMOS process. The nominal delay was designed to be 5 μs, in order to demodulate a PWM signal with 100 kHz carrier. This corresponds to a carrier period of 10 μs. The simulated variation of the variable length delay line is shown in Table 3.1, at three process corners: FF, typical, and SS. These represent the minimum, typical, and maximum delays that could be produced from identically designed delay lines given variation in the manufacturing process. The measured delay of the fabricated delay line is also shown. The measured delay is larger than the typical delay, possibly indicating a manufactured wafer near the SS corner, or possibly due to extra capacitance in the layout not accounted for in simulation. Note that the fabricated VLDL range of 3.7 μs to 9.9 μs includes the target delay of 5.0 μs, so this variation does not disrupt the operation of the circuit.

The operation of the VLDLs within the delay locked loop was also examined. A 100 kHz square wave was applied to the demodulator input for several cycles, during which the delay lock loop finds the appropriate control input for the VLDLs. The square wave was then removed, and the FR_CLK_EN signal brought high, to

activate the on-chip clock. This was repeated several times, and the resulting on-chip clock frequency was found to be either 103 kHz or 109 kHz. Locking to two possible frequencies is the expected behavior, as described in Sec. 3.2.2, but both frequencies being above the preamble frequency of 100 kHz was unexpected. This is likely due to a mismatch between the loading of the VLDLs in the delay-locked loop and the VLDL in the ring oscillator. Unequal loading can be mitigated in future designs by simply adding "dummy" gates to the output of the VLDLs to ensure that they are driving equal capacitances. The jitter of the on-chip clock was also measured to ensure its suitability for driving digital logic. At a frequency of 103 kHz ($T = 9.68\mu$s), the RMS jitter was found to be 45 ns, or approximately 0.5% of the period.

To ensure operation of the demodulator, the signal described in Fig. 3.1, consisting of several cycles of 100 kHz preamble, followed by a byte of PWM data, was applied to the demodulator input from an arbitrary waveform generator. The measured input and output waveforms are shown in Fig. 3.7, showing the reliable operation of the demodulator.

Power Consumption

The power consumption of the demodulator and clock generator was measured in several operating modes. Power consumption was determined by measuring the supply current using a pico-ammeter in series with the supply voltage. The results are shown in Table 3.2. A complication with this measurement was that several digital circuits unrelated to the demodulator shared the demodulator supply, and had substantial leakage current. Simulations show that the demodulator accounted for no more than a third of the leakage current. This can be corrected for by subtracting the simulated leakage current of the unrelated circuits from the measured current. Both the raw and corrected values of power consumption are shown in Table 3.2.

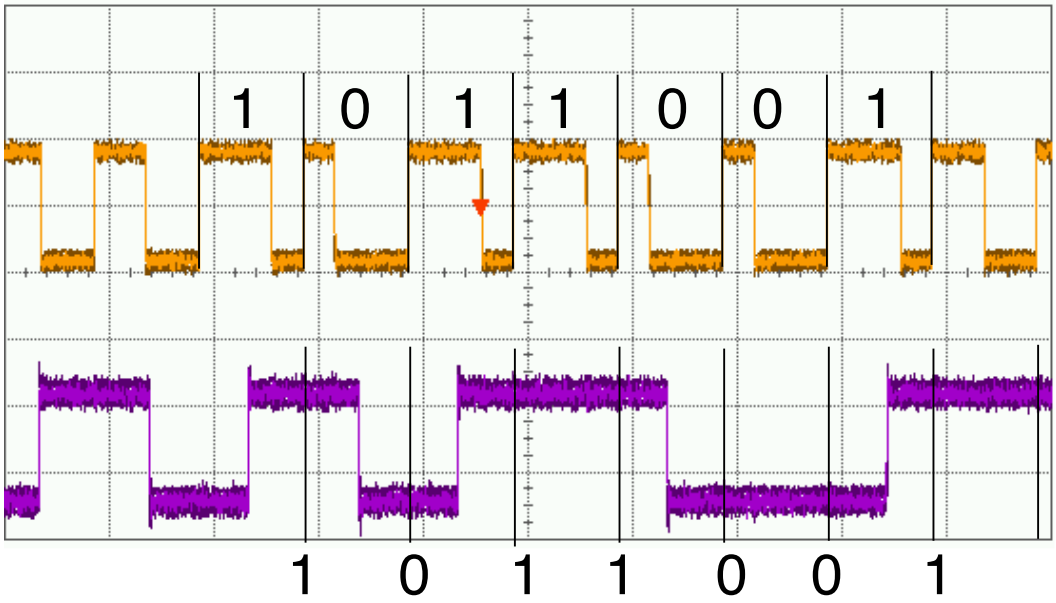


Figure 3.7: Demodulator operation, showing PWM input encoding a single random byte, '1011001', in orange. The demodulator output of the same value is shown in purple.

Table 3.2: Power consumption of demodulator and clock generator for different operating modes.

Operating Mode	Measured Power (nW)	Adjusted Power (nW)
Off (leakage)	380	130
Demodulation	470	220
CLK Generation	440	190

3.3 Improved Demodulator

After fabricating the demodulator and clock-generator, the analysis of the architecture was further developed and several methods were developed that would improve its performance. This includes a more detailed examination of delay-line performance and how to reduce its power consumption, as well as an analysis of how to control the precision of the clock generator. Additionally, some of the control logic was simplified, reducing the number of gates and the power consumption. The performance

of the demodulator and clock generator with these improvements was then explored in simulation.

3.3.1 Low-Power Delay Line

The two critical parameters for evaluating low-power delay lines are the delay-power and the accuracy. By accuracy, we mean the difference between the designed delay time and the delay time of the manufactured delay lines. By delay-power, we mean the energy consumed to generate one unit time of delay:

$$P_D = \frac{E_D}{t_D} \tag{3.10}$$

where P_D is the delay-power, E_D is the energy consumed, and t_D is the amount of delay. This metric allows for the performance of delay elements with unequal delays to be compared. For example, a delay element that consumes half the energy of some reference element, but only produces half the delay, has the same delay-power as the reference element. Two of these elements could be combined in series to produce a delay structure with the same performance as the reference element.

Of the time dependent electrical phenomenon that could serve as the basis of an electrical delay line, such as the voltage across a charging capacitor, the current through a charging inductor, or temperature change due to resistive heating, it is capacitive charging that is the simplest to construct in modern CMOS processes, and consumes the least amount of energy.

Consider the most general capacitive delay element shown in Fig. 3.8. In this circuit, the capacitor is initially discharged, $V_C = 0$, and at time $t = 0$ the switch closes and begins to charge the capacitor with current I. The voltage V begins to rise, and when it reaches some threshold V_{TH}, the comparator triggers and drives the

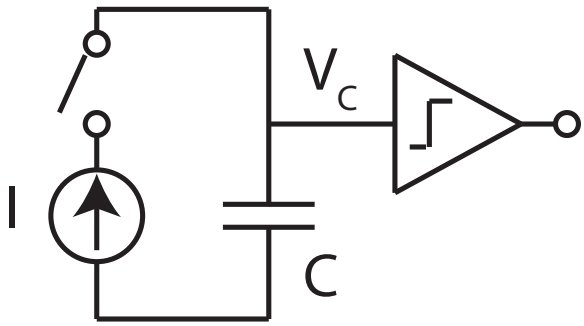


Figure 3.8: Canonical capacitive delay element.

output high. The energy consumed, the delay time, and the delay-power are:

$$E_D = \frac{1}{2} C V_{th}^2 \tag{3.11}$$

$$t_d = \frac{C V_{th}}{I} \tag{3.12}$$

$$P_D = \frac{E_D}{t_d} = \frac{1}{2} I V_{TH} \tag{3.13}$$

It can be seen from Eqn. 3.13 that only by reducing the charging current, I, and the switching threshold, V_{TH} can the delay-power be reduced. Reducing the value of the capacitance, C, has no effect.

Now consider a practical implementation of a capacitive delay structure: a chain of inverters. A single inverter is shown in Fig. 3.9. In this structure, the delay is produced by charging the parasitic capacitances at each node, composed primarily of gate capacitance of the following stage, by finite current sourced through devices that form the inverter. The inverter has a supply voltage V_{DD}, some load capacitance, C, and the devices that form the inverter have length, L.

The energy consumed by a single stage switching is simply:

$$E_D = C V_{DD}^2 \tag{3.14}$$

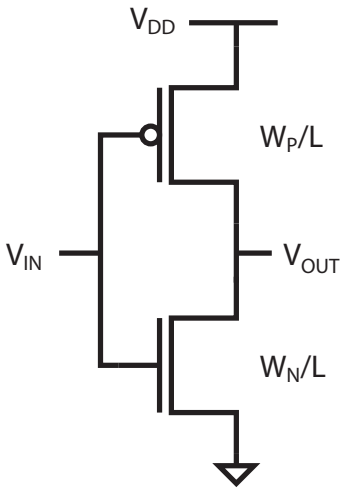


Figure 3.9: CMOS inverter used as basic delay element.

Assuming strong inversion and negligible output resistance, the current drawn from the supply when switching from low to high is:

$$I = \frac{\mu_p C_{ox}}{2} \frac{W}{L} (V_{DD} - V_T)^2 \tag{3.15}$$

This current will charge the inverter load capacitance, and the following inverter will switch when the output voltage is equal to $V_{DD}/2$. The time it takes for this is:

$$t_D = \frac{CV_{DD}}{2I} \tag{3.16}$$

The delay power can then be found by combining Eqns. 3.10, 3.14, 3.15, and 3.16:

$$P_D = \mu_p C_{ox} \frac{W}{L} V_{DD}^2 (V_{DD} - V_T)^2 \tag{3.17}$$

This model of the delay power is an approximation that assumes no channel-length modulation, and operation in strong inversion. The first approximation is valid, since channel-length modulation is reduced is devices with large L [40], and it will be shown shortly that inverters used as delay elements should be constructed from devices with

large L. The approximation of strong inversion may not always be valid. For the case of weakly inverted devices, when $V_{DD} < V_T$, Eqn. 3.15 can be replaced with an equation for sub-threshold drain current:

$$I = I_S \frac{W}{L} e^{n\frac{V_{GS}}{\frac{kT}{q}}} \tag{3.18}$$

Following the same derivation as for Eqn. 3.17, the delay-power in sub-threshold is

$$P_D = 2I_S \frac{W}{L} V_{DD} e^{n\frac{V_{DD}}{\frac{kT}{q}}} \tag{3.19}$$

It can be seen from Eqns. 3.17 and 3.19 that, irrespective of strong or weak inversion, the delay-power is a strong function of V_{DD}. This is because both the charging current and the switching voltage depend on V_{DD}. This shows that the most effective way to reduce the delay-power is simply to reduce the supply voltage. In actual practice the potential for this is often limited, since the supply is usually shared by many other circuits that dictate the minimum supply voltage, and creating additional supplies usually costs substantial power overhead.

Fig. 3.10 plots the relationship between the supply voltage and the delay power, and shows that the delay power falls rapidly with decreasing V_{DD}. Additionally, the delay power can be reduced by increasing the length, L, of the devices that form the inverter. This has the effect of reducing the charging current, but not the switching threshold. The relationship between the delay power and the device length is plotted in Fig. 3.11, and shows the delay power to be inversely proportional to the length, as indicated by Eqn. 3.17.

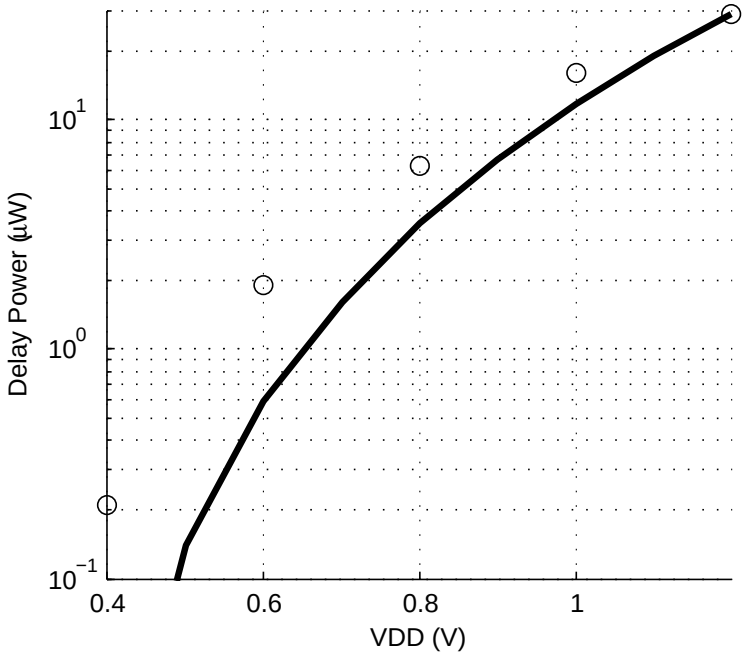


Figure 3.10: Simulated relationship between delay power and supply voltage plotted with Eqn. 3.17. Note Eqn. 3.17 diverges from the simulated results as V_{DD} approaches V_T. Device widths are scaled to provide equal rise and fall times, and lengths are both 1 μm. Delay power falls rapidly with decreasing V_{DD}

3.3.2 Clock Generator Precision

The discrete nature of the VLDL means that tuned delay will never exactly match the delay of the signal preamble. While this is unlikely to cause problems the demodulator, the finite precision of the VLDL implies finite precision of the clock generator. The effects of an on-chip clock frequency that differs from the nominal frequency vary depending on the circuit being clocked. For digital logic, small deviations in clock frequency are unlikely to cause errors. For A/D converters, a deviation in clock frequency will cause a deviation in the sampling rate, causing the acquired signal to appear to shift in frequency if reconstructed at the nominal sampling rate. If used to drive an RF transmitter, deviations in the clock frequency may cause deviations in the RF carrier and modulation frequencies, depending on the particular architecture.

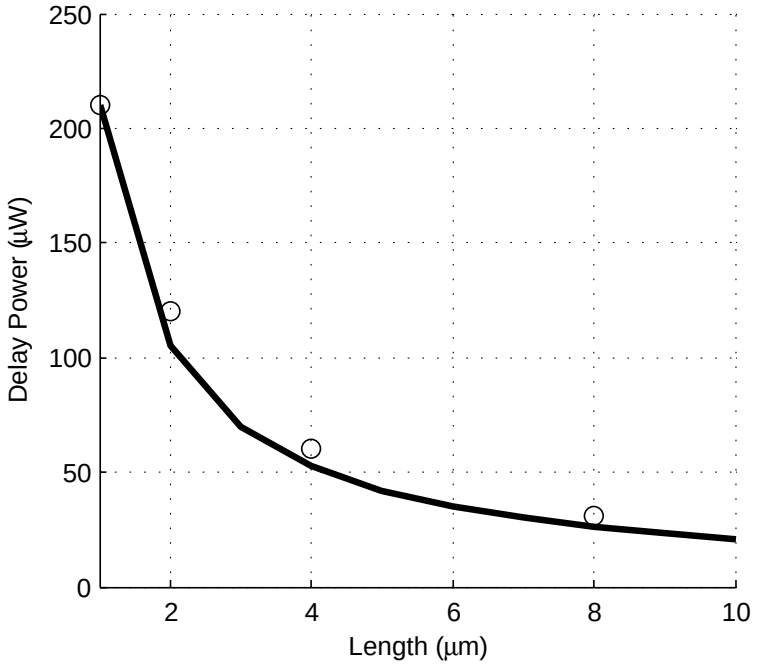


Figure 3.11: Simulated relationship between delay power and device length. $V_{DD} = 0.4$ V. Device widths are scaled to provide equal rise and fall times. Delay power is inversely proportional to device length.

Although these effects may be problematic if severe enough, there is likely a quantifiable tolerance below which they are insignificant. Once these tolerances are identified, they will constrain the precision required of the clock generator. The precision of the clock generator can then be set using the following process.

The precision of the clock generator frequency is controlled by the shortest delay element in the chain, shown in Fig. 3.2. The clock frequency is given by

$$f_C = \frac{1}{2\left(t_F + (2^{N-1})t_0\right)} \tag{3.20}$$

where f_C is the clock frequency, t_F is the delay of the fixed delay element, t_0 is the delay of the shortest element, and n is the control input.

Because the clock frequency is inversely proportional to the delay, uniform in-

creases in delay do not cause linear decreases in frequency. In other words, the change in frequency due to an increase in the control input is different for each control input. The largest decrease in frequency occurs in the transition from control input $N = 0$ to $N = 1$. This change in frequency is given by

$$\Delta f_C = \frac{1}{2t_F} - \frac{1}{2(t_F + t_0)} \tag{3.21}$$

This value of Δf_C is equal to the worst case precision, f_p. Eqn. 3.21 can then be rewritten, solving for t_0:

$$t_0 = \frac{2f_p t_F^2}{1 - 2f_p t_F} \tag{3.22}$$

where t_F is the minimum delay, occurring the control input $S = 0$. For $1 >> 2f_p t_F$, this is approximately

$$t_0 \approx 2f_p t_F^2 \tag{3.23}$$

In this way, once the precision, f_p, is specified, the appropriate delay for the smallest element, t_0, can be found.

3.3.3 Successive Approximation Register

In the original demodulator architecture, the control signal for the variable length delay line is generated by a regular up/down counter. This type of counter, however, requires a large number gates, and so was replaced by a successive approximation register (SAR), which are commonly used in SAR ADCs [61]. Rather than counting up or down by one LSB, the bits of this register are determined sequentially, from highest to lowest, by performing a binary search. Both the active and leakage power

of the SAR is much lower than that of the up/down counter. This has the additional advantage of establishing the control value after N cycles, compared to 2^{N-1} cycles for the up/down counter initialized to its middle value.

3.3.4 Simulated Results

The modifications to the demodulator described above were simulated to evaluate their performance. The variable delay line was constructed to lock to a 100 kHz clock and with a precision of 2 kHz (2%) of the clock frequency. This requires that the delay element have a nominal delay of 5 μs, which can be adjusted from 2.5 μs to 7.5 μs to account for device variation. This is formed by combining a fixed delay element of 2.5 μs in series with an adjustable delay element of 0-5.0 μs.

All delay elements are constructed from weak inverters, as in Fig. 3.9, where W_N and W_P are chosen to generate equal rise and fall times, and L is then chosen to generate the appropriate delay. To ensure the clock generator has a precision of at least 2 kHz, Eqn. 3.23 indicates a required unit delay, t_0 of 20 ns. Eqn. 3.9 then requires a value of $N \geq 8$, indicating an 8-bit delay line is required.

The delay of the 8-bit delay line was simulated across the SS, TT, and FF process corners, at the minimum (0x00), medium (0x10), and maximum (0xFF) control signal inputs. The results are shown in Table 3.3. Note that at all process corners, the range of delays includes the nominal delay of 5.0 μs, indicating it can lock to the nominal input frequency of 100 kHz at any process corner.

This was constructed and simulated in a 120 nm CMOS process, with a supply voltage of 0.4 V. The result was an average power consumption during demodulation of 128 nW. Power consumption during clock generation is 70 nW, and the leakage power is 25 nW.

Table 3.3: VLDL delay in μs.

Control Bits	FF	TYP	SS
0x00	1.8	2.4	3.2
0x10	3.8	4.9	6.5
0xFF	5.8	7.5	10.0

Table 3.4: Comparison of low-power demodulator publications, as well as the original version of the demodulator presented here (Orig.) and the improved version (Imp.)

	[45]	[46]	[47]	[62]	[48]	[63]	[49][4]	Orig.	Imp.
Tech. (nm)	500	500[2]	1000	250	90	180	130	130	130
Data Rate (kbps)	250	-	32	-	200	10000	100	100	100
Power (μW)	<2.25[1]	1[3]	-	115	0.217	8.05	0.05	0.22	0.13
VDD (V)	1.5	-	1.5	1.0	0.3	0.7	0.5	0.4	0.4
E/bit (pJ)	9	0.38	-	57.5	1.09	0.8	0.5	2.2	1.3

[1]Total system power while receiving.

[2]Silicon on sapphire.

[3]Simulated total system power while receiving.

[4]Requires 200 kHz clock for operation.

3.4 Comparison to Previous Work

The performance of the demodulator presented here is summarized in in Table 3.4, along with several other recently published low-power demodulator designs, as discussed in Sec. 3.1.2. Both the power consumption and energy per bit are included. Which of these is relevant depends on the particular application and method of energy harvesting. For wirelessly powered sensors, in which the high-power carrier used to deliver energy is modulated with data, it is the power consumption that is the relevant metric. This is because power transfer can still occur during demodulation, and the demodulator can run directly from the power delivered by the carrier. In this case, it is only important that the combined power consumption of the demodulator and other sensor components is less than the delivered power. Other methods of energy

Table 3.5: Comparison of low-power clock publications, as well as the original version of the clock generator presented here (Orig.) and the improved version (Imp.).

	[58]	[60]	[57]	[55][1]	[64]	[59]	Orig.	Imp.
Tech. (nm)	180	90	130	180	65	120	130	130
Freq. (kHz)	31.25	100	100	32.8	3200	18.5	100	100
VDD (V)	1.8	0.8	1.1	0.94	0.45	1.5	0.4	0.4
Power (nW)	360	280	100	5.58	423	120	190	70

[1]Requires external crystal.

harvesting may attempt to store the received power on a large capacitor or battery, and then use that energy to power the sensor during demodulation. In these cases, the amount of available energy is fixed, so the energy per bit is the relevant metric.

It can be seen that the demodulator designs presented here compare quite favorably with current literature. The power consumption of the improved demodulator is 130 nW and requires no additional clock, which is lower than other self-contained demodulator publications. The demodulator presented in [49] has a lower power consumption of 50 nW, but requires a 200 kHz clock for operation. As will be shown, this is likely to draw at least 100 nW of additional power, brining the total power to at least 150 nW, still larger than the power drawn by this work.

The performance of the clock generator, as well as several recently published low-power clock generators operating in the kHz range, is summarized in Table 3.4. The power consumption figures are all for unloaded clock generators. The improved clock-generator presented here has a power consumption of 70 nW, lower than previously published designs, with the exception of [55], which requires an external crystal for operation. Both the original and improved clock generators can operate from a 400 mV supply, which is lower than most other designs, with the exception of [64], and have the lowest reported power consumption of all fully-integrated designs that can operate below 1 V.

3.5 Conclusions

This work has presented a low-power, fully-integrated, combined demodulator and clock generator appropriate for use in low-power sensors. By observing that both demodulation and clock generation require an accurate time reference, this work is able to substantially reduce the power consumption of demodulation and clock generation operations by sharing a calibrated time reference between both the demodulator and clock generator circuits. The original circuit consumes 220 nW during demodulation and 190 nW during clock generation. An improved version is also presented, which consumes 130 nW during demodulation and 70 nW during clock generation.

Chapter 4

Wireless Power Transmission

4.1 Introduction

One of the primary difficulties in constructing millimeter sized sensors is finding a suitable small source of electrical power. Because nearly all circuit components can be integrated, and the density of modern integrated circuits is so high, it is rarely the silicon area that constrains the minimum size of wireless sensors. Rather, it is often the relatively large power source that composes the most volume and weight of wireless devices. In general, three categories of power sources for remote sensors can be discriminated, based on the location of electrical power generation and energy storage.

The first category is tethered devices, in which electrical power is generated at a distance from the sensor, and conducted to the sensor via wires. This can allow for the construction of very small sensors, since the wire diameter can be quite small, the connection between wire and chip is quite small, and the device generating electrical power is separated from the sensor. The tether can also be used to transmit information to and from the sensor. The primary disadvantage is that tethering greatly

restricts the sensor placement and movement. The wires are of finite length, and the connection of the tether to the sensor can be damaged or destroyed if the sensor is displaced too far from the external reader. Multiple tethers can easily become tangled in the case of multiple moving sensors. Additionally, for biological applications in which the sensor is implanted in tissue, a transcutaneous tether is required that presents many difficulties. These include motion artifacts, limited sensor placement locations, difficult surgical procedure for insertion, and the need for repeated cleaning of the insertion site to prevent infection [65].

The second category is unpowered, or energy storage devices, in which no power is delivered from an external source to the sensor. Instead, the sensor contains a battery that stores sufficient electrical energy to power the devices for its complete lifetime. This category is attractive because it is both completely untethered and highly reliable: the energy stored in the battery is a well controlled variable and there is no way that power delivery can be interrupted under normal operation. The downside is that this is almost sure to significantly increase the size of the sensor, since battery capacity is proportional to the battery volume, and the battery capacity will likely need to be large. Additionally, most modern batteries with high energy density will exhibit reduced capacity with age, requiring even larger batteries for devices with very long lifetimes. This degradation is accelerated at at increased temperature, which is a particular concern for implantable biological or medical sensors [66].

The third category is energy harvesting devices, in which some form of energy other than electrical is converted into electrical energy via a transducer inside the sensor. This may include an energy storage element, such as a rechargeable battery or capacitor which can serve as a buffer [67], since the power available from harvesting is often irregular. Many sources of energy for harvesting have been proposed and examined, including thermal [26] [27] [5], solar [10], mechanical [28], biochemical [29],

and electromagnetic [68].

Although the optimal energy source depends highly on the particular application, it is the harvesting of electromagnetic energy, or radio-frequency (RF) energy, that will be considered here. As discussed in Chapter 1, many other sources of energy have limitations proscribing their use in millimeter-sized wireless sensors for general applications. Because RF energy can be broadcast by a remote transmitting circuit, the harvested energy source can be placed under control of the complete sensor system, with the timing and power of the source adjusted to meet the needs of the system. This is in contrast to many of the previously discussed sources, which occur naturally but are not directly controllable. Furthermore, an RF source can be introduced in applications where these other sources are unavailable.

A further advantage of wireless power transmission is that the only constraint on sensor placement is the range from the transmitting circuit. The transmitter irradiates a volume with sufficient strength to energize the wireless sensor, and the sensor can operate reliably at any location within this volume. This is in contrast to other energy harvesting schemes, such as thermal harvesting from a temperature gradient, in which the energy source exists only at narrower range of locations. For applications deployed in large volumes of space, multiple transmitters can be deployed and arranged so that their combined fields irradiates a large volume.

The focus of this work is on a wireless power transmission system suitable for use in millimeter-scale medical sensing applications. The primary design tradeoff in such a system is the size of the receiving antenna versus the effective range of such a system: the desire for small sensor size restricts the area of the antenna, which at millimeter dimensions directly reduces the received power and consequently the effective range of such a system. It will be shown that despite this, power can be delivered to an area of 800 cm^2. This is suitable for monitoring multiple, small laboratory animals

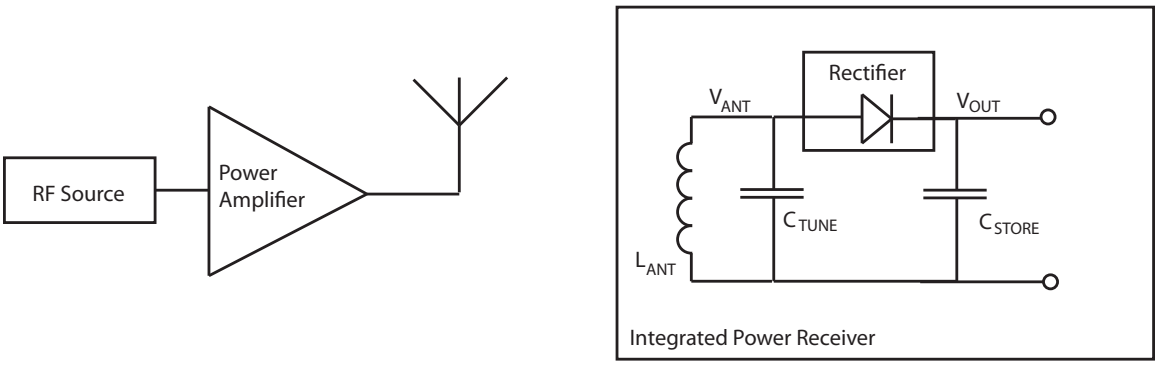


Figure 4.1: Block diagram of generic wireless power link.

in a cage, or other sensor applications in which the distance from the transmitter to the sensor is restricted to tens of centimeters.

4.2 Background and Previous Work

A diagram of a generic wireless power link is shown in Fig. 4.1. The link consists of an external transmitter, which is composed of an RF source, power amplifier and transmitting antenna. The sensor is composed of a coil of wire serving as an antenna, indicated by the inductor L_{ANT}, and frequently a tuning capacitor C_{TUNE}. The inductive antenna and capacitor form a resonant LC circuit with frequency

$$f = \frac{1}{2\pi\sqrt{L_{ANT}C_{TUNE}}} \tag{4.1}$$

C_{TUNE} is adjusted to set the LC resonant frequency equal to the transmitter frequency. The electromagnetic field generated by the transmitting antenna will excite the LC circuit, creating an AC voltage across the antenna, V_{ANT}. This AC voltage is converted to DC by a rectifier. Although indicated by a diode, there are a variety of more complex rectifier circuits suitable for this application. The rectifier then charges a storage capacitor, C_{STORE}. The output of this link is suitable for driving a DC/DC

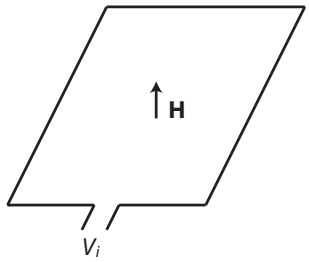


Figure 4.2: Receiving loop antenna in magnetic field, **H**.

converter or voltage rectifier, which can then be used to power the remaining sensor circuitry.

A central difficulty of this architecture is that rectifiers commonly require some minimum input voltage magnitude, V_{min}, to function. The power delivered from the link is therefore highly nonlinear with respect to the transmitter power: if the transmitter power is below the threshold required to generate V_{min}, no DC voltage is generated by the rectifier. Maximizing V_{ANT} is consequently a high priority in the power link design.

4.2.1 Basic Model

Fig. 4.2 is drawing of a simple receiving coil geometry within a field generated by the transmitting antenna. The magnetic field induces a voltage across the coil terminals given by Faraday's Law:

$$V_i = j\omega\mu \int_S \mathbf{H} \cdot d\mathbf{S} \tag{4.2}$$

where ω is the angular frequency, μ is the magnetic permeability of the medium, **H** is the time-harmonic magnetic intensity, and S is the surface bounded by the coil. When the receiving coil dimensions are much smaller than the wavelength of the

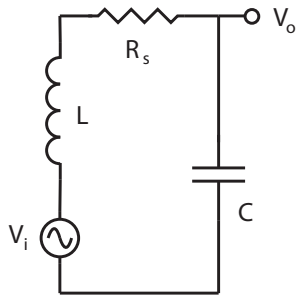


Figure 4.3: Series RLC circuit formed by coil and tuning capacitance. R_s is the series resistance of the coil.

transmitted frequency, there is little variation in **H** across S and Eqn. 4.2 becomes

$$V_i = j\omega\mu H A \cos\theta \tag{4.3}$$

where H is the magnitude of the magnetic intensity, A is the area of the coil, and θ is the angle between **H** and a vector normal to the coil.

By placing a capacitor in parallel with the receiving coil, the magnitude of the received voltage can be increased. This forms a series RLC circuit, as shown in Fig. 4.3. Driven by the induced voltage, V_i, the coil and capacitor form a voltage divider.

$$\frac{V_o}{V_i} = \frac{Z_C}{Z_L + Z_C} \tag{4.4}$$

$$Z_C = \frac{1}{j\omega C} \tag{4.5}$$

$$Z_L = j\omega L + R_s \tag{4.6}$$

To achieve resonance, the value of C is fixed by Eqn. 4.1 given particular values of L and ω. Combining Eqns. 4.1, 4.4, 4.5 and 4.6 gives

$$\frac{V_o}{V_i} = \frac{j\omega L}{R} = jQ \tag{4.7}$$

where Q is the inductor quality-factor, the ratio of the inductor's reactance to its series resistance. The magnitude of the voltage developed by the resonant LC circuit can then be found by combining Eqns. 4.3 and 4.7:

$$|V_o| = \omega \mu H A Q \cos\theta \tag{4.8}$$

Eqn. 4.8 indicates that to maximize the voltage developed by the receiving coil, the transmitter frequency, coil area, and coil quality factor should be maximized. As has already been discussed, increasing the coil area is often impossible since medical and biological sensors are almost always size constrained. Increasing the transmitter frequency is also difficult, since high frequency electromagnetic waves can be significantly attenuated by biological tissue, which would have the effect of decreasing H in Eqn. 4.3. Decreasing the series resistance, and thereby increasing the quality factor, is always desirable.

In general, there have been two approaches to choosing the transmitter frequency and designing the power transmission link [68]. The first approach is based on purely inductive coupling [69], in which the transmitter frequency is in the range of tens of MHz. The second, more recently explored approach, is to use a higher transmitter frequency in the GHz range, trying to balance the increased electromagnetic power density at high frequency with the increased attenuation through lossy material [70] [71].

4.2.2 Inductive Coupling

In inductively coupled systems, the transmit antenna geometry, receiving coil geometry, and separation between the two are all significantly smaller than the wavelength of the transmitter frequency. Inductive coupling has the potential for high efficiency, in part because the electrically small transmitter radiates little energy into space. It is frequently used in high power applications such as power distribution, as well as induction furnaces and stoves. Unfortunately, for a millimeter sized receiving coil separated by more than a few millimeters from the transmitter, the coupling between the transmit and receive coils is generally poor [68]. Electrically, this configuration can be modeled as a poorly coupled electrical transformer [69].

A modern version of this approach was constructed with a custom integrated receiving antenna, built from an electroplated gold coil stud bumped on top of a bulk CMOS IC [13]. The coil has an area of 2×2.5 mm^2 with 35 turns. The gold coil provides an unusually high quality factor, Q, of 6.7. When transmitting at 13.56 MHz, the rectifier achieves an output voltage of 3.19 V and transfers 250 mW of power when the receive and transmitting coils are coincident. Unfortunately, when the coils are separated by 1 mm the output voltage drops to 2.3 V, and no data beyond that range is provided.

Another inductively powered sensor has been constructed with an integrated coil using standard bulk CMOS metal layers [72]. The coil is 2×2.18 mm^2 with 4 turns constructed from 2 metal layers. The sensor is also unusual in that in contains a switchable bank of binary weighted tuning capacitors that are switched in parallel with the coil by successive approximation logic, to provide optimal tuning. Separated from the transmitting coil by 10 mm and transmitting at 160 MHz, a power link gain of approximately -20 dB was reported.

4.2.3 High Frequency Power Transfer

High frequency power transfer differs from inductive power transfer in that the geometry of the transmitting coil, or the distance between the transmitting and receiving coils, is a substantial fraction of the wavelength of the transmitter frequency. In this arrangement, the transmitting antenna may radiate significant power, and the electromagnetic field incident on the receiving coil is a combination of the inductive and radiative modes of the transmitting antenna [68]. Recent research suggests that the increased power at higher frequency outweighs the higher attenuation by biological material, and that the optimal transmission frequency may be in the GHz range [70].

Several wireless power transmission links have been published in GHz range, including a wirelessly powered glucose sensor has been presented that uses a circular, 1 cm diameter receiving antenna, operating at a frequency of 1.8 GHz [9]. The antenna is constructed from gold, with a trace width of 0.5 mm, and a thickness of 5 μm. The rectifier output is 1.2 V, and is able to deliver 3 μW of power at a distance 15 cm from the transmitter, operating at an effective isotropic radiated power (EIRP) of 40 dBm. The 1 cm antenna diameter is relatively large compared to other millimeter-sized wirelessly powered sensors.

Another wireless power link operating at 915 MHz has been presented, with a rectified output of 1.2 V and 140 μW of power delivered through 15 mm [73] of bovine muscle tissue. The transmitting antenna is a 2×2 cm^2 square loop, and the receiving antenna is a 2×2 mm^2 square loop constructed from a bond wire. A 3-stage, self-synchronous rectifier is used to generate the DC output voltage. Additionally, a bank of binary weighted switchable capacitors can be switched in parallel with the receiving antenna for adaptive tuning, similar to [72].

Small antennas can also be constructed from on a small PCB, and then flip-chip

bonded to a sensor IC. This approach has been presented with a 1.1 × 1.1 mm^2 antenna, operating at 533 MHz [32]. The transmitting antenna is a 15 mm loop. Over a distance of 13 mm, 660 nW of power transmission is achieved.

4.2.4 Receiving Coil Positioning

For applications in which the positioning of the receiving antenna relative to the transmitting antenna is fixed, or in which either can be moved by the operator to achieve optimal positioning, the power transmission link only needs to function at over a small volume of possible receiving coil positions. This would be the case, for example, for an implantable sensor and a handheld reader that contains the transmitting antenna, which could be waved by hand near the sensor. Other applications may have the transmitting antenna position fixed, with only limited control over the receiving antenna positioning. This is the case for automated systems that record data from sensors implanted or attached to freely moving animals. For example, small laboratory animals in a cage are free to move about an area on the bottom of the cage, and possibly the whole volume of the cage if there are structures on which they can climb.

4.2.5 Effect of Field Orientation

Another difficulty of wireless power transmission is that the received voltage varies with the receiving coil's alignment in the transmitted field. This can be seen from Eqn. 4.2, in which the induced voltage depends on the dot product of the magnetic field vector and a vector normal to the coil. Consequently there are null orientations, in which the normal vector of the coil is at a right angle to the magnetic field, which generate no induced voltage. This constrains the orientations of the receiving coil

under which power transmission can occur.

There are essentially two solutions to this problem. The first is to solve the problem on the receiving end, by constructing multiple coils with differing orientations. For example, by using three receiving coils that are mutually orthogonal, it is impossible for all three coils to be simultaneously orthogonal to the transmitted field [74]. The difficulty with this is that it increases the manufacturing complexity of the receiving chip, as well as substantially increasing the volume of the receiver.

The problem can also be solved on the transmitting end, by some mechanism which varies the orientation of the transmitted field. One possibility is to use multiple transmitting antennas placed at different locations [75]. At a particular position in space, p_0, a transmitting coil or linearly polarized antenna creates its own field vector, $\mathbf{H}(p_0)$. Multiple antennas energized sequentially will create corresponding fields, $\mathbf{H}_0(p_0)$, $\mathbf{H}_1(p_0)$, ..., $\mathbf{H}_n(p_0)$. With suitable selection and placement of the transmitting antennas, these fields can be guaranteed to have mutually orthogonal components, such that there is no orientation of the receiving coil that will generate zero induced voltage.

Another possibility for varying the orientation of the transmitted field is the use of a circularly polarized antenna [76]. With circular polarization, the direction of $\mathbf{H}(p_0)$ is not constant, but rotates in a plane with time:

$$\mathbf{H} = H_0[\hat{\mathbf{a}}\cos\omega t - \hat{\mathbf{b}}\sin\omega t] \tag{4.9}$$

where the orthogonal unit vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ define the plane of rotation. Note that there is still a third orthogonal vector, $\hat{\mathbf{c}} = \hat{\mathbf{a}} \times \hat{\mathbf{b}}$, along which if the receiving coil is aligned, no power transmission will occur. A single polarized antenna therefore reduces the possible null orientations, but does not eliminate the possibility entirely.

This approach could of course be combined with the preceding approach of using multiple antennas to eliminate this possibility.

4.3 Wireless Power Transmission Link

The design space for a wireless power transmission links is large, and the optimal design will depend on the specific application for which it is intended. The focus of this work is on wireless power transmission links for applications in which the receiving antenna is constrained to several square millimeters or smaller, typically for biological or medical applications.

Most similar work, such as that previously described in Sec. 4.2 tends to assume a particular application model in which the receiving antenna is some distance from transmitting antenna, but not free to move within an area or volume. That is, the receiving antenna position is constrained to a particular axis relative to the transmitting antenna, usually in the direction of maximum field strength. If not stated directly, this model is implied by only reporting performance of the link versus a single variable, the antenna separation distance. For example, [73] reports an effective range of 7.5 mm to 17.5 mm, and performance with just 1 mm of axial misalignment. This model is not unreasonable and in fact applies to a large number of potential applications: for example, applications in which the transmitting antenna is hand-held, or can be otherwise freely moved in relation to the receiving antenna to achieve the optimal axial alignment.

However, this assumption is not valid for many other potential applications. For example, consider an implantable sensor intended to monitor laboratory animals. The system would consist of one or more antennas mounted to the animal cage, in which the animals are free to move about the cage. In this application, the receiving antenna

cannot be constrained to a single axis relative to the transmitting antenna. The link must be characterized within either the whole volume of the cage, or at least within a plane in the cage (if the animals are restricted to moving on the cage floor). This type of characterization has been performed for other wireless power transmission systems [74] [75] [77] [76], although with the size of these receivers tends to be on the order of centimeters or more, with the exception of [76].

The system presented here has been characterized in both ways: with the receiver at variable separation from the transmitter along a single axis, and at a fixed height above the transmitter over a planar area. The results can then be compared for both fixed and freely-moving applications.

4.4 Rectifier

To convert the AC voltage developed by the receiving coil to a DC voltage, an 8-stage Dickson charge pump was used. A schematic of the rectifier is shown in Fig. 4.4. To understand the operation of the rectifier, imagine that a sinusoidal input with magnitude V_in is applied to $V_{RF,in}$. Initially $V_A < V_{DC,out}$, and the diode-connected PMOS device M2 is off. During the negative swing of $V_{RF,in}$, the voltage at V_A falls below $V_{DC,in}$, until the voltage drop across M1 becomes equal to M1's threshold voltage, V_T. At this point point M1 begins to conduct, and holds $V_A = V_{DC,in} - V_T$. This continues until V_{RF} reaches its most negative value and then begins to rise, at which point M1 turns off. This causes the voltage at V_A to rise an equal amount, since neither M1 or M2 are conducting and there is no current flowing through capacitor C1. V_{RF} increases $2V_{in}$ as it rises from its most negative to its most positive value, causing V_A to rise to $V_{DC,in} - V_T + 2V_{in}$. When V_A rises a threshold voltage above $V_{DC,out}$, M2 begins to conduct, charging the output to the value of V_A minus one

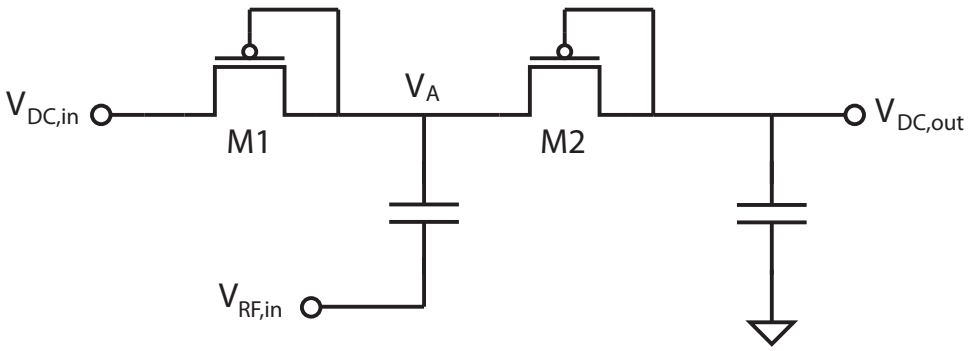


Figure 4.4: Single stage of a Dickson charge-pump type AC/DC rectifier.

threshold voltage:

$$V_{DC,out} = V_{DC,in} + 2(V_{in} - V_T) \tag{4.10}$$

Eqn. 4.10 illustrates an important property of rectifier circuits, which is that there is some rectifier threshold voltage below which no DC output is generated. In the case of this rectifier, the threshold is simply equal to the PMOS threshold voltage, V_T. This is a central difficulty of wireless power transmission, since it means no DC power transfer can occur without developing a minimum voltage across the receiving coil.

In actual practice, the diode connected devices M1 and M2 do not behave as ideal diodes, for which there is zero current for voltage biases below the threshold. Rather, some sub-threshold current still flows, allowing operation at input voltages below V_T. However, the rectifier output current, and thus the output power, will be much lower than if the input voltage is above V_T and M1 and M2 operate in strong inversion.

Active rectifier circuits have been proposed, in which the rectifier consists of some arrangement of actively driven devices. This can improve efficiency, and allow for the rectification of lower input voltages, but these rectifiers require some DC power

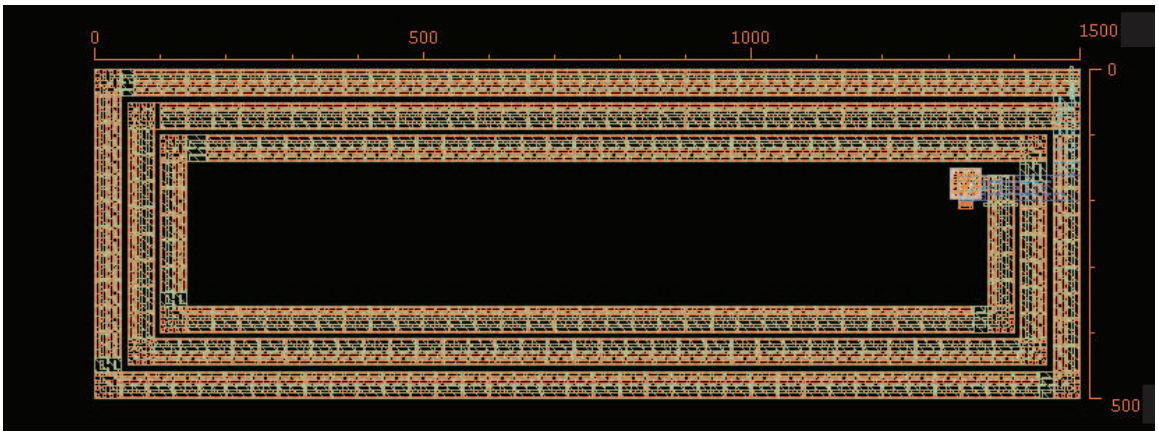


Figure 4.5: Physical layout of integrated receiving coil. Outer dimensions are 1.5 × 0.5 mm^2, with 3 turns. Trace width is 40 μm with 10 μm spacing between traces.

source for active operation. This creates a circular dependence in which the rectifier requires a DC voltage for operation, but no DC voltage is available until the rectifier begins operating. Therefore some passive rectifier design is still required to start the system, and the use of an active rectifier cannot directly remove the minimum input voltage constraint.

4.5 Integrated Antenna

The first receiving antenna design explored was a fully integrated design, constructed from the metal layers of a standard bulk CMOS process. The primary benefit of integrating the antenna is reducing manufacturing costs, since it requires no additional manufacturing steps to construct the antenna and attach it to the sensor. The second benefit is that the sensor thickness is likely reduced, since an external antenna will have to sit above the sensor IC. However, because the IC thickness is much less than the IC width and length, the benefits of reducing the thickness are not as significant as reducing the width or length.

The layout of an integrated coil is shown in Fig. 4.5. The outer dimensions are

1.5×0.5 mm^2. The trace width is 40 μm, with 10μm spacing between the traces. To reduce the series resistance of the coil, the traces are composed of 4 metal layers wired in parallel. The coil is composed of 3 turns.

4.5.1 Simulation

A 3D model of the coil was constructed in Ansoft HFSS, a full-wave electromagnetic simulator. Because the dimensions of the receiving coil are small, it can be modeled by lumped element circuit analysis. In this case, the model is simply an inductance, L_c, and resistance, R_c, in series. The purpose of the simulation is to extract the values of the parameters of the model.

The results of the simulation are the 1-port scattering (S) and impedance (Z) parameters of the coil. The circuit parameters are derived from the Z-parameters using the following relations:

$$L_c = \frac{\mathbf{Im}[Z_{11}]}{\omega} \tag{4.11}$$

$$R_s = \mathbf{Re}[Z_{11}] \tag{4.12}$$

The simulation is run over a frequency range of 750-1000 MHz, and the circuit parameters extracted over this range. Plots of L_c and R_c are shown in Fig. 4.5. Note that both parameters are relatively constant over the simulated range, which indicates the soundness of the two element model. Evaluated at a frequency of 750 MHz, the value of L_c is approximately 12.5 nH, and R_c is 7 Ω.

The value of the inductance is required prior to fabrication, so that a tuning capacitance can be added in parallel with the coil to set the resonant frequency. For

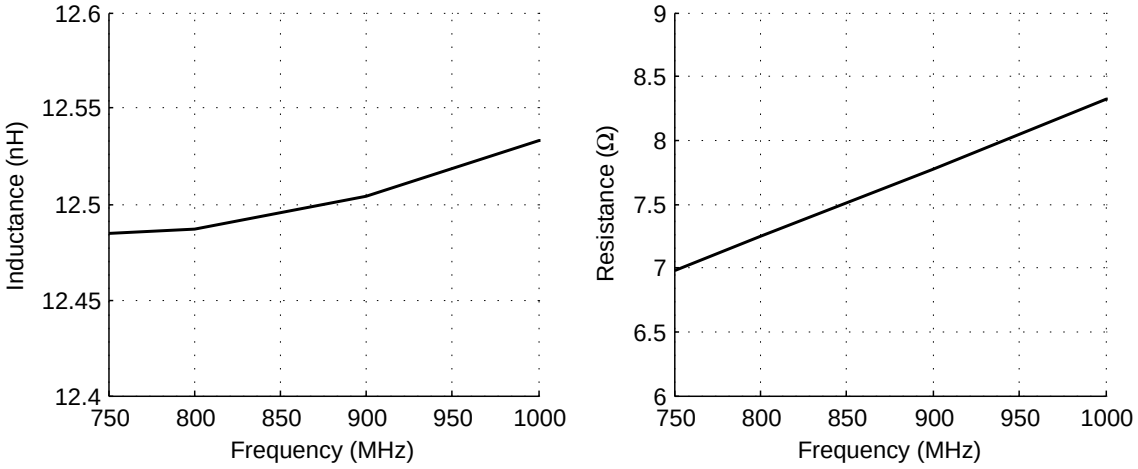


Figure 4.6: Simulated inductance and resistance of integrated coil antenna shown in Fig. 4.5.

this chip, a tuning capacitance value of 3.7 pF was chosen, and constructed using a high-quality metal-insulator-metal (MIM) integrated capacitor. Using Eqn. 4.1, this gives a resonant frequency of 740 MHz.

4.5.2 Experimental Results

The integrated antenna and rectifier was manufactured in a standard 0.13 μm bulk CMOS process. To evaluate the performance of the antenna, the chip was wire-bonded to a PCB rather than packaged to reduce parasitic inductances due to the package bond wires and pins. Because the coil geometry is so small, direct measurement of the voltage developed across the coil was not possible: the wires necessary to connect the coil to instruments for measurement would develop induced currents, just as the coil does, as well as shifting the resonance of the LC circuit. Therefore, the link performance is measured at the rectifier output. For this experiment, only the chip ground and rectifier output are wire-bonded to the PCB.

To test the antenna, the PCB was placed above a wide-band horn antenna. The

horn antenna was driven by a wide-band RF power amplifier, driven by an RF signal generator. Transmitting at +40 dBm, the frequency of the transmitter was swept from 500 MHz to 1 GHz in 1 MHz increments and the output of the rectifier observed with a DC voltmeter. Unfortunately, no significant DC output voltage was observed from the rectifier over this range, or for repeated experiments in which the placement and orientation of the chip were varied above the transmitting antenna.

The are several possible explanations for the null result of this experiment. The measured DC resistance of the coil is low, which suggests the result is not caused by break in the coil. It is possible, and in fact likely, that the actual values of L and C are different from those design for, which would shift the resonant frequency away from its nominal value of 740 MHz. However, the amount of deviation required to shift the resonant frequency outside the range of 500-1000 MHz is very large. The most likely explanation, which is supported by the measured performance of a larger antenna in the following section, is simply that the coil geometry of 1.5×0.5 mm^2 is too small to develop sufficient voltage at the transmitted power level. Compared to current publications, this coil area of 0.75 mm^2 is smaller than any of the millimeter-sized coils described in Sec. 4.2: the smallest is 1.21 mm^2 [32], which uses a flip-chip bonded antenna rather than an integrated antenna, likely having much lower series resistance.

4.6 Bond-wire Antenna

In addition to the integrated antenna, an antenna constructed from a bond-wire was also fabricated and measured. This was possible because the original chip contained a duplicate rectifier, not connected to the integrated coil. The chip was glued to a PCB, wire-bonding the rectifier input and ground to the PCB, and then forming a loop

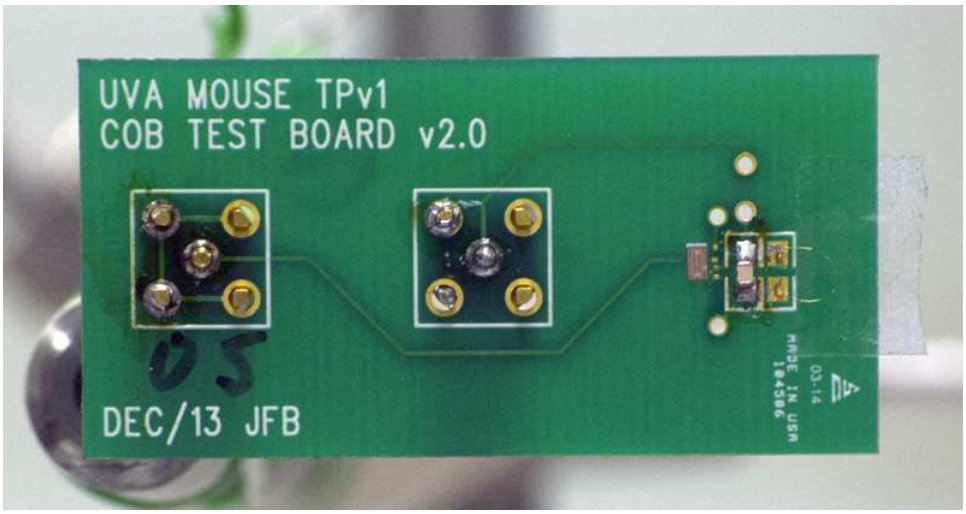


Figure 4.7: Measured rectifier output voltage variation with antenna separation.

antenna with a length of gold bond-wire. A 1.1 pF surface mount (SMT) capacitor is soldered in parallel with the bond wire loop to achieve a resonant frequency near 950 MHz. The SMT capacitor and bond-wire loop were placed as close to the chip as is possible to minimize parasitic inductances. Note that the use of an SMT capacitor was only necessary because it made the value of the capacitance adjustable for testing purposes. A capacitance of 1.1 pF is easily integrable using a metal-insulator-metal capacitor. A photograph of this PCB is shown in Fig. 4.7. From the photograph, the coil area was measured to be 12.3 mm^2.

The resonant frequency was found by placing the chip 10 cm above a wide-band horn antenna, and sweeping the transmitter frequency over 800-1000 MHz in 5 MHz increments, while transmitting at +40 dBm. The peak DC voltage at the rectifier output was 0.98 V at 920 MHz.

4.6.1 Patch Antenna

Although useful for testing duo to its large bandwidth, the horn antenna is impractical for use as the transmitter in most wireless sensor systems because of its significant

size, weight and expense. A rectangular patch antenna was chosen as a more practical alternative, since it is thin, light-weight and can be inexpensively manufactured on a 2-layer PCB.

The patch antenna is essentially formed by rectangle of metal, on the top layer of the PCB, situated above a ground plane on the bottom of the PCB. The bandwidth of a patch antenna is relatively narrow, with the center frequency depending on the length of the patch. For a chosen center frequency f, the patch length, L, is given by

$$L = \frac{c}{2f\sqrt{\epsilon}} \tag{4.13}$$

where c is the speed of light and ϵ is the electric permittivity of the dielectric [78]. A patch antenna was designed and fabricated with patch dimensions $L = 7.3$ cm and $W = 10.5$ cm. The total size of the PCB is 10×13 cm^2. The patch antenna thickness is quite small, less than 2 mm.

4.6.2 Experimental Results

Several experiments were conducted to evaluate the performance of the power transmission link with the bond-wire antenna. In these experiments the the patch antenna was oriented parallel to the ground, facing upwards. The PCB containing the chip and bond-wire antenna was suspended above the patch antenna with the loop oriented in the YZ plane. A diagram of this is shown in Fig. 4.8, which also indicates the coordinate system used in the following discussion.

It should be noted that because the angle of the receiving coil is fixed, this work does not directly address the worst case misalignment effects described in Sec. 4.2.5. This is the case for most other wireless power transmission publications as well. As an area of future research, the effects of misalignment may be mitigated by appropriate

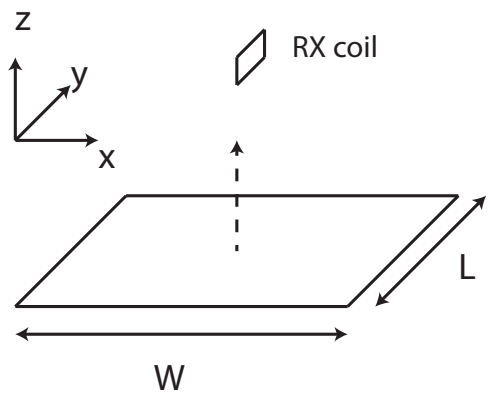


Figure 4.8: Placement and orientation of receiving coil relative to transmitting patch antenna (not to scale). The axes origin (0,0,0) is in the center of the patch antenna.

tiling of transmitting antennas, or using a circularly polarized rather than a linearly polarized patch antenna [78], similar to the methods described in Sec. 4.2.5.

In the first experiment, the antenna was positioned as shown in Fig. 4.8, at a distance 20 cm above the center of the patch antenna. The transmitter frequency was then varied over the narrow bandwidth of the patch antenna, while observing the DC output voltage of the rectifier. The maximum DC output level was found at a transmitter frequency of 925 MHz, and this frequency was used in the following measurements. The rectifier open-circuit voltage and short-circuit current was then measured by varying the transmitted power level. These are plotted in Fig. 4.9.

It can be seen from Fig. 4.9 that the rectifier output voltage and current vary non-linearly over the transmitted power levels. Both the rectifier output voltage and current increase sharply at approximately 35 dBm of transmitter power. This is likely due to the transition of the rectifier from sub-threshold to strongly inverted operation, as described in Sec. 4.4. As expected, the rectifier output current in sub-threshold is quite small.

It can also been seen from Fig. 4.9 that the rectifier output voltage seems to

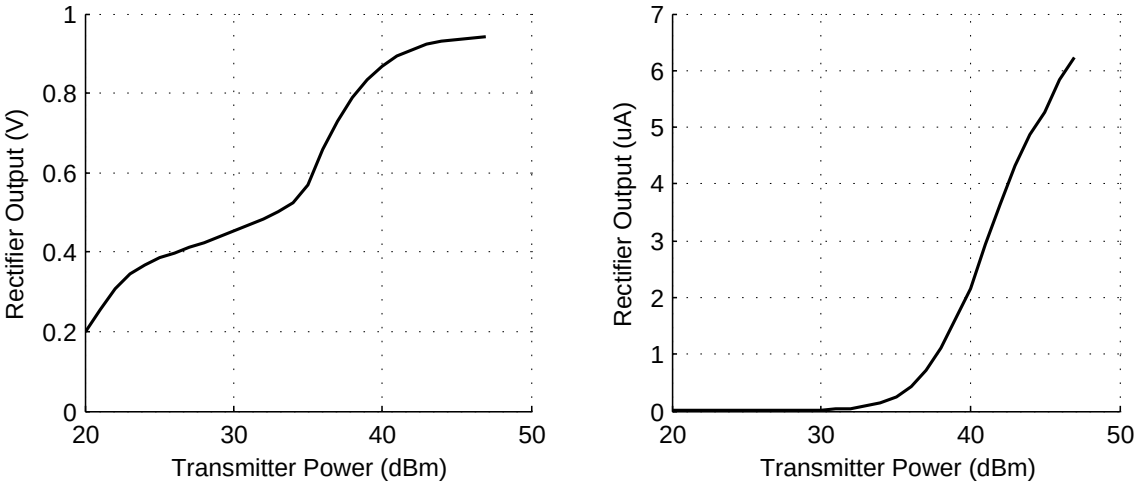


Figure 4.9: Rectifier open-circuit voltage and short-circuit current with bond-wire antenna, situated 20 cm above transmitting patch antenna. Transmitter frequency is 925 MHz.

saturate as the power level is increased, although the short-circuit output current does not. This behavior was not expected, and is due to on-chip leakage at the rectifier output node, possibly through ESD diodes in the output pad. That this saturation is due to leakage is confirmed by applying a DC voltage to the rectifier output and measuring the resulting current, in the absence of any applied field. This is shown in Fig. 4.10, which shows substantial current increase above 0.8 V, close to the saturation level seen in Fig. 4.9.

In the next experiment, the rectifier open-circuit voltage and short-circuit current were measured as the receiving coil was moved throughout a volume above the patch antenna. Because the magnitude of the fields produced by the patch antenna has symmetry across the X- and Y-axes [78], the volume was restricted to a single quadrant above the antenna. The receiving antenna was moved between 0-30 cm in both X and Y directions, and 5-25 cm in the z direction, in 5 cm increments. This forms a $30 \times 30 \times 25$ cm^3 volume with a 5 cm grid, resulting in 245 unique measurement points. The transmitter power was set to +40 dBm for these measurements.

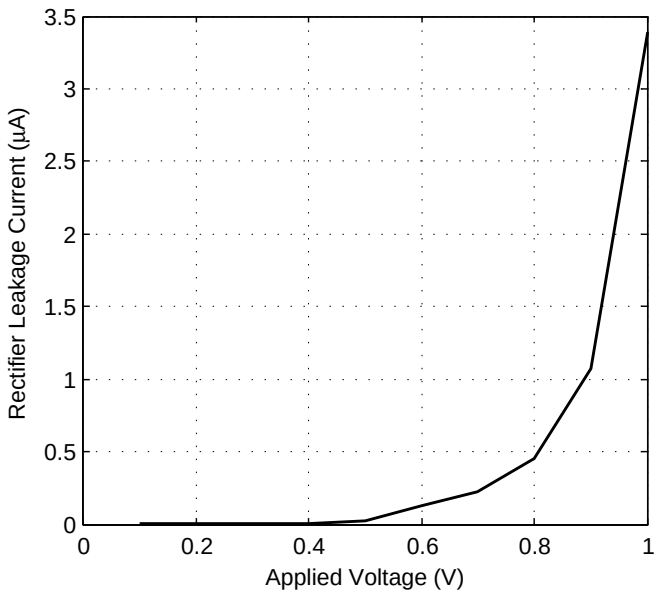


Figure 4.10: Measured rectifier output leakage current variation with rectifier output voltage.

A photograph of the experimental setup is shown in Fig. 4.11. The patch antenna as well as the PCB with chip and receiving coil can be seen. The positioning grid was aligned with the transmitting antenna such that positioning of the receiving coil in the XY plane can be accomplished by moving the base of the support along the grid. The accuracy of the positions reported here are likely within a few mm. It should also be noted that the measurements were not taken within an anechoic environment, and are likely influenced by reflections within the room.

Several curves can be extracted from the measured data to help illuminate the way in which the system performance varies as the receive coil position changes. Fig. 4.12 shows how the rectifier output voltage changes with separation along the Z direction. This is shown for the coil above the center of the antenna at (0 cm, 0 cm) in the XY-plane, as well as offset from center, at (25 cm, 0 cm) and (0 cm, 25 cm) in the XY-plane.

If the X and Y offset is varied while the height above the antenna is fixed, the

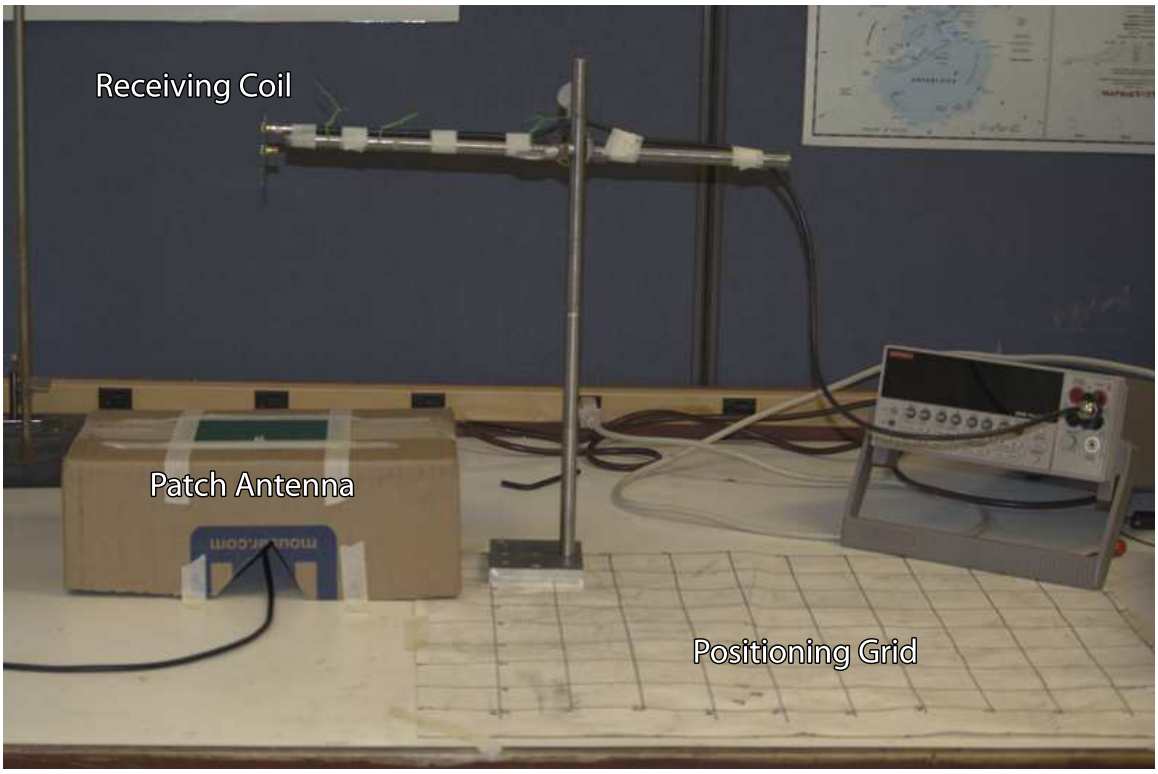


Figure 4.11: Experimental setup to measure power transmission.

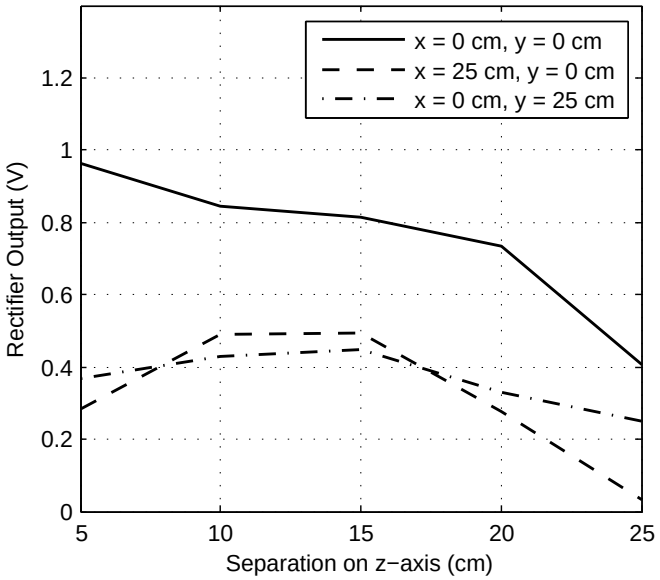


Figure 4.12: Measured rectifier output voltage variation with antenna separation.

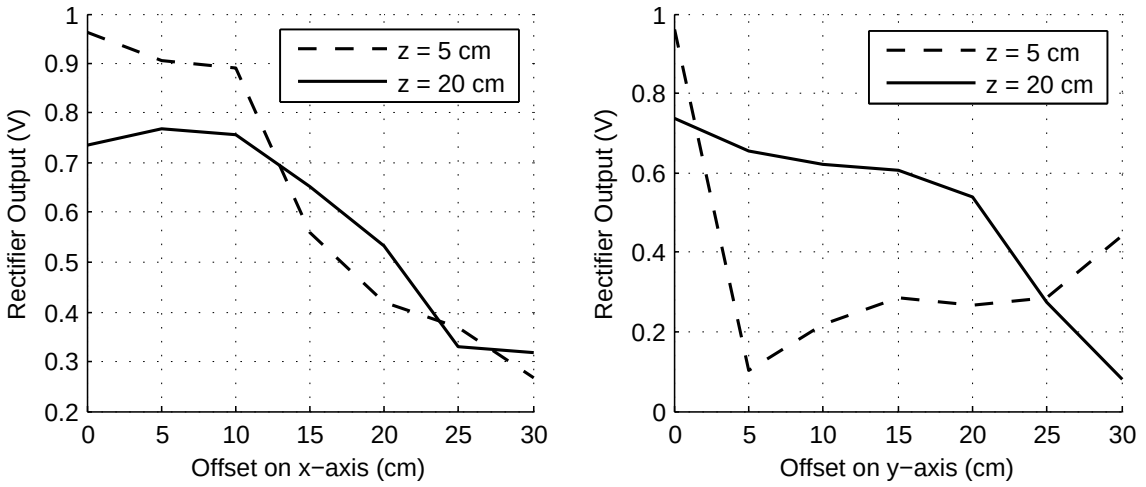


Figure 4.13: Measured rectifier output voltage 5 cm and 20 cm above patch antenna. Receiving antenna is displaced along x-axis in the left figure, and along the y-axis in the right figure.

curves in Fig. 4.13 are produced. Two heights, 5 cm and 20 cm above the patch antenna, are plotted. It can be seen that, 5 cm above the patch antenna, the rectified output voltage falls off sharply as the receiving coil is moved along the Y-axis. However, if the height above the patch is increased to 20 cm, the received voltage is much less sensitive to movement along the Y-axis. This is a reasonable result, since spatial derivatives of the electric and magnetic fields tend to be highest very close to the transmitting antenna, and decrease as separation from the antenna increases.

This suggests the interesting and somewhat counter-intuitive result that for applications in which the receiving antenna is constrained to move within a plane above the patch, for example in the case of the laboratory animals moving on the floor of a cage, placing the transmitting antenna as close as possible does not give the best system performance because. Rather, some distance should be placed between the transmitter and the target plane so as achieve more uniform coverage.

By examining the collected data of rectifier voltage versus position, it was found

Table 4.1: Measured power output of rectifier at 400 mV DC, with receiving antenna displaced along X- and Y-axes. Z-separation is 20 cm. Transmitter power is 10 W at 925 MHz.

		X disp.				
		0 cm	5 cm	10 cm	15 cm	20 cm
Y disp.	0 cm	2.2	1.9	1.9	1.9	0.8
	5 cm	1.8	1.7	3.0	2.1	1.7
	10 cm	0.9	1.8	2.9	2.3	1.2

that at 20 cm above the transmitting antenna, the receiving coil produces greater than 400 mV of DC rectified voltage when positioned within rectangle of width 20 cm in the X direction and 10 cm in the Y direction. The amount of power delivered within this area was then found by holding the output voltage of the rectifier fixed at 400 mV, and measuring the current flowing out of the rectifier. From this measurement the power was computed, and is shown in Table. 4.1.

These measurements show the transmitting power in only the positive X- and Y-directions. The total area covered by the antenna is thus four times this area: 20×40 cm^2. Because this area is significantly larger than the dimensions of the transmitting antenna, 10×13 cm^2, one can imagine a system in which multiple transmitting antennas are tiled to cover an arbitrarily large area. The antennas could share a single power amplifier with switching to energize a single antenna at a time. This would allow a large area to be scanned to search for a particular sensor, or to record measurements from a whole population of sensors.

Table 4.2: Comparison of wireless power transmission publications with receiver placement along a single axis.

	[13]	[9]	[72]	[32]	[73]	[79]	[80]	[81]	This work
Technology (nm)	350[A]	130	130	65	130	180	130	600	130
RX Antenna Area (mm^2)	5	78.5	4.36	1.21	4	3.3	31mm[B]	19.6	12.3
Frequency (GHz)	0.014	1.8	0.16	0.533	0.915	5.2	3.65	2.765	0.925
RX DC Voltage (V)	2.3	1.2	3.1	0.5	1.2	1	1.5	3.3	0.4
RX Power (μW)		3		0.66	140		2250	720	2.2
Range (mm)	1	150		13	15	35	100	30	200
TX Power (W)		10		> 1	0.25	5	1		10

[A] With Au electroplating.

[B] Monopole antenna.

Table 4.3: Comparison of wireless power transmission publications with receiver placement in a planar area at a height above the transmitter.

	[75]	[77]	[76]	This Work
Frequency (MHz)	0.606	0.120	1600	925
Receiver				
Antenna Size (mm^3)	113	47	32	12.3 $mm^{2,A}$
Plane Height (cm)	3		5	20
Plane Area (cm^2)	100	50	254	800
DC Voltage (V)	3			0.4
Minimum Power	20 mW	2 W	20 μW	0.8 μW
Transmitter				
Antenna Size (cm^2)		1135^B	12	130
Power (W)		40^C	0.5	10

^{A}Single-turn loop antenna. Others are cylindrical solenoids.

^{B}Total area of transmitting device. [82]

^{C}Total power consumption of transmitting device. [82]

4.6.3 Comparison to Previous Work

The performance of several recently published power transmission links intended for millimeter-sized sensors is shown in Table. 4.2. These values in the table assume optimal orientation and placement of the coil, with movement of the sensor along just a single axis. It can be seen that this work has the largest range among works with mm-scale receiving coils: 20 cm compared to the next largest range of 15 cm [9], with substantially smaller receiving antenna area, 12.3 mm^2 vs. 78.5 mm^2, and similar power delivery, 2.2 μW vs. 3 μW. The cost of this is low DC output voltage power. The 0.4 V developed by the rectifier in this work is lower than the DC output voltage of most other works, but still large enough to drive circuits operating in sub-threshold.

Additionally, the performance of several recent works for which the receiving an-

tenna is free to move in a planar area some height above the transmitting antenna are shown in 4.3. The amount of received power reported is the minimum power received over the total area, i.e. the received power under worst-case positioning. This work varies from others in that it uses a single-turn bond-wire loop, while the others use cylindrical solenoids that occupy a much larger volume and are more difficult to construct. This work is also able to cover a larger area at greater height than other works, 800 cm^2 vs. 254 cm^2 [76], albeit with lower DC voltage and output power.

4.7 Conclusions

Taken together, Tables 4.2 and 4.3 show the tradeoffs inherent to wireless power transmission systems. The major variables of the design space are the effective range, the receiving antenna size, and the received power. This works demonstrates that effective range and area of wireless power transmission systems can be significantly improved above current limitations, at the price of reduced power delivery. Whether this cost can be incurred of course depends on the particular application. However, while low, the received DC voltage and power developed by this system is above the level required for many circuits common to wireless sensors that operate in subthreshold, as has been demonstrated in the other chapters of this work. With a receiving antenna area of 12.3 mm^2, the system is able to deliver 2 μW of DC power at a range of 20 cm, and a minimum of 0.8 μW of power over an area of 800 cm^2, with a minimum DC voltage of 0.4 V.

Chapter 5

Conclusions

The central thesis of this work is that the effective range of millimeter-sized wirelessly powered sensors can be increased, and the size of the sensor reduced, by focusing on reducing the voltage and power requirements of the sensor. This work provides a substantial amount of evidence that this is case, by both demonstrating that several circuits common to wireless sensors can be designed to run at low power and voltage levels, and by demonstrating that the range of wirelessly powered sensors can be improved, while still maintaining sufficient received voltage and power to drive circuits operating in or near sub-threshold.

Chapter 2 describes a method for identifying individual sensors using low energy, just 39 fJ/bit while operating from a 0.4 V supply, based on variation inherent to the integrated circuit manufacturing process. Because methods of variation based identification may be unreliable, a circuit architecture is presented that allows for the characterization of individual ID bits as either reliable or unreliable using only two read operations. Several protocols for addressing individual chips are presented that use this reliability information to improve the reliability of the addressing system. A mathematical analysis of these methods is presented, and used to determine which has

the highest performance. A model for estimating the reliability of individual bits due to electrical noise is also presented, as well as an analysis of the effects of temperature variation. The circuit was fabricated in a 0.13 μm CMOS process and tested. The parameters of the addressing system for use with a hypothetical application are also presented, which can address 1000 unique chips using 31 bits per chip, and a total error rate of less than 10^{-6}. Unlike similar implementations of identification methods based on random variation, this method does not require the sensor to transmit its ID back to the external reader, which further reduces the energy required.

Chapter 3 describes a novel architecture for a low-power, combined demodulator and clock-generator. The demodulator is able to compensate for the effect of manufacturing variation by employing a calibrated delay line, and the same calibration is used for generating an accurate on-chip clock. The power consumption of delay lines is examined and it is shown that power consumption can be significantly decreased by decreasing the supply voltage and charging current. Based on this, two versions of a low-power, digitally controlled, variable-length delay line are presented, based on chains of current-starved and long-length inverters. The effect of the delay line resolution on the clock generator precision is also examined. The demodulator and clock-generator was fabricated in a 0.13 μm CMOS process, and consumes 220 nW during demodulation and 190 nW during clock generation. The simulated results of an improved demodulator and clock generator are also presented, which consumes 130 nW during demodulation and 70 nW during clock generation.

Chapter 4 describes a wireless power transmission system intended for use with millimeter-sized wireless sensors. The system is able to deliver 2 μW of DC power over a range of 20 cm, or 0.8 μW of power over an area of 800 cm^2, with a minimum output voltage of 0.4 V. This is accomplished with a loop antenna formed from a bond-wire, with 12.3 mm^2 of loop area. A low-cost patch antenna is used for

transmission, and because the area of the patch antenna is smaller than the area in which power transmission occurs, multiple transmitting antennas can be tiled to cover an arbitrarily large area.

As the miniaturization of energy harvesting, wireless medical sensors continues, it is likely to allow for many new applications that are currently unrealizable. For miniaturization to continue, both the size of the integrated components and the energy harvesting transducer must be reduced. And although the continued scaling of integrated circuits is likely to continue for some time, there is no guarantee that the size of most energy harvesting transducers can be reduced while providing the same levels of power and energy. In fact, the amount of energy produced by most energy transducers, whether thermal, electromagnetic, or otherwise, depends on the energy density of the source being exploited, creating an upper limit on how much energy can be produced by a transducer with a given size. To reduce the size of the transducer therefore requires reducing the energy consumption of the sensor. And although some of this energy reduction is likely to come from technology scaling, new circuit techniques and architectures will be required to continue to improve the efficiency of wireless sensors.

References

[1] S. Jocke, J. F. Bolus, S. N. Wooters, A. D. Jurik, A. C. Weaver, T. N. Blalock, and B. H. Calhoun, "A 2.6-uw sub-threshold mixed-signal ecg soc," in *Symposium on VLSI Circuits Digest of Technical Papers*, 2009.

[2] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag, and A. P. Chandrakasan, "A micro-power eeg acquisition soc with integrated feature extraction processor for a chronic siezure detection system," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, April 2010.

[3] L. Yan, J. Yoo, B. Kim, and H. Yoo, "A 0.5-uvrms 12-uw wirelessly powered patch-type healthcare sensor for wearable body sensor network," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 11, November 2010.

[4] H. Kim, R. F. Yazicioglu, S. Kim, N. van Helleputte, A. Artes, M. Konijnenburg, J. Huisken, J. Penders, and C. van Hoof, "A configurable and low-power mixed signal soc for portable ecg monitoring applications," in *Symposium on VLSI Circuits Digest of Technical Papers*, 2011.

[5] Y. Zhang, F. Zhang, Y. Shaksheer, J. D. Silver, A. Klinefelter, M. Nagaraju, J. Boley, J. Pandey, A. Shrivastava, E. J. Carlson, A. Wood, B. H. Calhoun, and B. P. Otis, "A batteryless 19 μ w mics/ism-band energy harvesting body sensor

node soc for exg applications." *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 199–213, 2013.

[6] S. B. Lee, H. Lee, M. Kiani, U. Jow, and M. Ghovanloo, "An inductively powered scalable 32-channel wireless neural recording system-on-a-chip for neuroscience applications," *IEEE Transactions of Biomedical Circuits and Systems*, vol. 4, no. 6, pp. 360–371, December 2010.

[7] Z. Xiao, C. Tang, C. M. Dougherty, and R. Bashirullah, "A 20uw neural recording tag with supply-current-modulated afe in 0.13um cmos," in *IEEE International Solid-State Circuits Conference*, 2010.

[8] M. Mark, Y. Chen, C. Sutardja, C. Tang, M. Gowda, D. Werthimer, and J. Rabaey, "A 1 mm3 2bmps 330fj/b transponder for implanted neural sensors," in *Symposium on VLSI Circuits Digest of Technical Papers*, 2011.

[9] Y. Liao, H. Yao, A. Lingley, B. Parviz, and B. P. Otis, "A 3-μw cmos glucose sensor for wireless contact-lens tear glucose monitoring," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 335–344, January 2012.

[10] G. Chen, H. Ghaed, R. Haque, M. Wieckowski, Y. Kim, G. Kim, D. Fick, D. Kim, M. Seok, K. Wise, D. Blaauw, and D. Sylvester, "A cubic-millimeter energy-autonomous wireless intraocular pressure monitor," in *IEEE International Solid-State Circuits Conference*, 2011.

[11] P. Cong, W. H. Ko, and D. J. Young, "Wireless batteryless implantable blood pressure monitoring microsystem for small laboratory animals," *IEEE Journal of Solid-State Circuits*, vol. 10, no. 2, February 2010.

[12] M. M. Ahmadi and G. A. Jullien, "A wireless-implantable microsystem for continuous blood glucose monitoring," *IEEE Journal of Solid-State Circuits*, vol. 3, no. 3, pp. 169–180, June 2009.

[13] M. Kuhl, P. Gieschke, D. Rossbach, S. A. Hilzensauer, T. Panchaphongsaphak, P. Ruther, B. G. Lapatki, O. Paul, and Y. Manoli, "A wireless stress mapping system for orthodontic brackets using cmos integrated sensors," *Solid-State Circuits, IEEE Journal of*, vol. 48, no. 9, pp. 2191–2202, 2013. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6528028

[14] F. Kocer and M. P. Flynn, "An rf-powered, wireless cmos temperature sensor," *IEEE Sensors Journal*, vol. 6, no. 3, June 2006.

[15] J. Yin, J. Yi, M. K. Law, Y. Ling, M. C. Lee, K. P. Ng, B. Gao, H. C. Luong, A. Bermak, M. Chang, W. Ki, C. Tsui, and M. Yuen, "A system-on-chip epc gen-2 passive uhf rfid tago with embedded temperature sensor," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 11, November 2010.

[16] N. Chaimanonart, M. Zimmerman, and D. Young, "Adaptive rf power control for wireless implantable bio-sensing network to monitor untethered laboratory animal real-time biological signals," pp. 1241–1244, 2008.

[17] E. G. Kilinic, G. Conus, C. Weber, B. Kawkabani, F. Maloberti, and C. Dehollain, "A system for wireless power transfer of micro-systems in-vivo implantable in freely moving animals," *IEEE Sensors Journal*, vol. 14, no. February, pp. 522–531, 2 2014.

[18] E. Nadimi, H. T. Sgaard, T. Bak, and F. W. Oudshoom, "Zigbee-based wireless sensor networks fro monitoring animal presence and pasture time in a strip of

new grass," *Computers and Electronics in Agriculture*, vol. 61, no. 2, pp. 79–87, May 2008.

[19] J. I. Huircn, C. Muoz, H. Young, L. V. Dossow, J. Bustos, G. Vivallo, and M. Toneatti, "Zigbee-based wireless sensor network localization for cattle monitoring in grazing fields," *Computers and Electronics in Agriculture*, vol. 74, no. 2, pp. 258 – 264, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169910001584

[20] S. Paasovaara, M. Paldanius, Saarinen, J. Hkkil, and K. Vnnen-Vainio-Mattila, "The secret life of my dog: Design and evaluation of paw tracker concept," pp. 231–240, 2011. [Online]. Available: http://doi.acm.org/10.1145/2037373.2037409

[21] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," pp. 88–97, 2002. [Online]. Available: http://doi.acm.org/10.1145/570738.570751

[22] D. J. Mennill, S. M. Doucet, K. A. Ward, D. F. Maynard, B. Otis, and J. M. Burt, "A novel digital telemetry system for tracking wild animals: a field test for studying mate choice in a lekking tropical bird," *Methods in Ecology and Evolution*, 2012.

[23] (2014, January) Bee sensors take flight to help farmers. Commonwealth Scientific and Industrial Research Organisation. [Online]. Available: http://www.csiro.au/Portals/Media/Bee-sensors-take-flight-to-help-farmers.aspx

[24] B. H. Calhoun, J. F. Bolus, A. D. Jurik, A. F. Weaver, and T. N. Blalock, "Sub-threshold operation and cross-hierarchy design for ultra low power wearable

sensors," in *International Symposium on Circuits and Systems*, 2009, pp. 1437–1440.

[25] J. F. Bolus, "Low-power wireless ecg sensor design," Master's thesis, University of Virginia, 2010.

[26] V. Leonov, T. Torfs, P. Fiorini, and C. van Hoof, "Thermoelectric converters of human warmth for self-powered wireless sensor nodes," *IEEE Sensors Journal*, vol. 7, no. 5, May 2007.

[27] Y. Yang, X. Wei, and J. Liu, "Suitability of thermoelectric power generator for implantable medical electronic devices," *Journal of Physics D*, vol. 40, 2007.

[28] P. Mitcheson, E. M. Yeatman, G. K. Rao, A. Holmes, and T. C. Green, "Energy harvesting from human and machine motion for wireless electronic devices," *Proceedings of the IEEE*, vol. 96, pp. 1457–1486, 2008.

[29] P. Mercier, A. Lysaght, S. Bandyopadhyay, A. P. Chandrakasan, and K. M. Stankovic, "Energy extraction from the biologic battery in the inner ear," *Nature Biotechnology*, vol. 30, pp. 1240–1243, 2012.

[30] G. Chen, S. Hanson, D. Blaauw, and D. Sylvester, "Circuit design advances for wireless sensing applications," *Proceedings of the IEEE*, vol. 98, no. 11, November 2010.

[31] A. Yakovlev, S. Kim, and Poon, "Implantable biomedical devices: Wireless powering and communication," *IEEE Communications Magazine*, April 2012.

[32] M. Mark, T. Bjrnin, L. Ukkonen, L. Sydnheimo, and J. M. Rabaey, "Sar reduction and link optimization for mm-size remotely powered wireless implants using

segmented loop antennas," in *BioWireleSS 2011 : IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems*, 2011.

[33] R. J. M. Vullers, R. van Schaijk, H. J. Visser, J. Penders, and C. Van Hoof, "Energy harvesting for autonomous wireless sensor networks," *IEEE Solid-State Circuits Magazine*, pp. 29–38, 2010.

[34] K. Lofstrom, W. R. Daasch, and D. Taylor, "Ic identification circuit using device mismatch," in *Solid-State Circuits Conference*, 2000.

[35] J. Hirase and T. Furukawa, "Chip identification using the characteristic dispersion of transistor," in *14th Asian Test Symposium'*, 2005.

[36] Y. Su, J. Holleman, and Otis, "A digitial 1.6 pj/bit chip identification circuit using process variations," *IEEE Journal of Solid-State Circuits*, vol. 43, pp. 69–77, 2008.

[37] D. Holcomb, W. Burleson, and K. Fu, "Power-up sram state as an identifying fingerprint and source of true random numbers," *IEEE Transactions on Computers*, vol. 58, no. 9, pp. 1198–1210, 2009.

[38] H. Fujiwara, M. Yabuuchi, H. Nakano, H. Kawai, K. Nii, and K. Arimoto, "A chip-id generating circuit for dependable lsi using random address errors on embedded sram and on-chip memory bist," in *Symposium on VLSI Circuits Digest of Technical Papers*, 2011, pp. 76–77.

[39] S. Rosenblatt, D. Fainstein, A. Cestero, J. Safran, N. Robson, T. Kirihata, and S. S. Iyer, "Field tolerant dynamic intrinsic chip id using 32 nm high-k/metal gate soi embedded dram," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 4, pp. 940–946, 2013.

[40] C. A. N. B. Rabaey, J. M., *Digital Integrated Circuits*, C. G. Sodini, Ed. Pearson Education, Inc., 2003.

[41] B. Dell, J. F. Bolus, and T. N. Blalock, "An automated unique tagging system using cmos process variation," in *Proceedings of the 17th ACM Great Lakes Symposium on VLSI*, 2007.

[42] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of mos transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 1433–1439, 1989.

[43] B. Razavi, *Design of Analog CMOS Integrated Circuits*, K. T. Kane, Ed. McGraw Hill, 2001.

[44] T. Meyer, R. E. Johanson, and S. Kasap, "Effect of 1/f noise in integrating sensors and detectors," *IET Circuits, Devices & Systems*, vol. 5, no. 3, pp. 177–188, 2011.

[45] U. Karthaus and M. Fischer, "Fully integrated passive uhf rfid transponder ic with 16.7- mu;w minimum rf input power," *Solid-State Circuits, IEEE Journal of*, vol. 38, no. 10, pp. 1602 – 1608, 2003.

[46] J.-P. Curty, N. Joehl, C. Dehollain, and M. Declercq, "Remotely powered addressable uhf rfid integrated system," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 11, pp. 2193 – 2202, 2005.

[47] V. Pillai, H. Heinrich, D. Dieska, P. Nikitin, R. Martinez, and K. Rao, "An ultra-low-power long range battery/passive rfid tag for uhf and microwave bands with a current consumption of 700 na at 1.5 v," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 7, pp. 1500 –1512, 2007.

[48] T. Kleeburg, J. Loo, N. J. Guilar, E. Fong, and R. Amirtharajah, "Ultra-low-voltage circuits for sensor applications powered by free-space optics," in *IEEE International Solid State Circuits Conference*, 2010.

[49] A. Shrivastava, J. Pandey, B. Otis, and B. Calhoun, "A 50nw, 100kbps clock/data recovery circuit in an fsk rf receiver on a body sensor node," in *VLSI Design*, 2013, pp. 72–75.

[50] Y. LIn, D. Sylvester, and D. Blaauw, "A sub-pw timer using gate leakage for ulta low-power sub-hz monitoring systems," in *Custom Integrated Circuits Conference*, 2007, pp. 397–400.

[51] Y. Lee, B. Giridhar, Z. Foo, D. Sylvester, and D. Blaauw, "A 660pw multi-stage temperature-compensated timer for ultra-low-power wireless sensor node synchronization," in *International Solid-State Circuits Conference*, 2011, pp. 46–48.

[52] E. Vittoz and J. Fellrath, "Cmos analog integrated circuits based on weak inversion operation," *IEEE Journal of Solid-State Circuits*, vol. 12, no. 3, pp. 224–231, June 1977.

[53] W. Thommen, "An improved low power crystal oscillator," in *European Solid State Circuits Conference*, 1999, pp. 146–149.

[54] D. Ruffieux, "A high-stability, ultra-low-power quartz differential oscillator circuit for demanding radio applications," in *European Solid-State Circuits Conference*, 2002, pp. 85 85–88.

[55] D. Yoon, D. Sylvester, and D. Blaauw, "A 5.58nw 32.768khz dll-assisted xo for real-time clocks in wireless sensing applications," in *IEEE International Solid-State Circuits Conference*, 2012.

[56] A. Shrivastava, "Mixed signal platform circuits for lifetime improvement of ultra low power systems," Ph.D. dissertation, University of Virginia, 2014.

[57] A. Shrivastava and Ca, "A 150nw, 5ppm/$^{\text{deg}}$c 100 khz on-chip clock source for ultra low power socs," in *IEEE Custom Integrated Circuits Conference*, 2012.

[58] J. Lim, K. Lee, and K. Cho, "Ultra low power rc oscillator for system wake-up using highly precise auto-calibration technique," in *European Solid State Circuits Conference*, 2010, pp. 274–277.

[59] A. Paidimarri, D. Griffith, A. Wang, A. P. Chandrakasan, and G. Burra, "A 120nw 18.5khz rc oscillator with comparator offset cancellation for ±0.25% temperature stability," in *IEEE International Solid-State Circuits Conference*, 2013.

[60] T. Tokairin, K. Nose, K. Takeda, K. Noguchi, T. Maeda, K. Kawai, and M. Mizuno, "A 280nw, 100khz, 1-cycle start-up time, on-chip cmos relaxation oscillator employing a feedforward period control scheme," in *Symposium on VLSI Circuits Digest of Technical Papers*, 2012, pp. 16 16–17.

[61] T. O. Anderson, "Optimal control logic for successive approximation analog-to-digital converters," *Deep Space Network Progress Report 13*, pp. 168–176, 1972.

[62] S. Song, N. Cho, and H. Yoo, "A 0.2-mw 2-mb/s digital transceiver based on wideband signaling for human body communications," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 9, pp. 2021–2033, September 2007.

[63] S. Kim, J. Woo, W. Shin, G. Hong, H. Lee, H. Lee, and S. Kim, "A low-power referenceless clock and data recovery circuit with clock-edge modulation for biomedical sensor applications," in *International Symposium on Low Power Electronics and Design*, 2011, pp. 347–350.

[64] D. Jee, D. Sylvester, D. Blaauw, and J. Sim, "A 0.45v 423nw 3.2mhz multiplying dll with leakage-based oscillator for ultra-low-power sensor platforms," in *IEEE International Solid-State Circuits Conference*, 2013.

[65] P. J. Schaeffer and D. J. Pierotti, "A transcutaneous wire interface for small mammals using an expanded ptfe patch," *Journal of Neuroscience Methods*, vol. 114, pp. 81–85, 2002.

[66] M. Broussely, P. Biensan, F. Bonhomme, P. Blanchard, S. Herreyre, K. Nechev, and R. J. Staniewicz, "Main aging mechanisms in li ion batteries," *Journal of Power Sources*, vol. 146, pp. 90–96, 2005.

[67] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Transactions on Embedded Computing Systems*, vol. 6, 2007.

[68] J. S. Ho, S. Kim, and A. S. Y. Poon, "Midfield wireless powering for implantable systems," *Proceedings of the IEEE*, vol. 101, no. 6, pp. 1369–1378, June 2013.

[69] W. J. Heetderks, "Rf powering of millimeter- and submillimeter-sized neural prosthetic implants," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 5, pp. 323–327, May 1988.

[70] A. Poon, S. O'Driscoll, and T. Meng, "Optimal frequency for wireless power transmission into dispersive tissue," *Antennas and Propagation, IEEE Transactions on*, vol. 58, no. 5, pp. 1739 –1750, May 2010.

[71] M. Zargham and P. G. Gulak, "Maximum achievable efficiency in near-field coupled power-transfer systems," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 6, no. 3, pp. 228–245, June 2012.

[72] ——, "A 0.13μm cmos integrated wireless power receiver for biomedical applications," in *IEEE European Solid State Circuits Conference*, 2013.

[73] S. O'Driscoll, A. S. Y. Poon, and T. H. Meng, "A mm-sized implantable power receiver wth adaptive link compensation," in *IEEE International Solid-State Circuits Conference*, 2009.

[74] B. Lenaerts and R. Puers, "Inductive powering of a freely moving system," *Sensors and Actuators, A: Physical*, pp. 522–530, 2005.

[75] D. Russel, D. McCormick, A. Taberner, P. Nielsen, P. Hu, D. Budgett, M. Lim, and S. Malpas, "Wireless power delivery system for mouse telemeter," in *IEEE Biomedical Circuits and Systems Conference*, 2009.

[76] A. J. Yeh, J. S. Ho, Y. Tanabe, E. Neofytou, R. E. Beygui, and A. S. Y. Poon, "Wirelessly powering miniature implants for optogenetic stimulation," *Applied Physics Letters*, vol. 103, 2013.

[77] C. T. Wentz, J. G. Bernstein, P. Monahan, A. Guerra, A. Rodriguez, and E. S. Boyden, "A wirelessly powered and controlled device for optical neural control of freely-behaving animals," *Journal of Neural Engineering*, vol. 8, 2011.

[78] R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, *Microstrip Antenna Design Handbook*. Artech House, 2000.

[79] M. H. Ouda, M. Arsalan, L. Marnat, A. Shamin, and K. N. Salama, "5.2-ghz rf power harvester in 0.18-μm cmos for implantable intraocular pressure monitoring," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 5, pp. 2177–2184, May 203.

[80] E. Y. Chow, A. L. Chlebowski, and P. P. Irazoqui, "A miniature-implantable rf-wireless active glaucoma intraocular pressure monitor," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 4, no. 6, pp. 340–349, December 2010.

[81] R. R. Harrison, R. J. Kier, C. A. Chestek, V. Gilja, P. Nuyujukian, S. Ryu, B. Greger, F. Solzbacher, and K. V. Shenoy, "Wireless neural recording with single low-power integrated circuit," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 4, pp. 322–329, August 2009.

[82] Ferro Solutions. [Online]. Available: http://www.ferrosi.com/products.htm