Electromagnetic Field Manipulation: Biosensing to Antennas

Thesis by
Constantine Sideris

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
(Defended November 4, 2016)

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Ali Hajimiri, whom I first met 8 years ago as a 3rd year undergraduate student in his Advanced Analog Circuits (EE114) class. His passion for teaching and research convinced me to pursue graduate studies in the field of analog integrated circuit design. He has helped me grow tremendously both personally and professionally, and this work would not have been possible without his guidance, encouragement, and support.

I am also greatly indebted to Professor Elad Alon, who hosted me in his group at the Berkeley Wireless Research Center as a visiting scholar for a year. He has also provided me with invaluable support and guidance and has continued to meet with me regularly to discuss research progress even since I returned to Caltech.

I would also like to thank Professor Oscar Bruno with whom I have been collaborating over the past year on various electromagnetic structure optimization projects and who has always been available to meet with me regularly and discuss ideas.

I would like to thank Professor David Rutledge and Dr. Sander Weinreb for sitting on my candidacy and defense committees and providing very insightful feedback and suggestions.

When I first joined the Caltech High Speed Integrated Circuits (CHIC) lab as a summer undergraduate research fellow, I worked very closely under the guidance and mentorship of Professor Hua Wang, who is now a professor at the Georgia Institute of Technology. Hua showed me the fundamentals of integrated circuit design, instilled in me a passion for research and taught me the virtues of hard work and perseverance.

I am grateful for all of the current and previous members of the Caltech High Speed Integrated Circuits (CHIC) and Mixed-Signal Integrated Circuits and Systems (MICS) labs whom I have had the privilege of crossing paths with, including Juhwan Yoo, Edward Keehr, Aydin Babakhani, Steven Bowers, Florian Bohn, Kaushik Sengupta, Kaushik Dasgupta, Joe Bardin, Amir Safaripour, Alex Pai, Tomoyuki Arai, Aroutin Khachaturian, Behrooz Abiri, Brian Hong, Reza Fatemi, Parham Porsandeh-Khial, Milad Taghavi, Aryan Hashemi, Abhinav Agarwal, Xavier Chen, Reza Hashemi, Manuel Monge, Matthew Loh, Saman Saeedi, Meisam Nazari, and Stefan Turkowski.

I am also extremely indebted to all of my close friends who have offered their constant support and motivation throughout this whole process.

Finally, I would be remiss not to thank my family, who have provided me with unconditional love and without whom I am sure I would not be at this point today. Mom, Dad, and Elias– I love you deeply and no words can truly convey my infinite gratitude for everything that you have done and that you do for me.

To my parents, Anastasia and Athanasios,
and my brother, Elias

ABSTRACT

We will explore how understanding and controlling electromagnetic fields can provide significant impact across a multitude of applications throughout the whole frequency spectrum from DC to daylight. Starting from the DC end of the electromagnetic spectrum, we motivate the design of a new integrated magnetic biosensing design as well as various improvements to the initial design based on spatial and temporal manipulations of the magnetic fields. Next, we look into the RF domain and develop maximal performance bounds for antennas and other electromagnetic structures. We develop rapid simulation techniques which when coupled with heuristic optimization algorithms can quickly and effectively produce new antenna structures with little to no manual intervention. We demonstrate the efficacy of these techniques in the context of on-chip antenna designs and a 3D printed coupling antenna for a dielectric waveguide communication link. We present the design of a 120GHz dual-channel 100Gbps QPSK/64QAM transceiver IC developed in a standard 28nm bulk CMOS process. Finally, we explore the higher THz regime in the context of photonic device optimization. We optimize compact photonic multiplexer devices which are fabricated in a standard foundry process and evaluate their performance against simulation results.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter I: Introduction

In this work, we will explore how understanding and controlling electromagnetic fields can lead to significant improvements in functionality and performance in a variety of different engineering applications.

There was much interest in the nineteenth century in the study of electromagnetism and electromagnetic fields, which were not well understood at that point. Starting with experimental observations and leading to the development of many mathemetical equations by numerous scientists including Charles-Augustin de Coulomb, Hans Christian Ørsted, Carl Friedrich Gauss, Jean-Baptiste Biot, Félix Savart, André-Marie Ampère, and Michael Faraday. [53] These equations were seemingly disconnected but were integrated into a concise, simple set of partial differential equations by James Clerk Maxwell. Maxwell was further able to show that this set of equations implies that light propagates in the form of electromagnetic waves. [70, 71, 36, 21]

Perhaps one of the most amazing aspects of Maxwell's equations is that they accurately describe the electromagnetic field across many orders of magnitude of frequency (Fig. 1.1)—essentially from DC (electro/magnetostatics) to visible light and even all the way up to Xrays and beyond. [1] In this work, we will actually traverse the electromagnetic frequency spectrum and demonstrate how careful design and manipulation of the fields can lead to advances in many different areas of engineering including biosensing [76, 107, 37, 56, 113, 115, 112, 93, 94], antenna design [6, 87, 95, 92, 91], and photonic device design [83, 55, 68] .

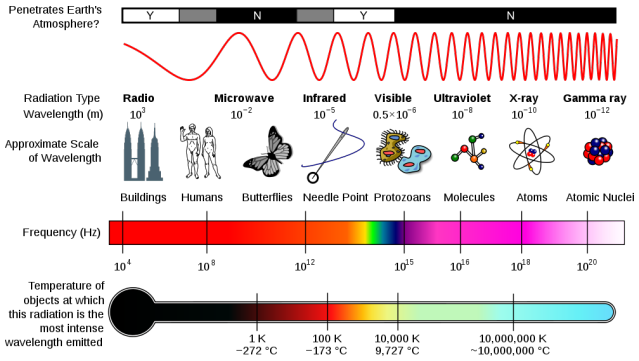

Figure 1.1: Illustration of the Electromagnetic Spectrum from "DC to Daylight".

We will begin with static magnetic fields and develop a novel magnetic biosensing principle [113]. By taking advantage of the spatial field profile, we will design a

magnetic sensor with spatially uniform transducer gain response across its active sensing area. [112] Furthermore, by taking advantage of the differing magnetization properties of different types of magnetic nanoparticle materials, we design a magnetic spectrometer sensor capable of differentiating among multiple types of magnetic labels concurrently on the sensing surface, enabling previously impossible magnetic-multiplexing biological experiments. [93, 94] Finally, we motivate the design of the latest version magnetic biosensor which utilizes a new frequency drift cancellation approach.

Next, we delve higher into the electromagnetic frequency spectrum and explore the design of antennas as well as other RF electromagnetic structures such as dielectric waveguide couplers. [95] We first develop an optimal maximum performance bound for arbitrary parameters of electromagnetic structures such as the directivity, gain, or radiation efficiency of antennas. [92] We also look into the optimization of real metallic antenna and coupling structures. We develop an ultrafast technique for evaluating new iterations of electromagnetic structure designs on millisecond timescales with modest computational power, achieving several orders of magnitude faster performance than optimized, commercial electromagnetic simulation software. [91] We present examples of optimizing an on-chip metal antenna on a realistic, lossy Silicon dielectric substrate. Finally, we present a design example of a waveguide coupling antenna which is optimized using the techniques developed, fabricated by 3D printing, and compared against simulation results. [95] We conclude the RF part of the thesis by presenting a 120GHz 100Gbps dual-channel QPSK/64-QAM transceiver IC designed in a 28nm bulk CMOS process with the intent of driving both optimized dual-polarization antenna structures for wireless communication and optimized dual-mode dielectric waveguides for a wireline communication modality.

Finally, we reach the Terahertz regime and explore photonic device optimization. We utilize a Binary Particle Swarm Optimization (BPSO) algorithm [52] for optimizing small footprint multiplexer devices which are designed for and fabricated in a standard photonics process. We present measurement results for the optimized devices, compare against simulations, and study the causes for deviation between the results by taking and analyzing SEM scans of the fabricated chip. [68]

Chapter II: Biosensors

2.1 Introduction to Biosensing

Biology is one of the most rapidly expanding fields in science today. It is a huge field with many subfields and much interest and funding has recently been invested into biology due to various advancements which have made the field considerably more quantitative. [76, 107, 37, 56, 113, 115, 112, 93, 94] Furthermore, fields such as mathematics, physics, and classical engineering (mechanics, circuit design, signal processing, communications, etc.) are quickly becoming mature while biology leaves countless questions and dilemmas left to be answered and solved.

Biology mainly deals with a study of all things living or organic and attempts to understand and explain how they work. It was not until recently that a quantitative aspect of biology began to emerge. Many of the recent quantitative discoveries are largely thanks to the new subfield of bioengineering. Bioengineering attempts to use the well formed traditional principles of engineering to invent and solve problems relating to any aspect of biology. Instead of simply studying biological phenomena, bioengineering allows engineers to design their own systems inspired by biology or using biology as their tool. [11]

One area of bioengineering which has become very prevalent due to its practical applications is biosensing. Although biosensing is a very broad term, the field essentially encompasses any method for detection (and often quantification) of biological molecules [116]. Biosensing approaches can allow a scientist to detect the presence of a specific strand of nucleotides (DNA/RNA) in a solution full of many contaminants or even to detect presence of larger and more complicated biomolecules such as proteins, antibodies, and even cells. The more sophisticated forms of biosensing allow the scientist to take these measurements several steps further and actually discover (to a certain degree of accuracy) the concentrations of these target biomolecules in a solution.

There are seemingly countless potential real-world applications for molecular biosensing. Applications which appear the most important today include:

Biomedical: detecting pathogens such as viruses from a blood sample.

Genetics: detecting predispositions to various genetic diseases by matching gene markers to sample DNA from a patient.

Criminal Forensics: matching trace amounts of (often contaminated) crime-scene DNA to suspects in a crime investigation.

These three areas are all multibillion dollar industries and their advancement directly benefits society.

Biosensors: From Transducers to Beyond

As described, the general problems to be solved in biosensing are detecting the presence of a biomolecular structure amidst contaminants and quantifying the concentration of such a structure present in a sample. The biomolecule being targeted is generally referred to as the analyte. The sensor which can detect and quantify the analyte is called the transducer. This name is given due to the fact that the sensor essentially converts signal from one form of information into another, such as from a biomolecule into an electrical signal carrying information about it. Ideally, a transducer would have a direct one-to-one linear mapping from input to output so that the information is trivially encoded and none is lost. However, this is never the case with real world transducers. Fig. 2.1 is a simulated diagram of a typical sensor response with respect to concentration of signal being measured (the signal region is still ideal since it is drawn as perfectly linear, which is again never the case for a real transducer):

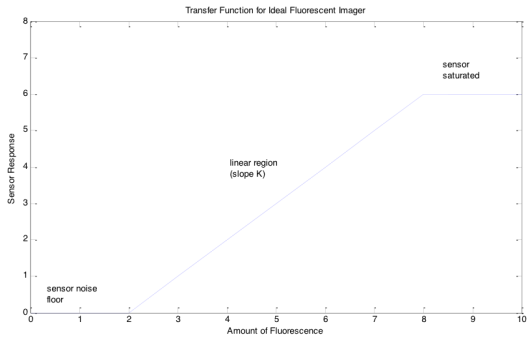

Figure 2.1: Simulated Transfer Function for an Ideal Fluorescent Imaging Sensor

This figure is not based on real data and is just meant for illustrative purposes. There are generally three main regions to note in the fluorescent imager response: the dead-zone, the linear region, and the saturation region. The dead-zone is due to the sensor's inherent noise floor. Every real world sensor or device suffers from a

variety of noise sources, such as Johnson (thermal) noise, Flicker noise, and drift. In order for a signal to be detectable by the sensor, it must have greater strength than these noise sources, which create what is often called the sensor noise floor. Signals above the noise floor generate a visible response and in general one hopes that the response is linear with respect to amount of signal above the noise floor. Furthermore, careless chemistry often results in parts of the background in a sample fluorescing that do not represent signal. This background fluorescence also directly contributes to the noise and degrades the signal-to-noise ratio (SNR). When too much signal is present, the sensor response can become saturated. This means that the sensor due to some physical limitation (such as reaching maximum camera luminosity for a fluorescent imager or reaching the maximum voltage supply rail for an amplifier circuit) cannot generate a larger output in response to more signal. Thus, it corresponds to a flattening of the response curve. The linear or active region is the only one of the three which allows recovery of the original signal based on the sensor output.

It is important to realize that all of this terminology is introduced in the most general sense possible since it applies to many different modes of biosensing and transduction. The signal input and sensor output can be arbitrary. For example, for a fluorescent imager, the signal input is light in a specific wavelength band and the output is a luminosity reading. A CCD based fluorescent imager with an appropriate fluorescent filter uses photodetector devices which convert incoming light at the correct frequency into charge stored onto a capacitor. The simulated figure above shows the response region as being perfectly linear. Perfect linearity is desirable, although it is generally not the case due to physical limitations such as nonlinearities of active devices in the sensing chain. The slope of the response line is labeled K. K is called the transducer gain. The units of K are the units of the type of sensor response over the units of the type of signal measured. For example, the CCD transducer may have volts / candela as the units for K. An initial intuition may tell us to maximize the value of K, since this means that we will get as large of a sensor response as possible for the same level of input signal, making the signal easier to discern from the noise floor. However, increasing K also makes the sensor saturate quicker, since multiplying K by a fixed value m implies that a signal which is m times weaker than before will now produce the same response. This means that the signal level which will put the sensor into saturation, making it useless for further measurements, has now become m times smaller for the system with K*m transducer gain than what it was for the original K system. Thus, varying K trades

off minimum detectable signal for maximum detectable signal. If the increase in minimum detectable signal exactly matches the decrease in maximum detectable signal, then the systems dynamic range remains the same. Dynamic range is defined as the ratio between the largest detectable signal to the smallest detectable signal. As can be seen in the example above, the dynamic range primarily depends on the noise floor and saturation strength. When the system is highly nonlinear and cannot be approximated accurately with a linear curve, K must become a function of s, the amount of fluorescent response, and we would write the transducer gain in the active/sensing region as K(s)*s instead of simply K*s.

A Specific Biosensing Approach: Fluorescent Imaging

In the 1960s, the GFP (green fluorescent protein) was isolated in jellyfish [90]. Since then, scientists have successfully been able to replicate this gene and the protein itself to create customized fluorescent proteins as well as fluorescent tags or markers at the molecular scale. Fluorescent markers are very useful due to their ability to absorb light of one wavelength and reemit light at a different wavelength. An imaging setup irradiates the sample with light of the excitation wavelength and concurrently detects light in the emission wavelength. The detection is often realized with a camera and a sharp filter blocking any wavelengths outside of the detection wavelength. The amount of emission wavelength light detected is directly proportional to the quantity and size of fluorescent markers present in the sample. If there are no markers present, then one expects to not receive any return signal in the emission band. Furthermore, since the fluorescent effect happens at the molecular level, single molecule structures can be designed which fluoresce [73]. These molecules can be as small as individual DNA nucleotides, and technology has even been developed for creating synthetic DNA strands that can have fluorescent markers inserted at any point within the strand including at the 5 and 3 terminal ends [41]. This is extremely powerful as it enables debugging or probing of biological systems and reactions down to the nucleotide level. Various types of fluorescent markers have been developed in the past decade, including markers which act as quenchers and can absorb the fluorescent emission of normal markers when in close proximity, effectively providing the bioengineer with an off switch element for debugging.

While the introduction and development of fluorescent imaging has largely been responsible for bringing about quantitative bioengineering in the past two decades, fluorescent methods suffer various drawbacks. First of all, fluorescent imagers and microscopes are physically large machines which can consume significant amounts

of lab space. Furthermore, fluorescent imaging systems can be extremely cost prohibitive and can range from a few $100K to over $1million [45]. These two drawbacks essentially take fluorescent imaging out of the picture for Point of Care (PoC) home diagnostics and may even hinder small research institutions or hospitals without ample funding from having access to state of the art imaging resources. Finally, most fluorescent markers suffer from signal decay, have a short lifespan, and must be primarily handled in dark environments.

Bioengineering meets Electrical Engineering: A Beautiful Union

CMOS chip fabrication processes have been around for decades. Barring a steep initial startup cost, modern CMOS chips generally cost on the order of a few cents per chip. CMOS technology has become so mature that billions of transistors can be fit on a chip of few square millimeters of area, all of which function properly. This allows for enormous, reliable computational power in a form factor smaller than a penny. Finding an effective way to design a CMOS chip which can be used as a biosensor would therefore solve both the cost and space issues mentioned in the previous section.

Just as markers in the form of beads exist which can fluoresce under light excitation, beads made up of a ferrous inner material have recently been developed which can become polarized under an external magnetic field [43]. Polarization implies that the beads can generate their own magnetic field when they are in the presence of an external magnetic field, but do not retain magnetization in the absence of an external field. These beads are not quite as small yet as fluorescent markers, but can be currently produced as small as 10nm in diameter. For comparison, a DNA nucleotide is 1-2nm wide and approximately 0.3nm long. The outer surface of these beads can also be manufactured to express many different kinds of functional groups. Common groups are amines (R-NH2), aldehydes (R-C(=O)H), and carboxyls (R-COOH). These functional groups allow for the beads to be readily attached to target biomolecules such as proteins and even the 5 or 3 ends of single or double stranded DNA. Thus, if a way to sense the presence of these magnetic beads exists, then they can essentially be as effective for detection of target biomolecules as fluorescent markers are today.

The workflow of medical diagnostics today requires patient sample collection at specialized facilities, followed by analysis at large, centralized labs. This process, however, is not very time or cost effective. Because of the inefficiencies of this

time consuming system, which can lead to delays in care, there is high demand for fast, low-cost, personalized medicine via at-home and point-of-care (PoC) medical diagnostics [106, 57]. By enabling on-site sample collection and analysis, PoC systems promise to play a crucial role in numerous biomedical applications, such as epidemiological disease control, biohazard detection, and environmental monitoring [47].

There has been a fair amount of progress on optical microarray technology for biomolecular assays [100]. Implementation of optically driven systems involves utilizing optical sources and filter setups, both of which greatly increase the cost and complexity of the system [49]. Alternatively, integrated biosensors based on magnetic labeling schemes offer higher sensitivity and lower cost due to elimination of optics, and have emerged as a viable alternative to assays that use fluorescence for biomolecular detection [113, 115, 112, 93, 94, 78, 33, 27, 102]. For instance, the frequency-shift sensor of [113] demonstrates an example of a highly sensitive, cost-effective magnetic particle biosensor in CMOS with no need for external magnets. The use of such a magnetic sensor for nucleic acid and protein targets has been successfully demonstrated recently [78].

We will first introduce and motivate the base magnetic biosensing approach. Next we will develop and analyze the design of a uniform gain magnetic biosensor. Finally, we will extend the magnetic sensing idea to a modality capable of concurrent multi-probe multiplex detection experiments.

2.2 II. Magnetic Biosensing Approach

We have designed a new kind of transducer which is capable of detecting (and recently also quantifying) these paramagnetic beads. The transducer can be implemented in a completely standard CMOS process, with absolutely no custom features or any post-processing beyond printing a surface chemistry of probe biomolecules [113]. The transducer we employ for detecting the presence of magnetic beads is a simple spiral on-chip inductor. As explained above, the beads to be sensed are made of paramagnetic material which becomes temporarily polarized in the presence of an external magnetic field. Passing any sort of current (whether it be static or alternating) through a spiral coil inductor generates a magnetic field due to Ampère's Law. This field magnetizes all beads sitting over the inductor surface, which causes an increase in the total magnetic energy stored in the system. Since effective inductance is proportional to the magnetic flux through the inductor surface, we can

deduce that the total inductance of the coil will also increase in the presence of these beads. We can mathematically quantify this physical process. Assume a magnetic particle with effective susceptibility and volume Vp is placed close to the sensing inductor. When a current I conducts through the inductor spiral, a local polarization magnetic field is produced. Assuming that the magnetic particle is of negligible size in comparison to that of the whole inductor, it will not alter this polarization field appreciably. Thus, the total magnetic energy increases by ΔE_m after introducing the magnetic particle:

$$\begin{aligned}\Delta E_m &= (E_{m'} - E_m) = \frac{1}{2}\iiint \vec{H}\cdot\vec{B'}dv - \frac{1}{2}\iiint \vec{H}\cdot\vec{B}dv \\ &= \frac{\mu_0}{2}\iiint_{V_p}\left[||\vec{H}||^2(1+\chi_{eff}) - ||\vec{H}||^2\right]dv = \frac{\chi_{eff}}{2\mu_0}\iiint_{Vp}||\vec{B}||^2 dv \\ &\approx \frac{\chi_{eff}}{2\mu_0}||\vec{B}||^2 Vp\end{aligned}$$

Furthermore, we know that:

$$\Delta L = \frac{2\Delta E_m}{I^2}$$

So change in inductance can be quantified as:

$$\Delta L = \frac{2\Delta E_m}{I^2} = \frac{\chi_{eff}}{I^2\mu_0}||\vec{B}||^2 Vp$$

Thus, a magnetic bead over an inductor will induce a change in inductance which is directly proportional its effective magnetic permittivity and volume [112]. Simplified steps for a viable magnetic biosensing approach are presented in Fig. 2.2:

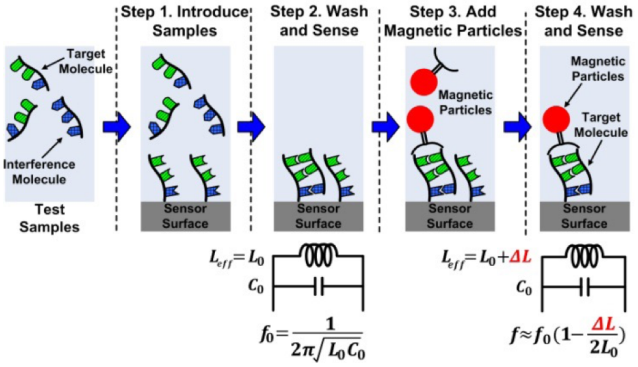

Figure 2.2: Magnetic Biosensing Scheme

The rest of the technique involves measuring this very tiny change in inductance as accurately as possible in order to maximize the signal to noise ratio (SNR). The simplest technique which can be implemented for measuring inductance involves adding a fixed known capacitance to the tank and characterizing the resulting impedance function. The complex impedance of an ideal inductor in parallel with a capacitor is: $Z_{tank} = j\omega L || \frac{1}{j\omega C} = \frac{j\omega L}{1-\omega^2 LC}$

We notice that when $\omega = \omega_{res} = \frac{1}{\sqrt{LC}}$, the magnitude of Z_{tank} is maximized and approaches infinity. Thus, if we experimentally find ω_{res} by sweeping frequency, measuring impedance magnitude, and finding the absolute maximum, we can deduce the inductance as: $L = \frac{1}{w_{res}^2 C}$.

Earlier, we derived that the change in inductance due to a tiny magnetic bead is $\Delta L = \frac{2\Delta E_m}{I^2} = \frac{\chi_{eff}}{I^2\mu_0} ||\vec{B}||^2 Vp$. Thus, taking as being the inherent device inductance before any perturbation, the new inductance is $L_n = L_o + \Delta L$.

We wish to find out what shift in resonance frequency this corresponds to, so we substitute into the ω_{res} equation:

$$\omega_{resN} = \frac{1}{\sqrt{(L_o + \Delta L)C}} = \frac{1}{\sqrt{(1 + \frac{\Delta L}{L})L_0 C}} = \frac{1}{\sqrt{L_0 C}\sqrt{1 + \frac{\Delta L}{L}}}$$
$$\approx \frac{1}{\sqrt{L_o C}}\left(1 - \frac{\Delta L}{2L}\right) \text{using a first order Taylor expansion since } \Delta L << L_o \quad = \omega_{res0} - \omega_{res0}\frac{\Delta L}{2L}$$

Thus, the frequency difference is: $\Delta\omega = \omega_{resN} - \omega_{res0} = \omega_{res0}\frac{\Delta L}{2L_o}$.

Substituting and expanding further we are left with the frequency shift $\Delta\omega = \omega_{res0}\frac{\frac{\chi_{eff}}{I^2\mu_0}||\vec{B}||^2 Vp}{2L_o}$ in the impedance transfer function of the LC tank with the addition of a single bead. Assuming all the beads are identical and the $||\vec{B}||$ is relatively uniform across the whole inductor surface (as we will see in a later section, inductor uniformity is NOT a well founded assumption!), then we can approximate the resonance frequency shift as a linear function of N:

$$\Delta\omega(N) = \omega_{res0}\frac{\frac{\chi_{eff}}{I^2\mu_0}||\vec{B}||^2 Vp}{2L_o}N = K * N \text{with } K = \omega)_{res0}\frac{\frac{\chi_{eff}}{I^2\mu_0}||\vec{B}||^2 Vp}{2L_o}$$

We have thus derived the transducer gain K for an ideal LC tank sensor device which has a uniform magnetic field distribution across its whole surface. For

identical beads (all with volume Vp), this results in a K which is constant meaning that our transfer function is perfectly linear. As we will later, for any standard inductor design, the planar XY magnetic field is extremely nonuniform and varies extensively, resulting in a position dependent K:

$$K(x,y) = \omega_{res0} \frac{\frac{\chi_{eff}}{I^2 \mu_0} ||\vec{B}(x,y)||^2 V_p}{2L_0}$$

Due to the structural complexity of a real inductor, $||\vec{B}(x,y)||^2$ should be computed by using numerical 3D EM simulations of the target inductor structure. This can be accomplished relatively easily by using popular simulation packages such as Ansoft HFSS or IE3D [39].

Using an Oscillator to Measure Frequency Shift

So far we have only considered the case of an ideal LC tank. In reality, inductors cannot be made out of perfect conductors and thus have considerable amounts of series resistance. A more realistic LC tank model looks like this:

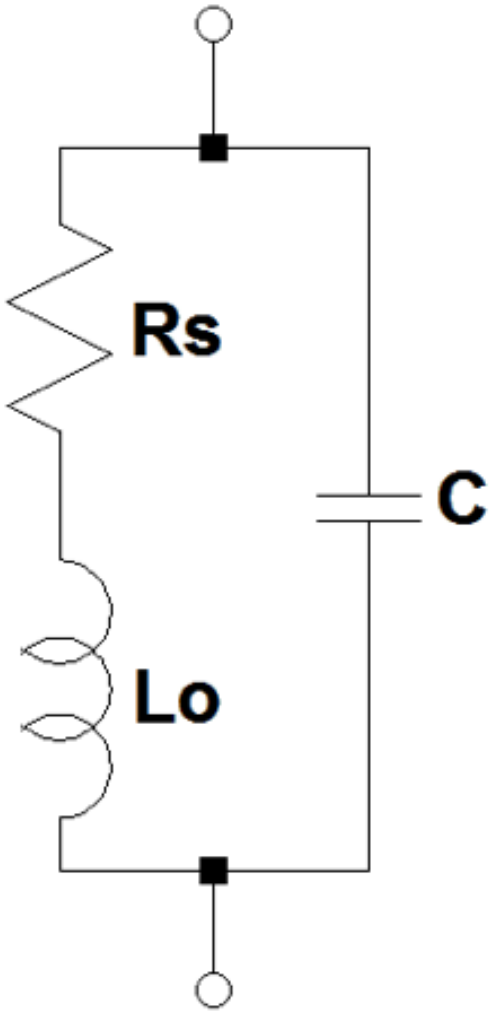

Figure 2.3: LC tank circuit model with parasitic series resistance Rs.

The complex impedance of this system is $Z = \frac{j\omega L_0 + R_s}{1 + j\omega C(j\omega L_0 + R_s)} = \frac{j\omega L_0 + R_s}{1 - \omega^2 L_0 C + j\omega R_s C}$. We are interested in the frequency where the system presents zero phase shift (a completely real impedance), which happens when $omega = frac1\sqrt{L_0 C}\sqrt{1 - \frac{R^2 C}{L}} = \frac{1}{L_0 C}\sqrt{1 - \frac{1}{Q}}$. With Q defined as the quality factor of the inductor at the ideal resonance frequency: $Q = \frac{\omega_{res} L_0}{R_s} = \frac{1}{Rs}\sqrt{\frac{L_0}{C}}$

In CMOS, a reasonable value of Q for operating frequencies around 1GHz is 10. For a Q of 10, the resonant frequency will be downshifted by approximately 5% from the ideal Rs=0 situation. This shift does not really concern us much, since it just means that the starting point in frequency is different and we are primarily concerned with relative frequency shifts. The major effect of series resistance on

the LC tank which does create concern is the widening of the impedance transfer function:

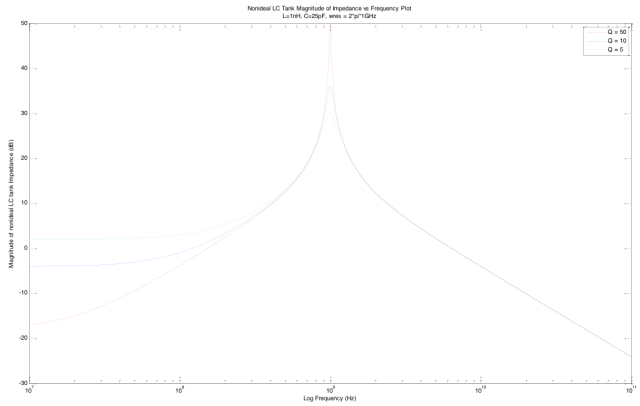

Figure 2.4: Magnitude of LC tank impedance around resonant frequency for various quality factors (Q).

As can be seen in this simulation, the smaller Q becomes, the wider the peak of the magnitude of the impedance becomes around its center point. This widening is highly undesirable as it makes small changes in resonant frequency (left or right shifts in that peak) much more challenging to detect in the presence of noise. Thus, we wish to have a transfer function to measure which has a very narrow resonance with respect to frequency. One way to achieve this is by maximizing Q. While maximizing Q is indeed an important design goal, achieving a quality factor in a typical CMOS process of greater than 10 at a resonant frequency of 1GHz is highly unlikely. Depending on the process type, even Qs as low as 5 may not be unreasonable. Near the frequency of resonance, we can perform an impedance transformation and convert the nonideal tank into an approximate parallel RLC tank:

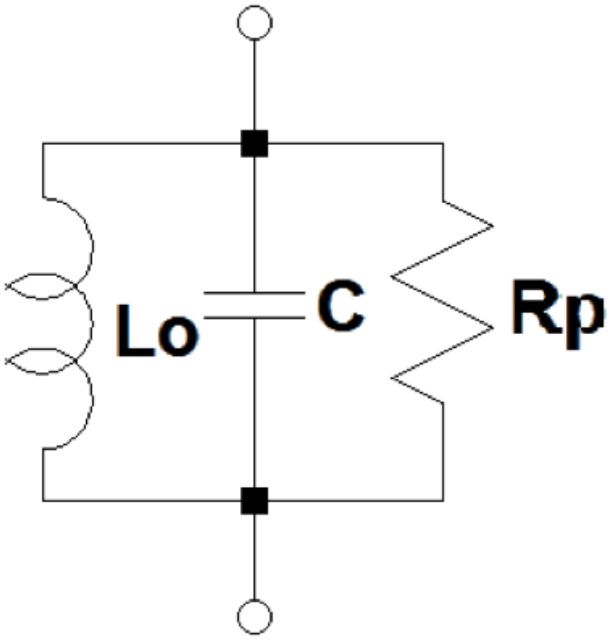

Figure 2.5: LC tank transformation to parallel parasitic resistance Rp.

with $R_p = (1 + Q^2)R_s$ being an equality at exactly the resonance frequency and deviating beyond that. If we can introduce a negative resistance in parallel with this RLC tank, we can effectively make Rp arbitrarily large, narrowing the impedance profile around the resonance frequency. This has been thoroughly studied in [32]. Incorporating the LC sensor tank into an oscillator configuration not only significantly narrows this profile, but also causes the tank to naturally oscillate at its resonant frequency (roughly the resonant frequency, in reality the parasitics of the active device and the other elements in the loop provide some additional phase shift, although it is usually negligible compared to that of the LC tank.) Since the oscillator causes the tank to oscillate at its resonant frequency, all we have to do to be able to approximate the value of the inductor in the tank (knowing all of the other parameters) is tap the oscillator voltage and measure the frequency. Fig. 2.6 is a simulated plot comparing the impedance profile of the nonideal LC tank with the line-width narrowed phase noise profile of the oscillator using the same tank:

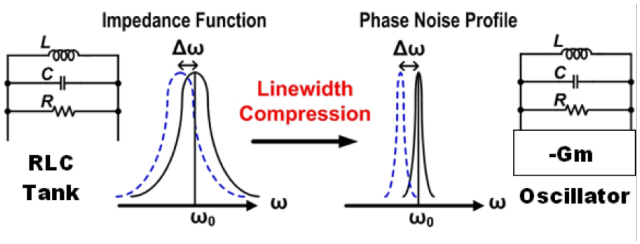

Figure 2.6: Comparison of LC tank impedance profile with LC oscillator phase noise profile.

We are close to describing a fully viable system, so it is a good idea to present a system block diagram at this point of what has been covered so far:

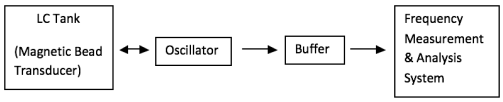

Figure 2.7: Simplified Magnetic Biosensor chip block diagram.

Noise Reduction Techniques

The main disadvantage brought about by using an oscillator sensing configuration is the considerable noise added to the system by the active devices of the oscillator. CMOS processes due to the imperfect junctions created between different materials (silicon, oxide, etc.) generally have very poor Flicker (1/f) noise performance. Although 1/f noise is dominant in lower frequencies (with Johnson/thermal noise being dominant at higher frequencies), the low frequency phase noise becomes upconverted and centered around the carrier tone (ω_{res}) due to nonlinearities in the system and the nonlinear Impulse Sensitivity Function (ISF) which is a time-sensitive transfer function from a noise source to phase or amplitude noise in the oscillator (treated by Hajimiri and Lee in [31].)

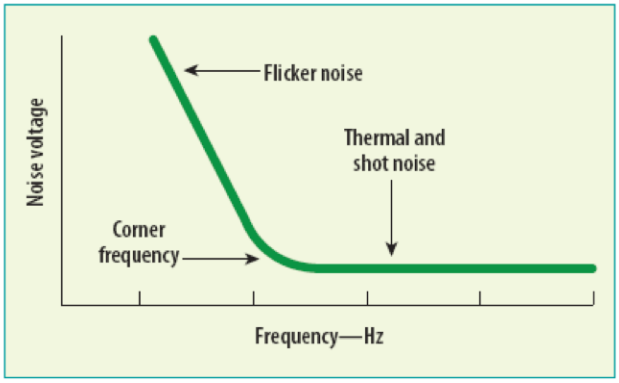

Figure 2.8: Oscillator Noise Profile.

As can be seen in fig. 2.8 above, thermal noise follows a flat noise power profile across all frequencies of interest. Below a certain frequency called the device 1/f corner, the Flicker noise becomes dominant and has noise power roughly proportional to 1/f. Since thermal noise is flat across all frequencies, it has an impulse as its autocorrelation function and therefore subsequent samples are not correlated at all. Thermal noise is zero-mean, which means that its effects can be largely reduced by taking enough samples of the signal with thermal noise and averaging them. Flicker noise, on the other hand, has a constantly rising noise profile as frequency decreases, so no matter how much averaging is done, the biggest contributors to the noise (lowest frequency components) will still remain untouched. Furthermore, all realistic systems are prone to frequency drift due to thermal and electrical fluctuations. Thus, there is a limit to how much averaging can be done since after a certain point significant drifting effects will become noticeable, offsetting the measurements and introducing considerable error into the results.

In our magnetic biosensing system, averaging can be effectively realized and controlled by varying the counting time per measurement. The oscillation frequency is estimated by counting the number of rising edges of the oscillator signal for a time window of predetermined length T. The predicted frequency is thus the number of edges divided by T. The fact that frequency counting using digital counters is being utilized to measure the frequency of the oscillator implies that we will experience frequency uncertainty with a 1/T noise profile due to the fact that the initial phase difference between the reference clock used for counting and the oscillator being counted is unknown. Counting for long enough divides this initial segment (which

stays constant length) by increasing values of T, making its effects negligible after a certain T in comparison to the oscillators internal noise sources. Fig. 2.9 is a plot (actually taken from real chip measurements), demonstrating these principles:

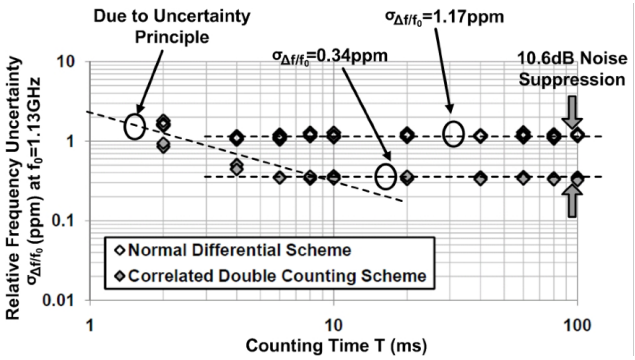

Figure 2.9: Frequency uncertainty comparison of Correlated Double Counting vs. Normal Differential over varying counting time intervals.

The decreasing thermal noise profile due to increasing T (counting window and effective averaging time period) cannot be seen since it is essentially completely dominated by the counting frequency uncertainty. As expected, past a certain T the noise profile flattens out due to having reached the region where the device Flicker noise becomes the dominant noise source. Longer counting intervals are unnecessary, but had they been measured one would expect the noise profile to eventually start increasing again due to the effects of drift. Drift is extremely difficult to model since it is heavily dependent on the specifics of the circuit, process, and environment, and thus we always avoid operating our circuit in that region. To further reduce the effects of drift, a realistic sensor implementation should always have a reference LC tank which is operated concurrently with the actual sensor tank. The reference LC oscillator should be strategically placed in very close proximity on-chip to the sensor oscillator so that thermal fluctuations affect the dynamics of both tanks (and more importantly the active devices in both circuits) roughly in the same manner causing big contributors of drift to be correlated between sensor and reference. Leaving the reference inductor always clear of magnetic beads, we can cancel out significant thermal drifting effects by counting the frequency of each oscillator and dividing the reference frequency by the sensor frequency. Assuming that thermal fluctuation mainly affects the parasitic capacitances of the active devices and for the most part leaves the inductance unchanged, we can mathematically show the effectiveness of this approach:

$$\Delta f_{sense} = \frac{1}{2\pi\sqrt{(L_0 + \Delta L) * (C_{tank} + \Delta C_{drift})}}$$
$$\Delta f_{ref} = \frac{1}{2\pi\sqrt{(L_0) * (C_{tank} + \Delta C_{drift})}}$$
$$X = \frac{\Delta f_{ref}}{\Delta f_{sense}} = \sqrt{\frac{L_0 + \Delta L}{L_0}} = \sqrt{1 + \frac{\Delta L}{L_0}} \approx 1 + \frac{\Delta L}{2L_0}$$

which cancels out fluctuations in the effective tank capacitance. It is important to note that this exact cancellation is achieved due to both the reference and sensor tanks having exactly the same intrinsic capacitance, C_{tank}. If we have a slight variation, between the two tanks in tank capacitance due to process variation (let the difference between the two C_{tank}s be $C_{mismatch}$), then the result X becomes:

$$X = \frac{\Delta f_{ref}}{\Delta f_{sense}} \approx (1 + \frac{\Delta L}{2L_0})\sqrt{\frac{C_{tank} + C_{mismatch} + \Delta C_{drift}}{C_{tank} + \Delta C_{drift}}}$$
$$\approx (1 + \frac{\Delta L}{2L_0})\left(1 + \frac{C_{mismatch}}{2(C_{tank} + \Delta C_{drift})}\right)$$

which shows that the larger the mismatch in tank capacitance, the more one will notice the effects of thermal capacitance fluctuation. Although this has not been verified experimentally yet, mismatch in tank capacitance can be measured by using both inductors as references (no magnetic beads on either) and used as a compensation factor for achieving almost the same (the accuracy of cancellation depends on the accuracy by which the difference in tank capacitance can be quantified) cancellation as in the perfect match scenario. It is important to keep in mind that this approach can ONLY be used to cancel capacitive mismatch between the sensor and reference cells due to systematic process variation and cannot account for other variations such as electrical and thermal drift gradients which may not affect both cells in the same manner.

2.3 Uniform Transducer Gain Sensor

The main issue with the sensor design presented in [115] is nonuniformity of the transducer gain. The design in [115] makes use of a standard spiral inductor, rendered as a 3D model here:

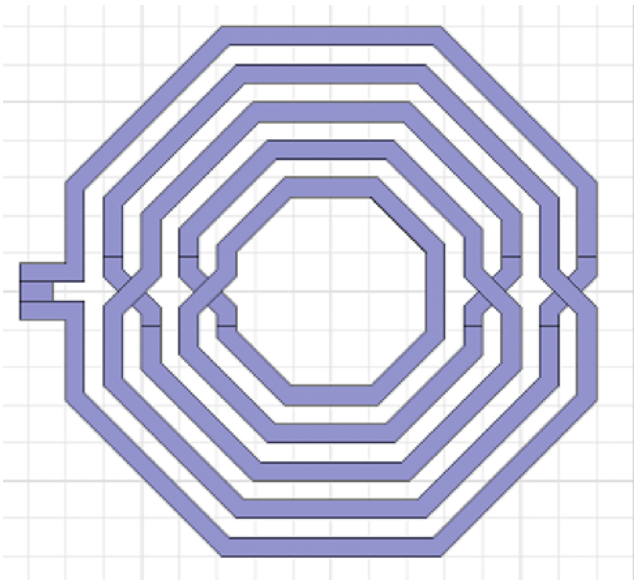

Figure 2.10: Top-view of standard on-chip spiral inductor (rendered using Ansoft HFSS 10.)

In order to understand the effect of the surface magnetic field of the sensing inductor on the uniformity of the sensor gain, we recap the expression for the sensor transducer gain. The physical process, which relates the presence of magnetic particles with frequency down-shift, can be modeled as follows. As before, assume a magnetic particle with effective susceptibility χ_{eff} and a volume Vp is placed close to the on-chip sensing inductor.

When a current I conducts through the coil, the local polarization magnetic field is $\vec{H}$. Assuming the presence of the magnetic particle will not alter this $\vec{H}$, the total magnetic energy therefore increases by ΔE_m after placing the particle,

$$\begin{aligned}\Delta E_m = (E_{m'} - E_m) &= \frac{1}{2}\iiint \vec{H}\cdot\vec{B'}dv - \frac{1}{2}\iiint \vec{H}\cdot\vec{B}dv \\ &= \frac{\mu_0}{2}\iiint_{V_p}\left[||\vec{H}||^2(1+\chi_{eff}) - ||\vec{H}||^2\right]dv = \frac{\chi_{eff}}{2\mu_0}\iiint_{Vp}||\vec{B}||^2 dv \\ &\approx \frac{\chi_{eff}}{2\mu_0}||\vec{B}||^2 Vp\end{aligned}$$

where $\vec{B'}$ and $\vec{B}$ are the local magnetic flux density with and without the magnetic particle. The approximation holds when the particle is small enough so that the polarization field is homogenous across its volume.

The sensor transducer gain can be defined as the relative frequency-shift (due to the inductance change) per particle as:

$$TransducerGain = \left(\frac{\Delta f}{f_0}\right)_{perparticle} = -\frac{\Delta L}{2L_0} = -\frac{1}{2}\cdot\frac{2\Delta E_m/I^2}{L_0}$$
$$= -\frac{1}{2}\cdot\frac{2\frac{\chi_{eff}}{2\mu_0}||\vec{B}||^2 Vp/I^2}{L_0} = -\frac{1}{2}\cdot\frac{\chi_{eff}Vp}{\mu_0 L_0}\cdot\frac{||\vec{B}||^2}{I^2}$$

This expression shows that the sensor transducer gain is location-dependent on the sensor surface and is proportional to the field quantity $||\vec{B}||^2/I^2$.

The inductor of [115] has been optimized for high quality factor at an operating frequency of 1GHz with nominal inductance of approximately 2nH. A heat-plot generated with HFSS 3D electromagnetic simulation software reveals the strength of the magnetic field along the surface of the sensor:

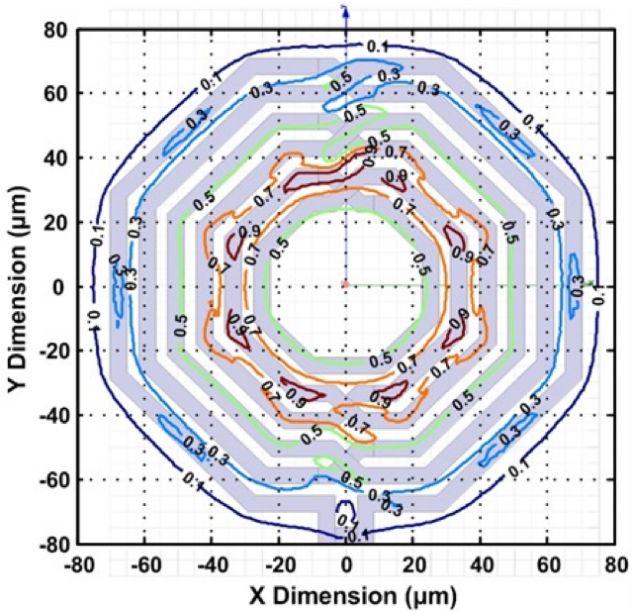

Figure 2.11: Simulated heat map plot of standard spiral inductor magnetic field profile normalized to 1.0.

As can be seen, the center region of the inductor (inside all of the spirals) has relatively uniform gain. However, outside of this small center region, the transducer gain varies drastically, even when traversing small distances in the radial direction.

It can be seen that the sensitivity varies almost 10:1 from the most to the least sensitive regions. This directly implies that while we may have good sensitivity in select regions of the sensor inductor, at the same time other regions suffer from very poor sensitivity. Thus, when using this system as a method to detect magnetic beads bound onto target DNA strands, we may get varying levels of signal depending on where the bead-bound DNA strands happen to pass over the sensor physically. This is certainly not an ideal situation as we wish to have the ability of receiving a consistent electrical response for fixed solutions with DNA molecules. Receiving varying responses with the same input stimulus easily leads to erroneous data and based on calibration may either lead to false positives or false negatives depending on where the threshold of detection has been adjusted.

Furthermore, this nonuniform sensor is completely nonviable if we wish to incorporate a sensing scheme which can not only sense the presence or absence of a specific target DNA strand but is also capable of detecting approximate concentration of the strand in the solution. This is that due to the fact that we do not know where the beads travel over the sensor, and that each bead may produce drastically different electrical responses. Thus, we cannot simply divide the overall sensor response by a normalized response per unit bead factor (as mentioned in an earlier section, the transducer gain, K, is now position dependent instead of being constant across the sensor area) to determine approximately how many beads are present over the sensor area. A sensor with very good uniformity is required for this purpose.

To design a uniform gain sensor we must first analyze which aspects of the inductor lead to the varying sensitivity with location. As derived previously, the square of the magnitude of the magnetic field right above the sensor surface (approximately the same vertical location of the beads) is directly proportional to the sensitivity. This implies that in order to design a sensor with uniform gain across its whole surface, we must now not only be concerned with the quality factor and inductance of the inductor but also with the magnetic field profile. The main challenge is the fact that we are restricted to relatively planar structures since the inductor is on-chip and each metal layer is at most a couple microns thick. We must first comprehend why the original sensor has largely nonuniform magnetic field. Studying an overlay of the inductor layout with a heat-map of the magnetic (B) field strength, we notice that the greatest variations in field strength are due to some of the area being above the inductor metal and some being above empty silicon dioxide. From classical electromagnetics we know that the interior of an ideal solenoid contains a very

uniform B field. Being confined to an almost completely planar structure, we can only create an extremely crude approximation to the solenoid in CMOS. We call this approximation the "bowl" shaped inductor due to its resemblance to a bowl:

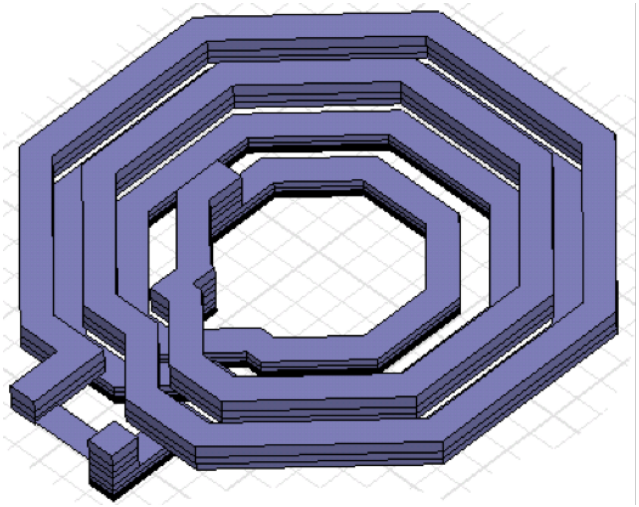

Figure 2.12: 3D rendering of "bowl" shaped inductor topology.

As can be seen, we essentially have two interleaved vertical layers of spiral which work to compensate each other's weak" spot regions. Furthermore, almost the whole area of the inductor is now covered with metal from either the top or bottom layer. The fields also add constructively creating an effective mutual inductance so that the total inductance of the sensor is greater than just the individual sum of the inductance of the each of the two spirals. While this drastically helps the uniformity of the sensor, we can still do better.

Noticing that the main problem region in uniformity of the stacked inductor now is an arc over which neither of the two spiral layers covers, we see that the uniformity can be greatly improved by increasing the metal width of the inner, top spiral to cover this area of weaker B-field. After performing this modification, we indeed notice a major smoothing out of the B-field strength as can be seen in design 1 of Fig. 2.13.

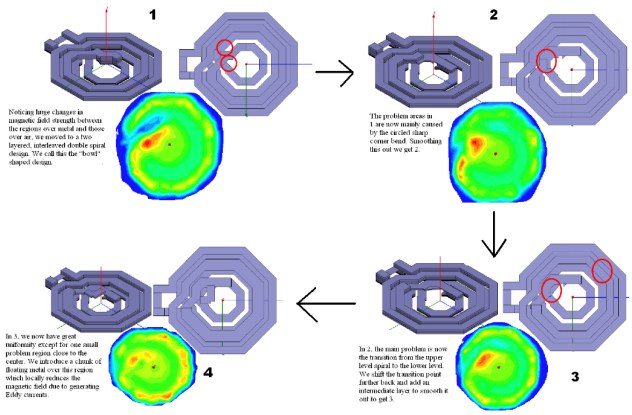

Figure 2.13: Uniform gain inductor design process.

Two small regions are now left with non-uniform B-field. One has considerably weaker sensitivity than the rest of the inductor (which can be seen as the blue area in the heat-map of design 1), whereas the second has considerably stronger sensitivity (red area in the heat-map.) We realize that these two regions are mainly due to the sharp corner bend in the inductor metal in that position, so we smooth this out and are left with design 2 in Fig. 2.13. We now realize that the main problem area now is exactly over the transition from the upper layer metal spiral to the lower layer, as shown with a red circle in design 2. We move the transition point further back and add an intermediate layer in an effort to distribute out the effect of the B-field directional change to reach design 3 shown in Fig. 2.13 As can be seen in design 3, this results in much improved field uniformity, leaving only a small spot of stronger B-field. We correct for this by inserting a small segment of floating metal directly above this area. Eddy currents can now form in the floating metal, directly over the stronger B-field region of the sensor, causing the magnetic field of that specific region to become degraded. The overall effect is a tiny degradation in overall inductance and quality-factor (due to shielding a small fraction of the inductor), but a substantial improvement in linearity. The final design along with its corresponding field magnitude heat-map implemented in the sensor chip is shown in Fig. 2.13 as design 4.

Fig. 2.14 illustrates the slight degradation in inductance and quality factor versus frequency resulting from adding the floating piece of metal:

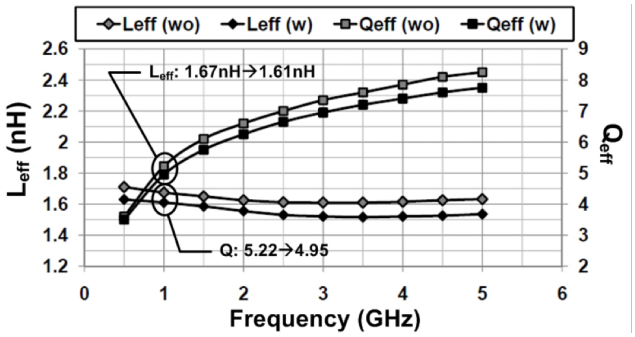

Figure 2.14: Comparison of inductance and quality factor before and after addition of floating metal.

Measurement Results

Fig. 2.15 shows a system diagram of our uniform gain magnetic biosensor chip. This project was done in collaboration with Professor Hua Wang. [112]

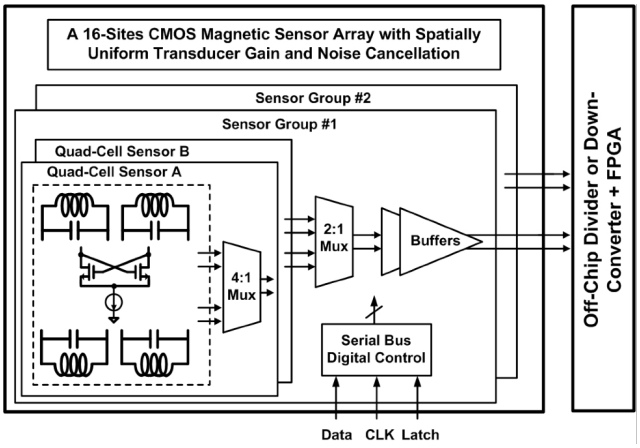

Figure 2.15: Block diagram of uniform gain biosensor chip.

Fig. 2.16 is an actual die photograph of the chip manufactured in IBMs 45nm SOI process:

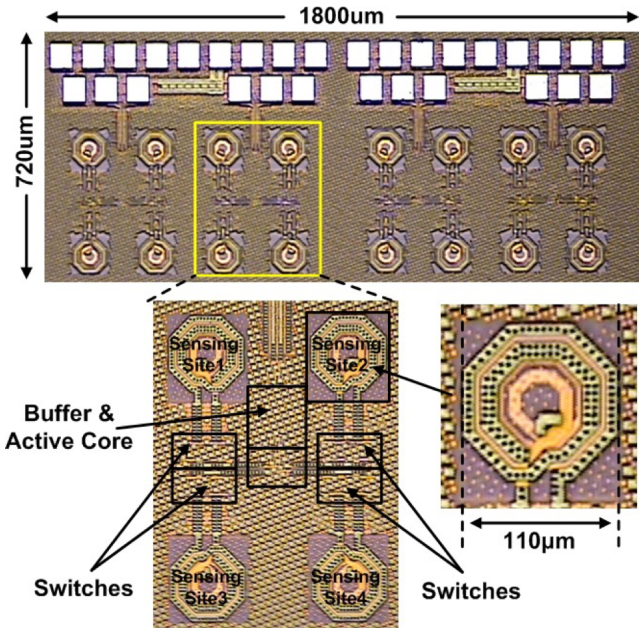

Figure 2.16: Uniform gain biosensor chip die photograph.

The sensing oscillator operates at a nominal frequency of 1.13GHz. Its phase noise profile is measured with an RDL phase noise analyzer, achieving -47.2dBc/Hz and -120.3dBc/Hz at 1kHz and 1MHz offsets, respectively. To verify the Correlated Double Counting (CDC) noise cancellation functionality, the frequency counting results for two different arrangements (both with counting duration of 0.1s) are shown in Fig. 2.17 The normal differential measurement approach described earlier suppresses the common-mode frequency drift, while an additional noise reduction (from σ=1179Hz to σ=391Hz) is achieved after enabling CDC. This is tested by taking measurements from two cores which are right next two each other but do not share common active devices (no CDC), followed by taking measurements from two cores which are approximately the same distance apart as the no-CDC scenario but do share an active oscillator core (CDC):

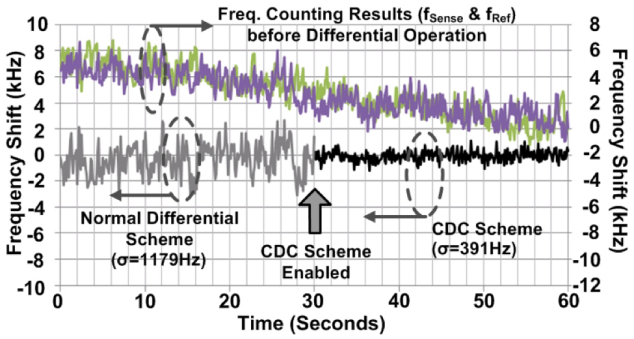

Figure 2.17: Time domain frequency stability comparison between CDC and Normal Differential schemes.

Fig. 2.18 shows measurements taken which portray the frequency uncertainty for both CDC and no-CDC schemes with respect to counting interval time. As expected, due to frequency counting uncertainty, we see a 20dB/decade drop in uncertainty for short counting times. The frequency uncertainty curve overshadows the theoretical noise floor caused by thermal noise so this is not seen in the results. However, after a certain point, increasing the counting time of each sensor does not yield better results. This is due to the fact that we have hit the region where we have averaged out a significant amount of frequency uncertainty and thermal noise causing the upconverted Flicker noise to become dominant. Furthermore, we notice a 10.6dB reduction in the noise-floor when enabling Correlated Double Counting due to the 1/f noise cancellation. This is actually better than the theoretical maximum improvement. We believe that this is because the original 1/f noise profile model used for the theoretical derivation may not accurately represent this IBM process.

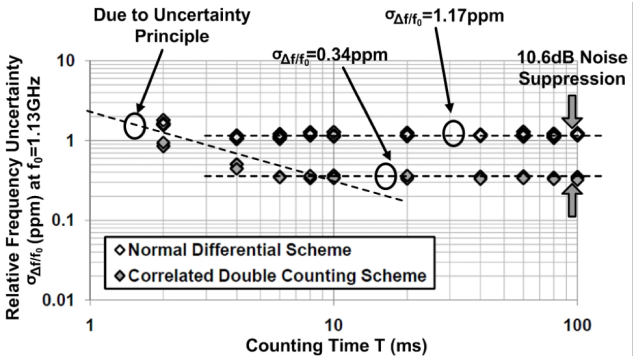

Figure 2.18: Frequency uncertainty comparison of Correlated Double Counting vs. Normal Differential over varying counting time intervals.

To verify the sensor gain uniformity, two sets of magnetic sensing experiments were

performed. Magnetic particles, DynaBeadső M450-Epoxy (Diameter=4.5μm), are used as the test samples in both measurements due to their ease of handling. First, a single particle is randomly placed onto the sensing surface and the normalized sensor responses together with the particles positions are recorded and plotted:

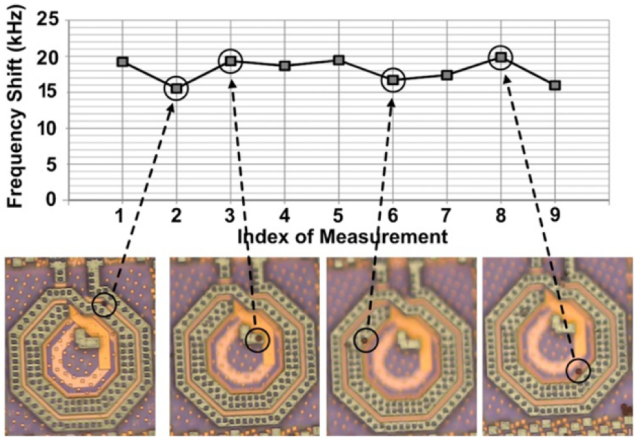

Figure 2.19: Measuring sensor uniformity by evaluating frequency shift for a single 4.5u magnetic bead placed in varying locations.

The consistent frequency-shift reading with an average value of 18.2kHz per particle verifies the uniform sensor transducer gain. In the second experiment, different numbers of magnetic particles are deposited onto the sensor surface and their corresponding output frequency-shifts are shown:

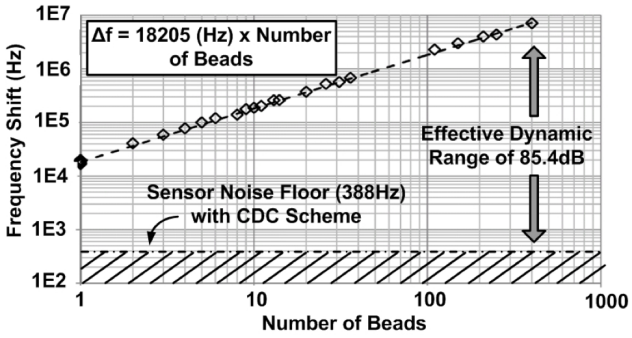

Figure 2.20: Sensor Linearity and Dynamic Range Measurement.

Note that with a noise floor of 388Hz after CDC operation, a single 4.5μm magnetic particle is still far above our sensing limit. The measured linear response (up to 409 beads) indicates an effective dynamic range of at least 85.4dB. The sensor surface

area is 80 x 80 μm, so we were not physically able to carry out measurements with more than 409 beads.

2.4 Multiplexed Magnetic Biosensor: A Magnetic Spectrometer

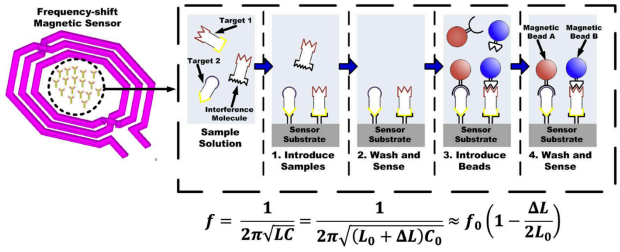

Figure 2.21: A frequency-shift magnetic biosensor functionalized with a two-probe affinity-based sandwich assay. The sensor used in this experiment must be capable of differentiating between the two types of magnetic beads utilized in order to correctly quantify concentration of each target analyte.

The frequency-shift sensor of [112] demonstrates an example of a highly sensitive, cost-effective magnetic particle biosensor in CMOS with no need for external magnets. The use of such a magnetic sensor for nucleic acid and protein targets has been successfully demonstrated recently [78]. Despite their cost and sensitivity advantages, however, the magnetic biosensors mentioned lack multi-probe diagnostics in contrast to the different color fluorophores which are available and can be utilized for simultaneous single-site multiple target differentiation in fluorescent-based approaches. Multi-probe detection schemes not only significantly increase the amount of targets which can be detected in a given sensing area, but are crucial for experiments requiring signal colocation such as detecting multiple biomarkers on a single cell. The lack of multi-probe magnetic diagnostics is primarily due to the current approaches simply measuring changes in the magnetic susceptibility, either at low frequencies [33, 27, 12, 102] or at fixed RF frequency [113, 115]. Consequently, these approaches do not provide a clear path for differentiating between a large number of small magnetic particles and a smaller number of larger size magnetic particles with similar total magnetic content (or anything in between). Recent developments in the field of magnetic biosensing have begun paving the way for enabling efficient multi-probe diagnostics. Namely, [48, 82] utilize Hall effect magnetic sensors and strong external biasing magnets to differentiate among different magnetic markers. This is accomplished by measuring the sensor signal response at different external biasing field conditions and making use of the fact that different magnetic labels exhibit unique magnetization profiles. [48, 82] show promising results, however, they

require very strong external biasing fields which can only be provided by expensive permanent magnets. Furthermore, the field strength must be varied, implying that a mechanical system must be employed to modify the relative distance between the permanent biasing magnet and the sensing apparatus. [79] presents an alternate approach to distinguish among different magnetic marker types by using AC susceptometry to characterize the frequency response of the magnetic susceptibility of the markers. This approach does not require an external magnet, however, [79] use a discrete setup with separate excitation and sensing coils and bulky and expensive test equipment for driving and detection such as an RF frequency synthesizer and a lock-in amplifier.

We propose, design, implement, and demonstrate a frequency shift magnetic sensor capable of multi-probe detection by designing our chip to differentiate between different kinds of magnetic material that can be used as labels for tagging biomolecules of interest. Similar to [79], our chip achieves this differentiation by performing AC susceptometry. However, our approach is fully integrated in a standard CMOS process and does not require any external test equipment or external biasing fields for operation. This is unlike GMR sensors which require exotic post-fabrication steps and Hall effect sensors which require external magnets and/or suffer sensitivity issues due to saturation of the sensor by the biasing field [110]. In Section Section II, we extend the frequency-shift methodology to a novel technique capable of performing multiplexed detection by way of magnetic spectroscopy. Section IV analyzes critical design areas and tradeoffs for sensor optimization. Finally, the results of a proof-of-concept implementation of such a multiplex-capable magnetic sensor design implemented in a standard CMOS 65 nm process are presented in Section V.

II. Multiplexed Operation

As mentioned previously, magnetic material placed over the inductor of the sensing LC tank causes a shift in inductance, ΔL. It is instructive to repeat expression for ΔL here in order to arrive at subsequent developments in the analysis: Assume a magnetic particle placed on the inductor surface has an effective magnetic susceptibility χ and a volume V_p. A current I flowing through the inductor generates a local magnetic field, $\vec{H}$. Under the assumption that the magnetic particle is small enough so as to not significantly perturb this $\vec{H}$, the difference in total magnetic energy ΔE_m can be computed:

$$\begin{aligned}\Delta E_m &= E_{m'} - E_m \\ &= \frac{1}{2}\iiint \vec{H}\cdot\vec{B}'dv - \frac{1}{2}\iiint \vec{H}\cdot\vec{B}dv \\ &= \frac{\mu_0}{2}\iiint_{Vp}\left[||\vec{H}||^2(1+\chi) - ||\vec{H}||^2\right]dv \\ &= \frac{\chi}{2\mu_0}\iiint_{Vp}||\vec{B}||^2dv \approx \frac{\chi}{2\mu_0}||\vec{B}||^2V_p\end{aligned}$$

Note that the volume integration is performed only over the volume of the magnetic particle (V_p) since the fields remain unchanged elsewhere. $\vec{B}'$ and $\vec{B}$ represent the local magnetic flux density in the volumetric region V_p in the presence of the bead and without the bead respectively. The approximation made is that the magnetic bead is small enough such that the magnetic field is uniform inside its whole volume. As before, substituting ΔE_m into the expression for inductance yields:

$$\Delta L = \frac{2\Delta E_m}{I^2} \approx \frac{\chi}{\mu_0}\frac{||\vec{B}||^2}{I^2}V_p$$

Finally, substituting the expression for ΔL into the frequency-shift equation, we arrive at the expression for the transducer gain (as derived in the previous sections):

$$\frac{\Delta f}{f_0} = -\frac{\Delta L}{2L_0} = -\frac{\chi}{2L_0\mu_0}\frac{||\vec{B}||^2}{I^2}V_p$$

It can be readily seen from this expression that the transducer gain is proportional to the magnetic susceptibility of the particle χ, the local magnitude of the magnetic field squared $||\vec{B}||^2$, and the inverse of the current through the inductor squared, I^2.

$\vec{B} = \mu_0 NI/l$ for an infinite solenoidal inductor and to first order $\vec{B} \propto I$ for finite solenoids as well [16]. Therefore, we expect $||\vec{B}||^2$ and I^2 in (4) to generally cancel each other out when the transducer gain is spatially averaged over the whole inductor surface area. The local changes in $||\vec{B}||^2$ on the other hand are highly critical for sensor gain uniformity as will be discussed in the subsequent section on sensor optimization.

Of note in (4) is the direct dependence of the transducer gain on the magnetic susceptibility of the material present over the sensor surface, χ. For low frequency

measurements, it is often assumed that χ does not exhibit any frequency dependency. However, this is not the case at radio frequency (RF), where for typical magnetic beads used in biosensing experiments, χ actually shows strong frequency dependence [22]. Beads used in magnetic biosensing experiments consist of an outer polystyrene matrix encasing single-domain super-paramagnetic nanoparticles (Fig. 2.22). The outer matrix is nonmagnetic, and any measurable magnetic behavior of the beads is completely due to the single domain particles inside [22].

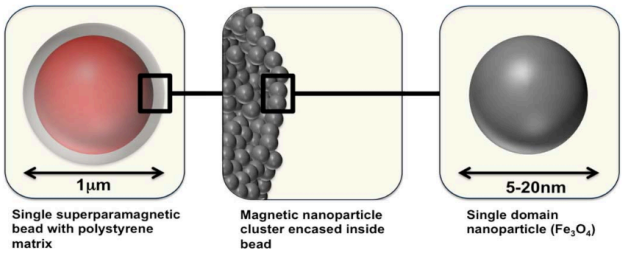

Figure 2.22: Construction of super-magnetic beads from encased iron oxide nanoparticles. The nanoparticles are single-domain and super-paramagnetic.

The paramagnetic property of the particles implies that they become magnetized and exhibit a magnetization field only when subjected to an external magnetic field. Different magnetic beads suitable for biomolecular tagging utilize different size single-domain magnetic nanoparticles in their matrices. The magnetic activity of the nanoparticles is caused by their magnetic domains rotating to align with the external magnetic field.

Indeed, in the presence of an external alternating magnetic field, the rate at which magnetic domains can reorient to align with the external field is dependent on the temporal response of the domain, with larger domains typically reorienting more slowly than smaller ones. This effect leads to the frequency-dependent response in the magnetic susceptibility, χ. As the frequency increases, the dipoles are limited by their finite reorientation speed, resulting in a decrease in the real part of χ due to a phase mismatch between the magnetic polarization vector **M(t)**, and the external magnetic field vector, **H(t)**. In fact, there exists a null frequency f_{null}, at which the delay of the magnetic dipole is such that it completely cancels the field effect, resulting in $\mathrm{Re}(\chi) = 0$. This is not due to Néel or Brownian relaxation, both of which are typically low frequency phenomena [22]. The null is caused by the magnetic dipole orientation always being in time-quadrature with respect to the external field at that frequency. Interestingly, beyond f_{null}, the real part of χ becomes negative and it asymptotically approaches 0. Fig. 2.23 shows the shape of the real part of χ

for a typical superparamagnetic nanoparticle along with magnetization state of the particle in each frequency range.

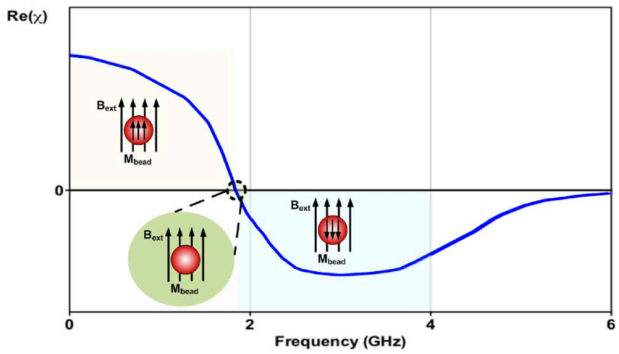

Figure 2.23: Typical frequency dependence curve of magnetic susceptibility, χ, for magnetic nanoparticles. The frequency shift of the sensor at each frequency is directly proportional to χ at that frequency. The phase of the magnetization vector in each frequency range for the nanoparticle is also shown.

A crucial observation for extending the frequency-shift sensor is that f_{null} and the magnetic frequency spectrum (signature) of the magnetic domains depend on the domains size and hence can be used as a spectroscopic means of differentiating between different kinds of beads that could be used for detecting different analytes. The sensitivity of the resonance-based sensor was derived as being directly proportional to χ. Therefore, if the LC tank can be tuned to vary its center frequency, f_0, across a wide range, measuring the relative frequency shift caused by magnetic material on the surface at each frequency in the range yields spectroscopic information on the magnetic susceptibility. This phenomenon provides an effective means of single-site multiplexing in a CMOS magnetic sensor by using the spectroscopic data to discern between various beads, similarly to optical systems.

The center frequency of the sensing LC resonator can be shifted by tuning either the inductance or the capacitance of the tank. Integrated inductors are very difficult to tune electrically, however. Furthermore, the transducer gain would not only be changing due to χ at each frequency but would also be affected by the absolute value of the inductance as L is in the derived expression for transducer gain. This complicates the measurement process, as the measured data would have to be adjusted to remove the dependency on the varying inductance. Tuning the capacitance to achieve frequency shifting is a more viable option. As shown by the susceptibility measurements in [22], an interesting range for obtaining spectroscopic data of χ is between 1 GHz and 6 GHz. Varactor tuning can only provide minimal amounts of

extra capacitance and therefore the preferred technique is one where a switchable passive capacitor bank is used. Fig. 2.4 shows a schematic representation of a simplified sensor cell in which biasing and switch circuitry have been omitted.

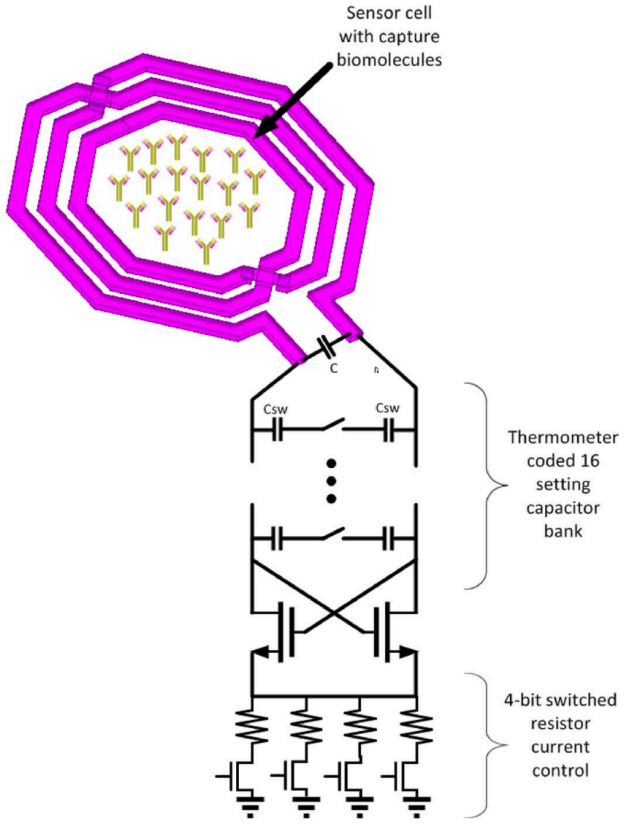

Figure 2.24: Sensor cell schematic. The switchable capacitor bank for tuning the frequency between 1.1 GHz and 3.3 GHz is thermometer encoded and layed out symmetrically for minimum variability among the frequencies switched. Switched resistors are used to control oscillator current optimized for maximum phase noise performance.

III. Sensor Optimization

Depending on the target application, there are various parameters that should be taken into consideration when designing a magnetic sensor cell. Perhaps the most important parameter is the sensor Signal-to-Noise Ratio (SNR), which directly quantifies the sensitivity of the cell. The overall sensor SNR is determined by a combination of the electrical SNR and the biochemical SNR (determined by the biological experimental assay and the surface chemistry involved). The goal of the circuit designer should be to optimize the electrical SNR such that the overall SNR is dominated by biochemical SNR. The biochemical SNR is independent of

the multiplexing to first order and is addressed elsewhere empirically with bioassay experiments [78]. In the rest of this section, we focus only on the factors determining the electrical SNR component of the overall SNR.

As derived in (4) of Section II, the frequency shift of the resonant LC tank used for sensing is proportional to $-(\chi_{eff}/2L_o\mu_0)(||\vec{B}||^2/I^2)V_p$. The sensor measurement is performed by counting the electrical oscillation frequency and detecting shifts in frequency. Thus, the frequency shift directly corresponds to the signal part of the electrical SNR. The noise is determined by the frequency instability (phase noise) of the oscillator, which in turn is determined by the electrical noise of the LC resonator and of the CMOS transistor active devices. Specifically, the inductor has an effective real series wiring resistance that contributes thermal noise to the system. The majority of the oscillator phase-noise, however, is due to the active devices. Thus, in order to maximize our effective electrical SNR, we aim to maximize the frequency shift at each frequency point as well as concurrently minimize the oscillation phase-noise.

As discussed previously, perhaps the most effective method which is controllable by the designer for maximization of the frequency shift is to minimize the total inductance. Assuming that a typical spiral configuration is used for the inductor design, there is a strong correlation between inductor physical area and amount of inductance. Thus, to achieve a smaller inductance, the active sensor area must be reduced. This poses a disadvantage for the subset of biological experiments that seek to detect large quantities of bigger biomolecules such as cells. The target bio-experiments must therefore also be taken into consideration when determining the inductor to use. Simply picking the smallest inductance possible that allows for good phase-noise performance and provides enough oscillator loop gain is therefore not necessarily the best option since a sensor which has a sensor area that is physically too small for large-object detection can negatively impact the biochemical SNR significantly.

To make matters more quantitative, a simple analysis making several simplifying assumptions is done regarding the electrical SNR. Per the analysis of [111], the majority of the $1/f^2$ noise is averaged out by the counting measurement process, leaving $1/f^3$ noise as the main contributor to the oscillator phase-noise. For a given process technology, assuming a fixed biasing current density for the core devices and a constant tank voltage amplitude V_{DD} (limited by the supply rail), the transistors DC current I_d and width W are related to the inductance L by the following relation:

$$I_d \propto W \propto \frac{V_{DD}}{R_{tank}} = \frac{V_{DD}}{\omega_0 LQ}$$

with R_{tank} and Q as the equivalent parallel resistance and quality factor of the LC tank respectively at the tank resonant frequency ω_0, $I_d \propto V_{DD}/R_{tank}$ is derived assuming that the oscillator has been designed for biasing at the onset of the voltage-limited operating regime [31]. At fixed biasing current density, the current I_d is directly proportional to the device width W, and therefore the Power Spectral Density (PSD) of the drain flicker noise current can be modeled as $i^2_{n,1/f}(\omega) = \beta/\omega$. Using the Linear Time-Varying (LTV) phase-noise model from [31], the $1/f^3$ jitter coefficient can be determined as

$$\begin{aligned}\sigma^2_{\Delta f/f_0} &\propto S_\phi(\omega) \cdot \omega^3 \\ &= \frac{c_0^2}{q_{max}^2} \cdot \frac{i^2_{n,1/f}(\omega)}{2\omega^2} \cdot \omega^3 = \frac{c_0^2}{q_{max}^2} \cdot \frac{\beta}{2} \\ &\propto \frac{c_0^2}{V_{DD}^2 C^2} \cdot \frac{V_{DD}}{\omega_0 LQ} \propto \frac{L}{Q}\end{aligned}$$

where c_0 is the DC term of the Impulse Sensitivity Function (ISF), $\Gamma(t)$, and q_{max} is the maximum tank charge swing. Thus, the noise floor $\sigma^2_{\Delta f/f_0}$ is directly proportional to the quantity L/Q which is completely determined by the sensing inductor design. It was earlier derived that the sensitivity $\Delta f/f_0 \propto ||\bar{B}||^2/I^2 L_o$ where $||\bar{B}||^2$ is the spatially averaged magnetic field over the inductor surface squared. Thus, for an optimally biased cross coupled CMOS oscillator with $1/f^3$ phase noise being dominant, the SNR can be computed:

$$\text{SNR} = \frac{\frac{\Delta f}{f_0}}{\sigma_{\Delta f/f_0}} \propto \frac{||\bar{B}||^2}{I^2 L} \cdot \sqrt{\frac{Q}{L}} = \frac{||\bar{B}||^2}{I^2} \cdot \sqrt{\frac{Q}{L^3}}$$

As discussed earlier, the spatially averaged $||\bar{B}||^2$ is proportional to I^2 to first order for most inductor topologies and thus the two quantities mostly cancel each other out, resulting in: SNR $\propto \sqrt{Q/L^3}$. This expression confirms the intuition that inductor quality factor (Q) should be maximized and also reveals a strong dependence on the value of the inductor. Namely, the inductance should be minimized in order to achieve the highest SNR possible. As mentioned before, there are various reasons,

however, why a designer would not want to select the smallest inductor possible such as reduced biological SNR and degradation of quality factor of on-chip inductors which are too small. Fig. 2.25 plots simulated SNR values for different spiral inductor configurations of a standard 3-turn differential inductor. Three turns were selected to achieve a large enough inductance within reasonable area constraints and with acceptable quality factor. Furthermore, an odd number of turns forces the center tap to be on top of the inductor, allowing for easier layout of the VCO and power distribution.

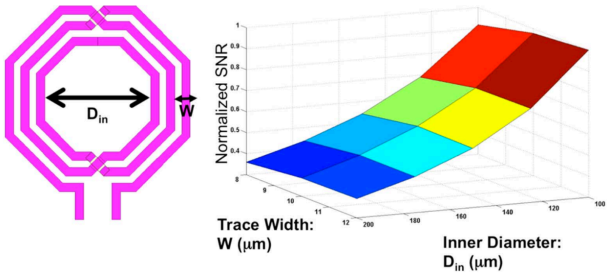

Figure 2.25: Simulated sensor SNR values for a 3-turn symmetric spiral inductor with different inner diameters (effective sensing area) and trace width sizes (f_0 = 1GHz). The plotted SNR is normalized by the maximum achieved SNR.

As detailed in the previous section on improving the gain uniformity of the sensor cell design, another important factor that must be considered is the uniformity of the cells sensitivity. Looking again at the expression for the sensor frequency shift, the shift is also directly proportional to the local magnitude of the spatial magnetic field squared, $||\vec{B}||^2$. Thus, if the magnetic field profile across the surface of the sensing inductor varies significantly, the same magnetic material will result in different frequency shifts depending on its physical location. This detection behavior is highly undesirable because the only sensing mechanism is measurement of the frequency shift which indicates that one would not be able to differentiate between more magnetic material in a weaker sensitivity spot versus less material in a stronger spot. Therefore, an inductor with irregular magnetic field magnitude profile, while suitable for qualitative biological experiments (detection of presence vs. absence of signal), is unsuitable for quantitative biological experiments which require accurate quantification of the amount of magnetic material present over the sensing surface. In the previous section [112], a custom inductor layout was designed, starting from a "bowl" shape, modifying trace spacing and thicknesses, and adding a shimming piece of metal with the objective of minimizing the change of $||\vec{B}||^2$ over the whole

inductor area. Defining a uniformity metric as the ratio of the largest $||\vec{B}||^2$ within the sensing area to the smallest $||\vec{B}||^2$, a 200:1 improvement in the uniformity was achieved over the initial design with minimal degradation in inductance and quality factor as a tradeoff.

The major causes of abrupt changes in the magnetic field profile over the area of the inductor occur in proximity to current carrying traces. On the other hand, the center region inside the traces of a spiral inductor exhibits fairly uniform magnetic field strength throughout the whole region due to spatial averaging of the fields generated from the outer circular wires. This is shown in Fig. 2.26 along with a linearity measurement performed using $4.5\mu m$ Dynabeads and the sensor presented in Section IV.

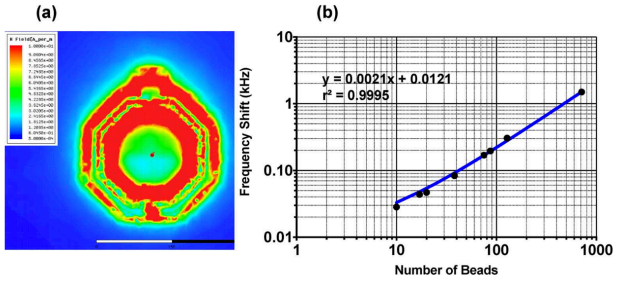

Figure 2.26: (a) Simulated magnetic field distribution over surface of the on-chip spiral inductor used in the implemented magnetic spectrometer prototype. The strongest points in field occur over and next to metal wires with the center of the inductor having uniform field strength. Thus, the center of this inductor is apt for use as the effective sensing region of the magnetic biosensor. (b) Log-log plot of measured frequency shift vs. number of beads for $4.5\mu m$ Dynabeads over sensor surface using the prototype spectrometer with corresponding linear regression. Great linearity is achieved throughout the whole measured range that spans over two orders of magnitude indicating that the sensing uniformity is adequate for quantitative experiments.

Therefore, another approach for achieving sensor cell uniformity is to use a spiral inductor with a sufficiently large empty center opening that can effectively be used as a sensor capable of quantification provided that magnetic material is only permitted to be located within this center region and not over the outer traces. This approach suffers from slightly reduced total effective sensing area, but significantly simplifies the design process by not requiring special layout modifications such as extra or thinner wiring and shimming pieces of metal that result in Eddy currents and can possibly reduce the quality factor.

IV. Multi-Frequency Operation

The sensor SNR and uniformity performance metrics have been discussed thus far. These two metrics apply to all resonance shift magnetic biosensors and are not specific to our magnetic spectrometer. In order to measure the magnetic susceptibility spectrum of magnetic content with a resonance shift sensor, it must be capable of operating across a range of different center oscillation frequencies. A wide tuning range with many operating points in between is desirable for maximizing the amount of magnetic information that can be measured by the device. The magnetic field profile remains unchanged across the whole operating range, since the sensor size is physically very small compared to the operating wavelength. The SNR, on the other hand, is a strong function of frequency since both the sensitivity and noise change at different frequencies. The frequency shift is proportional to the magnetic susceptibility at the sensing frequency, and the phase noise performance of the oscillator can be significantly different from the lowest to the highest operating frequency. The phase noise is mainly affected due to addition of extra capacitance to the resonant tank, the switch transistors being in different operating states, and the changing bias conditions of the active cross-coupled devices. Therefore, designing a multi-frequency sensor can be considerably more challenging than its single-frequency counterpart since close attention must be paid to the SNR at all operating frequencies rather than at a single point.

Such a wide tuning range can be very challenging to implement in practice for a single oscillator due to the nature of the parasitics of the switching devices required for tuning. As described in Section III, the most interesting frequency range for typical magnetic nanoparticles used in biological experiments is between 1 GHz and 6 GHz. Varactor tuning is therefore inadequate for such a broad range, and passive capacitances must be switched in order to achieve tuning. In a standard CMOS process, CMOS transistors must be used as the switching elements, which are subject to an on-resistance vs parasitic capacitance tradeoff. Therefore, great care must be taken in designing the switching network in order to achieve a wide tuning range while maintaining acceptable phase noise performance across the whole range. Furthermore, dynamic DC biasing techniques should be used for biasing the switch transistors in order to ensure that they are fully turned on or off and do not experience overvoltage conditions. Keeping the switching devices fully in the on or off states is crucial to prevent distortion of the oscillator ISF, which can very easily led to signifi cant phase noise degradation. Fig. 2.27 shows the schematic of a single switching branch utilizing inverters with custom supply voltages for setting the DC

voltages of the drain and gate nodes of the switch transistors.

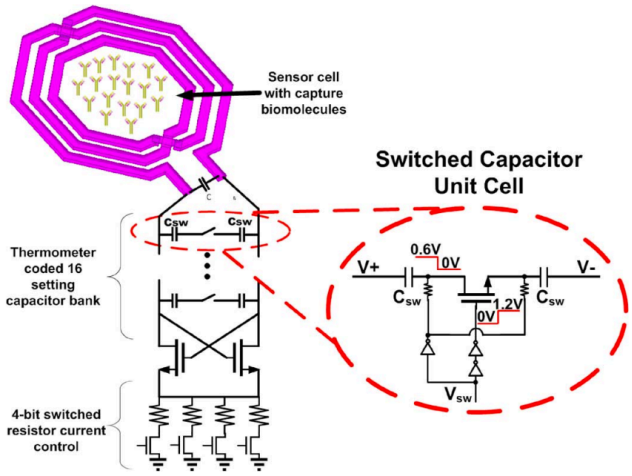

Figure 2.27: Zoom-in of switching cell of magnetic spectrometer sensing circuit. The drain/source nodes of the switching transistor must be biased to 0 V DC when ON and 0.6 V DC when off and the gate at 1.2 V DC when ON and 0 V DC when OFF to ensure full turn on and turn off of the switch transistor. The switch transistor not fully turning on or not being fully off when required results in a distorted large signal oscillation, which in turn can severely degrade the ISF of the oscillator leading to inferior phase noise performance. Inverter circuits with custom supplies are used to achieve these biasing conditions.

V. Sensor Implementation and Results

To enable multi-color magnetic sensing schemes and to take full advantage of the rich frequency-dependent information available, we have designed and demonstrated a CMOS magnetic spectrometer operating between 1.1 GHz and 3.3 GHz [93, 94]. Similarly to the single-frequency sensor presented in [115], the spectrometer does not require an external biasing magnet or any exotic post-fabrication processing. Unlike previous integrated designs, the sensor operates at multiple frequencies over a diverse range. Furthermore, under certain conditions and where at least one degree of freedom (number of different bead types which can be differentiated) can be sacrificed, a self-referencing scheme can be implemented with some simple signal processing. This relies on the fact that the same oscillator core is being used for all operating frequencies and therefore the sensor noise and thermal variation is strongly correlated among all of the frequencies. Therefore, if more measurement points exist than number of beads to be separated, the resulting system is over-defined and information can be extracted about the noise and used for drift stabilization. The information can either be extracted using black-box methods such as machine

learning or linear regression or by modeling the circuit in order to determine how variations in one frequency af fect the other frequencies of operation. Operating the device in this self referencing mode further increases its spatial multiplex density. Alternatively, the traditional approach of designating one sensor as a dedicated reference may be used as well, taking care that this cell remains clear of magnetic content over its surface.

As described in Sections II and III, a frequency-shift sensing scheme is used to detect and measure magnetic materials. The sensor core is a stable free-running oscillator with an on-chip LC resonator. The inductor of the LC tank is the magnetic sensing cell, where its value changes due to the changes of the χ of the magnetic material. Thus, the full frequency dependent information of χ is preserved by measuring frequency shift due to magnetic materials on the sensor at frequencies of interest. As discussed in Section III, the center region of a spiral inductor exhibits the greatest magnetic field uniformity, and therefore the effective useable sensing area of the $260\mu m$ x $260\mu m$ inductor in our implementation is $150\mu m$ x $150\mu m$. The frequency tuning from 1.1 to 3.3 GHz is achieved with a thermometer-coded switched capacitor bank, which allows for 16 unique frequencies of operation within the range. Thermometer coding is implemented to guarantee uniform frequency steps across the range (Fig. 2.28)

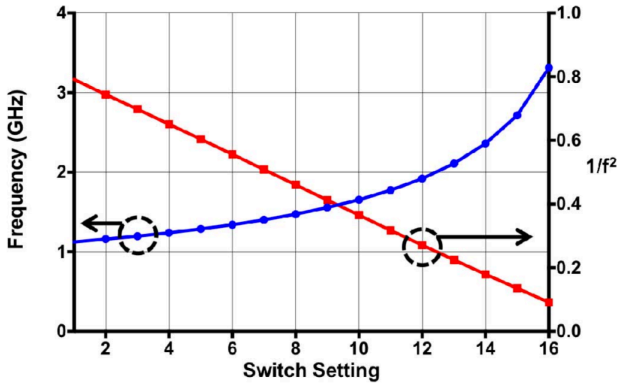

Figure 2.28: The blue curve plots all 16 operating frequencies from 1.1 to 3.3 GHz of the prototype magnetic spectrometer. The red curve plots $1/f_0^2$ (f in GHz) which should be completely linear with respect to switch setting since each step adds a fixed amount of capacitance. The great linearity shows the efficacy of using thermometer encoding for the passive capacitor bank.

An NMOS-only differential cross-coupled topology was chosen and optimized for phase-noise and power efficiency at 1.1 GHz, since beads generally exhibit the

largest response at this frequency [22]. The oscillator phase-noise was optimized by appropriately sizing the LC tank and active devices as well as the DC biasing points [31]. Care was taken to ensure that acceptable phase noise performance was achieved throughout the whole range up to 3.3 GHz, however. Furthermore, the chip has two on-board 36-bit frequency counters and a digital divide-by-64 multiplexed output. We have implemented a 4-sensor cell array as a demonstration of the ease of scalability of the system to a larger array. Extending a frequency-shift system to a large array has already been demonstrated by [115, 78] and would incur almost no extra area overhead since the sensor outputs are buffered and can be readily combined in the current domain. As also shown by [115], sensing cells can be placed at the same spacing as cell size or less without cross-coupling issues. This is due to the fact that the magnetic field rolls off very rapidly outside the vicinity of the sensing cell. Fig. 2.29 [93, 94] shows a block diagram of one half of the prototype sensing system.

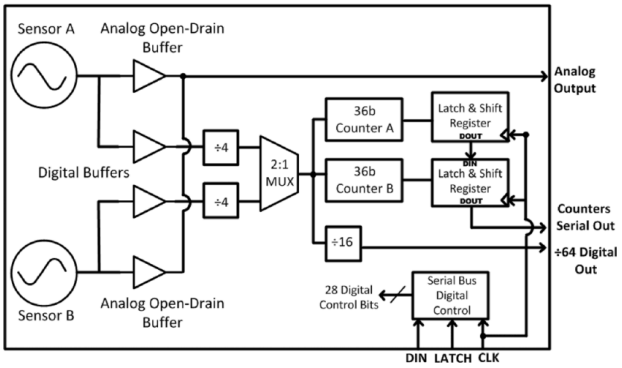

Figure 2.29: Block diagram of magnetic spectrometer prototype. The fabricated chip consists of two copies of this core unit and has a total of 4 sensing cells.

The oscillator achieves a state-of-the-art phase-noise of -133.7 dBc/Hz and -66.4 dBc/Hz at 1MHz and 1kHz offsets respectively at 1.1 GHz, and -124.8 dBc/Hz and -43.2 dBc/Hz at 1 MHz and 1kHz offsets respectively at 3.3 GHz (Fig. 2.30).

The sensor core draws between 3 and 10 mA depending on the adjustable current bias setting from a 0.6 V supply in order to allow optimal biasing for good phase noise performance. The cell consumes as little as 1.8 mW at the lowest current setting. The frequency signature of $1\mu m$ and $4.5\mu m$ Dynabeads magnetic beads from Invitrogen [42] and $1\mu m$ beads from Bangs Laboratories [40] were measured, as shown in Fig. 2.31 [93, 94].

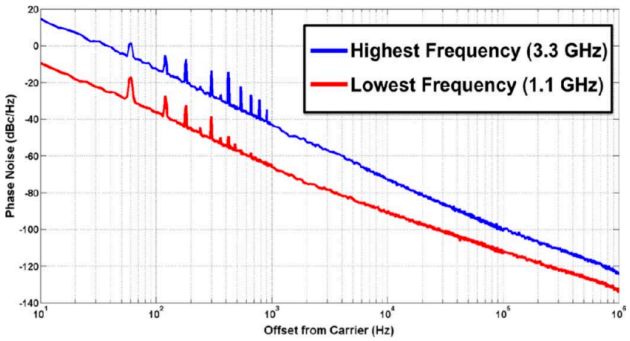

Figure 2.30: Plot of oscillator phase-noise performance at lowest (1.1GHz) and highest (3.3 GHz) frequencies of operation. The oscillator achieves a phase-noise of -133.7 dBc/Hz and -66.4 dBc/Hz at 1 MHz and 1 kHz offsets respectively at 1.1 GHz, and -124.8 dBc/Hz and -43.2 dBc/Hz at 1 MHz and 1 kHz offsets respectively at 3.3 GHz.

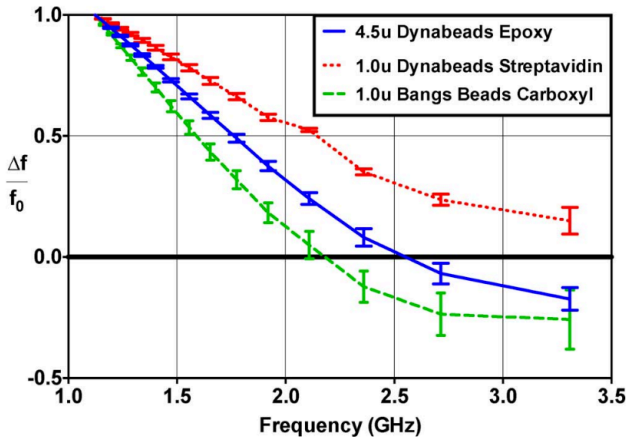

Figure 2.31: Frequency response measurement of three different kinds of beads. Data is normalized to the lowest frequency (1.1 GHz) to be independent of concentration. Three measurements have been done for each bead type. Error bars represent one standard-deviation around the mean.

As explained previously, these magnetic beads are composed of magnetic nanoparticles within a polystyrene matrix and coated with a biologically relevant functional group. These magnetic bead sizes are suitable for performing many different biologically relevant assays including DNA [78], protein [78], and cell [114] detection experiments. The sensor can be used with any paramagnetic nanoparticles, however, including ones with considerably smaller outer diameter. The shape of the measured frequency response depends on the type of bead being used and the scale of the measured curve depends on the quantity of beads present on the sensor surface. Thus, measuring 100 1μm Dynabeads will result in the same curve as measuring a

single $1\mu m$ Dynabead multiplied up by 100. The response at each frequency was normalized to that of the lowest frequency (1.1GHz) in order to render the measured data scale-invariant for establishing a concentration-independent standardized scale. Bead samples were prepared by washing aliquots from stock bead solution with deionized water followed by resuspending them in deionized water at a concentration of 10 mg/ml. They were deposited over the surface of the sensor by using a sharp pick and allowed to dry before taking measurements. No reference was used for these measurements as each reading was taken over a short time period during which the sensor drift was negligible. The reference techniques described earlier may be applied, however, to further improve the SNR. Fig. 2.31 [93, 94] plots the mean of three measurements for each type of bead and demonstrates that there is greater variability in response at the higher frequencies mainly due to larger phase noise leading to a lower electrical SNR for these frequencies. Of note is that the frequency response of each bead is dependent on the size and configuration of nanoparticles inside the polymer matrix, and not on the external size of the bead itself. Thus, it is not surprising that the $1\mu m$ Bangs beads have a significantly different response from the $1\mu m$ Dynabeads. Since larger magnetic domains are expected to move slower and thus crossover the null point at lower frequencies, this suggests that the Bangs Laboratories beads have the largest internal nanoparticle size, while the $1\mu m$ Dynabeads have the smallest. This is consistent with the known nanoparticle sizes: Bangs Laboratories estimates that their bead nanoparticle size is between 15 and 20 nm [7], which is considerably larger than the Dynabeads, which contain 78 nm nanoparticles [24]. The Dynabeads exhibit less than 2% standard deviation in their outside diameter [24] whereas variability in the properties of the Bangs beads is unknown. The variability in the diameter of the internal nanoparticles is also unknown, however, each $1\mu m$ bead contains hundreds of thousands to millions of nanoparticles, so minimal deviations in nanoparticle size become spatially averaged out on a per bead basis. Furthermore, two different batches of Dynabeads were used throughout the measurement process and the bead measurements were done over a time period of several days. All measurements were taken at standard room temperature conditions and no temperature control was used. Nanoparticle size variability and thermal effects are therefore both taken into account in the error bars of the measurement curves.

Multiplexed detection may be readily achieved by using beads with frequency signatures that are sufficiently orthogonal. An evaluation of the sensors multiplexing capability was performed by measuring the frequency responses of a series of

density-constant colloids containing differing ratios of $1\mu m$ Dynabeads to $1\mu m$ Bangs Laboratories beads. Seven mixtures of Dynabeads and Bangs beads prepared at a concentration of 10 mg/ml were evaluated at the following proportions: [D% : B%]: 100:0, 87.5:12.5, 75:25, 50:50, 25:75, 12:87.5, 0:100. Three measurements were taken of each sample and averaged to reduce the effects of variability in spotting of beads on the sensor cells. Fig. 2.32 [23] shows the averaged frequency signatures of all the mixtures.

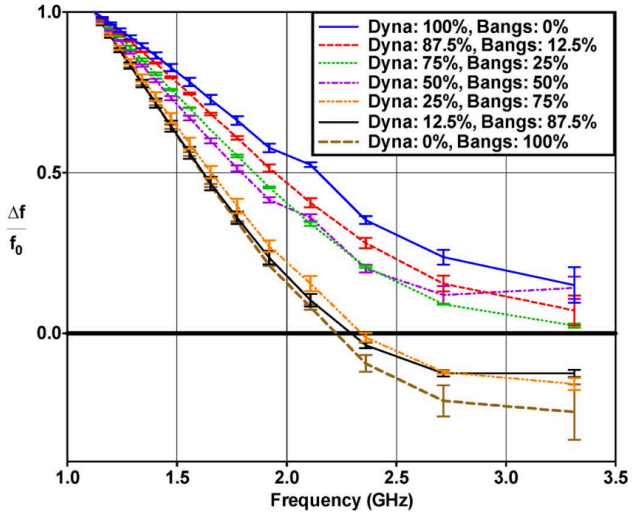

Figure 2.32: Frequency response measurements of mixtures of various proportions of Dynabeads and Bangs $1\mu m$ beads. Three measurements have been done for each mixture. Error bars represent one standard-deviation around the mean.

Using the data from Fig. 2.31 as basis vectors and taking the different standard deviations at each frequency into account, we performed an inverse-σ weighted least squares analysis to predict the proportions of each bead present in each sample and plot the results against the known exact proportions in Fig. 2.33 [93, 94].

In the absence of noise the number of different beads which can be used is equal to the rank of the V matrix. Therefore, up to 16 different bead types can theoretically be distinguished by the prototype sensor since it can take measurements at 16 distinct frequency points. However, in the presence of noise, systems which are more over-determined (number of measurements exceeds number of bead classes) tend to yield more accurate results. Intuitively, this is due to the fact that the extra degrees of freedom can be utilized to learn more information regarding the noise rather than to differentiate extra bead types. Furthermore, all bead types used in a multiplex experiment must exhibit sufficiently orthogonal frequency signatures

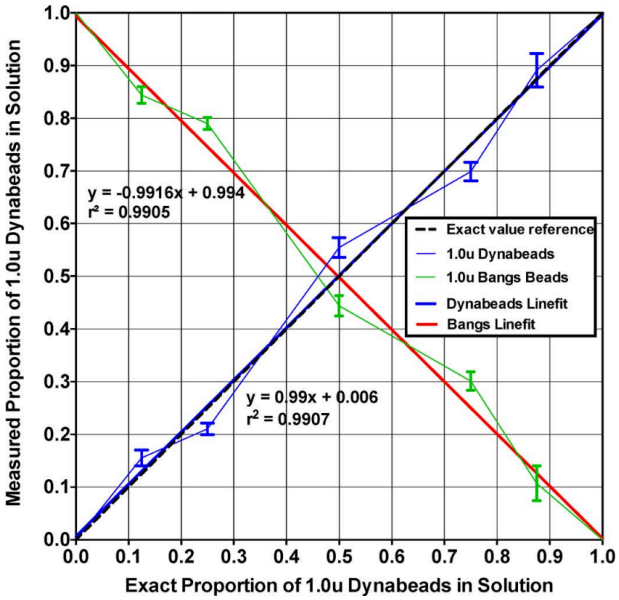

Figure 2.33: Weighted least squares analysis of measured solutions from Fig. 2.32 with errorbars. The y-axis represents the exact proportion of Dynabeads of the mixture of Dynal and Bangs beads and the x-axis represents the best estimate of a simple weighted least squares algorithm based on the measurement data. The weights used are the inverse of the variance of the control measurements of pure Dynabeads and pure Bangs solutions. This demonstrates the viability of the sensor for multiplexed biosensing.

in order to minimize reconstruction error. In an ideal, noiseless system, the only criterion necessary for successful reconstruction is that the V matrix which contains the frequency signatures of each of the two beads is full column rank. However, the presence of noise, the rank of V alone may not be enough to guarantee an acceptable reconstruction. In this case, the matrix condition number provides a good indicator of how close the matrix is to being singular, with larger condition numbers indicating higher likelihood of larger reconstruction error. The condition number of the V matrix in our experiment was computed to be 5.28 implying that the relative error in the bead reconstruction was bound at most by roughly a factor of 5 times the error in the measurements of the sample solutions. Finally, a chip die photo is shown in Fig. 2.34 [93, 94] next to a USB powered handheld PCB test bed for performing in-vitro bio-experiments. The results clearly demonstrate the sensor's ability to concurrently detect different magnetic beads, indicating the viability of the sensor for complicated bioassays requiring multiple marker types.

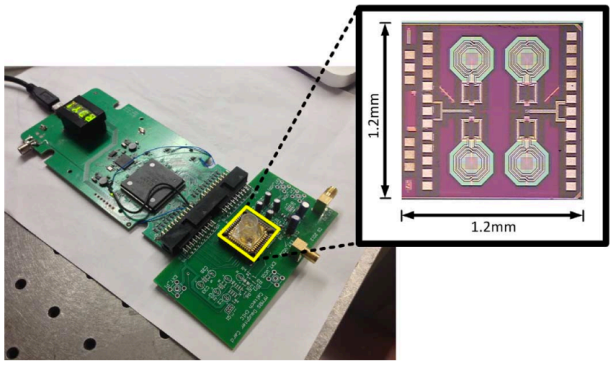

Figure 2.34: Chip die photo of the full 2 x 2 magnetic spectroscopy system with prototype USB powered handheld testbed system.

Conclusion

The frequency-shift based magnetic detection methodology was reviewed and extended to magnetic spectroscopy in order to enable multiplex experiments with magnetic beads. The technique is fully compatible with the standard CMOS processes of today and is capable of achieving high sensitivity while maintaining ultra-portability and low power consumption. Furthermore, important design parameters that should be optimized when designing frequency-shift sensors such as the electrical SNR and transducer gain uniformity were analyzed. Finally, a magnetic spectrometer prototype was implemented in a 65 nm CMOS process as a proof of-concept. Each sensing sites occupies 260 μm x 260 μm (with 150 μm x 150 μm active sensing area), a surface area sufficient for protein or DNA detection experiments w ith detectable bead binding capacity. The full size of the four sensor chip is 1.2 x 1.2mm and the whole system can be powered from a laptop USB port to demonstrate portability of the design. Previous work in magnetic multiplexing [48, 82, 79] significantly bridged the gap between optical and magnetic sensing technologies. Our work further narrows this gap by utilizing a standard CMOS process, consuming ultra-low power, and not requiring any extra mechanical parts such as moving rare-earth magnets.

2.5 Next Generation Design: A Drift Cancellation Approach

In the previous sections of this chapter, we have developed a magnetic biosensing methodology capable of quantifying magnetic nanoparticles with uniform transducer gain and have extended it to a multiplex-capable system by utilizing magnetic spectroscopy.

One major remaining issue is longterm frequency-drift of the sensor due to electrical and thermal changes. The resonant frequency of the sensor LC tank depends on both the values of the tank inductance and capacitance. As described earlier, magnetic material within the vicinity of the inductor's magnetic field causes a shift in the effective inductance which results in a change in the resonant frequency.

Under the base assumption that the capacitance remains constant and the inductance varies only due to magnetic nanoparticles on its surface, simply measuring the resonant frequency of the tank (or similarly the oscillation frequency of the tank embedded in an LC oscillator configuration) allows direct quantification of the magnetic material. Unfortunately, the tank reactances may vary also due to thermal and electrical fluctuations causing corresponding shifts or drifts in the resonant/oscillation frequency.

Many realistic biological experiments occur over time periods of hours and thus these electrical/thermal frequency instabilities present a series issue to the sensor SNR. In this section, we develop a solution to these issues in the form of a new sensor design with a frequency drift cancellation loop.

A Fourth-Order Transformer Based Tank

We analyze the following circuit:

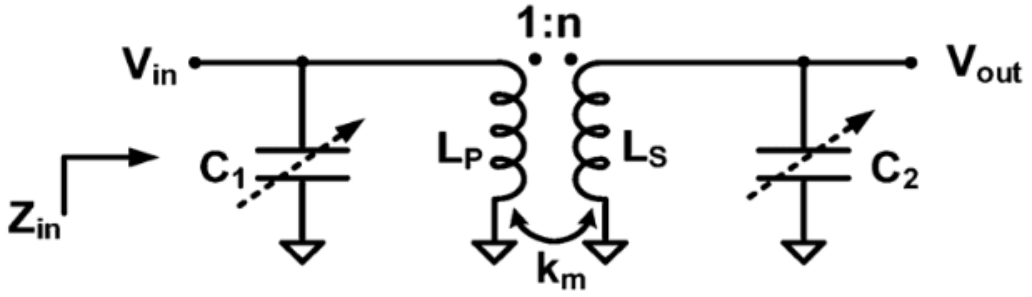

Figure 2.35: Fourth order transformer-based LC Tank

Assuming ideal reactive components, the system in Fig. 2.35 has two possible resonant frequencies [4]:

$$\omega_{1,2}^2 = \frac{1 + \left(\frac{L_s C_2}{L_p C_1}\right) \pm \sqrt{1 + \left(\frac{L_s C_2}{L_p C_1}\right)^2 + \left(\frac{L_s C_2}{L_p C_1}\right)(4k_m^2 - 2)}}{2L_s C_2(1 - k_m^2)}$$

Our first remark is that the ratio $\frac{\omega_1^2}{\omega_2^2}$ depends only on the ratios $\frac{L_s}{L_p}$ and $\frac{C_2}{C_1}$. Thus, if the tank is layed out in such a way where process variation or thermal changes affect L_s / L_p and C_1 / C_2 in the same way, such external fluctuations will not have much of an effect on the ratio $\frac{\omega_1^2}{\omega_2^2}$.

Furthermore, if we design the tank such that $L_sC_2 = L_pC_1$, the expression for the resonant frequencies simplifies to:

$$\omega_{1,2}^2 = \frac{1 \pm k_m}{L_sC_2(1-k_m)(1+k_m)}$$

and thus: $\frac{\omega_1^2}{\omega_2^2} = \frac{1+k_m}{1-k_m}$.

Now, we examine what happens to first order if only one of the capacitances (C_2) varies while all the other components remain the same (again with $L_s = L_p = L$ and $C_1 = C_2 = C$):

$$\begin{aligned}
\omega_{1,2\Delta}^2 &= \frac{1 + \left(\frac{C+\Delta C}{C}\right) \pm \sqrt{1 + \left(\frac{C+\Delta C}{C}\right)^2 + \left(\frac{C+\Delta C}{C}\right)(4k_m^2 - 2)}}{2L_sC(1-k_m^2)} \\
&\approx \frac{2 + \frac{\Delta C}{C} \pm \sqrt{2 + 2\frac{\Delta C}{C} + (1 + \frac{\Delta C}{C})(4k_m^2 - 2)}}{2L_sC(1-k_m^2)} \\
&= \frac{2 + \frac{\Delta C}{C} \pm 2k_m\sqrt{1 + \frac{\Delta C}{C}}}{2L_sC(1-k_m^2)} \\
&= \frac{1 + \frac{\Delta C}{2C} \pm k_m(1 + \frac{\Delta C}{2C})}{L_sC(1-k_m^2)} \\
&= \frac{(1 + \frac{\Delta C}{2C})(1 \pm k_m)}{L_sC(1-k_m)(1+k_m)}
\end{aligned}$$

And thus, the ratio:

$$\frac{\omega_{1\Delta}^2}{\omega_{2\Delta}^2} = \frac{1+k_m}{1-k_m}$$

remains exactly the same as before (to first order), even though only one of the two capacitances was perturbed. This implies that even if the parasitics do not drift completely in the same way, the ratio of the two frequencies still remains constant to first order.

The next key to the puzzle is the fact that the effect of the magnetic particles on the inductance of the sensor is frequency dependent, as explained in the previous section on multiplexing and magnetic spectroscopy. If the operating frequency is high enough, in fact, the magnetic beads appear transparent and do not modify the effective inductance of the sensor. The optimal operating point for maximizing signal is $\omega_1 = 2\pi$ 1GHz and the second operating frequency should ideally be above 6 GHz for obtaining electrical drift information without being affected by the magnetic material present.

For $\frac{\omega_1}{\omega_2} = 6$, we would need k_m to be: $\frac{1+k_m}{1-k_m} = 36$, or $k_m = 35/37 \approx 0.945$ which is an unrealistically large coupling factor for planar on-chip transformers.

Assuming a maximum realistic coupling factor $k_m = 0.8$ can be achieved, this would force $\omega_2 = 3\omega_1$ or 3GHz if the lower frequency is set to 1GHz. Looking at our measured magnetic bead data from the magnetic spectrometer chip, we observe that in most cases the response of the beads at 3GHz is in the opposite direction, i.e. the effective inductance decreases instead of increasing. This actually works to our advantage and serves to increase the effective signal when the two measured frequencies are divided while still canceling out (at least to first order) electrical/thermal drift effects.

2.6 Concluding Remarks

Biosensing is a crucial and rapidly expanding research field. New advancements in detection technology and electronics has allowed the miniaturization and significant cost reduction of detection modalities.

In this chapter, we have motivated and demonstrated several versions of a biosensing device capable of quantifying magnetic nanoparticles which can be conjugated onto a variety of different biomolecules for developing biosensing assays. Every version of the magnetic biosensor is developed in a standard CMOS process without any post-process modifications required besides surface chemistry functionalization for biological experiments.

We have studied the spatial, static magnetic fields over the sensing inductor's surface to build insight and design a new inductor with modified shape and fields which are 200 times more uniform than the original, spiral based design. Furthermore, we have studied the temporal dependence of the polarization of magnetic nanoparticles (due to their finite size magnetic domains) and have taken advantage of the effect in order to design and demonstrate a multiplex-capable magnetic sensing system able

to differentiate among 16 different types of magnetic labels concurrently present on the transducer active sensing area.

Finally, we have analyzed a new approach for solving the sensor frequency drift issue due to electrical noise and thermal fluctuations. We are currently in the process of designing the latest iteration of the magnetic biosensor with the new frequency drift cancellation approach. Once implemented and verified, this approach will significantly improve the longterm frequency stability of the sensor by several orders of magnitude, making longer-run biological experiments viable.

In the previous sections, we developed and discussed the concept of magnetic biosensing and how judicious manipulation of the magnetic fields in the vicinity of a sensor can lead to significant improvements in sensitivity as well as functionality. In the following sections of this thesis, we will continue exploring the potential unlocked by "designing" electromagnetic field distributions in the RF, millimeter-wave, and even photonics frequency regimes.

3.1 Maximum Performance Bounds: Introduction

Electromagnetic structure design has been a very important area of research over the past century. Analytical solutions to Maxwell's equations, with the exception of simple examples [6, 87, 103], often do not exist making designing structures to meet given target specifications a very difficult problem. As an example, experienced antenna designers usually begin with a canonical design such as a dipole or loop antenna and make adjustments or tweaks to the structure in an effort to meet desired performance metrics. While these "tweaks" are often guided by experience and understanding of the large-scale effects on the electromagnetic fields, they are ultimately heuristic based efforts and no generalized approach for designing antennas in arbitrary environments exists.

In fact, even designing electromagnetic structures computationally is usually extremely challenging due to the fact that most relevant design optimization problems are NP-hard with complex solution spaces that admit many local maxima. Thus, for the structure design problem, a good heuristic algorithm must be employed which can hopefully converge to an acceptable solution. [81, 85] We will show, however, that if various constraints of the optimization problem are relaxed, a global optimum can be readily found for many different systems which represents a maximum performance bound for the original system. While the relaxed system may not lead to a feasible solution of the original system, it can provide considerable design insight such as maximum achievable performance as well as possibly an initial seed solution for a heuristic solver of the constrained (realistic) system.

The examples presented will be in the context of antennas, however, it is important to note that the framework developed is general and can be applied to numerous areas in electromagnetics. To motivate the derivation of the relaxation and the bound, we will begin with a typical metal antenna structure design problem:

Suppose we wish to design a planar, metal antenna structure on a dielectric substrate (such as a printed circuit board or silicon IC). We must define a design region over the substrate inside which the antenna will be fabricated as well as a minimum feature size. The minimum feature size is typically determined by the design rules of the fabrication process or by the total amount of computational power/time available to the designer. This is due to the fact that a finer feature size will result in a finer design grid within the constrained design region which directly translates to a larger amount of unknowns that must be determined. Without loss of generality, we also assume that the design grid is square and size NxN. Each square can either be left empty or filled with a solid patch of metal (which at RF frequencies is well approximated by a Perfect Electrical Conductor [PEC] boundary condition.) (Fig. 3.1)

The center point of the grid will be forced empty as this will be the feed-point for the antenna and will have a fixed current value across it in order to excite the designed metal structure surrounding it. The design objective will be set as maximizing the gain in the far-field for $\theta = 0, \phi = 0$ or in the +z direction. It is important to note that specific conditions are being presented in order to arrive at a well defined design problem– the method presented is general and can deal with any number of antenna feed points or optimization objectives (such as maximizing overall radiation efficiency instead of directional gain, for example.)

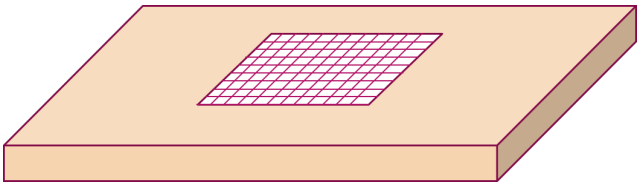

Figure 3.1: A representative dielectric substrate with antenna design grid indicated on its surface

Before writing down the description of the problem in mathematical terms, it is instructive to observe that there are $N^2 - 1$ design unknowns each which can take a binary value of empty or metal. Due to the binary or 0-1 nature of the problem, it is known to be an NP-hard problem [75] which implies that there are no known polynomial time algorithms for arriving at the globally optimum solution and in order to determine the global optimum, every single feasible solution in the design space must be evaluated. In this design example, there are 2^{N^2-1} possible designs which can be implemented. Even for a small N, such as $N = 10$, which would lead to a rather modestly sized antenna problem, this results in $2^{99} = 6.3x10^{29}$

possible designs. Assuming it takes 5 minutes for a powerful CPU to complete an electromagnetic simulation [39] of a single design variation, it would take $2^{99} * 5 = 6x10^{24}$ years to evaluate the whole solution space, which is clearly not a feasible approach! In a later chapter, we will develop heuristics for attacking the design problem directly, but as stated, we will first develop a relaxation in the constraints which makes the global solution of the new (relaxed) problem easily solvable.

Due to the well known current-equivalence theorems for Maxwell's equations [88, 84], given fixed excitation/source conditions, any dielectric (or metal) material can be replaced with an equivalent current distribution which generates the same electric (E) and magnetic (H) fields everywhere in space. Furthermore, a boundary condition enforcing zero tangential E-field is equivalent to a PEC or perfect metal material. Utilizing these two insights, we can write the design problem as a quadratically constrained quadratic programming (QCQP) problem in terms of surface currents as the unknowns:

$$\max_{J(x)} \frac{U_{(0,0)}(J(x))}{P_{in}(J(x))}$$

subject to:

$$J(feed) = 1$$
$$J^*(x) * E(x) = 0 \quad \forall x \in A$$

where:

$J(x)$ is the surface current distribution to be designed over the antenna design region A.

$U_{0,0}(J(x))$ is a function which maps the surface current distribution $J(x)$ over the design region A to the radiation intensity in the far-field at $\theta = 0, \phi = 0$ by way of Maxwell's equations.

$P_{in}(J(x))$ computes the total input power from the surface current distribution $J(x)$ over the region A into the system. Note that because of the PEC enforcing constraint ($J^*(x) * E(x) = 0$) that only the feed source current element can contribute nonzero input power since any other surface region with nonzero current must correspondingly have zero tangential electric field. Thus, for this design problem $P_{in}(J(x))$ is equivalent to the incident power from the feed source into the antenna.

$J(feed) = 1$ defines an excitation condition for the antenna.

$J^*(x) * E(x) = 0$ which must be satisfied over the whole design surface A ensures that any nonzero area of current $J(x)$ is on a PEC (metal) boundary condition.

To show that the design problem that we defined above is actually a QCQP, we must elaborate on the specific expressions for the radiation intensity ($U_{(\theta,\phi)}$) and input power (P_{in}).

Radiation Intensity of a current distribution is a quadratic form

From [103] we have the following expression for the Radiation Intensity (U) of a current distribution:

$$
\begin{aligned}
U(\theta,\phi) &= \frac{k^2}{32\pi^2\eta_0}\left(\left[L_\phi + \eta_0 N_\theta\right]^2 + \left[L_\theta - \eta_0 N_\phi\right]^2\right) \\
N_\theta &= \iint_S \left(J_x cos\theta cos\phi + J_y cos\theta sin\phi - J_z sin\theta\right) e^{+jkr' cos\psi}\, ds' \\
N_\phi &= \iint_S \left(-J_x sin\phi + J_y cos\phi\right) e^{+jkr' cos\psi}\, ds' \\
L_\theta &= \iint_S \left(M_x cos\theta cos\phi + M_y cos\theta sin\phi - M_z sin\theta\right) e^{+jkr' cos\psi}\, ds' \\
L_\phi &= \iint_S \left(-M_x sin\phi + M_y cos\phi\right) e^{+jkr' cos\psi}\, ds'
\end{aligned}
$$

WLOG we assume that the magnetic current terms (M_x, M_y, and M_z are 0), which makes $L_\theta = 0$ and Ł$_\phi = 0$. Further, it can be seen that N_θ and N_ϕ are both linear in the driving terms J_x, J_y, and J_z, so we assume $J_x = 0$ and $J_y = 0$ making J_z the only non-zero term for the purposes of cleanliness of the proof. This makes $N_\phi = 0$ and leaves the expression for N_θ as:

$$
N_\theta = \iint_S \left(-J_z sin\theta\right) e^{+jkr' cos\psi}\, ds'
$$

Substituting into the expression for the radiation intensity, $U(\theta,\phi)$, we have:

$$U(\theta,\phi) = \frac{k^2\eta_0}{32\pi^2}\overline{\left(\iint_S (-J_z sin\theta)\, e^{+jkr'cos\psi} ds'\right)}\left(\iint_S (-J_z sin\theta)\, e^{+jkr'cos\psi} ds'\right)$$

$$= \frac{k^2\eta_0}{32\pi^2}\left(\iint_S \iint_S \left(J_z^*(ds) J_z(ds') sin^2\theta\right) e^{-jkrcos\psi} e^{+jkr'cos\psi} ds ds'\right)$$

$$= \left(\iint_S \iint_S J_z^*(ds) K(ds, ds') J_z(ds') ds ds'\right)$$

$$\text{with} \quad K(ds, ds') = \frac{k^2\eta_0 sin^2\theta}{32\pi^2} e^{-jkrcos\psi} e^{+jkr'cos\psi}$$

Incident/Input Power is a quadratic form

It is well known that the total power in a system with only electric sources (J) can be computed as a volume integral over the region containing non-zero J of $J^* \cdot E$, where E is the collocated electric field:

$$P_{in} = real\left(\iiint_V J^* \cdot E dV\right)$$

$$= real\left(\iint_S J^* \cdot E dS\right) \quad \text{for a surface current distribution}$$

Due to the linearity of Maxwell's equations, there exist Green's functions for any system mapping the electric (J) and magnetic (M) currents to the electric (E) and magnetic (H) fields:

$$E(r) = \iiint_V \left(J(r')G^{EE}(r,r') + M(r')G^{EH}(r,r')\right) dV$$

$$H(r) = \iiint_V \left(J(r')G^{HE}(r,r') + M(r')G^{HH}(r,r')\right) dV$$

In the expression above for P_{in}, we assume that we only have electric currents that are nonzero and the magnetic field is not required, so we have:

$$E(r) = \iiint_V J(r')G^{EE}(r,r')dV$$

and substituting into P_{in}:

$$\begin{aligned} P_{in} &= real\left(\iint_S \iint_S J^*(r)J(r')G^{EE}(r,r')dSdS'\right) \\ &= \iint_S \iint_S J^*(r)J(r')real[G^{EE}(r,r')]dSdS' \\ &= \iint_S \iint_S J^*(r)G(r,r')J(r')dSdS' \\ &\text{with} \quad G(r,r') = real[G^{EE}(r,r')] \end{aligned}$$

Now we can rewrite the optimization problem, substituting for the radiation intensity and incident power:

$$\max_{J(x)} \iint_S \iint_S J^*(r)K(r,r')J(r')dsds'$$

subject to:

$$\iint_S \iint_S J^*(r)G(r,r')J(r')dsds' = 1$$

$$J^*(x) * \iint_S J(x')G^{EE}(x,x')dV = 0 \quad \forall x \in S$$

$$J(feed) = 1 \quad \text{(excitation source condition)}$$

If we could solve this optimization problem for the globally optimal $J(x)$ distribution over the design surface S, we would equivalently have the metal antenna structure which maximizes the gain in the upwards ($\theta = 0, \phi = 0$) direction. Unfortunately,

the $J^*(x) * \iint_S J(x')G^{EE}(x,x')dV = 0 \quad \forall x \in S$ constraint is a "0 - 1" type of constraint and leads to the problem being NP-hard [75]. In the following section, we will show that if we ignore this constraint, the globally optimum solution to the resulting problem can be found. Since the solution space of the problem without the constraint is a **superset** of the original problem, the resulting solution will always be a **maximum** performance bound for the original design problem. The constraint is important in the sense that it enforces that any region of the surface with non-zero electric current flowing must have 0 tangential electric field and therefore be metallic. Without the constraint, we allow for arbitrary current distributions to be designed. Due to the current equivalence of Maxwell's equations, just as any dielectric can be represented by equivalent current source distributions, currents can be replaced with dielectrics which will scatter in such a way that the resulting electromagnetic field distribution is identical. Theoretically, if we could design antennas containing arbitrary dielectrics (as opposed to just metals, for instance) then we could realize the solution found from the relaxed problem. Unfortunately, most real dielectrics are lossy and the dielectric equivalent formulation may require dielectrics with gain rather than loss depending on relative signs of the currents and their collocated electric fields. Nonetheless, the optimal solution found for the relaxed problem may serve as a guide for the real design problem both as a sense of maximum performance that can be expected as well as providing insight as to the structure or current distribution that may be required in order to approach that performance bound.

Now we derive the solution to the optimization problem without the troublesome "0 - 1" constraint.

3.2 Statement of the Problem

Most generally, we pose an integral equation optimization problem:

$$\min / \max_{J(x)} \int_V \int_V J^*(x)K(x,x')J(x')dxdx'$$

subject to:

$$\int_V \int_V J^*(x)G(x,x')J(x')dxdx' = 1$$

where K must have the property: $K(x,x') = K^*(x',x)$ and the function $J(x)$ is the unknown and must be determined over some surface S or volume V.

3.3 Derivation

We begin by forming the Lagrangian and taking derivative with respect to λ and the functional $J(x)$:

$$\mathcal{L}(J(x),\lambda) = \int_V \int_V J^*(x)K(x,x')J(x')dxdx' - \lambda\left(1 - \int_V \int_V J^*(x)G(x,x')J(x')dxdx'\right)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 = 1 - \int_V \int_V J^*(x)G(x,x')J(x')dxdx'$$

$$\frac{\partial \mathcal{L}}{\partial J(x)} = \int_V K(x,x')J(x')dx' - \lambda \int_V G(x,x')J(x')dx' = 0$$

3.4 Discretization / Numerical Solution

The resulting expression is a continuous eigenvalue expression similar to a Fredholm integral equation of the second kind. Except in very special circumstances (such as when the limits of the integral extend to infinity and a Fourier transform can be taken), closed form solutions do not exist and a numerical approach must be taken for arriving at a solution. The two most common methods for discretizing and solving the problem are the method of moments and the Nystrom method. In this work, we focus on the method of moments as it is typically more suitable for design problems defined on a fixed grid or mesh.

Method of Moments

In this section we briefly review the method of moments. We can approximate the unknown current function J(x) as an expansion of known basis functions:

$$J(x) = \sum_{k=0}^{N-1} \alpha_k \phi_k(x)$$

substituting into our eigenvalue integral equation:

$$\int_V K(x,x') \sum_{k=0}^{N-1} \alpha_k \phi_k(x')dx' = \lambda \int_V G(x,x') \sum_{k=0}^{N-1} \alpha_k \phi_k(x')dx'$$

and switching order of integration and summation (which is allowed in all cases because the summation is over a finite number of terms and each of the individual integrals of the basis functions against the kernels K and G are assumed to be finite):

$$\sum_{k=0}^{N-1} \alpha_k \int_V K(x,x')\phi_k(x')dx' = \lambda \sum_{k=0}^{N-1} \alpha_k \int_V G(x,x')\phi_k(x')dx'$$

This expression is still continuous in the x variable, so we must test with a set of weighting functions, $\psi_m(x)$, in order to arrive at a linear system of N equations with N unknowns:

$$\int_V \psi_m(x) \sum_{k=0}^{N-1} \alpha_k \int_V K(x,x')\phi_k(x')dx'dx = \lambda \int_V \psi_m(x) \sum_{k=0}^{N-1} \alpha_k \int_V G(x,x')\phi_k(x')dx'dx$$

(Note: The weighted integral approach against a "testing" function is named as such because the solution is being evaluated N times in order to arrive at N independent equations for solving for the α coefficients.)

Again we switch the order of summation and integration:

$$\sum_{k=0}^{N-1} \alpha_k \int_V \int_V \psi_m(x)K(x,x')\phi_k(x')dx'dx = \lambda \sum_{k=0}^{N-1} \alpha_k \int_V \int_V \psi_m(x)G(x,x')\phi_k(x')dx'dx$$

for m=0,...,N-1, which results in N equations.

We can clean this up further by defining matrices A and B as follows:

$$A_{m,k} = \int_V \int_V \psi_m(x)K(x,x')\phi_k(x')dx'dx$$
$$B_{m,k} = \int_V \int_V \psi_m(x)G(x,x')\phi_k(x')dx'dx$$

and the problem in more concise can equivalently be written as:

$$A\alpha = \lambda B\alpha$$

which is known in mathematics as a generalized eigenvalue problem.

This can be solved using any standard numerical method. The smallest eigenvalue, λ_{min} represents the minimum of the initial objective function with its corresponding eigenfunction α_{min} representing the coefficients for the basis expansion approximation of the optimal current distribution, $J_{min}(x)$, which achieves the global minimum

of the problem. Similarly, the largest eigenvalue, λ_{max} represents the maximum of the objective with α_{max} being the coefficients of the basis expansion approximation of the globally maximizing current distribution, $J_{max}(x)$. The proof that the extremal eigenvalues and their respective eigenvectors correspond to the optimal minimizing and maximizing solutions of the QCQP presented is beyond the scope of this work; however, it is readily available in the literature about generalized Rayleigh quotients. [8]

The testing functions, $\psi_m(x)$ are usually picked to be the same as the basis expansion functions $\phi_k(x)$, which results in minimizing the error of the approximation to the continuous solution in a least-squares sense and is known in the mathematical literature as the Galerkin method. [46]

In the most general case, the A and B matrices can be computed by performing a numerical electromagnetic simulations for each basis function utilized. i.e. N basis functions would require N separate simulations to gather the data required for computing A and B. Fortunately, each of these simulations is independent from the rest and therefore this step can be **completely** parallelized without any communication. Recently large virtual compute clusters (such as Amazon Elastic Compute Cloud [38] and Google) are available for affordable hourly rates which allow the instantiation of thousands of compute cores simulataneously. Since this computation step can be completely parallelized, the computation time of the A and B matrices can be linearly reduced with increasing number of CPU cores.

In many cases, there are considerably faster approaches to computing the extremal eigenvalues λ_{max} and λ_{min} and their corresponding eigenvectors. For example, in the situation where $G(x, x') = \delta(x - x')$, B reduces to the identity matrix and the generalized eigenvalue problem becomes a standard eigenvalue problem: $A\alpha = \lambda * \alpha$. The well known Power Iteration method can be used in this situation for solving for λ_{max} by starting with a random α vector and applying the A matrix operator to it (matrix-vector multiplication) until it converges to the maximal eigenvector α_{max}. At this point, $\lambda_{max} = \frac{\alpha_{max}^H A \alpha_{max}}{\alpha_{max}^H \alpha_{max}}$. Computation of Ax usually does not require the A matrix to be fully computed. Indeed, Ax for arbitrary vector x usually involves a single electromagnetic simulation step since the x vector corresponds to excitation current sources. Thus, if the Power Iteration method converges to a satisfactory approximation of λ_{max} in fewer than N steps, where N is the number of basis expansion functions being used, then it can potentially be significantly faster than fully computing A and solving the eigenvalue problem with a standard approach.

Often the A matrix turns out to be extremely low rank compared to its size. For example, if the K and G kernels are chosen such that the optimization is maximizing the gain of an antenna at a given far-field position (θ, ϕ), the rank of A is at most 2.

In the next section, we will touch on linear current distributions in free-space. Following this, we will explore linear current distributions on lossy dielectric substrates as well as arbitrary current distributions on lossy dielectric substrates. The utility of this bound and how closely it can be approached with real antenna designs will also be explored.

3.5 Linear Current Distributions

In this section we will explore linear current distributions and identifying optimal distributions for maximizing various different objective functions.

Maximum Directivity of Linear Current Distribution

It has been well known since 1948 [105, 58, 14] that in the absence of loss or other constraints, a linear current distribution can be designed which yields arbitrarily large directivity in the far-field. This is a phenomenon known as superdirectivity. For realistic antennas, superdirective current distributions are almost never practical due to the huge opposing currents required which require extremely accurate phase control and result in unacceptably low radiation efficiencies due to the finite loss of metals and I^2R resistive losses. Since the directivity rapidly approaches infinity as the order of basis functions is increased (as expected), the problem is poorly conditioned and it is difficult to perform an accurate numerical analysis. Instead, we briefly touch on superdirective line current distributions working with the analytical expressions. In the next section, we will present a brief sketch of a proof that an arbitrary length line current distribution can achieve infinite directivity in the far-field. Although this has been known for over half a century [14], most of the proofs that exist in the literature are considerably more complex.

Proof: Superdirectivity of a Linear Wire Distribution

We assume an arbitrary line current distribution $J_z(z)$ on the Z-axis which can be nonzero from $-z_0$ to z_0. (Fig. 3.2)

Due to judicious alignment of the current distribution along the z-axis, the far-field pattern of the antenna is invariant of ϕ and only depends on θ. Thus, the far-field RCS for the distribution is given exactly by [6]:

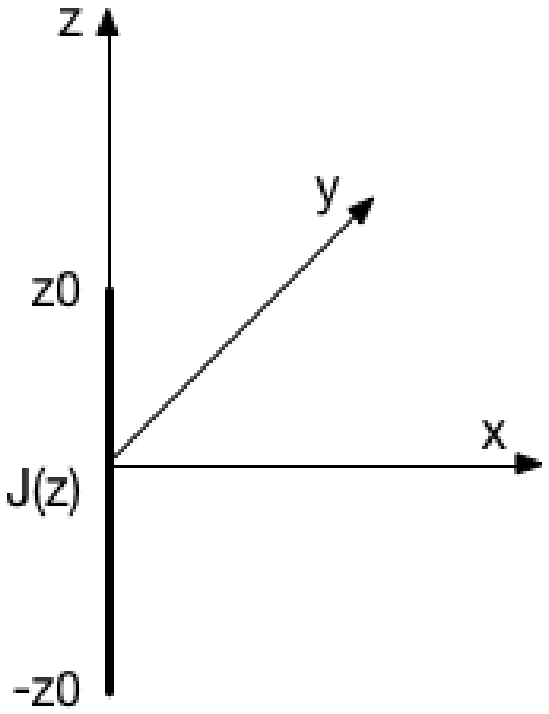

Figure 3.2: Line current distribution $J_z(z)$ from $-z_0$ to z_0 centered on the z-axis

$$RCS(\theta) = \frac{\eta_0 k^2}{32\pi^2}\left\| \int_{-z_0}^{z_0} J_z(z) sin(\theta) exp(jkzcos(\theta)) dz \right\|^2$$

To calculate all of the power radiated into the far-field, we must integrate across the surface of the far-field sphere:

$$P_{rad} = 2\pi \int_0^{\pi} RCS(\theta) sin(\theta) d\theta = 2\pi \frac{\eta_0 k^2}{32\pi^2} \int_0^{\pi} \left\| \int_{-z_0}^{z_0} J_z(z) sin(\theta) exp(jkzcos(\theta)) dz \right\|^2 sin(\theta) d\theta$$

Factoring out the $sin(\theta)$ term from the inner integral:

$$P_{rad} = 2\pi \frac{\eta_0 k^2}{32\pi^2} \int_0^{\pi} \left\| \int_{-z_0}^{z_0} J_z(z) exp(jkzcos(\theta)) dz \right\|^2 sin^3(\theta) d\theta$$

We are interested in maximizing the gain (or directivity since the system is lossless) for θ = 90 degrees which is defined as:

$$D(\pi) = \frac{4\pi RCS(\pi)}{Prad}$$

Writing out the expression explicitly:

$$D(\pi) = \frac{2\left\|\int_{-z_0}^{z_0} J_z(z)dz\right\|^2}{\int_0^{\pi}\left\|\int_{-z_0}^{z_0} J_z(z)exp(jkzcos(\theta))dz\right\|^2 sin^3(\theta)d\theta}$$

The $exp(jkzcos(\theta))$ term is troublesome because it does not allow for easy separation of the z and θ variables. However, we consider the case where z_0 is very small (i.e. the antenna is electrically small). The maximum achievable gain of an electrically small antenna intuitively is a lower bound for that of a longer distribution since the longer current distribution (larger z_0) is a superset of the solution space of the small one.

$cos(\theta) \leq 1$, so for $kz_0 << 1$, we can approximate $exp(jkzcos(\theta)) \approx 1 + jkzcos(\theta)$.

Let's now work with the denominator of $D(\pi)$:

$$\begin{aligned}
\int_0^{\pi}\left\|\int_{-z_0}^{z_0} J_z(z)exp(jkzcos(\theta))dz\right\|^2 sin^3(\theta)d\theta &\approx \int_0^{\pi}\left\|\int_{-z_0}^{z_0} J_z(z)(1+jkzcos(\theta))dz\right\|^2 sin^3(\theta)d\theta \\
&= \int_0^{\pi}\left\|\int_{-z_0}^{z_0} J_z(z)dz + jkcos(\theta)\int_{-z_0}^{z_0} J_z(z)zdz\right\|^2 sin^3(\theta)d\theta \\
&= \int_0^{\pi}\left(\left\|\int_{-z_0}^{z_0} J_z(z)dz\right\|^2 - k^2cos^2(\theta)\left\|\int_{-z_0}^{z_0} J_z(z)zdz\right\|^2\right)sin^3(\theta)d\theta \\
&\quad + \int_0^{\pi}\left(2kcos(\theta)Re\left(j\int_{-z_0}^{z_0}\int_{-z_0}^{z_0} J_z^*(z')J_z(z)zdzdz'\right)\right)sin^3(\theta)d\theta \quad (3.1)
\end{aligned}$$

$\cos(\theta)sin^3(\theta)$ integrates to 0 from 0 to π, so the cross-term falls out and we are left with:

$$\int_0^\pi \left(\left\| \int_{-z_0}^{z_0} J_z(z)dz \right\|^2 - k^2 cos^2(\theta) \left\| \int_{-z_0}^{z_0} J_z(z) z dz \right\|^2 \right) sin^3(\theta) d\theta$$

$$= \int_0^\pi sin^3(\theta)d\theta \left\| \int_{-z_0}^{z_0} J_z(z)dz \right\|^2 - k^2 \int_0^\pi sin^3(\theta)cos^2(\theta)d\theta \left\| \int_{-z_0}^{z_0} J_z(z) z dz \right\|^2$$

$$= \frac{4}{3} \left\| \int_{-z_0}^{z_0} J_z(z)dz \right\|^2 - k^2 \frac{4}{15} \left\| \int_{-z_0}^{z_0} J_z(z) z dz \right\|^2 \quad (3.2)$$

We can now rewrite our expression for directivity as:

$$D(\pi) = \frac{2 \left\| \int_{-z_0}^{z_0} J_z(z)dz \right\|^2}{\frac{4}{3} \left\| \int_{-z_0}^{z_0} J_z(z)dz \right\|^2 - k^2 \frac{4}{15} \left\| \int_{-z_0}^{z_0} J_z(z) z dz \right\|^2}$$

We first note that if the function $J_z(z)$ is purely even, then $\int_{-z_0}^{z_0} J_z(z) z dz = 0$ and

$$D(\pi) = \frac{3}{2} = 1.5$$

which agrees with the known directivity for a Hertzian dipole antenna which is assumed to have a uniform current distribution. It is interesting to note that as long as $J_z(z)$ is symmetric across $z = 0$ that the directivity of the resulting antenna in the $(0, 0)$ solid angle will approach 1.5 as the the length approaches 0, regardless of the actual distribution. On the other hand, if $J_z(z)$ is odd, then $\int_{-z_0}^{z_0} J_z(z)dz = 0$ and thus:

$$D(\pi) = 0$$

Let's now assume $J_z(z) = a + bz$, which has both an odd and even component:

$$\int_{-z_0}^{z_0} J_z(z)dz = \int_{-z_0}^{z_0} a + bz dz = 2az_0$$

$$\int_{-z_0}^{z_0} J_z(z) z dz = \int_{-z_0}^{z_0} az + bz^2 dz = \frac{2bz_0^3}{3}$$

We now see that if we judiciously choose the coefficients a and b, we can make the denominator of $D(\pi)$ go to 0 while maintaining a finite, nonzero numerator:

$$\frac{4}{3}(2az_0)^2 - k^2\frac{4}{15}\left(\frac{2bz_0^3}{3}\right)^2 = 0$$

$$\frac{4}{3}(4a^2z_0^2) - k^2\frac{4}{15}(\frac{4b^2z_0^6}{9}) = 0$$

$$b = \frac{3\sqrt{(}5)a}{kz_0^2}$$

Thus, if we let $J_z(z) = a(1 + \frac{3\sqrt{(}5)}{kz_0^2}z)$, the denominator of $D(\pi)$ is 0 while the numerator is nonzero, making $D(\pi)$ tend to infinity as z_0 gets smaller. Note that the magnitude of the odd term of $J_z(z)$ increases with the square of z_0 as the length of the antenna is made smaller. This implies that while indeed there is no theoretical bound for the directivity of a linear current antenna of any length, the currents must be made larger and of opposing phase to achieve superdirectivity. In practice, this highly impractical due to ohmic (resistive) wire losses and difficulty and controlling the precision of the phases of adjacent currents.

In the next section, we introduce a resistive loss term and compute a theoretical bound for the maximum radiation efficiency and gain achievable for such a realistic antenna.

Maximum Radiation Efficiency of Lossy Linear Current Distribution

As before, we begin with an arbitrary line current distribution $J_z(z)$ on the Z-axis which can be nonzero from $-z_0$ to z_0. Ohmic loss is proportional to $I^2 * R$, so given a loss resistivity R (Ohm/m), the total power burned due to resistive losses can be computed:

$$P_{res} = R\int_{-z_0}^{z_0} I_z^2(z)dz = \frac{R}{dx^2}\int_{-z_0}^{z_0} J_z^2(z)dz$$

Again, due to judicious alignment of the current distribution along the z-axis, the far-field pattern of the antenna is invariant of ϕ and only varies with θ. Thus, the far-field RCS for the distribution is given exactly by:

$$RCS(\theta) = \frac{\eta_0 k^2}{32\pi^2}\left\|\int_{-z_0}^{z_0} J_z(z)sin(\theta)exp(jkzcos(\theta))dz\right\|^2$$

and

$$P_{rad} = 2\pi \frac{\eta_0 k^2}{32\pi^2} \int_0^{\pi} \left\| \int_{-z_0}^{z_0} J_z(z) exp(jkzcos(\theta)) dz \right\|^2 sin^3(\theta) d\theta$$

The radiation efficiency of the antenna is now defined as:

$$\eta_{rad} = \frac{P_{out}}{P_{in}} = \frac{P_{rad}}{P_{rad} + P_{res}} = \frac{1}{1 + \frac{P_{res}}{P_{rad}}}$$

Thus, maximizing η_{rad} is equivalent to maximizing the ratio $\frac{P_{rad}}{P_{res}}$. Unfortunately, in its current state, the integral for computing P_{rad} cannot be computed in closed form for arbitrary distribution $J_z(z)$ due to the $exp(jkzcos(\theta)$ term. We proceed by Taylor expanding this term:

$$exp(jkzcos(\theta)) \approx 1 + jkzcos(\theta) - \frac{k^2 z^2 cos^2(\theta)}{2!} + ...$$

In fact, if we assume that the line current distribution $J_z(z)$ is an even function (which is a valid assumption in maximizing radiation efficiency due to symmetry), only the even terms of the expansion are important since the odd ones cancel out of the integration:

$$exp(jkzcos(\theta)) \approx 1 - \frac{k^2 z^2 cos^2(\theta)}{2!} + \frac{k^4 z^4 cos^4(\theta)}{4!} + ...$$

We will later provide a more accurate expression, but to obtain a simple closed-form bound, we proceed by using the triangle inequality:

$$\begin{aligned} P_{rad} &= 2\pi \frac{\eta_0 k^2}{32\pi^2} \int_0^{\pi} \left\| \int_{-z_0}^{z_0} J_z(z) exp(jkzcos(\theta)) dz \right\|^2 sin^3(\theta) d\theta \\ &<= 2\pi \frac{\eta_0 k^2}{32\pi^2} \int_0^{\pi} \int_{-z_0}^{z_0} \|J_z(z)\|^2 \|exp(jkzcos(\theta))\|^2 \, dz sin^3(\theta) d\theta \end{aligned} \quad (3.3)$$

Thus:

$$P_{rad} <= 2\pi \frac{\eta_0 k^2}{32\pi^2} \int_0^{\pi} \int_{-z_0}^{z_0} \|J_z(z)\|^2 \, sin^3(\theta) dz d\theta = 2\pi \frac{\eta_0 k^2}{32\pi^2} \int_{-z_0}^{z_0} \|J_z(z)\|^2 \, dz \int_0^{\pi} sin^3(\theta) d\theta$$

Further:

$$\int_0^{\pi} sin^3(\theta)d\theta = \frac{4}{3}$$

so:

$$P_{rad} <= 2\pi\frac{\eta_0 k^2}{32\pi^2}\frac{4}{3}\int_{-z_0}^{z_0}\|J_z(z)\|^2\,dz$$

Thus, we can easily compute an upper bound to the ratio $\frac{P_{rad}}{P_{res}}$:

$$\frac{P_{rad}}{P_{res}} <= \frac{2\pi\frac{\eta_0 k^2}{32\pi^2}\frac{4}{3}\int_{-z_0}^{z_0}\|J_z(z)\|^2\,dz}{\frac{R}{dx^2}\int_{-z_0}^{z_0}\|J_z(z)\|^2\,dz} = 2\pi\frac{\eta_0 k^2}{32\pi^2}\frac{4}{3}\frac{dx^2}{R}$$

And therefore we have an upper bound to the radiation efficiency:

$$\eta_{eff} <= \frac{1}{1+\frac{1}{2\pi\frac{\eta_0 k^2}{32\pi^2}\frac{4}{3}\frac{dx^2}{R}}}$$

$$\eta_{eff} <= \frac{\frac{\eta_0 k^2 dx^2}{12\pi}}{\frac{\eta_0 k^2 dx^2}{12\pi}+R}$$

$$\frac{12\pi R}{\eta_0 k^2 dx^2}$$

It turns out this bound is quite loose, so we go back to get a better bound:

$$P_{rad} = 2\pi\frac{\eta_0 k^2}{32\pi^2}\int_0^{\pi}\left\|\int_{-z_0}^{z_0}J_z(z)exp(jkzcos(\theta))dz\right\|^2 sin^3(\theta)d\theta$$

$$\left\|\int_{-z_0}^{z_0}J_z(z)exp(jkzcos(\theta))dz\right\|^2 \approx \left\|\int_{-z_0}^{z_0}J_z(z)(1-\frac{k^2z^2cos^2(\theta)}{2!}+\frac{k^4z^4cos^4(\theta)}{4!}+...)dz\right\|^2$$

and: Real$(exp(jkzcos(\theta))$ <= 1, so:

$$\left\|\int_{-z_0}^{z_0}J_z(z)exp(jkzcos(\theta))dz\right\|^2 <= \left\|\int_{-z_0}^{z_0}J_z(z)dz\right\|^2$$

and thus we have another upper bound for P_{rad}:

$$P_{rad} <= 2\pi\frac{\eta_0 k^2}{32\pi^2}\int_0^{\pi}\left\|\int_{-z_0}^{z_0} J_z(z)dz\right\|^2 sin^3(\theta)d\theta = 2\pi\frac{\eta_0 k^2}{32\pi^2}\frac{4}{3}\left\|\int_{-z_0}^{z_0} J_z(z)dz\right\|^2$$

Note that since $\left\|\int_{-z_0}^{z_0} J_z(z)dz\right\|^2 <= \int_{-z_0}^{z_0}\|J_z(z)\|^2\,dz$, this is a tighter bound than the first one.

As before, we assemble $\frac{P_{rad}}{P_{res}}$:

$$\frac{P_{rad}}{P_{res}} <= \frac{2\pi\frac{\eta_0 k^2}{32\pi^2}\frac{4}{3}\left\|\int_{-z_0}^{z_0} J_z(z)dz\right\|^2}{\frac{R}{dx^2}\int_{-z_0}^{z_0}\|J_z(z)\|^2\,dz}$$

Unfortunately the top and bottom integrals do not cancel out in this case, but it is well known that the function which maximizes the ratio $\frac{\left\|\int_{-z_0}^{z_0} J_z(z)dz\right\|^2}{\int_{-z_0}^{z_0}\|J_z(z)\|^2 dz}$ is the constant function $J_z(z) = \alpha_0$. Substituting that in:

$$\frac{\left\|\int_{-z_0}^{z_0}\alpha_0 dz\right\|^2}{\int_{-z_0}^{z_0}\|\alpha_0\|^2\,dz} = \frac{4z_0^2}{2z_0} = 2z_0$$

Thus, substituting in:

$$\frac{P_{rad}}{P_{res}} <= \frac{2\pi\frac{\eta_0 k^2}{32\pi^2}\frac{4}{3}2z_0}{\frac{R}{dx^2}}$$

and:

$$\eta_{eff} <= \frac{\frac{\eta_0 k^2 z_0 dx^2}{6\pi}}{\frac{\eta_0 k^2 z_0 dx^2}{6\pi} + R}$$

This bound actually turns out to be quite tight as can be seen in Fig. 3.3 which compares the bounds against the globally optimal current distribution for varying values of R computed numerically using the aforementioned generalized eigenvalue optimization technique without any approximation.

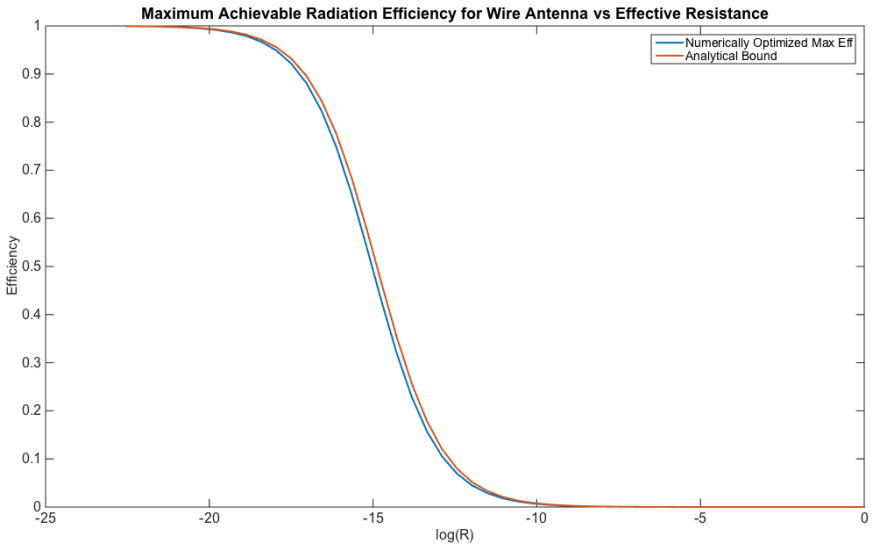

Figure 3.3: Max Radiation Efficiency Bound of Linear Current Distributions vs Optimal Maximizing Current Distribution

Note that this bound is more accurate for small z_0 because the approximation made for the $exp(jkzcos(\theta)$ term becomes more accurate as z_0 approaches 0; however, the error in the bound remains acceptable for reasonable lengths (such as $\frac{\lambda}{2}$.)

3.6 Optimizing the Gain of a Linear Current Distribution over a Lossy Silicon Substrate

Up until this point, we have focused our attention on the performance and efficiency of line current distributions in free-space. Most modern embedded systems involve some sort of dielectric substrate on which the antenna sits. For example, this substrate can be a Printed Circuit Board (PCB) made of FR-4, Rogers [44], or other ceramic laminate materials or even silicon in the case of integrated on-chip antennas. In this section, we will study maximum performance bounds of a line current distribution on such a dielectric substrate (Fig. 3.4). In all of the following numerical examples, a silicon substrate (ϵ_r = 11.9, conductivity [σ] = 10 S/m) of dimensions 1.74 x 1.74 x 0.217 mm is utilized with a PEC ground plane on its bottom (-$\hat{z}$) face and a current distribution is optimized sitting on the top (+$\hat{z}$) surface. A custom Finite Difference Frequency Domain (FDFD) [103, 101] solver is used and all of the numerical results are computed at an operating frequency of 100GHz. This makes the effective size of the silicon substrate 2λ x 2λ x $\frac{\lambda}{4}$ in terms of the wavelength λ in silicon, which is $\frac{c_0}{4\sqrt{11.9}} \approx 869\mu m$. Note that any type

of substrate of arbitrary dimensions could have been used as well as any operating frequency. In fact, had a time-domain solver (such as Finite Difference Time Domain [FDTD]) been used, judicious use of Fourier Transforms could yield rich broadband frequency information with a single simulation. We chose 100GHz as the frequency of interest because it is a typical design frequency for on-chip silicon radiators and used the dimensions specified as they are also common silicon IC dimensions used in standard bulk CMOS processes. Finally, a PEC ground plane is used on the bottom surface of the substrate since most IC's are soldered against a PCB on top of a ground plane. Since the index of silicon is much larger than air, most of the radiated energy is absorbed into the silicon substrate and the bottom ground plane serves to reflect that back out through the top surface into free-space in the $+\hat{z}$ direction.

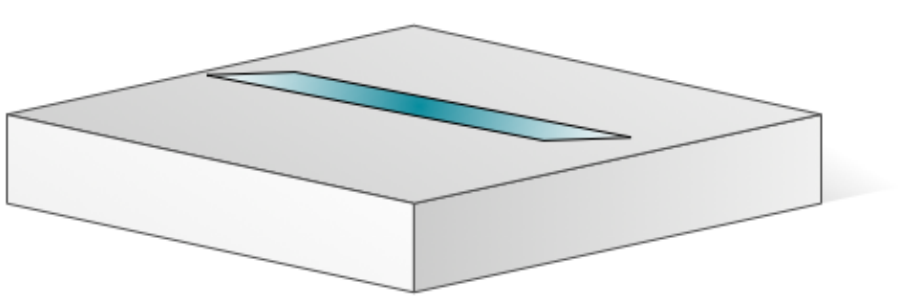

Figure 3.4: Illustration of a hypothetical line current centered on a dielectric substrate

For our first experiment, we will study the maximum gain of a line current distribution along the $\hat{x}$-axis on the surface of the silicon dielectric descrbed above centered at the origin. Since a closed-form Green's function does not exist for the finite, lossy silicon substrate with PEC back-plate grounding, we must use a computational electromagnetic simulation method to obtain numerical approximations to the required Green's functions. For the current experiment, and everything that follows in this thesis, we will be using Finite-Difference methods as our simulation workhorse. In order to ensure 2nd order convergence, all of the E and H fields for the FD method are located in different positions on a lattice called the Yee cell (Fig. 3.5). The Yee cell is the unit cell building block for the full system which consists of a cascade of cells in the 3D dimensions [103, 101]. Often fields must be collocated to a central position and this can be done with simple linear interpolation (averaging.)

An absorbing boundary condition named the Perfectly Matched Layer (PML, Fig. 3.6) [10] is used on the outer boundaries of the Yee cell simulation domain lattice to

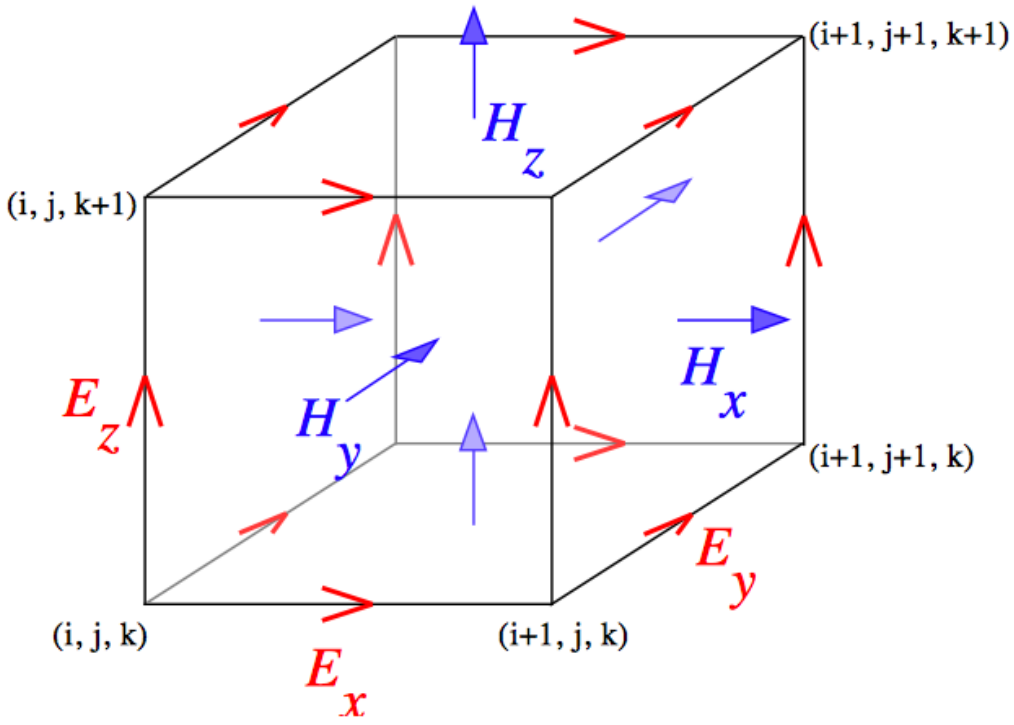

Figure 3.5: A single Yee cell illustrating the relative positions of th E and H fields

absorb any incident electromagnetic fields. This is required due to the fact that the numerical simulation must be truncated to a finite domain size, however, the initial antenna problem is an open-boundary problem. In order to convert the problem to a feasible closed-boundary problem, a PML is used which is a fictitious lossy material that is perfectly impedance matched to free-space around the outer boundaries of the simulation domain in order to absorb incident radiation instead of erroneously scattering it back inside the volume.

The fields on the surfaces of a virtual cube within the simulation domain (Fig. 3.7) are used to perform a Near-Field to Far-Field (NF2FF) transformation to obtain the far-fields of the antenna structure being simulated. The NF2FF makes use of Love's equivalence principle (ref) and the free-space Green's function to project the E/H fields on this virtual surface onto the far-field sphere. [103]

Once the silicon dielectric geometry and PEC ground are defined within the constructs of the numerical FDFD simulation, electric current excitations ($\mathbf{J}_{\mathbf{src}}$) can be defined and the resulting electromagnetic fields can be resolved everywhere in the computational domain. In order to optimize for maximizing the gain of a current distribution within this setup, the far-field data is required for each discrete excitation

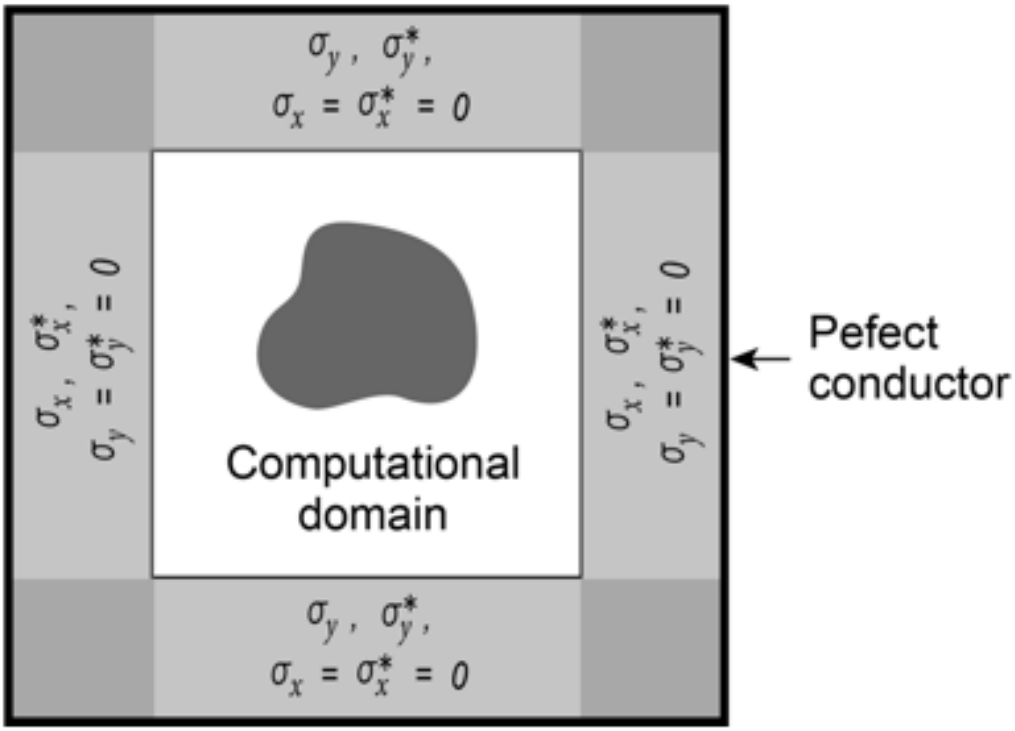

Figure 3.6: Cartoon illustrating the Perfectly Matched Layer surrounding the computational domain to absorb incident EM radiation

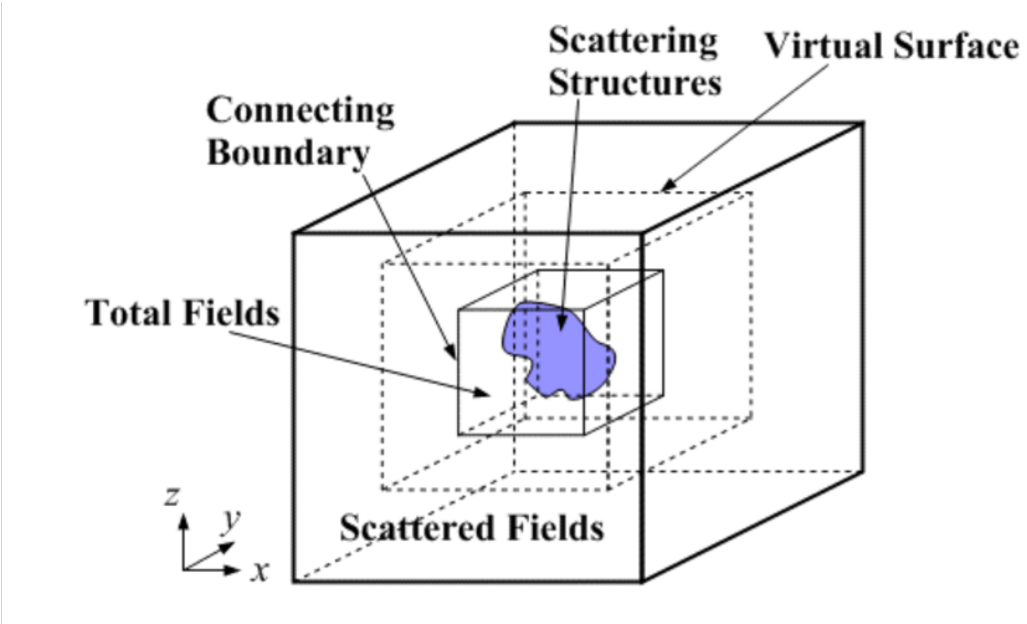

Figure 3.7: Diagram showing virtual surface used within computational domain for calculation of far-fields

source as well as the electric field data collocated with every electric current source that may be nonzero. The need for the far-field data is apparent, and the near-field collocated electric field data is required for computing the incident (input) power by calculating and summing the quantities $E^* * J$ (proportional to $V * I$, which is voltage times current and has units of power) for every nonzero source J.

Fig. 3.8 illustrates a simplified top down view of the substrate surface with X and Y directed sources as they are located on the Yee lattice. The actual simulated surface is discretized at $\frac{\lambda}{40}$ and so there are 80x81 X directed sources and 81x80 Y directed sources. The line of sources which need to be computed for evaluating maximum gain of a centered line distribution on the substrate surface is highlighted in cyan. This implies that 80 separate FDFD simulations need to be performed in order to compute the Numerical Green's Function (NGF) from each current source element on the line to the resulting EM fields which will then be used to solve the optimization problem. However, because the problem is symmetric about the Y axis (as well as the X axis), only half of the excitations need to be simulated and the other half can be computed simply by appropriately flipping and mirroring the resulting fields.

For each simulation, the electric fields collocated with every electric current source position on the line segment of interest are saved and due to linearity an impedance matrix, Z, can be formed mapping current densities to electric fields.

Further, it is well known that:

$$V = -\int E \cdot dl$$

so over one Yee cell of dimensions dx x dy x dz, the voltage due to a Jx excitation would be: $Vx = -Exdx$.

Similarly, current is defined as:

$$I = \int JdA$$

and thus over one Yee cell, the current due to a Jx excitation is defined as $Ix = Jxdydz$, so the impedance matrix mapping currents to voltages is defined as:

$$\tilde{Z} = Zdxdydz$$

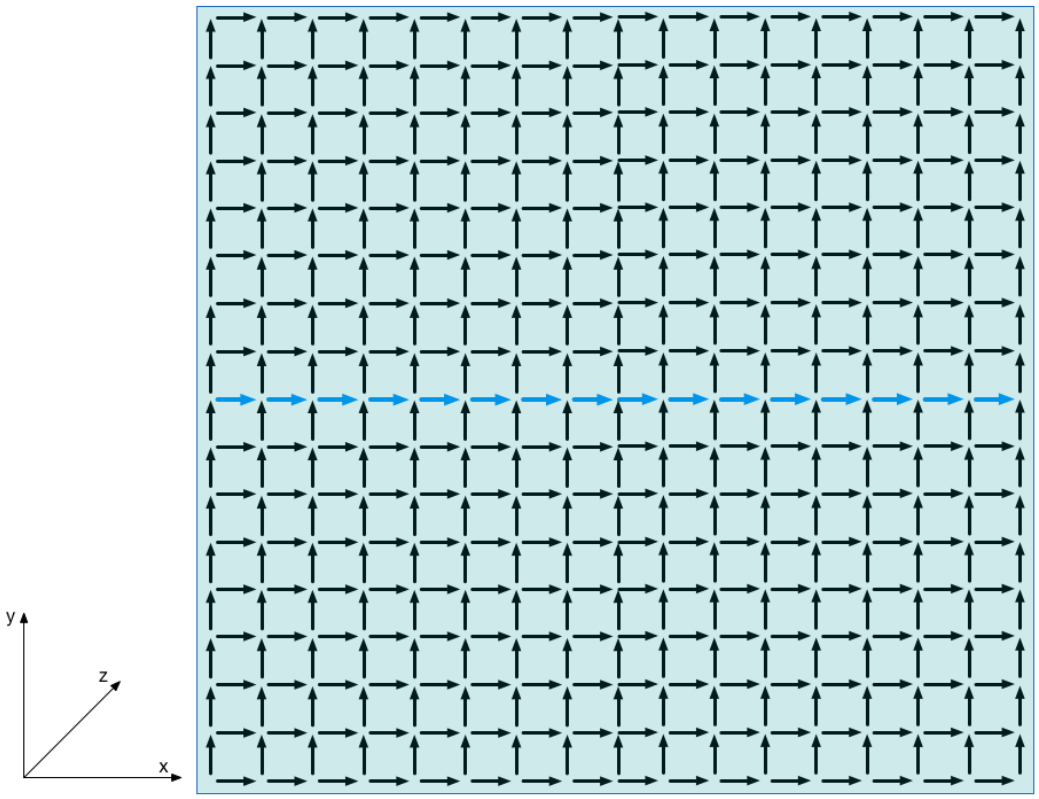

Figure 3.8: Diagram showing virtual surface used within computational domain for calculation of far-fields

Complex power is defined as $P_{complex} = \frac{V*I^*}{2}$ thus the quantity $x^H Z \frac{dxdydz}{2} x$ represents the complex (real input power + reactive energy stored) power produced by the current source excitations defined by the discrete vector x. In order to maximize antenna gain, we require the real part of this quantity and it is apparent that: $real(x^H Ax) = x^H \frac{A+A^H}{2} x$ so by defining the new matrix $\hat{Z} = \frac{(Z+Z^H)dxdydz}{2}$ the quantity $x^H \hat{Z} x$ now represents the real input power into the system from the electric current source excitations represented by the vector x.

Next, we store the tangential surface E and H field data on the faces of a virtual cube surrounding the substrate and sources in another matrix B. Again, due to linearity, Bx represents the E and H tangential surface fields on this virtual surface due to any arbitrary current excitation vector x. As mentioned in the previous derivation section:

$$U(\theta,\phi) = \frac{k^2}{32\pi^2\eta_0}\left(\left[L_\phi + \eta_0 N_\theta\right]^2 + \left[L_\theta - \eta_0 N_\phi\right]^2\right) = \frac{k^2\eta_0}{32\pi^2}\left(N_\theta^2 + N_\phi^2\right)$$

for zero values magnetic current sources. N_θ and N_phi are both linear integrals with respect to the electric current excitation Jx and by using an appropriate approximation rule for integration (we use Simpson rule in this work; however, any suitable integration rule can be utilized) we can relate the quantitites as follows:

$$N_\theta = c_1 * B * x$$
$$N_\phi = c_2 * B * x$$
$$\begin{pmatrix} N_\theta \\ N_\phi \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} * B * x = C * x$$
$$C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} * B$$

and thus, directivity at a specific solid angle on the far-field sphere can be defined as:

$$D_{\theta,\phi} = x^H \frac{4\pi k^2}{32\pi^2\eta_0} C^H C x = x^H \hat{C}^H \hat{C} x$$

with $\hat{C} = \sqrt{\frac{4\pi k^2}{32\pi^2\eta_0}} C$, where $\hat{C}$ is a 2 x N matrix with N representing the number of individual current source excitations on the line of interest.

Now, we must solve the optimization problem:

$$\max_x \frac{x^H \hat{C}^H \hat{C} x}{x^H \hat{Z} x}$$

As derived in the previous section, this is a standard, generalized Rayleigh quotient and the globally optimal solution can be found by solving the generalized eigenvalue problem:

$$\hat{C}^H \hat{C} x = \lambda \hat{Z} x$$

for the maximal eigenvalue λ_{max}. λ_{max} corresponds to the maximum gain achievable with the ideal line current distribution on the dielectric substrate and its corresponding eigenvector x_{max} contains the coefficients for the electric current sources on that line which achieves that gain.

Since the matrix $\hat{C}$ has two rows, the N x N square matrix $\hat{C^H}\hat{C}$ is at most rank 2. We can define a projection matrix:

$$W = \hat{Z^{-1}}\hat{C^H}$$

and then multiply the generalized eigenvalue problem on the left by W and use a change of variables: $x = Wy$

$$W^H\hat{C^H}\hat{C}Wy = \lambda W^H\hat{Z}Wy$$

$W^H\hat{C^H}\hat{C}W$ and $W^H\hat{Z}W$ are now both 2x2 and the generalized eigenvalue problem is trivially solvable. The eigenvalues do not change under this transformation, and $x_{max} = Wy_{max}$ where y_{max} is the 2x1 eigenvector of the transformed problem corresponding to the largest eigenvalue λ_{max}.

Results

First we optimize the max gain possible for a line current of varying length from $2dx$ to $80dx$ where dx is the unit step and is $\frac{\lambda}{40\sqrt{11.9}}$ ($\frac{\lambda_{Si}}{40}$) = 21.7 μm:

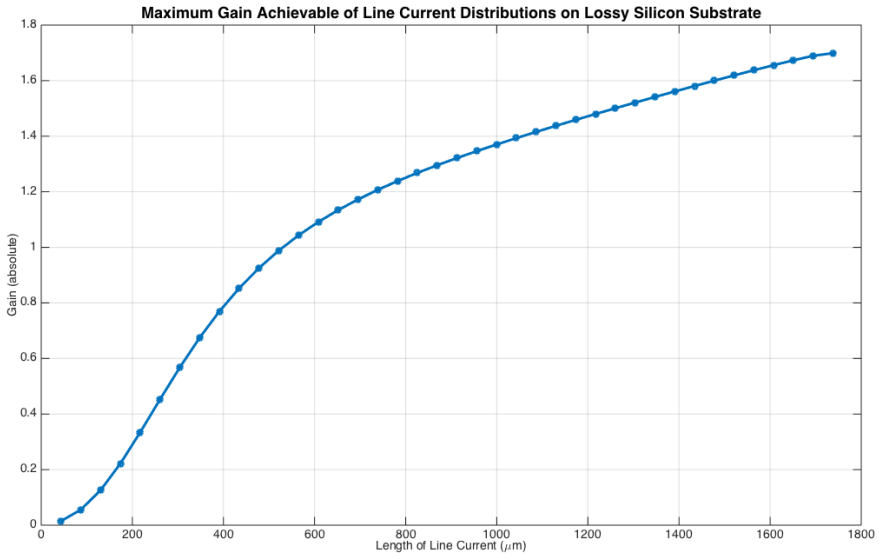

Figure 3.9: Diagram showing virtual surface used within computational domain for calculation of far-fields

As expected, the plot is monotonically increasing since current distributions which are longer have a solution set which is a superset of all short length ones. At this point, it's instructive to compare the maximum gain achievable of an ideal line current distribution against the gain of an equivalent length PEC dipole antenna fed by a current source excitation of length $2dx$ in the center:

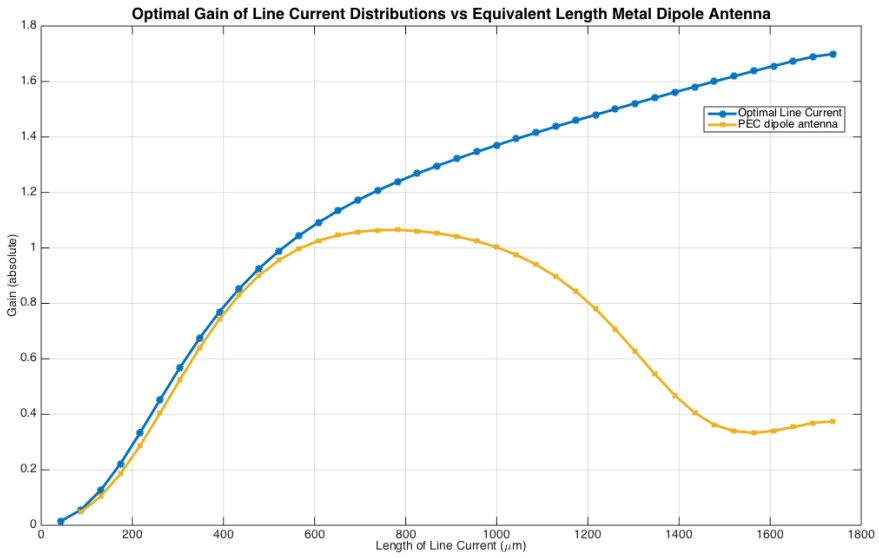

Figure 3.10: Diagram showing virtual surface used within computational domain for calculation of far-fields

Looking at the plot of the gain of the metal dipole (which is a real antenna) it is apparent that up to around an antenna of length 600 μm the gain of the dipole antenna approaches the maximum possible gain of a line current very closely. Beyond around 600 μm of length, the gain of the metal dipole starts falling significantly short of the predicted optimal bound. In fact, the gain of the metal dipole is non-monotonic and begins to decrease beyond a length of 800 μm.

To better understand why this phenomenon is occuring, it is useful to observe the current distributions of both the optimal line current as well as the metal antenna. Here we plot the real part of the current distribution of the optimal current for varying line lengths:

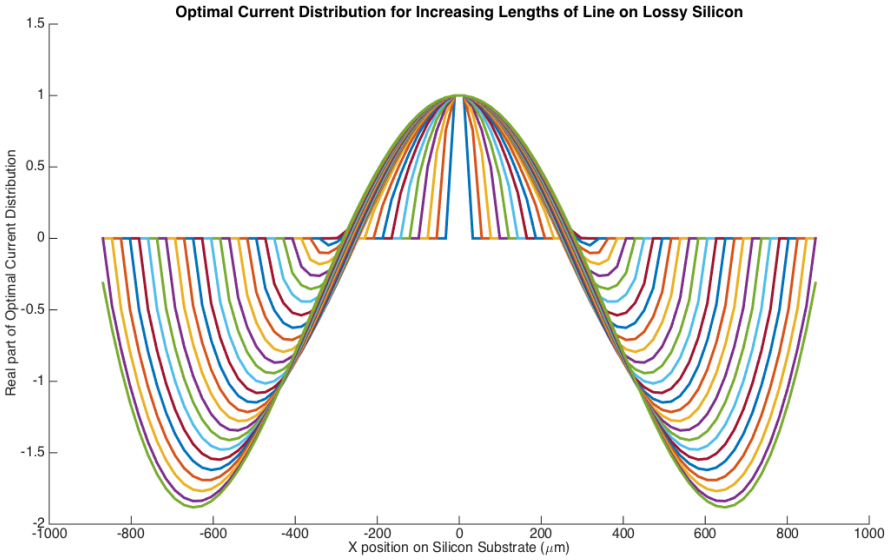

Figure 3.11: Optimal line current distributions for increasing line lengths

To better understand why the gain of the metal dipole antenna starts to fall short of the optimal gain, we compare the optimal current distribution against the current distribution generated by the realistic metal dipole for increasing lengths of line length:

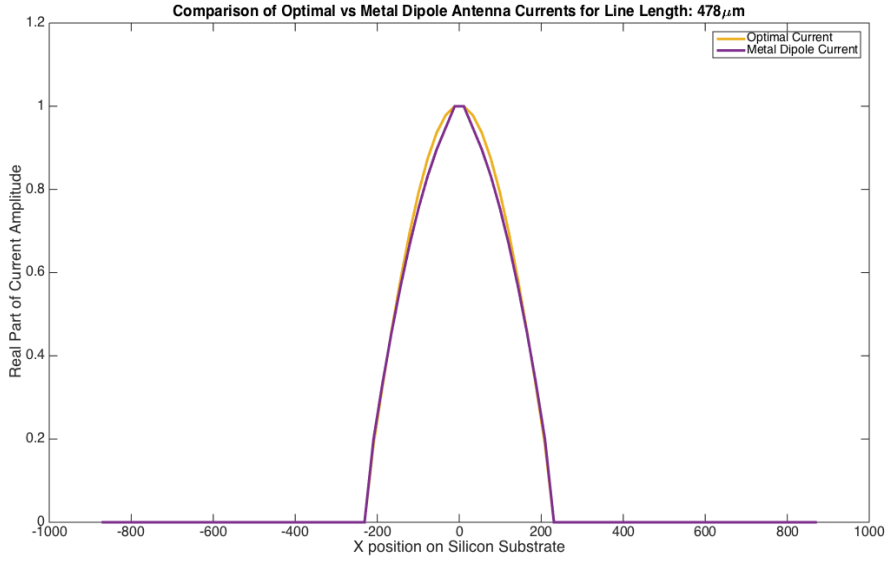

Figure 3.12: Optimal line current distributions for increasing line lengths

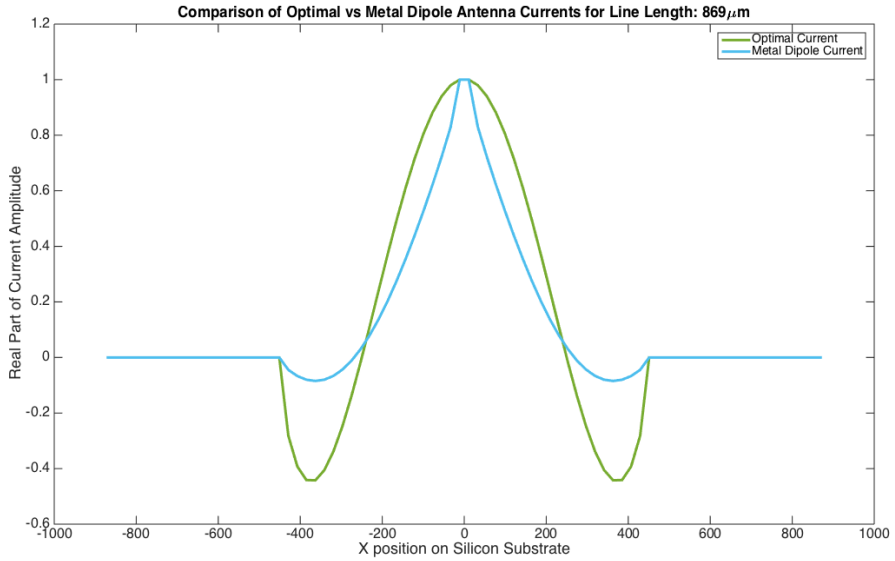

Figure 3.13: Optimal line current distributions for increasing line lengths

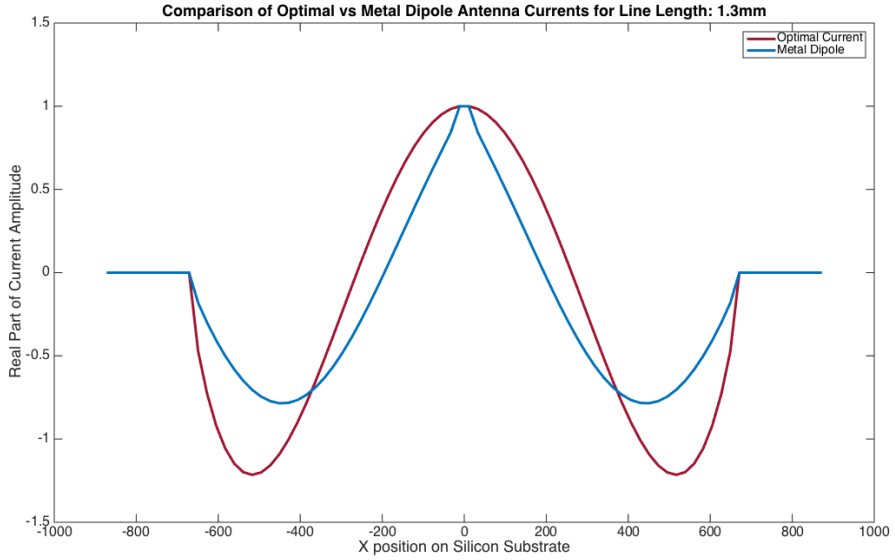

Figure 3.14: Optimal line current distributions for increasing line lengths

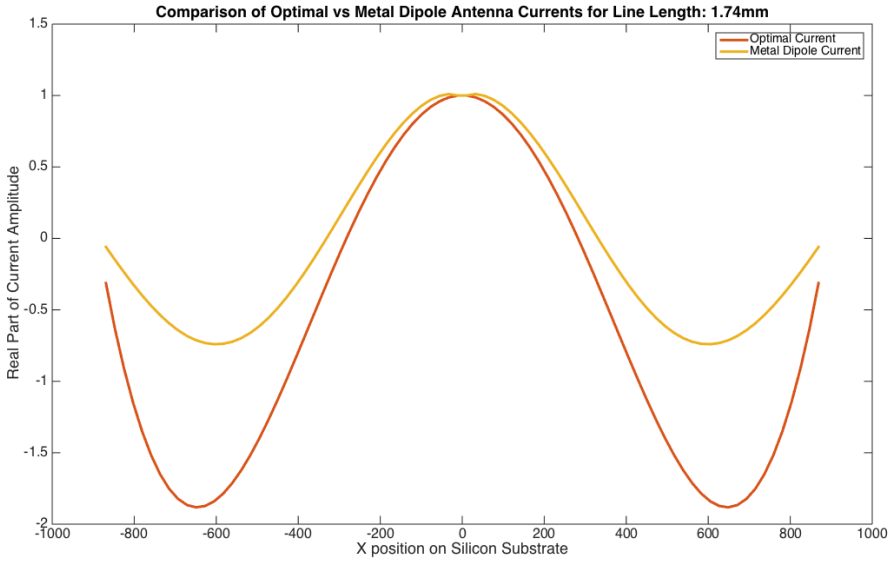

Figure 3.15: Optimal line current distributions for increasing line lengths

As can be seen, the current distribution of the real metal dipole for the 478 μm length antenna matches up very closely with the optimal current distribution, and thus the gain achieved by the metal dipole is quite close to the maximum gain achievable for any arbitrary current distribution along that length of line.

As the length of antenna becomes longer, the deviations between the optimal current and current along the metal dipole become increasingly evident. Most importantly, the the bottom "lobes" of the ideal current distribution dive deeper and deeper with increasing length compared to the metal antenna.

Thus, a simple metal dipole antenna cannot achieve close to the maximum gain possible of a line current on the substrate when it is longer than around 600 μm in this example. In order to obtain better performance (closer to the maximum bound) for a real antenna, multiple excitation sources at different points along the line should be investigated.

Before we shift our focus to 2D current distributions, we also plot the max gain achievable for varying lengths (centered on the X axis) of current versus position of the line on the Y axis:

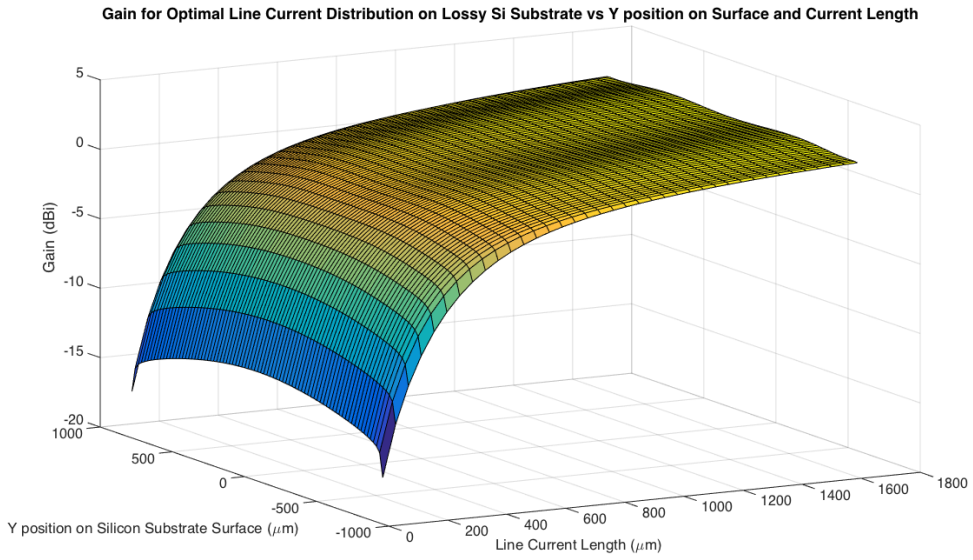

Figure 3.16: Optimal line current distributions for increasing line lengths

The max achievable gain of an optimal line current is symmetric about the Y-axis as expected since the silicon substrate exhibits mirror symmetry across the X and Y axes. Further, the max achievable gain is a strictly increasing function with respect to the length of current distribution allowed due to the fact that the design space of a longer length line current is a superset of all currents which are shorter since any arbitrary current distribution is allowed along the line.

Next, we plot the gain achieved in the broadside direction of a real PEC dipole antenna and sweep the same parameters as before: position on Y-axis along substrate surface and length of dipole antenna (it is always centered on X-axis):

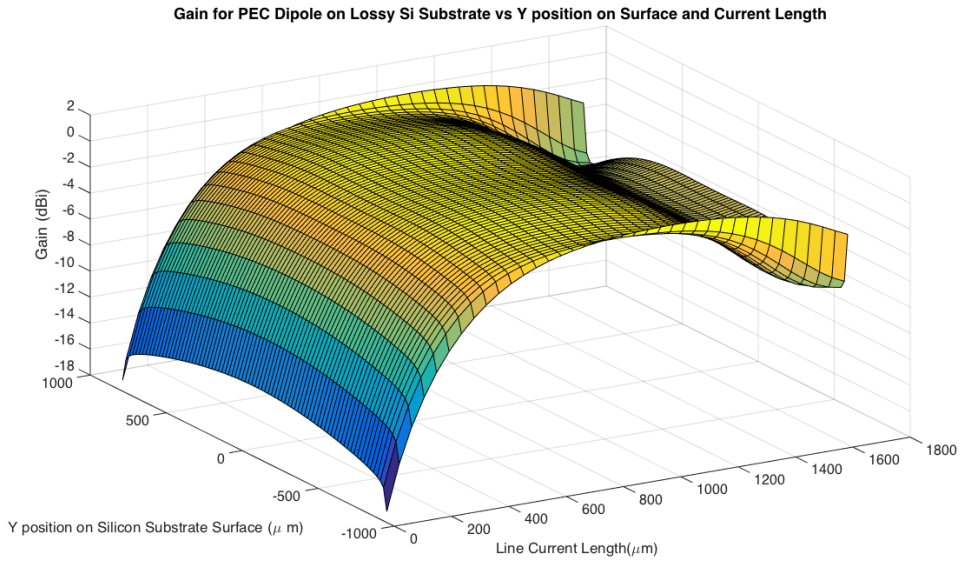

Figure 3.17: Optimal line current distributions for increasing line lengths

As discussed and analyzed earlier, beyond a certain length of dipole antenna, the maximum gain begins to decrease. This is due to the fact that the current distribution imposed by the PEC material begins to significantly deviate from the optimal current distribution for maximizing broadside gain beyond a certain length of antenna.

3.7 Bounding the Maximum Possible Achievable Gain of any Abritrary Antenna Structure over a given design region on a Lossy Silicon Substrate

Now we shift our focus to 2D patches of current on the silicon substrate surface. A real on-chip antenna or group of antennas does not have to be confined to a single dimension and thus we explore the maximum gain achievable by patches of current on the same lossy, grounded silicon substrate as in the previous section.

Fig. 3.18 plots the maximum gain achievable in the broadside ($\theta = 0, \phi = 0$) direction of an optimal square patch of current centered on the silicon surface vs the edge length of the patch. As expected, the plot is monotonically increasing with increasing patch size. The maximum gain achievable for this substrate in the broadside direction (by utilizing the whole substrate surface area) is 5.17 (or 7.1dBi). Note that this is a maximum bound and cannot necessarily be achieved by a real metallic antenna.

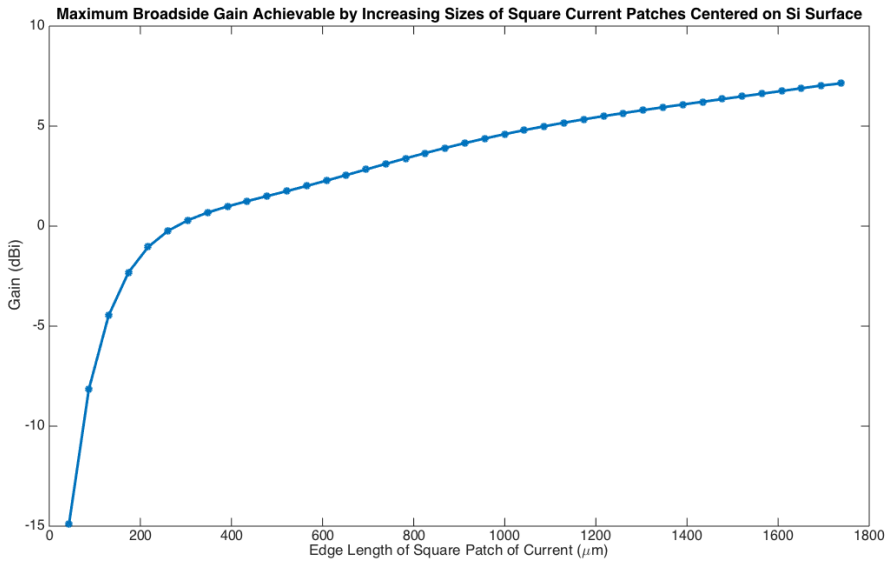

Figure 3.18: Maximum Broadside Gain Achievable by Increasing Size of Optimized Square Patch of Current Centered on Origin of Lossy Silicon Substrate Surface

Now we plot 2D heatmaps of the optimal current distributions for patch edge sizes $435\mu m$, $869\mu m$, $1.3mm$, and $1.74mm$:

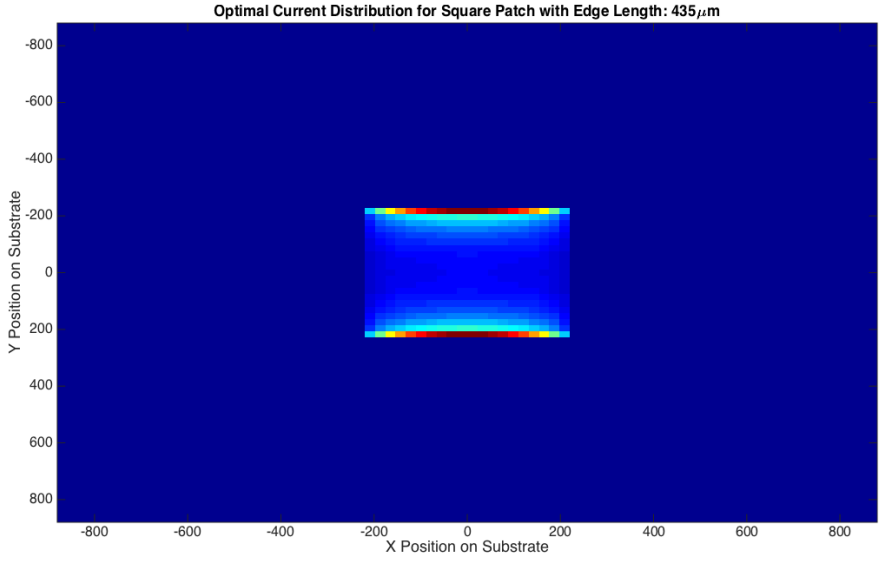

Figure 3.19: Heatmap of Square Current Distribution of Size $435\mu m$ x $435\mu m$ centered on Silicon Surface which Maximizes Broadside Gain

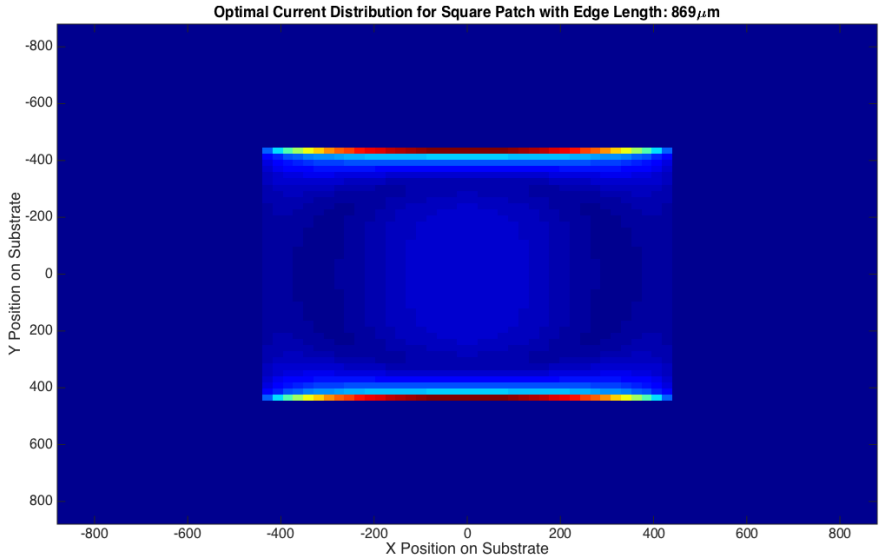

Figure 3.20: Heatmap of Square Current Distribution of Size 869μm x 869μm centered on Silicon Surface which Maximizes Broadside Gain

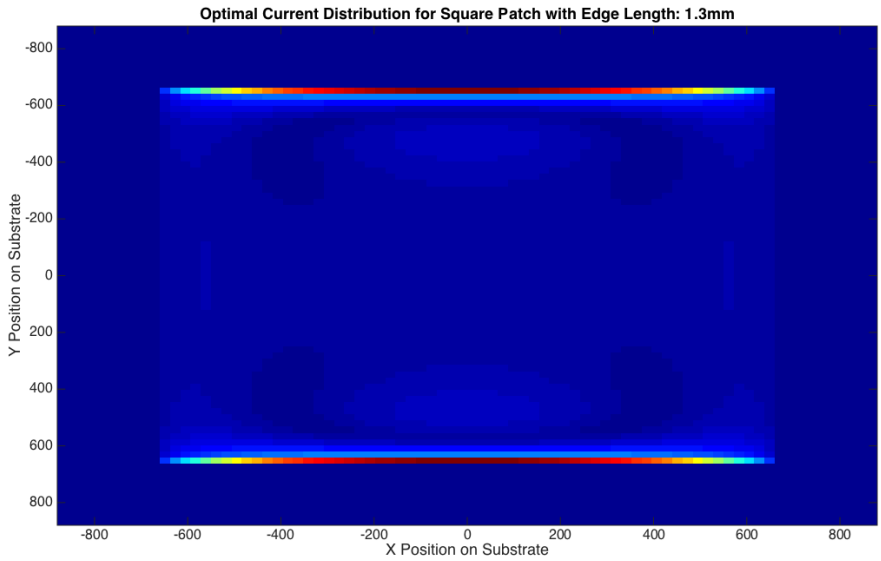

Figure 3.21: Heatmap of Square Current Distribution of Size 1.3mm x 1.3mm centered on Silicon Surface which Maximizes Broadside Gain

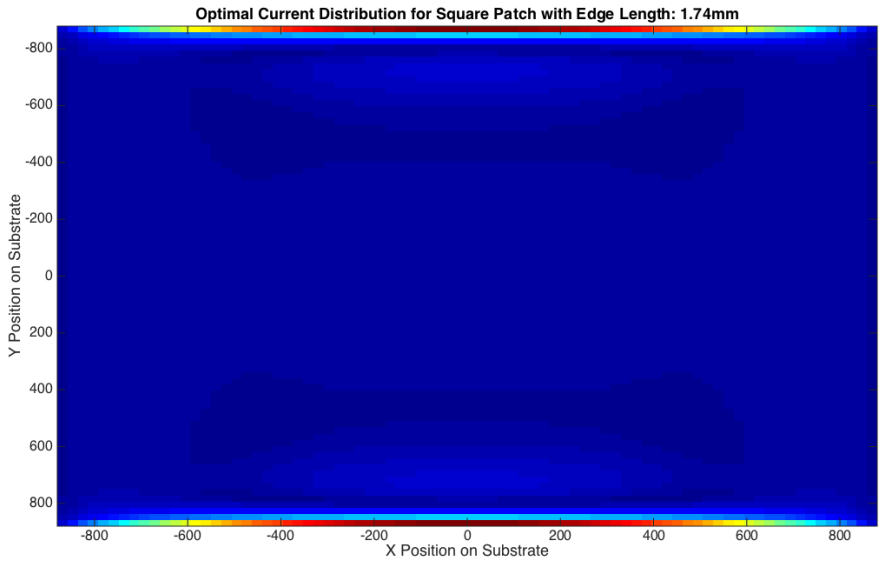

Figure 3.22: Heatmap of Square Current Distribution of Size 1.74mm x 1.74mm centered on Silicon Surface which Maximizes Broadside Gain

It is quite interesting to observe that all of the optimal distributions consist mainly of two horizontal strips of current on opposing edges of the square patch. The substrate is 90 degree rotationally symmetric, so there are actually two optimal distributions (90 degree rotated versions of each other) which achieve the maximum broadside gain.

Up until this point we have been solving the relaxed optimization problem for optimal current distributions over a given surface or volume that maximize a desired performance metric (such as antenna directivity, gain, or radiation efficiency.) While, as demonstrated with various examples, the results from this exercise often provide useful and tight upper performance bounds as well as good initial design insight, they unfortunately do not directly lead to a manufacturable antenna solution. In this chapter, we attack the design problem of real antennas. We will see that one of the most critical factors of a successful antenna optimization lies within the electromagnetic simulation step– specifically how rapidly it can be performed. After developing an extremely efficient and fast simulation framework, we close the loop with an optimization algorithm and show the performance of the algorithm on designing radiating antennas on lossy dielectric slabs as well as coupling antennas for dielectric waveguides (in the following chapter.) In the case of the waveguide coupler, we compare simulation results of the best resulting design with real measurements of the fabricated coupler. Finally, we describe the design of a dual-channel 50Gbps (per channel) 120GHz transceiver developed in a bulk CMOS 28nm process that will be used to demonstrate the complete multi-mode waveguide link.

4.1 Introduction to Heuristic Optimization Approaches

The most simplistic flow-chart of operations which characterizes any structure optimization algorithm is shown in Fig. 4.1. The electromagnetic simulation step is usually performed numerically by utilizing any number of methods such as the Finite Difference Time Domain (FDTD) [101], Finite Difference Frequency Domain (FDFD) [103], Finite Element Method (FEM) [108], and Integral Equation (IE) [109] techniques. Solving Maxwell's equations computationally in three dimensions is quite challenging and processing intensive, however, so the EM simulation step is almost always the rate-limiting step in the whole process. Almost all optimization algorithms require numerous iterations to converge to a solution, making the speed efficiency of the inner EM simulation + optimization loop critical for the success of the overall structure design problem. Therefore, we will mainly focus on the EM simulation block in the following sections of this chapter, but we will first touch briefly on the optimization block.

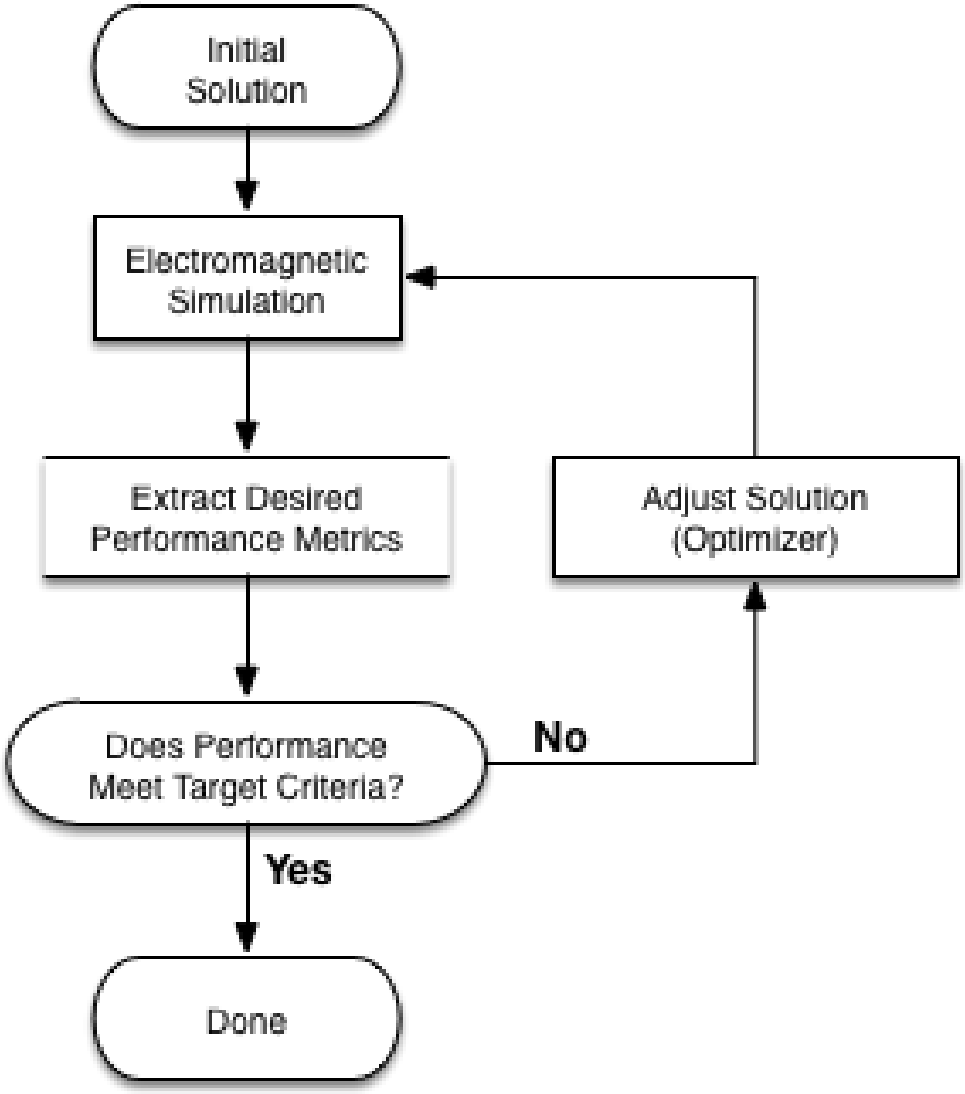

Figure 4.1: Simplified Flow Chart of Typical Structure Optimization Process

As mentioned in the previous chapter, the real antenna design problem is NP-hard [75] and the globally optimal solution unfortunately cannot be guaranteed to be found without an exhaustive search of the whole design space. Since even in the simplest of scenarios, an exhaustive search would be completely time prohibitive, we must look towards alternative, "heuristic" based, techniques to use in order to optimize real designs.

Most of these techniques can be classified under two distinct categories: local or global methods. Local methods begin with an initial (usually randomized) solution

and perform a variation of gradient descent to arrive at the closest local minimum (or maximum).

A local method will improve on the solution with every iteration until the local min (or max) is reached; however, once it has reached that point it will stagnate. Therefore, local methods often need to be seeded with many different initial solutions and run to completion in an attempt to explore the design space as thoroughly as possible. Due to this, the quality of the solution often depends on the initial guess and local methods tend to perform best when given reasonable or decently performing initial starting points.

On the other hand, global methods employ various heuristics to avoid getting "stuck" in local minima/maxima. Well known global methods include Genetic Algorithms (GA) [81, 85] and Particle Swarm Optimization (PSO) [85, 54, 13, 52, 67]. Although global methods may not always get stuck within local optima like local methods, they are less principled in their approach and rarely have any mathematical theoretical guarantees on their convergence or performance. Furthermore, they often contain critical internal parameters which may require meticulous fine-tuning in order to produce worthwhile optimization results– such as number of generations, number of mutations per generation, and cross-over approach for GA's and number of particles and particle velocity for PSO's. Due to these shortcomings, local methods are often preferred over global methods. Another common approach that some designers may take is to use a global method for the initial optimization stage in order to reach a decently performing initial solution, followed by a local optimizer to further optimize and finetune the design. [80]

In the next sections we will give a brief overview of local gradient descent and even simpler local method which we have named random coordinate ascent as well as the two global methods described: Genetic Algorithms and Particle Swarm Optimization.

Local Methods

The most common local method is first order gradient descent/ascent. Sensitivities of every pertubable unknown must be determined at each iteration. The sensitivity is essentially a gradient and quantifies the change of the objective function with respect to a physical perturbation of a parameter. Optimizing the "tooth" spacing of a grating coupler for maximizing light coupled into a waveguide from an external fiber provides a good example of this. Fig. 4.2 illustrates a simplified grating coupler

with four adjustable parameters (d_0, d_1, d_2, and d_3.)

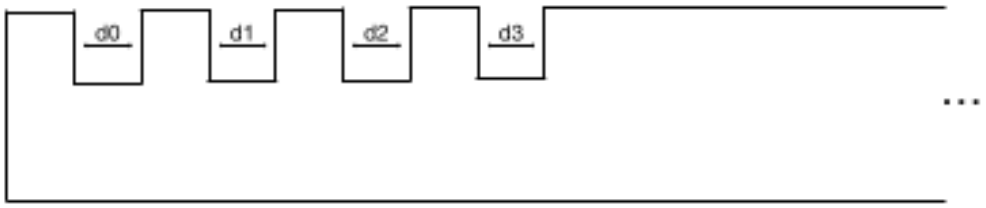

Figure 4.2: Simplified Grating Coupler Diagram with 4 Perturbable Parameters for Adjusting Tooth Spacing

To use a local method for optimizing the spacings, the effect of perturbing each of the d_k parameters on the coupling efficiency must be determined. The most basic approach would be to slightly adjust each parameter, one at a time, and simulate the effect on the objective function (coupling efficiency.) Then, the sensitivity is computed as follows (η is coupling efficiency):

$$\frac{\partial \eta}{\partial d_k} \approx \frac{\eta_{d_k - perturbed} - \eta_{orig}}{d_k - perturbed - d_k}$$

This effectively computes the gradient of the parameters numerically by using finite differences. Once the sensitivities have been approximated for all of the parameters, a small (a linear assumption is being made by taking only the first derivative into account, so care must be taken to adjust the step size as to not lead to approximation error) step can be taken in the direction of the gradient for each parameter to arrive at a new solution with improved coupling efficiency. This process is repeated until a local maximum is reached where all the sensitivities compute to 0 and the algorithm stagnates. This algorithm forms the foundation of most local based search methods.

The main fallback of the method as described is the fact that the simulation complexity scales linearly with the number of parameters. For a structure design problem with N perturbable parameters, N+1 electromagnetic simulations must be performed at each optimization step. As mentioned in the introduction of this chapter, 3D electromagnetic simulation can be quite time consuming even with the availability of significant computational power. A clever technique called the Adjoint Sensitivity Method exists which overcomes this issue and requires **only** 2 simulations (the forward simulation + the adjoint system simulation) per optimization step regardless of the number of perturbable unknowns. This technique was first discovered in the

context of Computational Fluid Dynamics [29, 30] and more recently was applied to Maxwell's equations for electromagnetic structure optimization [ref]. The details of the Adjoint approach are beyond the scope of this thesis; however, the reader may refer to [29, 30, 2, 60, 74, 28, 19, 17, 18, 50, 62, 77, 59, 51, 34] for an in-depth derivation.

Global Methods

There are numerous different heuristic-based global algorithms for structure optimization. Two of the most common ones are Genetic Algorithms (GA) and Particle Swarm Optimization (PSO). We utilized Binary Particle Swarm Optimization (BPSO) to optimize photonic multiplexer devices in Chapter VII. The reader is referred to the following references for in-depth analysis and discussion on the GA, PSO, and BPSO algorithms: [81, 85, 85, 54, 13, 52, 67].

4.2 Ultrafast Electromagnetic Structure Evaluation

As mentioned in the introduction, the success of any optimization approach strongly depends on the speed and efficiency of the forward (electromagnetic simulation) step. In this section, we develop an approach which allows the evaluation of any arbitrary solution within the optimization design space orders of magnitude faster than any standard electromagnetic simulation approach. In fact, many designs can be evaluated on a single CPU core of a standard machine in the millisecond timescale.

We first develop the approach for designs which can consist only of perfect metal (PEC) material and then extend it more generally to support electromagnetic structures designed with any dielectric materials.

Ultrafast PEC Structure Evaluation

While superposition of the resulting electromagnetic fields works for electric and magnetic current excitations, this approach does not work unfortunately for dielectric materials or metal conductors as illustrated in Fig. 4.3. This is due to the fact that dielectrics and metals can scatter fields and therefore have interdependence on each other. Thus, simply summing the resulting fields from two different pieces of metal cannot capture the interaction which would occur from the effect of the metals on each other. This interaction may actually be captured fairly readily, however, using an impedance matrix which maps currents to fields (similar to the previous chapter on current optimization).

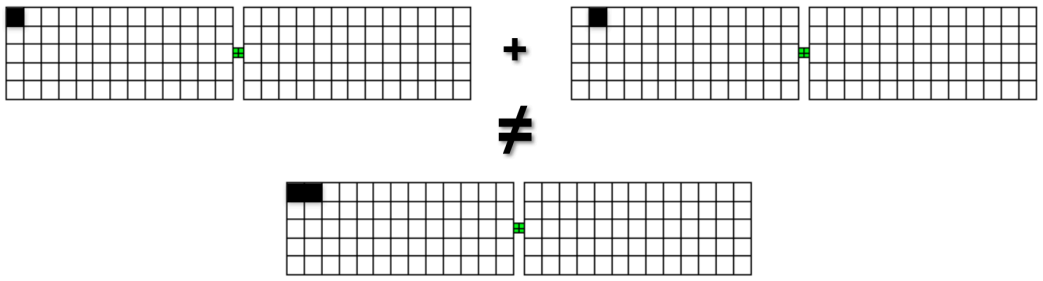

Figure 4.3: Simplified Antenna Design Diagram Illustrating the Failure of Superimposing the Resulting Fields from Two Pieces of Metal

Assume we have a design surface grid made of X and Y directed Yee cells in our simulation domain as in Fig. 3.4. In order to enforce a PEC condition on a Yee cell element, its collocated tangential electric field component (Ex for a Jx current and Ey for a Jy current) must be forced to 0. Instead of doing this at this point, we calculate the transfer function numerically (i.e. Numerical Green's Function) from each current source element on the grid to all of the Ex and Ey fields. For an NxN square substrate, this would normally require N^2 individual electromagnetic simulations. However, due to symmetry across the X and Y axes as well as 90 degree rotational symmetry, the number of simulations required can be cut by a factor of 8 to $\frac{N^2}{8}$. As with the current optimization problem, ALL of these simulations are completely **independent** of each other and can be run on separate cores/machines completely in parallel without any communication.

Once the data has been acquired from the simulations, we organize it into a matrix structure mapping J's to E's:

$$\begin{pmatrix} Ex \\ Ey \end{pmatrix} = x = Zy = \begin{pmatrix} Jx \\ Jy \end{pmatrix}$$

There are $M = N^2$ total E and J elements since the substrate grid is NxN as described earlier and thus Z is an MxM matrix.

Now, we assume one element in the grid is a source which provides input power for the antenna: $J_p = 1$ with $0 \leq p \leq M - 1$, and so:

$J_k^{src} = 1$ for k = p, 0 otherwise.

The resulting E fields on the substrate surface due to this source element are now:

$$E^{src} = ZJ^{src}$$

In order to enforce a PEC condition, the tangential E field must be 0 wherever there is PEC. This condition must be enforced by surface currents flowing on these PEC surfaces which generate their own fields. Thus, if we can find a current distribution over the locations where there should be PEC that exactly cancels the electric fields produced by the source element over the locations where there is PEC by the uniqueness property of Maxwell's equations we will have arrived at the same solution as if we had ran the initial simulation with PEC materials and the source.

To accomplish this, we define a matrix P which is used to extract the columns and rows of the Z matrix corresponding to the indices on the grid which should be PEC. Assume there are q PEC elements out of the M total elements and the set representing all of the indices of PEC elements is defined in a vector v_{pec}.

In Matlab notation:

$$J^{pec} = Z(v_{pec}, v_{pec}) \backslash E^{src}(v_{pec})$$

And the fields of the full system with the PEC and the source together are simply:

$$E^{total} = ZJ^{total}$$

where

$$J_k^{total} = J_k^{pec} \text{if k} \in v_{pec}, J^{src} \text{if k = p, 0 otherwise}$$

Again, because of uniqueness, the fields E^{total} are identical to the ones that would have been computed by performing an electromagnetic simulation with J^{src} and all of the indices defined by v_{pec} forced to be perfect conductors.

However, after the calculation of the Z matrix, only a single matrix solve of a small, dense matrix is required per evaluation of any arbitrary structure within the optimization design space. The largest that this matrix can be is MxM, which happens in the extreme case where every possible element on the design grid is forced to be a PEC material. Even in this situation, the system which needs to be solved is much smaller than the huge system which must be solved for determining

the solution to the full 3D Electromagnetic problem. In practice, standard solution techniques take on the order of milliseconds on a personal computer for each design evaluation, allowing an optimization loop to perform many iteration passes very rapidly.

Ultrafast Arbitrary Dielectric Structure Evaluation

In this section, we generalize the previous approach to allow the evaluation of arbitrary dielectric elements within the design grid. We begin by going back to the differential form of Maxwell's equations:

$$\begin{aligned} \nabla \times \mathbf{E} &= -j\omega\mu\mathbf{H} \\ \nabla \times \mathbf{H} &= j\omega\epsilon\mathbf{E} + \mathbf{J} \end{aligned} \tag{4.1}$$

These are continuous-space time-harmonic equations. In order for them to be solved numerically they must be discretized. Thus, E and H, and ϵ and J become vectors and the curl operators are also discretized into matrices:

$$\begin{aligned} \nabla \times \mathbf{E} &\longrightarrow D^E x_E \\ \nabla \times \mathbf{H} &\longrightarrow D^H x_H \end{aligned}$$

where D^E represents the discretized curl operator on the electric fields and D^H the discretized curl operator acting on the magnetic fields. In the case of the finite-difference method, these are usually simple centralized differencing operators. We can rewrite Maxwell's equations now as a linear system of equations:

$$\begin{aligned} D^E x_E &= -j\omega\mathbf{diag}(\mu)x_H \\ D^H x_H &= j\omega\mathbf{diag}(\epsilon)x_E + b_J \end{aligned} \tag{4.2}$$

Solving the first equation of eq.4.2 for x_H and substituting into the second one, we are left with a single equation for x_E:

$$\left[D^H \frac{\mathbf{diag}(\mu)^{-1}}{j\omega} D^E + j\omega\mathbf{diag}(\epsilon) \right] x_E = -b_J \tag{4.3}$$

Dividing every term by $j\omega$:

$$\left[-D^H \frac{\mathbf{diag}(\mu)^{-1}}{\omega^2} D^E + \mathbf{diag}(\epsilon)\right] x_E = -\frac{b_J}{j\omega} \tag{4.4}$$

Eq. 4.4 is a linear system of equations to solve for the electric fields, x_E, for a given electric current source excitation vector, b_J, for the system with permittivity and permeability vectors, ϵ and μ respectively. The system matrix $A = -D^H \frac{\mathbf{diag}(\mu)^{-1}}{\omega^2} D^E + \mathbf{diag}(\epsilon)$ and thus the electric fields can be found by solving: $x_E = A^{-1} \frac{b_J}{j\omega}$.

Let's start with the assumption that the system matrix A can be directly inverted. This assumption is usually not true as the system matrix A tends to be extremely large and sparse, but its inverse is usually dense, often making direct methods such as LU or LDL factorization intractable for even moderately sized 3D problems. However, starting with the assumption that we can obtain A^{-1} will motivate the approach for dealing with the situation where only iterative solves are feasible.

Assume now that $B = -D^H \frac{\mathbf{diag}(\mu)^{-1}}{\omega^2} D^E$ and thus $\left[B + \mathbf{diag}(\epsilon)\right] x_E = A x_E = -\frac{b_J}{j\omega}$. Now we have:

$$x_E = \left[B + \mathbf{diag}(\epsilon)\right]^{-1} \left(-\frac{b_J}{j\omega}\right)$$

In a typical optimization loop, a small amount of dielectric is changed from one step to the next, mathemtically implying that only a few elements of the ϵ vector must be modified. This motivates the use of the Woodbury Matrix Identity [ref], which efficiently allows small rank corrections to be applied to the inverse of a matrix:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

Define a matrix U which is a tall indicator matrix that maps the elements of the ϵ vector that we wish to modify (let's name this vector ϵ_Δ to the full-size ϵ vector which contains all of the values of permittivity for the full system (including the ones which remain unchanged.) Note that now: $A_{new}^{-1} = (A + U\mathbf{diag}(\epsilon_\Delta)U^T)^{-1}$ and thus substituing into Woodbury's matrix identity we have:

$$(A + U\mathbf{diag}(\epsilon_\Delta)U^T)^{-1} = A^{-1} - A^{-1}U(\mathbf{diag}(\epsilon_\Delta)^{-1} + U^T A^{-1}U)^{-1}U^T A^{-1}$$

Note that this correction or update to the matrix inverse for the modified ϵ vector only requires a matrix inverse of a matrix the size of the number of elements being modified. Assuming this number of elements is much smaller than the total size of the original system matrix (which for any realistic system should be a valid assumption), this approach allows updatingthe original solution to the new solution for the updated permittivity vector extremely efficiently as compared to re-solving the full system.

Now we consider the case where we can only obtain iterative solutions to the system of equations, i.e. we know x_E for a finite number of right hand side (RHS) excitation vectors, b_J. In the case of the current optimization and ultrafast PEC optimization algorithm, if we have N unknowns on the design grid, we would need to perform N individual solves (as mentioned before, these can all be run completely in parallel without communication) in order to obtain the Numerical Green's Functions required for defining arbitrary current distributions on the design grid or forming perfect metals. This same simulation data can also be used to define any arbitrary dielectric material over the same design region.

We define a matrix G which contains the set of column vectors $[b_{J0}, b_{J1}, ..., b_{JN-1}]$ solved for the optimization problem at hand. Thus, by performing N iterative solves of the original system, we have the matrix:

$$A^{-1}G = [x_{E0}, x_{E1}, ..., x_{EN-1}]$$

In the case of the impedance matrix required for determining quantities such as input power into the system for an antenna structure, only the field quantities collocated with the excitations are required and thus: $Z = G^T A^{-1} G$ is a square matrix.

Now, we revisit Woodbury's identity:

$$(A + U\mathbf{diag}(\epsilon_\Delta)U^T)^{-1} = A^{-1} - A^{-1}U(\mathbf{diag}(\epsilon_\Delta)^{-1} + U^T A^{-1}U)^{-1}U^T A^{-1}$$

Multiplying by G from the right and G^T from the left we have:

$$G^T(A+U\mathbf{diag}(\epsilon_\Delta)U^T)^{-1}G = G^T A^{-1}G - G^T A^{-1}U(\mathbf{diag}(\epsilon_\Delta)^{-1} + U^T A^{-1}U)^{-1}U^T A^{-1}G$$

U was defined as being a subset of the ϵ elements which can be modified, so by definition it can be written as a product of G multiplied by another arbitrary matrix, W: $U = GW$ and substituting back in:

$$G^T(A+U\mathbf{diag}(\epsilon_\Delta)U^T)^{-1}G = G^TA^{-1}G - G^TA^{-1}GW(\mathbf{diag}(\epsilon_\Delta)^{-1}+W^TG^TA^{-1}GW)^{-1}W^TG^TA^{-1}G$$

and since $Z = G^TA^{-1}G$:

$$Z_{new} = Z - ZW(\mathbf{diag}(\epsilon_\Delta)^{-1} + W^TZW)^{-1}W^TZ$$

And we have thus derived an update to the impedance matrix for any arbitrary change in the dielectric vector, ϵ_Δ which does not require ever having the inverse of the full system A^{-1} and just works on the original impedance matrix, Z.

Note that in the case of a perfect metal, $\epsilon_\Delta \to \infty$ and thus $\mathbf{diag}(\epsilon_\Delta) \to 0$, so the expression simplifies to:

$$Z_{new} = Z - ZW(W^TZW)^{-1}W^TZ$$

which is the same expression derived in the previous section for PEC's.

Design of a Real Metal Antenna Structure for Maximizing Broadside Gain on a Lossy Silicon Substrate

In this section, we use the ultrafast PEC structure simulation technique coupled with the coordinate ascent optimization algorithm to optimize an antenna on a realistic lossy silicon substrate for maximizing gain in the broadside direction. The same 1.74 x 1.74 x 0.217 mm 10 S/m silicon substrate with metal ground plane used for the current optimization examples in the previous chapter is utilized. This is because all of the Numerical Green's Function data acquired for those calculations can be reused here without any modication and also for comparison purposes to determine how close our heuristic optimization algorithm can approach the optimal maximum bound predicted. We will see that the resulting design closely approaches the maximum bound, especially when multiple excitation drive points are allowed.

The coordinate ascent algorithm [3] is perhaps one of the simplest local optimization algorithms, yet in many situations can yield very effective results. The algorithm

has the advantage of being very simple and fast. A random pixel on the optimization grid is selected at the beginning of each iteration and flipped. If the performance of the objective function increases, the change is accepted otherwise it is flipped back to the original position. The algorithm is run until stagnation, which occurs once no single flip of any element on the optimization grid results in an improvement in performance.

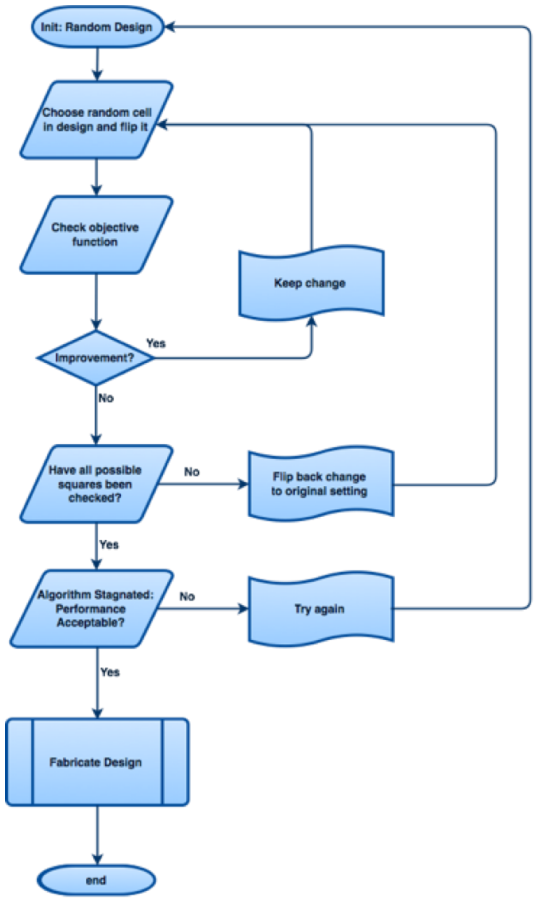

Figure 4.4: Flow Chart Illustrating Coordinate Ascent Optimization Algorithm

Improvement Factor

In our implementation of the coordinate ascent algorithm, we add an extra parameter named the "improvement factor." The improvement factor starts at 1 and is decreased by 1% every iteration. When it is 1, flipping a grid cell must improve the gain by 100% (2x) in order for the change to be accepted. Once it has reached 0, any change which improves the gain, even slightly, is taken. The motivation behind the improvement factor parameter is to reduce noise in the resulting design by first preferring design choices which improve the objective function the most.

Source/Drive Point Selection

From the current optimization section, we observed that the maximum possible broadside gain for this silicon substrate configuration was achieved with two horizontal line current distributions on the top and bottom edge of the substrate. The maximum current magnitude is in the center of each of these horizontal line currents and they are both in the same phase with respect to each other, so we postulate that these two points would be good feed points for a real metallic antenna design. Thus, we start the optimization code with an empty design grid and these two sources at the prescribed points as shown:

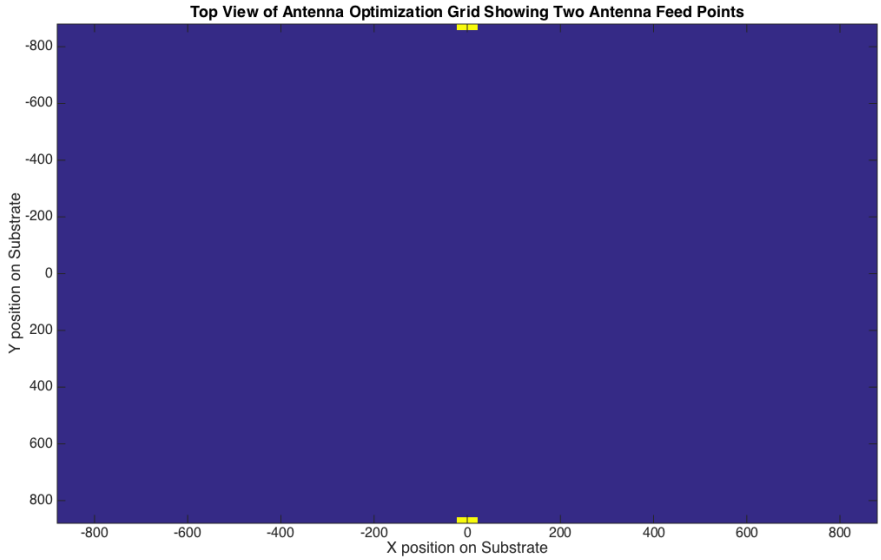

Figure 4.5: Top View Illustration of Si Substrate Surface Indicating Initial Empty Design Grid with Two Antenna Feed Sources

We run the coordinate ascent with improvement factor modification optimization algorithm with two different approaches for selecting grid cells to flip:

1) Uniform distribution: Every cell on the grid has exactly the same probability of being selected to be flipped.

2) Weighted by globally optimal current distribution: The globally optimal current distribution is first determined using the approach in the previous chapter and the grid cell to be flipped is drawn from a probability distribution weighted by the magnitude of the optimal current strength at each point.

(1) is simply the "base" algorithm and the intuition behind (2) is that areas with higher current density in the optimal current distribution are more likely to have a larger effect on the objective function than low or empty areas.

Fig. 4.6 compares these two random cell selection approaches:

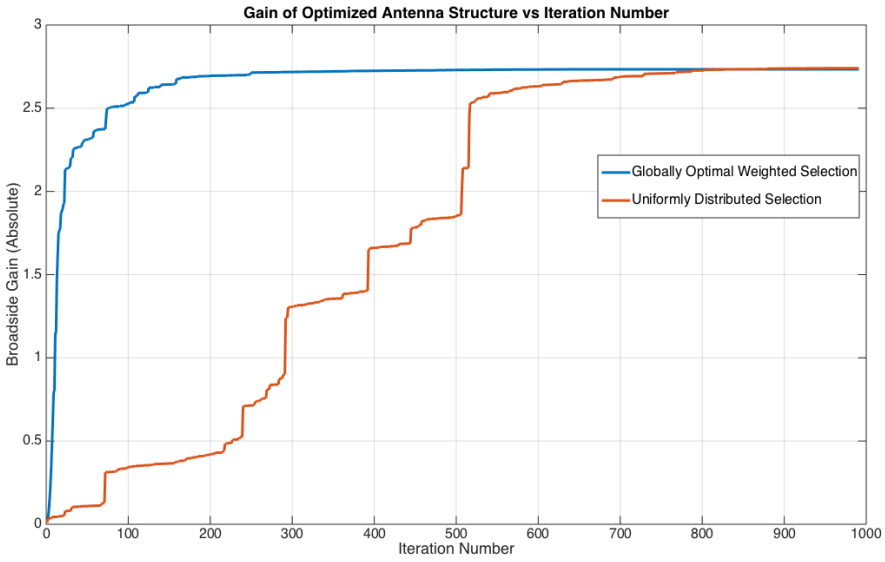

Figure 4.6: Absolute Broadside Gain vs Iteration Number for Metallic Antenna Optimization Utilizing Coordinate Ascent Algorithm with Two Different Random Cell Selection Approaches

It is readily apparent that the selection approach which weighs the probability distribution with the magnitude of the optimal current distribution converges much faster (about 4 times or 200 iterations vs 800) than the uniformly distributed approach. Both approaches eventually approach almost the same performance before algorithm stagnation and the resulting structures appear very similar in shape:

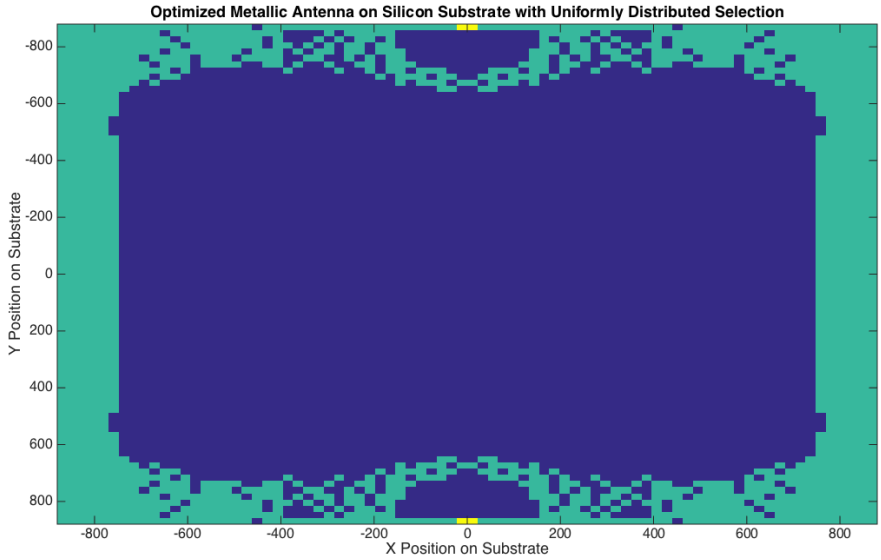

Figure 4.7: Result of Coordinate Ascent Algorithm Used to Optimize Metal Antenna over Lossy Silicon Substrate for Maximizing Broadside Gain using Uniformly Distributed Random Cell Selection Approach

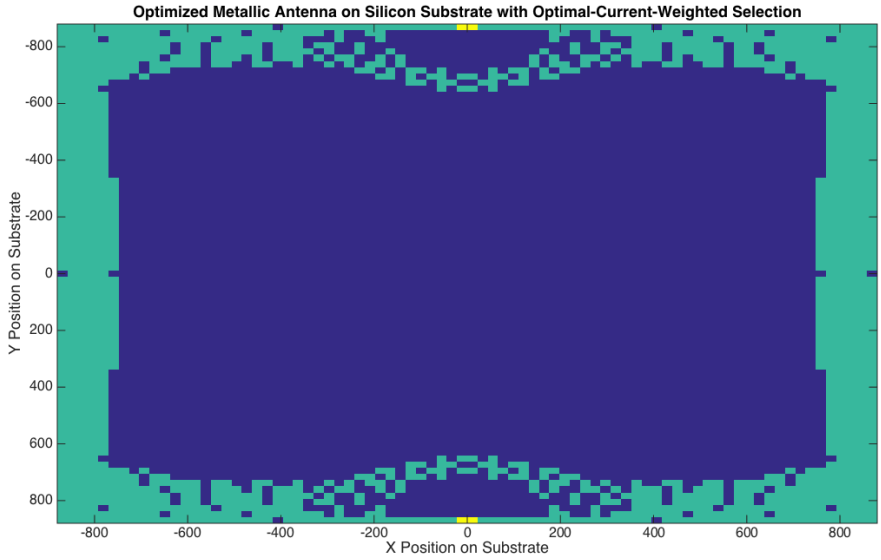

Figure 4.8: Result of Coordinate Ascent Algorithm Used to Optimize Metal Antenna over Lossy Silicon Substrate for Maximizing Broadside Gain using Optimal-Current-Distribution Weighted Random Drawing Approach

Antenna Cleanup Algorithm

While the antenna designs generated by the coordinate ascent algorihtm in the previous section can yield impressive performance, the resulting structures often look "noisy" and may contain superfluous blocks of metal which are not critical towards the performance of the final design. This was the main motivation behind the introduction of the "improvement factor" parameter; however, even the improvement factor does not completely solve the issue.

In order to clean up the final solution of an antenna structure designed using the algorithm in the previous section, we implement a simple post-processing algorithm. Unlike the design algorithm, the cleanup algorithm is only allowed to **remove** pieces of metal and cannot flip cells or insert more metal into the design. The algorithm proceeeds by selecting a random piece of metal to remove each iteration. If the performance of the objective function degrades by less than a prescribed amount (we found empirically that 0.1% works well) by removing the metal, the change is kept and the algorithm proceeeds. The algorithm stagnates when there are no more pieces of metal in the design grid which can be removed that lead to a reduction in performance of less than the desired amount (0.1% in our work.)

Fig. 4.9 illustrates the slight degradation in the gain due to running the cleanup algorithm. The gain starts at 2.74 (4.4dBi) and drops to 2.67 (4.3dBi) after the cleanup process has been completed:

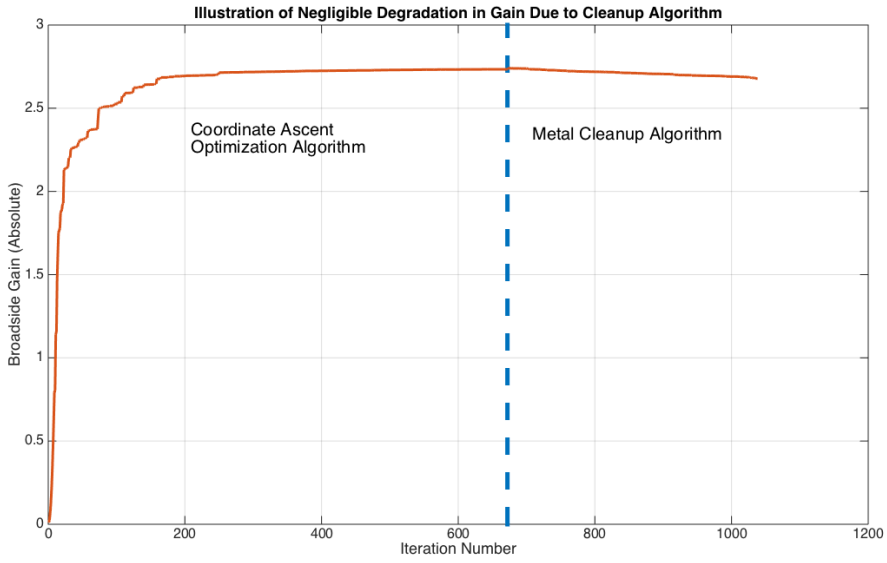

Figure 4.9: Objective Function vs Iteration Number

For only a slight penalty in gain, the resulting antenna structure is significantly simpler, consumes considerably less chip surface area, and is more readily fabricated as seen in Fig. 4.10:

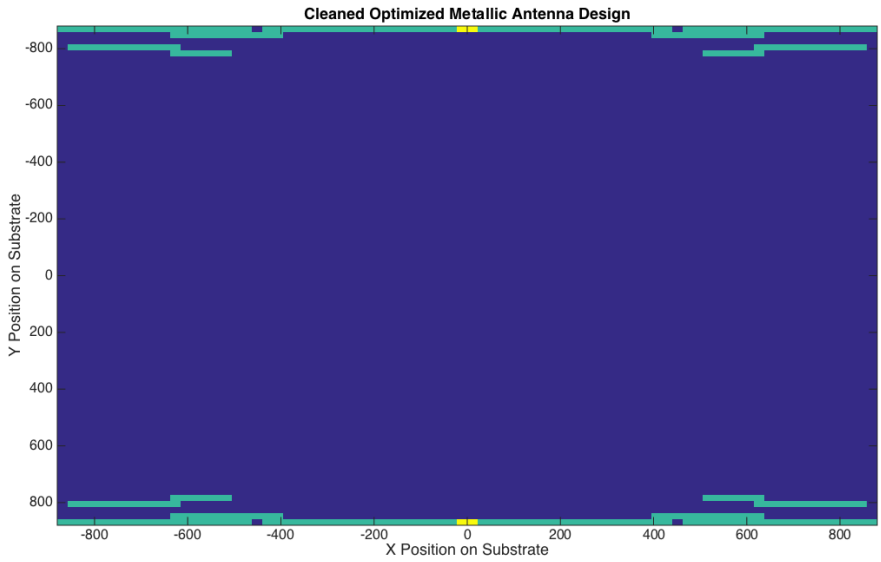

Figure 4.10: Result of Metal Cleanup Algorithm

Realtime Painting of Real Metal Antenna Structures with Instant Gain Evaluation

In this section, we develop a tool which allows designing metal antennas on a dielectric substrate surface with instantaneous evaluation of performance metrics after each structural change. This tool ultimately demonstrates the speed of the Ultrafast PEC Structure Evaluation method as compared to standard electromagnetic simulation techniques and allows rapid design insight to be developed by a human designer by allowing them to quickly iterate and experiment with many different design iterations.

The tool displays a top view of the antenna dielectric substrate and allows both PEC material and current sources (antenna feed points) at any point on the design grid in realtime by using the mouse. After each change to the design, the peformance of the modified antenna is computed in realtime and displayed to the human designer on a millisecond timescale. Furthermore, the algorithm computes the resulting gain of the antenna for the situation where all of the sources are driven with the same amplitude and phase and also when the sources are all driven at their optimal

amplitude and phase conditions (the optimal drives are computed using the approach in the previous chapter.)

We take the final resulting, clean antenna structure from the cleanup algorithm of the previous section and manually adjust it and add extra antenna feed points to improve the performance further. The resulting antenna has a gain of 6dBi in the broadside direction and a radiation efficiency of 60% (verified with Ansys HFSS) which is quite close to the theoretical maximum gain achievable (7dBi.) The optimal drive amplitudes are the same for all of the antenna ports and phases of 0, 45, and 90 degrees are required, which can be readily generated on-chip using phase-rotator circuitry. Fig. 4.11 illustrates this design:

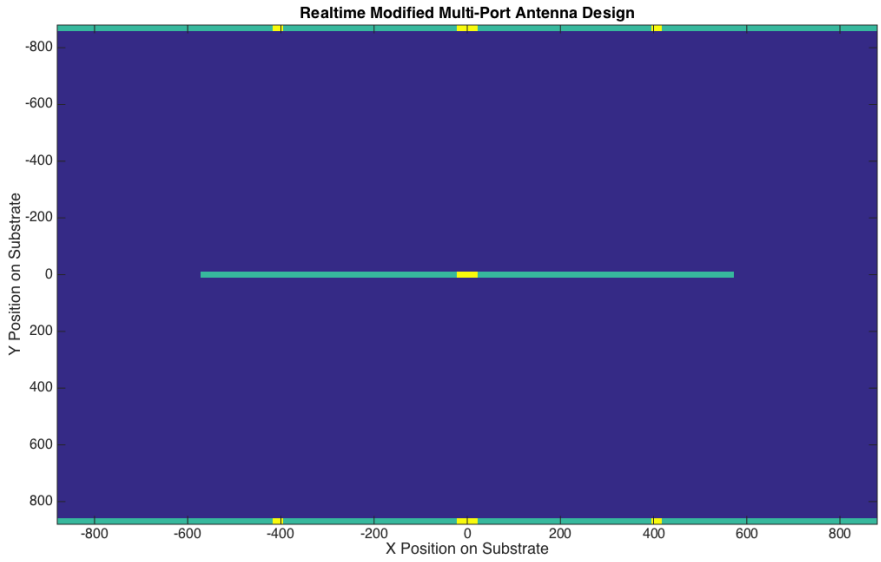

Figure 4.11: Manually Realtime Adjusted Multi-Port Metal Antenna Achieving a Broadside Gain of 6dBi and Radiation Efficiency of 60%.

Up to this point, we have focused all of our efforts on on-chip metallic antenna design and optimization. In the next chapter, we explore the versatility of the techniques and algorithms developed in this chapter by applying them to optimize metallic coupling antennas for coupling energy into and out of dielectric waveguides.

Utilizing the optimization techniques developed in the previous chapters, we implement an automated optimization algorithm which generates de-novo electromagnetic structures under specified design objectives. We demonstrate a particular design of a surface coupling antenna which is capable of launching energy efficiently into the main mode of a plastic waveguide. We use a standard 3D printer to print the designed system in order to verify our simulation results. The maximum available power gain of the system was measured as 0.128, implying that the coupler achieves an efficiency of at least 36% at the designed center frequency of 9.6GHz. The actual coupler efficiency is believed to be much higher because the polymer used is very lossy in this frequency range.

5.1 Introduction

Plastic waveguide based serial links are recently emerging as efficient, cost-effective alternatives to traditional copper wire based systems. [26, 104] Wireline links typically become extremely inefficient for distances approaching and surpassing 1m. While optical fibers can be efficient over long distances, optical links can be very expensive, do not integrate easily with standard CMOS processes, and require expensive alignment protocols. In contrast, plastic waveguides offer the advantages of low loss across medium distances (1-10m) and low integration costs since they can be excited electrically with metal coupling antennas. Unless the link is very long, the majority of the loss is dominated by the transmit and receive coupling losses. Furthermore, the overall power efficiency is the square of the individual coupler efficiency due to the TX and RX couplers. Therefore, it is crucial to utilize efficient couplers to transfer incident power effectively into the fundamental waveguide mode. In this chapter, we discuss the design of a custom, optimized surface coupler, which is 3.5x more efficient than a standard dipole antenna, leading to a greater than 10x improvement in link efficiency.

5.2 Optimization Algorithm

Our optimization algorithm works on a structured grid of unit cells which can either be filled in to represent a patch of metal or left empty. Due to mode profile symmetry, a good coupler should also be symmetric. We reduce the design space by optimizing one “wing” of the coupler and mirroring it for the other side. Since each cell can

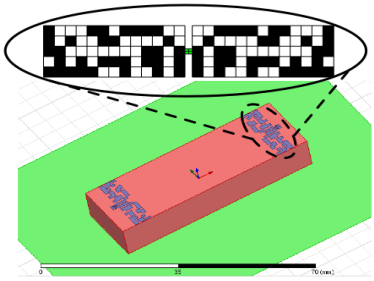

Figure 5.1: Model of Waveguide Link with Zoomed-in Diagram of Coupler Design

either be filled with metal or left empty, there are 2^N possible configurations for a grid of N cells. Given fixed waveguide dimensions, the number of cells that can be used is ultimately limited by the fabrication resolution. For our specific implementation we utilized a 13×5 cell grid per coupler wing which results in $2^{65} \approx 3.8 \times 10^{19}$ possible configurations. Due to the $0-1$ nature of the problem, it can be shown that it is NP hard, meaning that no known algorithms exist which can guarantee finding a global minimum faster than an exhaustive search. Even for our relatively "small" design problem, going through 2^{65} designs would require an intractable number of EM simulations. Although it is a very difficult problem, many heuristic algorithms exist for attacking $0-1$ problems including gradient based approaches and genetic algorithms. For our specific problem, finding a suitable gene crossover mechanism is tough and while we tried various techniques, none performed nearly as well as an even simpler optimization strategy: random coordinate ascent. At each iteration, we flip a random cell and keep the change if the performance improves, otherwise we revert the flip and repeat until convergence has been reached. We found that seeding the algorithm with a random solution performs best, although starting with "empty" and "full" designs yields acceptable results as well. We used our own custom FDTD-based 3D EM simulator coupled with the ultra-fast PEC simulation technique designed specifically for the specific problem at hand to optimize the ratio of the power coupled into the fundamental waveguide mode to the input coupler power at a single frequency. The resulting coupler is resonant because the optimization metric was enforced at a single frequency. Broadband couplers can be generated with multi-frequency objective functions. Fig. 5.2 shows the ascent of the coupling efficiency vs. iteration for a typical run.

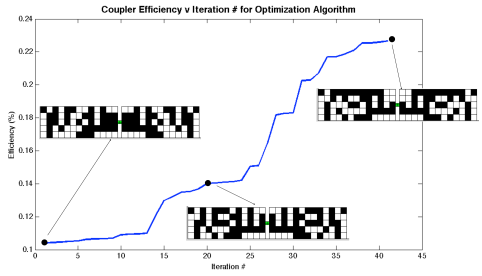

Figure 5.2: Coupler Efficiency vs Iteration with Intermediate Designs Overlayed

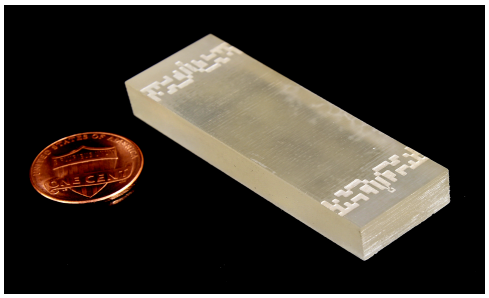

Figure 5.3: Photograph of 3D Printed Design

5.3 Design Implementation

We used a Projet HD 3000 3D printer to print the link including the guide polymer (VisiJet Crystal EX 200) and the grooves for the coupler, which were filled with silver paste ($\sigma \approx 2.6 \times 10^6$) as described in [118]. We optimized for a guide with cross-section close to $\lambda \times \lambda/2$ where λ is the wavelength inside the material. The max resolution of the printer is $\sim 200\mu m$, so we chose guide dimensions of $18.8 \times 9.4mm$. This corresponds to a center frequency near $9GHz$ in a guide with $\epsilon_r = 3.2$. The dielectric properties are not provided by the manufacturer for the polymer used, but we were able to achieve a good match between measurement and simulation data by assuming values of $\epsilon_r = 3.2$ and $\tan\delta = 0.04$. These values seem reasonable according to measurements of other printer polymers at 10GHz. [20] 4-port S-parameters were measured with calibrated RF GSSG differential probes from 25MHz to 25GHz using a Agilent N5242A VNA.

5.4 Measurement Results

The final design measured $19.2 \times 9.5 \times 56mm$ and the couplers were spaced $4mm$ from each end and measured $18.8 \times 4.7mm$. Short feedlines connect the GSSG pads to the coupler feed-points. The measured S-parameters are converted to 2-port differential parameters and plotted in Fig. 5.4. Since we targeted only coupling

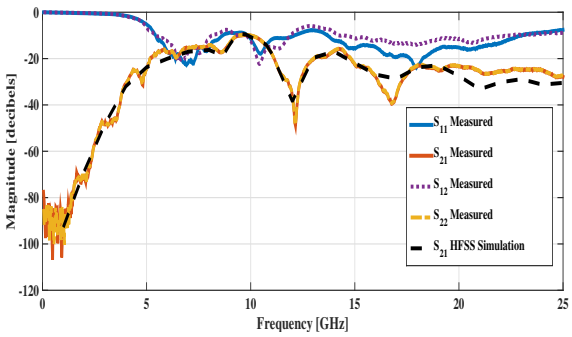

Figure 5.4: Measured S-Parameters of 3D Printed Device

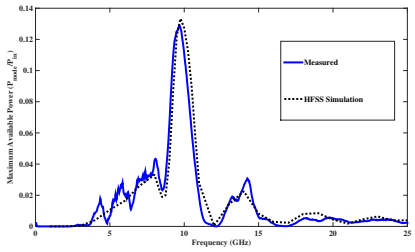

Figure 5.5: Measured and Simulated Maximum Available Power

efficiency at $9.6GHz$ in our algorithm, we compute the maximum available power gain at each measured frequency from the measurements and compare against HFSS simulation (Fig. 5.5) The peaks between measurement and simulation line up closely. Discrepancies can be explained by uncertainty in the material properties. The link efficiency peaks to 13% at $9.65GHz$. In a lossless system, this corresponds to a 36% coupling efficiency. However, the polymer and silver paste used are very lossy and we have noticed from our simulations that the majority of the losses come from the guide material, suggesting that the coupler itself is significantly more efficient and that a different plastic should be used for the next prototype. HDPE, for example, has been measured to have $\tan\delta < 0.002$ at these frequencies and would make a good candidate for a second prototype.

5.5 Conclusion

We have developed a fast, yet simple optimization algorithm to design surface coupling antennas for dielectric waveguides de-novo and briefly detailed the main loop of the algorithm. In order to test the validity of the generated designs, we used a 3D printer to print the waveguides using polymer material and deposited silver paste for the coupler structures. Our measurement results match well with the

HFSS simulated data, and based on our simulations we expect that we can achieve significantly improved link efficiency simply by switching to a different polymer material for the waveguide.

In this chapter, we discuss the implementation of a 120GHz center frequency transceiver in TSMC's 28nm bulk CMOS process. This is an on-going work in progress and is being designed in collaboration with Prof. Ali Niknejad, Andrew Townley, and Sashank Krishnamurthy of the Berkeley Wireless Research Center (BWRC). The transceiver is designed to be capable of driving off-chip antennas for wireless communication as well as waveguide couplers for a wireline based link. Two identical Transmit(TX)+Receive(RX) channels are included per chip to enable multi-polarization (for antennas) and multi-mode (for waveguides) communication. Using multiple modes or polarizations essentially enables doubling of the communication rate within the same form-factor. More than two modes can be utilized for waveguides, although care must be taken to avoid mode to mode coupling/interference and this has not been studied much yet or covered within this work. The chip design has not yet been completed or fabricated and thus this chapter contains information detailing the design architecture fundamentals and a few simulation results.

6.1 Design Architecture

Fig. 6.1 shows a system level block diagram of the whole chip design. A direct conversion approach is used for both upconversion and downconversion. An injection locking frequency chain generates a 120GHz oscillator signal for both TX and RX modules which is locked to an external 30GHz reference input. The following sections will describe the design of the TX, the RX, and the frequency locking network.

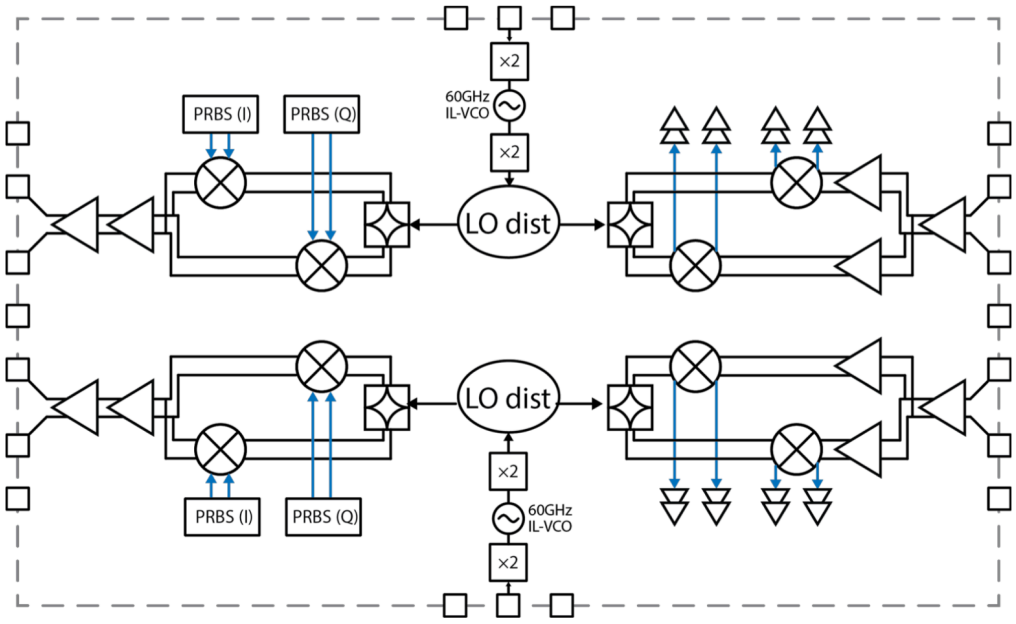

Figure 6.1: System Level Block Diagram of Dual Channel 120GHz Transceiver IC

Frequency Locking Network

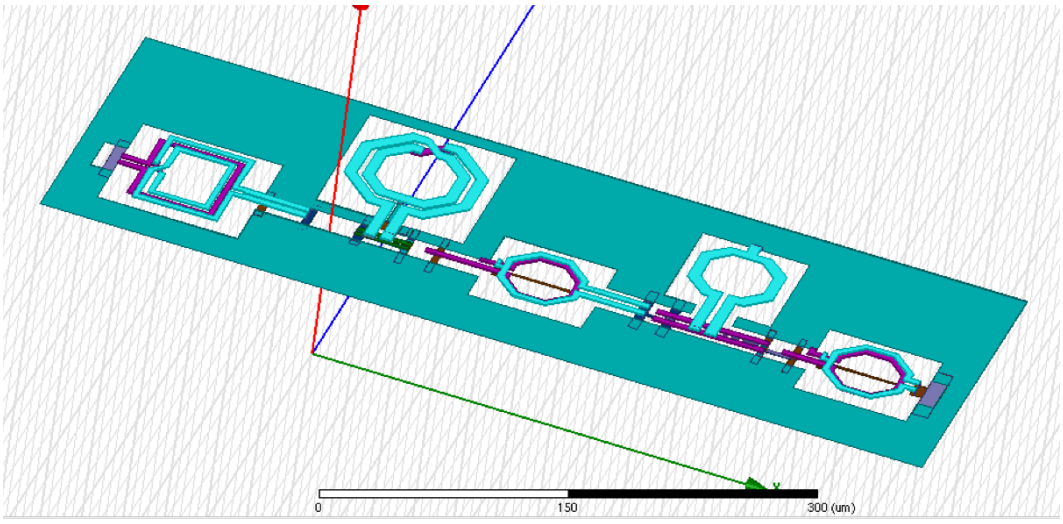

Figure 6.2: 3D Rendering of Frequency Chain Layout

An off-chip, nominally 30GHz single-ended reference signal is brought in externally through a bump. Fig. 6.2 shows a circuit level diagram of the locking network from input to 120GHz output. A 1:2 transformer is used as a balun to convert the input reference to a differential signal and to obtain some voltage gain. The secondary also has a center-tap for biasing the gate inputs of the 30GHz amplifier stage directly

after the input transformer. The 30GHz amplifier is a simple capacitively neutralized common source differential amplifier with a drain inductor for resonating and maximizing the gain at 30GHz. The 3dB bandwidth of this amplifier is +-5GHz around the center point (Fig.[bleh]) and it consumes 5mW of DC power.

The outputs of the 30GHz input amplifier stage drive a differential push-push doubler circuit which produces a 60GHz output. The gates of the doubler devices are DC biased at 275mV, below their threshold, and sized optimally to maximize second harmonic output current. A 1:1 transformer is used to resonate the drain capacitance of the doubler circuit, provide a differential 60GHz output, and match to the inputs of the injection devices of the next stage. The 30 to 60GHz doubler consumes 6mW of DC power and generates nominally a 500mVpp differential 60GHz signal.

The 60GHz differential signal generated by the doubler is fed into the gates of a differential pair which injects current into a 60GHz differential cross-coupled VCO. The 60GHz uses a 90pH inductor and 4 bits of a digitally switched capacitors for tuning the frequency from 57GHz to 65GHz. The injection differential pair injects the 60GHz reference current into the VCO tank in order to injection lock the oscillation frequency to the input reference.

The VCO differential output signal is fed into another doubler circuit, which is identical to the first doubler, that doubles the 60GHz input to 120GHz. An output transformer is used, as before, to resonate the drain capacitance of the doubler and generate a differential 120GHz signal to be fed to LO buffers and distributed to the TX and RX blocks of the transceiver system. A 200mVpp 120GHz signal is nominally generated at this point across the expected LO buffer input.

Transmitter

Fig. 6.3 illustrates a circuit level block diagram of the transmitter section. A 120GHz differential LO signal is taken from the output of the frequency locking chain, buffered with a common-source differential pair and fed into a passive differential quadrature hybrid.

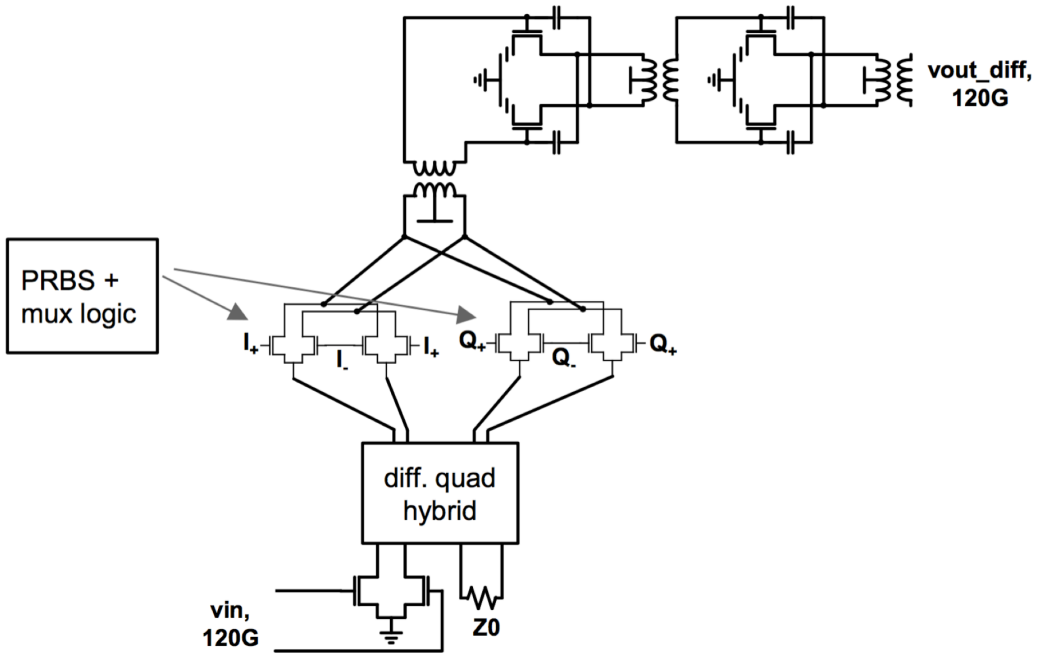

Figure 6.3: Circuit Level Block Diagram of 120GHz Transmitter

The quad hybrid (Fig. 6.4 generates 0 and 90 degree differentially and uses some lumped capacitance so that the design can be shrunk and fit into a smaller layout form factor.

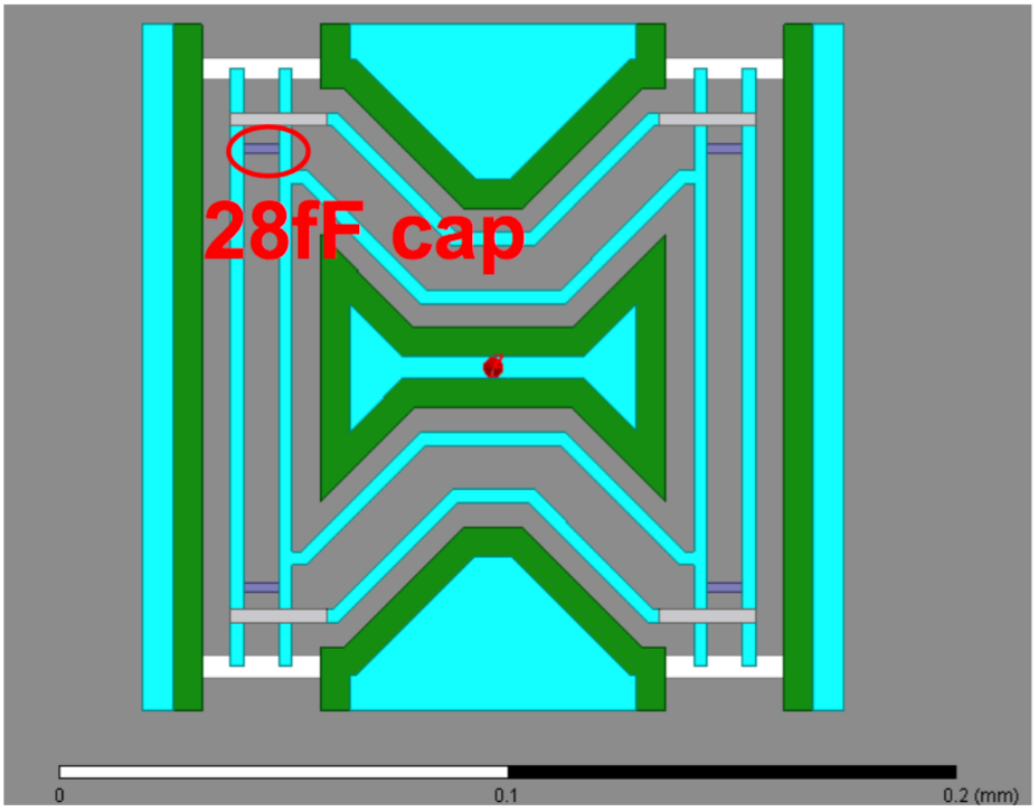

Figure 6.4: Differential Quadrature Hybrid for 0 and 90 Degree LO Phase Generation

The resulting 0 and 90 degree phase differential LO signals are fed into the sources of segmented modulator switch devices. The modulator devices are segmented to support QPSK, 16QAM, and 64QAM operating modalities based on the patterns provided by the PRBS circuit block. Finally, the modulated 120GHz output signal is transformer coupled into a 2-stage 120GHz Power Amplifier (PA) circuit. The two PA stages are identical and are common-source differential pairs which are capacitively neutralized to maximize stability. The output of the first PA stage is transformer coupled into the input of the second stage and the transformer is designed to resonate the parasitic cap of the devices at 120GHz and provide adequate inter-stage impedance matching. Finally, the output of the second PA stage is transformer coupled to differential SGS bump pads designed to drive an off-chip 100ohm differential load. Again, the output transformer is designed to resonate the drain parasitics of the second PA stage and provide good impedance matching conditions for maximizing power transfer to the output load. The PA consumes 16mW of total DC power off a 1V supply, and provides 13dB of nominal gain with a 3dB bandwidth of 40GHz centered around 120GHz. The nominal output power

of the TX block is 5dBm.

Receiver

The first stage of the receiver block after the input is the Low Noise Amplifier (LNA). The 100 ohm differential input is transformer coupled into the gates of the first LNA stage. The LNA consists of two capacitively-neutralized common source differential amplifier stages with resonated drain capacitance for centering the gain at 120GHz. As with the PA stage of the TX block, the output of the first stage is transformer coupled into the gate inputs of the second LNA stage. The transformer enables impedance matching the output of the first stage to the input of the second stage in a compact manner and further simplifies biasing of the stages by utilizing a center-tap tied to the supply rail. The LNA has a small signal gain of 18dB centered at 120GHz with a 40GHz 3dB bandwidth. Both stages together consume 33mW of power from a 1V supply. The second stage as can be seen in Fig. 6.5 is segmented into two separate stages with the inputs tied together but with separate drain outputs. This is to generate and provide isolation for the I and Q signals which feed into the demodulator/downconverter circuit.

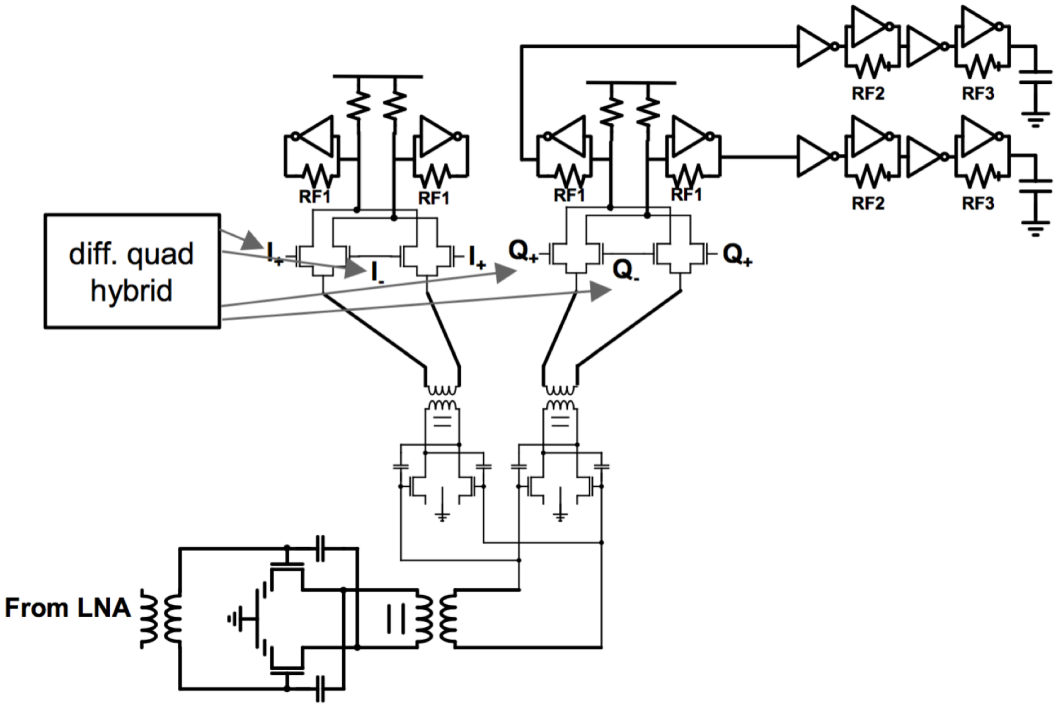

Figure 6.5: Circuit Level Block Diagram of 120GHz Receiver

The outputs of the I and Q LNA input stages are transformer coupled into the sources

of the I and Q mixer switches respectively. In this case, transformer coupling allows for the sources of the mixer devices to be DC grounded. The same differential quadrature hybrid design is used as in the TX block for generating 0 and 90 degree 120GHz LO signals from the LO buffer and fed into the gates of the demodulator switch devices.

The downconverted output is then amplified by a cascade of Cherry-Hooper broadband baseband amplifiers. The final amplifier stage is a buffer capable of driving an off-chip 100 ohm differential load. Each Cherry-Hooper stage is AC coupled to the subsequent one in order to avoid common-mode offset issues and to avoid the need of a common-mode offset cancellation feedback loop.

Compact power splitters designed ab initio using binary particle swarm optimization in a two dimensional mesh for a standard foundry silicon photonic platform are studied. Designs with a 4.8 $\mu m \times 4.8\ \mu m$ footprint composed of 200 nm × 200 nm and 100 nm × 100 nm cells are demonstrated. Despite not respecting design rules, the design with the smaller cells had lower insertion losses and broader bandwidth and showed consistent behaviour across the wafer. Deviations between design and experiment point to the need for further investigations of the minimum feature dimensions. This work has been done in collaboration with Prof. Joyce Poon and Jason Mak from the University of Toronto.

7.1 Introduction

Foundry fabricated silicon (Si) photonics seeks to implement highly sophisticated photonic integrated circuits (PICs) at low cost by using the mature manufacturing process of microelectronics [99, 61, 5]. The high refractive index contrast in Si photonic platforms not only allows for compact device footprints, but also makes possible device concepts that can take advantage of the strong optical confinement and scattering (e.g., grating couplers, micro-resonators, photonic crystals) [35, 23, 64]. The growing availability of foundry Si photonics, in combination with expanded computation capabilities for detailed electromagnetic simulations, increases the opportunity to explore device designs that cannot be implemented in traditional, lower index contrast PIC platforms such as silica and compound semiconductors, and are yet volume manufacturable.

Device design performed by topology optimization without any a priori assumptions on the geometry has recently generated much interest [98, 25, 89, 65]. As opposed to conventional design methodologies in which a few critical geometric parameters are tuned on a fixed geometry, topology optimization can find unexpected solutions with good performance within demanding constraints by exploring much larger parameter spaces. An example is the polarization beam splitter of [89], which had a design footprint constraint of 2.4 Œ 2.4 μm^2. However, usual optimization approaches (e.g., in [25, 89, 65]) rely on high resolution rendering of intricate geometric features, such as through electron-beam lithography, which can result in designs that are incompatible with the design rules and minimum feature sizes in

foundry processes which use deep ultraviolet (DUV) photolithography.

In this work, we investigate foundry fabrication of an optimization designed 2 Œ 2 3-dB power splitter. Power splitters area common building block in PICs, and in Si photonic platforms are typically implemented as multi-mode interference (MMI) couplers, directional couplers, and adiabatic couplers. Typical footprints in a standard Si photonic platform have footprints around 39 Œ 5.2 μm^2 for a 3-dB directional coupler, and 158 Œ 4.1 μm^2 for an MMI coupler [66]. Because power splitters may be instantiated many times in a PIC, a size reduction of the 2 Œ 2 splitter can save substantial circuit area. Here, we explore the design and implementation of 2 Œ 2 power splitters with a footprint constraint of 4.8 μm Œ 4.8 μm designed through optimization that accounted for the minimum feature size of the foundry process. We will first briefly describe the Si photonic platform and the optimization problem setup. Two device variants and their measurements are presented, followed by discussions on directions for improvement.

7.2 Process Description

The designs to be described were implemented in the A*STAR IME Baseline Silicon photonics platform [63, 86], which provides a 220 nm thick silicon layer with a 2.1 μm top oxide cladding and a 2 μm buried oxide, and partial etches for rib waveguides and grating couplers. The process uses 248 nm DUV photolithography on 8" silicon-on-insulator (SOI) wafers. The 2 Œ 2 power splitter is designed for the fully etched 220 nm thick Si layer. To impose the symmetry expected of a 2 Œ 2 3-dB power splitter, the design region is a quadrant of the device with a feed waveguide which is reflected horizontally and vertically to constitute the complete device with 4 ports, as illustrated in Fig. 7.1(a).

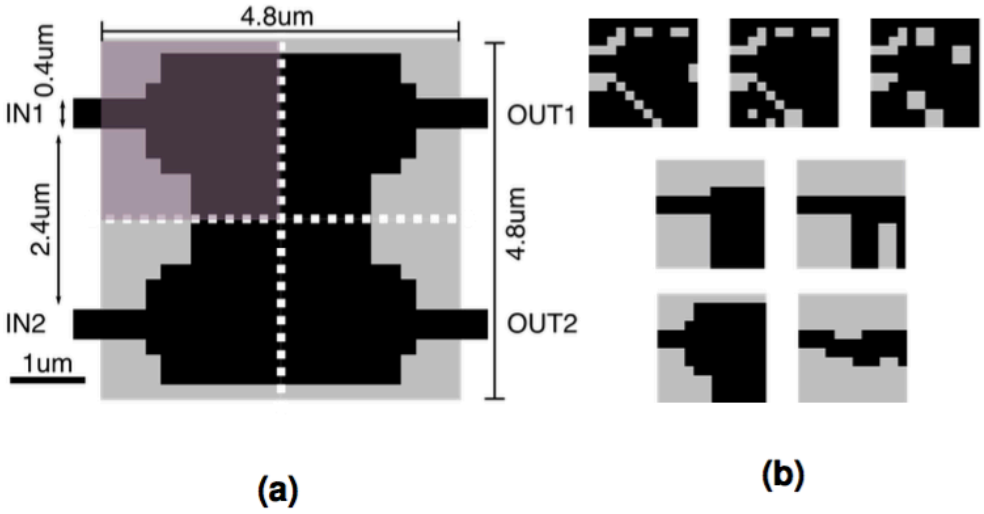

Figure 7.1: (a) Schematic of the design problem. The design area (highlighted, top left) is reflected vertically and horizontally along the dotted white lines along with an input port. Black and grey cells represent the presence and absence of material respectively. In FDTD simulations, a TE0 mode is launched into IN2 and the TE0 mode power is monitored at OUT1 and OUT2. (b) Initial configurations for the optimization of the 200 nm cell design. The initial population is based on these shapes with randomly initialized particle swarm velocity.

A square mesh is applied on the design region and a binary variable is assigned to each square mesh cell to indicate the presence of Si or SiO_2 in the cell. This parameterization of the design region is commonly used in structural optimization [9, 96], and has also been applied to electromagnetic design problems [98, 89, 95]. Because the smallest feature in such a design will be a single cell, we use the heuristic of taking the cell side length to be the minimum manufacturable feature/spacing size. The platform had design rules similar to [35], which had a nominal minimum feature size of 180 nm. Therefore, we first chose to optimize based using a mesh composed of cells with a size of 200 nm Œ 200 nm to satisfy design rules.

7.3 Optimization and Design Approach

We used the binary particle-swarm optimization (BPSO) algorithm [54] to optimize the binary-valued configurations of the cells. The particle swarm is transformed from a continuous configuration space to a discrete configuration space through treating each binary variable as a dimension and applying thresholding. BPSO was chosen for its ease of implementation, and for its reported applicability to similar

problems [13, 52, 67]. To design the device, we used BPSO to optimize the figure of merit:

$$f(x) = \min\left[P1(x, 1550nm), P2(x, 1550nm)\right]$$

where x is the configuration of cells of the design and P1(x, λ) and P2(x, λ) are the transmitted power in the fundamental mode of the transverse electric polarization (TE0) at the output ports, OUT1 and OUT2, for a TE0 input at IN2 for a given configuration and wavelength, λ. This is shown in Fig. 7.1(a). Eq. 1 heuristically encourages 3-dB power splitting and minimization of the insertion loss by improving the worse performing output. Each evaluation of the objective function entailed a three dimensional (3D) finite difference time domain (FDTD) simulation of the design. A FDTD simulation mesh size of 50 nm Œ 50 nm Œ 40 nm with a power cutoff condition of 0.001% of initial energy was used. As a compromise between the size of the configuration space and tractability of the simulations, we constrained the optimization to a 4.8 μm Œ 4.8 μm area, using 12 Œ 12 cells. Each iteration required approximately 10 seconds on an Intel i7-3770 CPU computer with 16 GB of RAM.

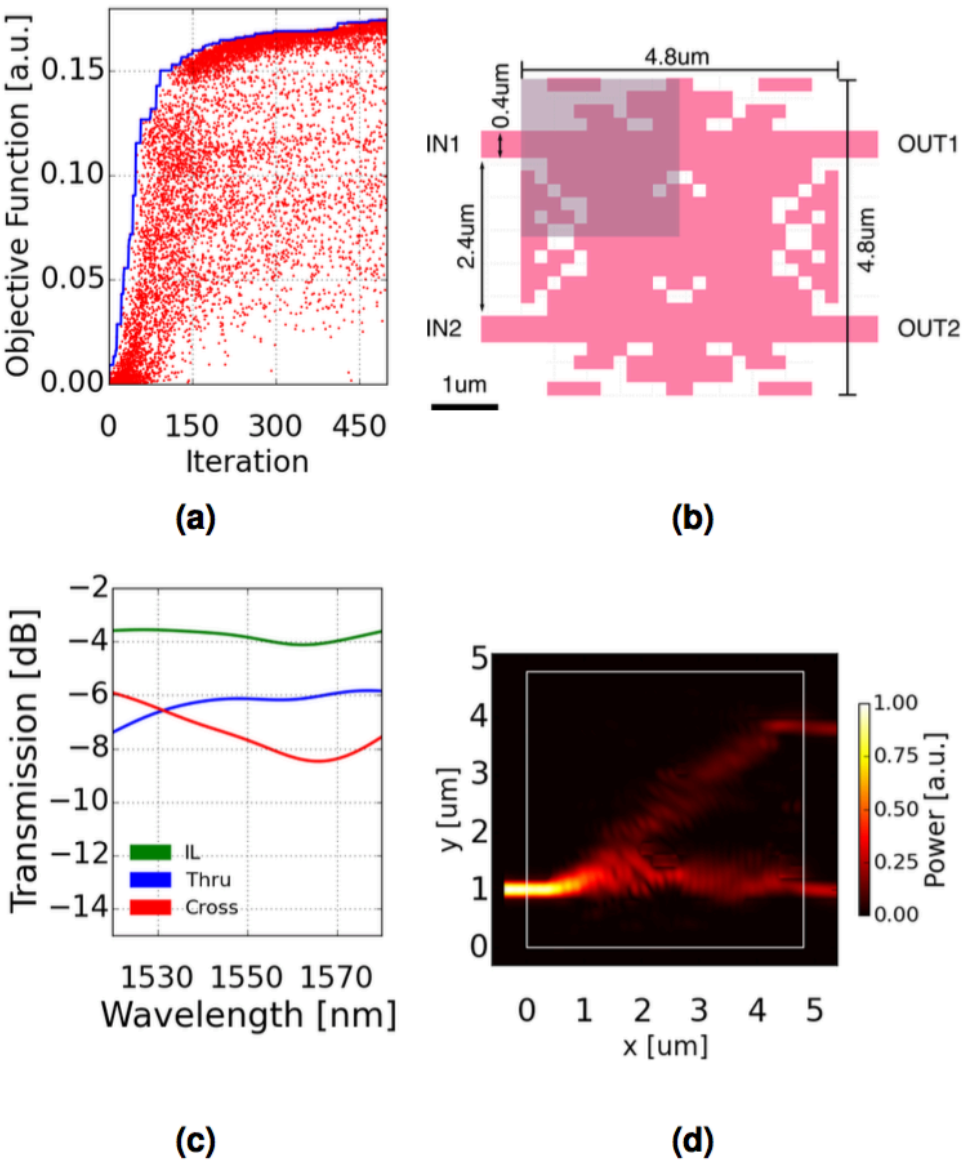

Figure 7.2: (a) The figure of merit at each iteration for the device with the 200 nm cells. Red dots are the individual values, and the blue line is the best value obtained by that iteration. (b) The device design with the highest figure of merit value of 0.174. (c) Simulated transmission and loss spectra of (b). (d) Normalized intensity profile of the device.

The design for 200 nm cells was optimized based on an initial population of 21 using 3 repetitions of the heuristic configura- tions in Fig. 7.1(b) with randomized velocities and run over 500 iterations. Fig. 7.2(a) shows the convergence to a figure of merit value of 0.174 and the resulting design is shown in Fig. 7.2(b).

7.4 Results

The transmission spectrum of the simulated device in Fig. 7.2(c) shows a wavelength dependent splitting ratio, crossing over at 1531 nm with insertion loss of 3.56 dB and worst insertion loss of 4.11 dB. The optimization resulted in a design with an unconventional geometry, confined within the 4.8 μm Œ 4.8 μm design region. The normalized power profile in Fig. 7.2(d) shows that the device did not operate as an purely as an MMI or a directional coupler. This design is highly compact since extra bend-in waveguides are not required because of the large spacing between the input ports.

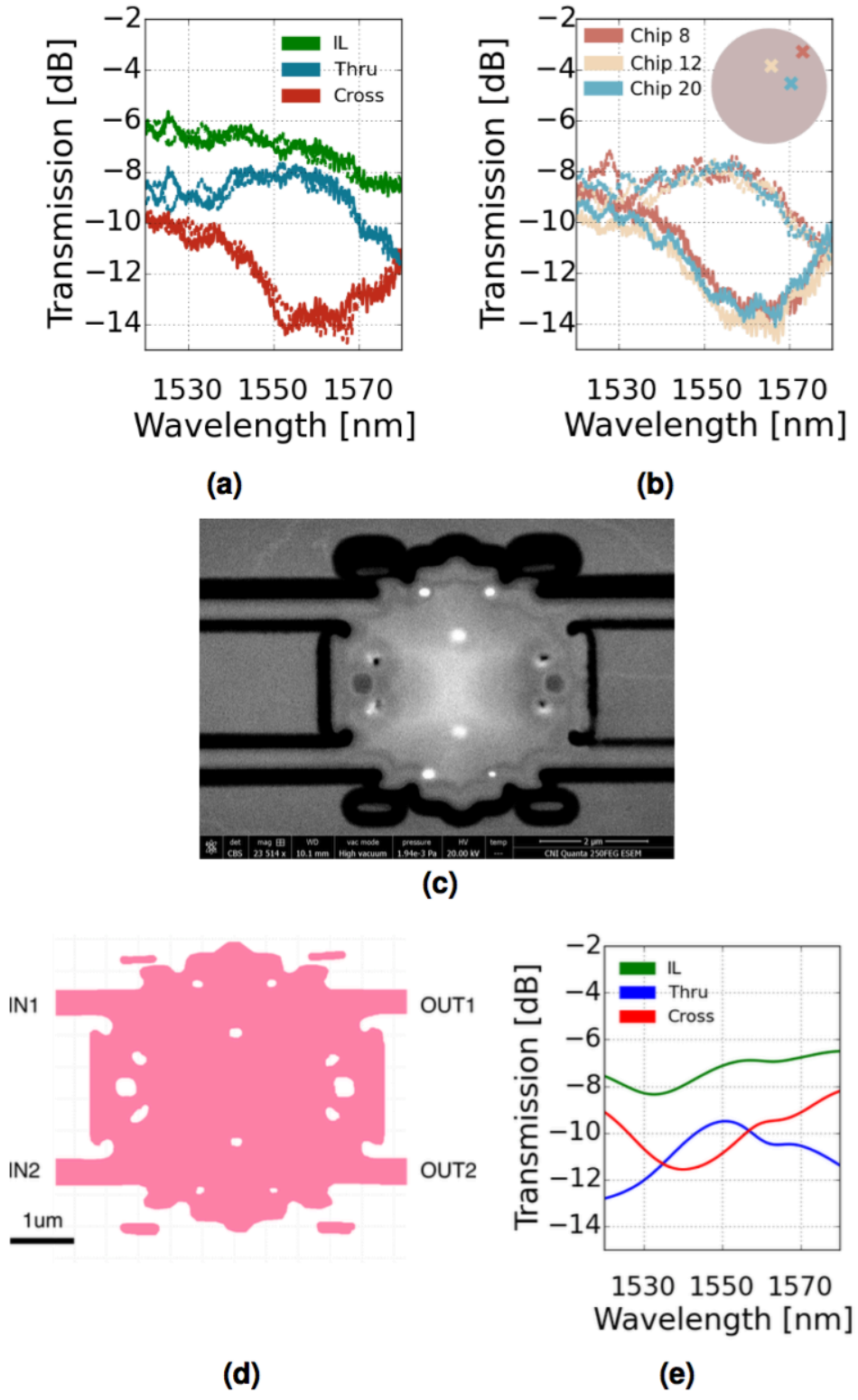

Figure 7.3: (a) Measured through (blue) and cross (red) port trans- mission spectra and the insertion loss of the 200 nm cell device for input from IN1 (solid) and IN2 (dashed). (b): Measured spectra from several devices from across the wafer. The die locations are shown in the inset wafer-map. (c) Backscattered electron SEM of the device with most of the cladding removed. (d) Trace of (c) used for FDTD simulations. (e) Simulated transmission spectram based on (d).

This design was implemented in the A*STAR IME Baseline Si photonics process, with the input/output ports connected to a row of TE grating couplers for coupling out of chip with a fiber array with an 8° angle polish. An additional pair of grating couplers connected with a waveguide was included in the row to optimize the fiber-to-chip alignment. A swept tunable laser and optical power monitor were used for spectral measurements. The measured spectra Fig. 7.3(a) show a symmetric response as expected from the symmetry of the device. Fig. 7.3(b) shows the spectra of a few devices from across the wafer were similar, confirming inter-die fabrication reproducibility.

However, these measurements greatly differed from the simulation, with insertion loss now at best around 6 dB, and less balanced power splitting. To determine the cause of the discrepancy, we obtained the scanning electron micrographs (SEMs) of the fabricated devices, shown in Fig. 7.3(c), by partially etching away the SiO_2 cladding layer in a hydrofluoric acid solution. We imaged with back-scattered electrons to view the features under the residual SiO_2. The SEMs revealed that many of the smaller isolated features in the nominal design were absent and the features were rounded. Although the rounding was expected [117], it was unanticipated that the smaller features would be absent since the minimum feature size design rule was nominally satisfied.

To simulate the expected transmission of the fabricated device, we traced the SEM as in Fig. 7.3(d) and imported as 220 nm thick Si with the corresponding planar geometry into a 3D FDTD simulation. The computed transmission is plotted in Fig. 7.3(e). Insertion losses were closer to that of measurements, and the difference suggests other manufacturing non-idealities, such as sloped side-walls or partial etching, were present. In our measurements, light which is converted into the transverse magnetic (TM) polarization (due to sloped side walls) would have been filtered out by the TE grating couplers.

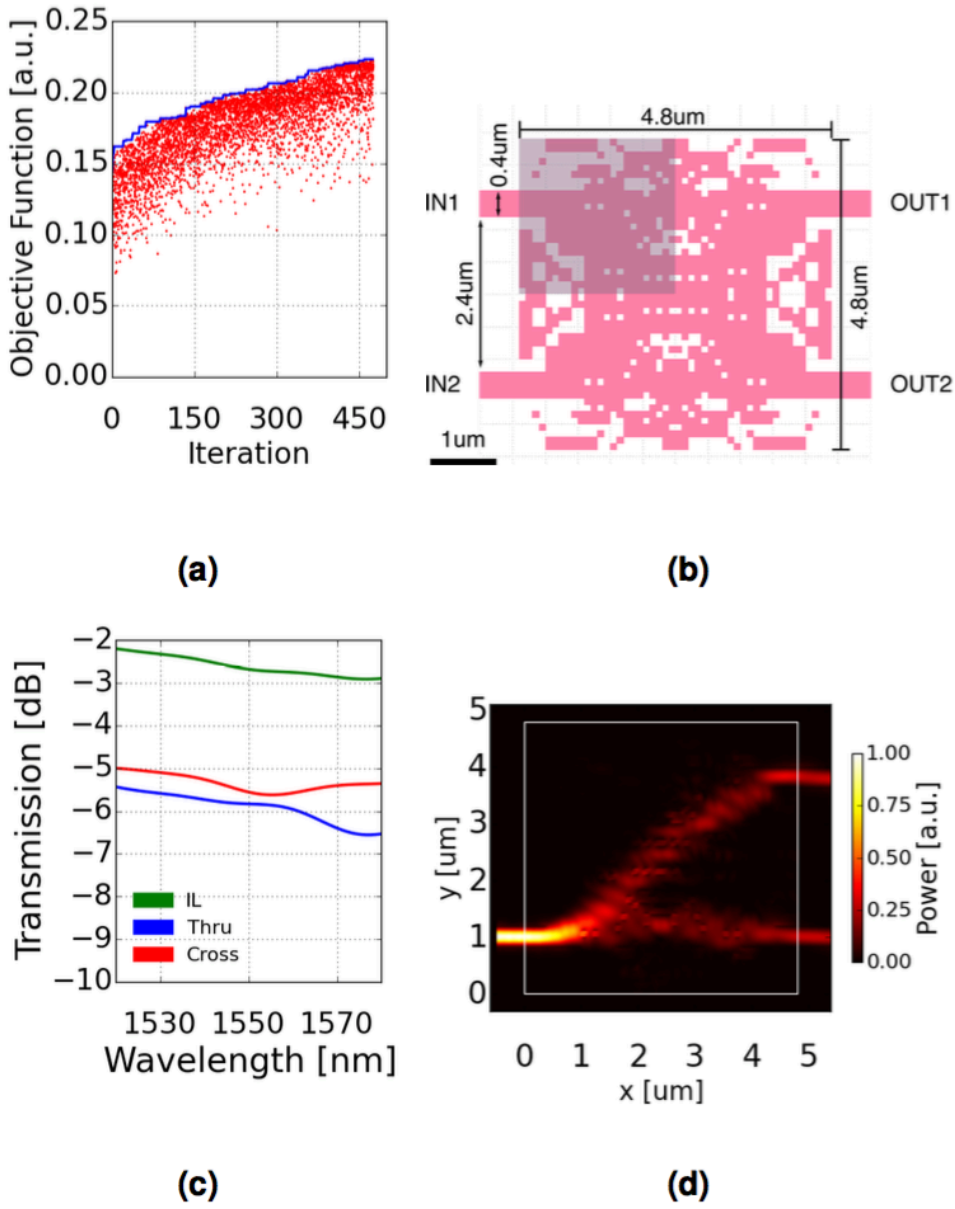

Figure 7.4: (a) Plot of the figure of merit at each iteration for the 100 nm cell device. Red dots are the individual values, and the blue line is the best value at each iteration. (b) The device with the best figure merit of 0.223. (c) Simulated transmission and insertion loss (IL) spectra of (b). Insertion loss is calculated as the sum of transmitted powers of the two arms with respect to input power. (d) Normalized intensity profile of the device.

We also implemented a second design using 100 nm Œ 100 nm cells, which are smaller than the allowed minimum feature size of the foundry process. The design was a variant of the 200 nm cell design and used 24 Œ 24 cells. The design was

developed through BPSO with an initial population size of 10 initialized to the nominal 200 nm cell design with random velocities and run over 478 iterations. Convergence is shown in Fig. 7.4(a), with improvement of the figure of merit from 0.174 to 0.223. The design is shown in Fig. 7.4(b). The simulated transmission spectrum Fig. 7.4(c) shows a more constant power splitting ratio without a crossover point with a lower worst case insertion loss of 2.9 dB than the 200 nm cell design.

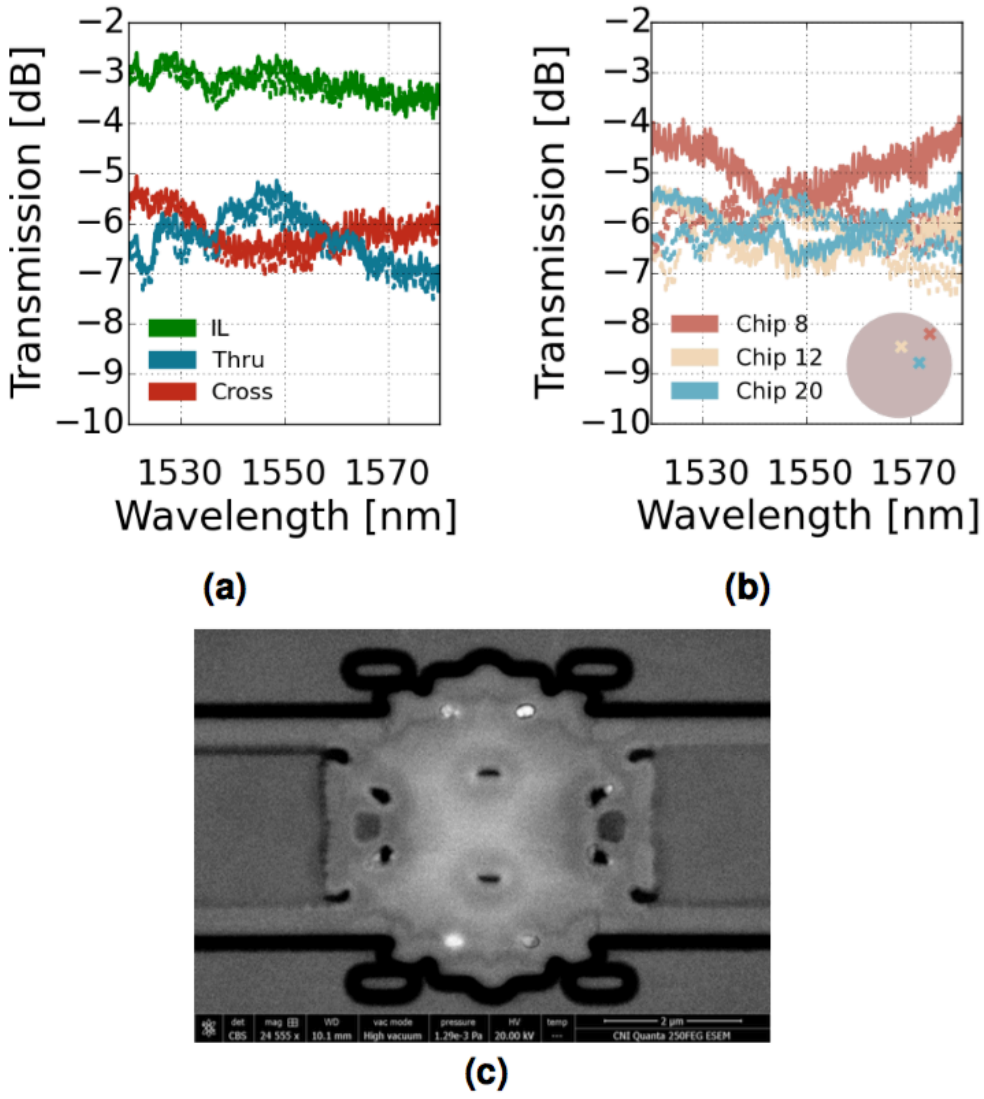

Figure 7.5: (a) Measured through (blue) and cross (red) port trans- mission, and insertion loss (green) of the 100 nm cell device for input from IN1 (solid) and IN2 (dashed). (b) Measured through and cross port spectra from a few dies across the wafer. The die locations are indicated in the inset wafer-map. (c) Backscattered electron SEM of the fabricated device with most of the oxide cladding removed.

Measurements in Fig. 7.5(a) show that the splitter had improved insertion loss and power splitting behavior compared to the measurement results of the 200 nm design. Although the spectra did not match the simulation, the cross port received power comparable with the through port and an equal power splitting ratio was obtained at wavelengths of 1535 nm and 1560 nm, and worst insertion loss was around 3.5 dB.

The SEM of the fabricated device, Fig. 7.5(c), is similar to the 200 nm cell device but has larger openings in holes which were enlarged during the optimization. This suggests that having a refinement in the mesh is useful for improving performance by allowing the boundaries of larger features to be re-positioned and not be constrained by the original mesh grid, even if the smaller features are not fabricated. Thus, in future designs, if small features could be filtered out of the design during the optimization, such as with filters proposed in [97], a minimum feature size may not necessarily need to be linked with the mesh size and having a larger problem space to work with may improve designs.

Our work highlights directions for improving ab-initio optimized photonic devices in foundry platforms. First, to overcome the high insertion losses of the devices demonstrated here, rather than fixing an area for the optimization, one may start with a conventional design and apply optimization while attempting to shrink the area in a multi-objective optimization [69]. It is possible that there is no physical mechanism to perform the power splitting within the imposed areal constraint. Second, to address the problem that using cells slightly above the minimum allowable feature size was insufficient to guarantee manufacturability, further studies on the minimum cell size are needed. In conventional devices, the minimum feature size are usually applicable to edges along one dimension, for waveguide gaps, tapers, and gratings. However, the optimization based device has features which approach the minimum feature size in two dimensions. Systematic characterization of the feature sizes using two-dimenionsal patterns, such as checkerboards, can provide better bounds for the minimum cell size. In addition, lithography simulation [117] can be included during the optimization. Finally, transformations to the device mask to compensate for lithography effects, as in [72], can enable more accurate fabrication.

7.5 Concluding Remarks

In conclusion, we have presented the first foundry Si photonic 2 x 2 power splitters designed by binary particle swarm optimizaton of a design area of 4.8 μm x 4.8 μm parameterized by square cells that have sizes chosen according to design rules. Power

splitting is demonstrated but to reduce the insertion loss and to have better match with the design, multi-objective optimization, deeper studies of the 2D minimum feature size, and lithography simulations can be incorporated in future work.

BIBLIOGRAPHY

[1] *Adapted from https://mynasadata.larc.nasa.gov/images/EM_Spectrum3-new.jpg*.

[2] H Akel and JP Webb. "Design sensitivities for scattering-matrix calculation with tetrahedral edge elements". In: *IEEE transactions on magnetics* 36.4 (2000), pp. 1043–1046.

[3] Elad Alon et al. "Equalization of modal dispersion in multimode fiber using spatial light modulators". In: *Global Telecommunications Conference, 2004. GLOBECOM'04. IEEE*. Vol. 2. IEEE. 2004, pp. 1023–1029.

[4] Masoud Babaie and Robert Bogdan Staszewski. "A Class-F CMOS Oscillator". In: *IEEE Journal of Solid-State Circuits* 48.12 (2013), pp. 3120–3133.

[5] Tom Baehr-Jones et al. "Myths and rumours of silicon photonics". In: *Nature Photonics* 6.4 (2012), pp. 206–208.

[6] C.A. Balanis. *Antenna theory: analysis and design*. Harper & Row series in electrical engineering. Wiley, 1982. ISBN: 9780471603528. URL: https://books.google.com/books?id=wARTAAAAMAAJ.

[7] *Bangs Laboratory Representative, private communication.*

[8] Klaus-Jürgen Bathe and Edward L Wilson. "Numerical methods in finite element analysis". In: (1976).

[9] Martin P Bendsøe. "Optimal shape design as a material distribution problem". In: *Structural optimization* 1.4 (1989), pp. 193–202.

[10] Jean-Pierre Bérenger. "Perfectly matched layer (PML) for computational electromagnetics". In: *Synthesis Lectures on Computational Electromagnetics* 2.1 (2007), pp. 1–117.

[11] S.A. Berger, W. Goldsmith, and E.R. Lewis. *Introduction to bioengineering*. Oxford University Press, 1996. ISBN: 9780198565161. URL: https://books.google.com/books?id=IX1RAAAAMAAJ.

[12] Pierre-A Besse et al. "Detection of a single magnetic microbead using a miniaturized silicon Hall sensor". In: *Applied Physics Letters* 80.22 (2002), pp. 4199–4201.

[13] Daniel W Boeringer and Douglas H Werner. "Particle swarm optimization versus genetic algorithms for phased array synthesis". In: *IEEE Transactions on antennas and propagation* 52.3 (2004), pp. 771–779.

[14] CJ Bouwkamp and NG De Bruijn. "The problem of optimum antenna current distribution". In: *Philips Res. Rep* 1.2 (1946), pp. 135–158.

[15] J. K. B. Cahn et al. “Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases”. In: *Biochemical Journal* 468.3 (2015), pp. 475–484. DOI: 10.1042/BJ20150183.

[16] Edmund E Callaghan and Stephen H Maslen. “The magnetic field of a finite solenoid”. In: (1960).

[17] Young-Seek Chung et al. “Optimal design method for microwave device using time domain method and design sensitivity analysis. i. fetd case”. In: *IEEE transactions on magnetics* 37.5 (2001), pp. 3289–3293.

[18] Young-Seek Chung et al. “Optimal design method for microwave device using time domain method and design sensitivity analysis. II. FDTD case”. In: *IEEE transactions on magnetics* 37.5 (2001), pp. 3255–3259.

[19] Young-Seek Chung et al. “Optimal shape design of microwave device using FDTD and design sensitivity analysis”. In: *IEEE Transactions on Microwave Theory and Techniques* 48.12 (2000), pp. 2289–2296.

[20] Paul I Deffenbaugh, Raymond C Rumpf, and Kenneth H Church. “Broadband microwave frequency characterization of 3-D printed materials”. In: *IEEE Transactions on Components, Packaging and Manufacturing Technology* 3.12 (2013), pp. 2147–2155.

[21] A. Einstein. “Zur Elektrodynamik bewegter Körper”. In: *Annalen der Physik* IV (1905), p. 15.

[22] PC Fannin et al. “Investigation of the complex susceptibility of magnetic beads containing maghemite nanoparticles”. In: *Journal of Magnetism and Magnetic Materials* 303.1 (2006), pp. 147–152.

[23] Shaoqi Feng et al. “Silicon photonics: from a microresonator perspective”. In: *Laser & photonics reviews* 6.2 (2012), pp. 145–177.

[24] Geir Fonnum et al. “Characterisation of Dynabeadső by magnetization measurements and Mössbauer spectroscopy”. In: *Journal of Magnetism and Magnetic Materials* 293.1 (2005), pp. 41–47.

[25] Lars H Frandsen et al. “Topology optimized mode conversion in a photonic crystal waveguide fabricated in silicon-on-insulator material”. In: *Optics express* 22.7 (2014), pp. 8525–8532.

[26] Satoshi Fukuda et al. “A 12.5+ 12.5 Gb/s full-duplex plastic waveguide interconnect”. In: *IEEE Journal of Solid-State Circuits* 46.12 (2011), pp. 3113–3125.

[27] Simone Gambini et al. “A CMOS 10kpixel baseline-free magnetic bead detector with column-parallel readout for miniaturized immunoassays”. In: *2012 IEEE International Solid-State Circuits Conference*. IEEE. 2012, pp. 126–128.

[28] Natalia K Georgieva et al. “Feasible adjoint sensitivity technique for EM design optimization”. In: *IEEE transactions on microwave theory and techniques* 50.12 (2002), pp. 2751–2758.

[29] Michael B Giles and Niles A Pierce. “Adjoint equations in CFD: duality, boundary conditions and solution behaviour”. In: *AIAA paper* 97 (1997), p. 1850.

[30] Michael B Giles and Niles A Pierce. “An introduction to the adjoint approach to design”. In: *Flow, turbulence and combustion* 65.3-4 (2000), pp. 393–415.

[31] Ali Hajimiri and Thomas H Lee. *The design of low noise oscillators*. Springer Science & Business Media, 1999.

[32] Donhee Ham and Ali Hajimiri. “Virtual damping and Einstein relation in oscillators”. In: *IEEE Journal of Solid-State Circuits* 38.3 (2003), pp. 407–418.

[33] Shu-Jen Han et al. “A high-density magnetoresistive biosensor array with drift-compensation mechanism”. In: *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*. 2007.

[34] Paul Hansen and Lambertus Hesselink. “Accurate adjoint design sensitivities for nano metal optics”. In: *Optics express* 23.18 (2015), pp. 23899–23923.

[35] Li He et al. “A high-efficiency nonuniform grating coupler realized with 248-nm optical lithography”. In: *IEEE Photonics Technology Letters* 25.14 (2013), pp. 1358–1361.

[36] Oliver Heaviside. *Electromagnetic Theory*. ’The Electrician’ Printing and Publishing Company, 1893.

[37] Carol A Holland and Frederick L Kiechle. “Point-of-care molecular diagnostic systems past, present and future”. In: *Current Opinion in Microbiology* 8.5 (2005). Antimicrobials / Edited by Malcolm Page and Christopher T Walsh ů Genomics / Edited by Stephan C Schuster and Gerhard Gottschalk, pp. 504–509. ISSN: 1369-5274. DOI: http://dx.doi.org/10.1016/j.mib.2005.08.001. URL: http://www.sciencedirect.com/science/article/pii/S1369527405001177.

[38] *https://aws.amazon.com/ec2*.

[39] *http://www.ansys.com*.

[40] *http://www.bangslabs.com*.

[41] *http://www.idtdna.com*.

[42] *http://www.invitrogen.com*.

[43] *http://www.invitrogen.com/site/us/en/home/brands/Dynal.html*.

[44] *http://www.rogerscorp.com*.

[45] *http://www.uvp.com*.

[46] Thomas JR Hughes, Leopoldo P Franca, and Gregory M Hulbert. "A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations". In: *Computer Methods in Applied Mechanics and Engineering* 73.2 (1989), pp. 173–189.

[47] Krzysztof Iniewski. *Integrated microsystems: electronics, photonics, and biotechnology*. CRC Press, 2011.

[48] David Issadore et al. "Ultrasensitive clinical enumeration of rare cells ex vivo using a micro-hall detector". In: *Science translational medicine* 4.141 (2012), 141ra92–141ra92.

[49] Byungchul Jang et al. "A CMOS fluorescent-based biosensor microarray". In: *2009 IEEE International Solid-State Circuits Conference-Digest of Technical Papers*. 2009.

[50] Jakob S Jensen. "Topology optimization of nonlinear optical devices". In: *Structural and Multidisciplinary Optimization* 43.6 (2011), pp. 731–743.

[51] Jakob Søndergaard Jensen and Ole Sigmund. "Topology optimization for nano-photonics". In: *Laser & Photonics Reviews* 5.2 (2011), pp. 308–321.

[52] Nanbo Jin and Yahya Rahmat-Samii. "Advances in particle swarm optimization for antenna designs: Real-number, binary, single-objective and multiobjective implementations". In: *IEEE Transactions on Antennas and Propagation* 55.3 (2007), pp. 556–567.

[53] Joseph F. Keithley. *The story of electrical and magnetic measurements: from 500 B.C. to the 1940s*. Wiley-IEEE Press, 1999.

[54] James Kennedy and Russell C Eberhart. "A discrete binary version of the particle swarm algorithm". In: *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*. Vol. 5. IEEE. 1997, pp. 4104–4108.

[55] L. Khriachtchev. *Silicon Nanophotonics: Basic Principles, Current Status and Perspectives*. Pan Stanford Publishing, 2016. ISBN: 9789814241137. URL: https://books.google.com/books?id=KmrvBQAAQBAJ.

[56] Frederick L. Kiechle and Carol A. Holland. "Point-of-Care Testing and Molecular Diagnostics: Miniaturization Required". In: *Clinics in Laboratory Medicine* 29.3 (), pp. 555–560. DOI: 10.1016/j.cll.2009.06.013. URL: http://dx.doi.org/10.1016/j.cll.2009.06.013.

[57] Gerald J Kost. *Principles & practice of point-of-care testing*. Lippincott Williams & Wilkins, 2002.

[58] Lincoln La Paz and Geoffrey A Miller. “Optimum current distributions on vertical antennas”. In: *Proceedings of the IRE* 31.5 (1943), pp. 214–232.

[59] Christopher M Lalau-Keraly et al. “Adjoint shape optimization applied to electromagnetic design”. In: *Optics express* 21.18 (2013), pp. 21693–21701.

[60] Hong-bae Lee and Tatsuo Itoh. “A systematic optimum design of waveguide-to-microstrip transition”. In: *IEEE transactions on microwave theory and techniques* 45.5 (1997), pp. 803–809.

[61] Andy Eu-Jin Lim et al. “Review of silicon photonics foundry efforts”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 20.4 (2014), pp. 405–416.

[62] Zin Lin et al. “Cavity-enhanced second-harmonic generation via nonlinear-overlap optimization”. In: *Optica* 3.3 (2016), pp. 233–238.

[63] Tsung-Yang Liow et al. “Silicon modulators and germanium photodetectors on SOI: monolithic integration, compatibility, and performance optimization”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 16.1 (2010), pp. 307–315.

[64] Marko Lonar et al. “Design and fabrication of silicon photonic crystal optical waveguides”. In: *Journal of lightwave technology* 18.10 (2000), p. 1402.

[65] Jesse Lu and Jelena Vukovi. “Objective-first design of high-efficiency, small-footprint couplers between arbitrary nanophotonic waveguide modes”. In: *Optics express* 20.7 (2012), pp. 7221–7236.

[66] Zeqin Lu et al. “Comparison of photonic 2× 2 3-dB couplers for 220 nm silicon-on-insulator platforms”. In: *2015 IEEE 12th International Conference on Group IV Photonics (GFP)*. IEEE. 2015, pp. 57–58.

[67] Guan-Chun Luh, Chun-Yi Lin, and Yu-Shu Lin. “A binary particle swarm optimization for continuum structural topology optimization”. In: *Applied Soft Computing* 11.2 (2011), pp. 2833–2844.

[68] Jason CC Mak et al. “Binary particle swarm optimized 2 x 2 power splitters in a standard foundry silicon photonic platform”. In: *Optics Letters* 41.16 (2016), pp. 3868–3871.

[69] R Timothy Marler and Jasbir S Arora. “Survey of multi-objective optimization methods for engineering”. In: *Structural and multidisciplinary optimization* 26.6 (2004), pp. 369–395.

[70] J. C. Maxwell. “A Dynamical Theory of the Electromagnetic Field”. In: *Philosophical Transactions of the Royal Society of London* 155 (1865), pp. 459–512.

[71] J. C. Maxwell. *A Treatise on Electricity and Magnetism*. 1873.

[72] Karan K Mehta et al. “High-Q CMOS-integrated photonic crystal microcavity devices”. In: *Scientific reports* 4 (2014), p. 4077.

[73] WE Moerner and David P Fromm. “Methods of single-molecule fluorescence spectroscopy and microscopy”. In: *Review of Scientific Instruments* 74.8 (2003), pp. 3597–3619.

[74] D Nair and JP Webb. “Optimization of microwave devices using 3-D finite elements and the design sensitivity of the frequency response”. In: *IEEE transactions on magnetics* 39.3 (2003), pp. 1325–1328.

[75] George L Nemhauser and Laurence A Wolsey. “Integer programming and combinatorial optimization”. In: *Wiley, Chichester. GL Nemhauser, MWP Savelsbergh, GS Sigismondi (1992). Constraint Classification for Mixed Integer Programming Formulations. COAL Bulletin* 20 (1988), pp. 8–12.

[76] James H. Nichols. “Point of Care Testing”. In: *Clinics in Laboratory Medicine* 27.4 (), pp. 893–908. DOI: 10.1016/j.cll.2007.07.003. URL: http://dx.doi.org/10.1016/j.cll.2007.07.003.

[77] Armand CR Niederberger et al. “Sensitivity analysis and optimization of sub-wavelength optical gratings using adjoints”. In: *Optics express* 22.11 (2014), pp. 12971–12981.

[78] Alex Pai et al. “A handheld magnetic sensing platform for antigen and nucleic acid detection”. In: *Analyst* 139.6 (2014), pp. 1403–1411.

[79] Kyoungchul Park et al. “Multiplexed sensing based on Brownian relaxation of magnetic nanoparticles using a compact AC susceptometer”. In: *Nanotechnology* 22.8 (2011), p. 085501.

[80] Alexander Y Piggott et al. “Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer”. In: *Nature Photonics* 9.6 (2015), pp. 374–377.

[81] Yahya Rahmat-Samii and Eric Michielssen. “Electromagnetic optimization by genetic algorithms”. In: *Microwave Journal* 42.11 (1999), pp. 232–232.

[82] Adam M Rauwerdink, Andrew J Giustini, and John B Weaver. “Simultaneous quantification of multiple magnetic nanoparticles”. In: *Nanotechnology* 21.45 (2010), p. 455101.

[83] G.T. Reed and A.P. Knights. *Silicon photonics: an introduction*. John Wiley & Sons, 2004. ISBN: 9780470870341. URL: https://books.google.com/books?id=t-dRAAAAMAAJ.

[84] Sembiam R Rengarajan and Yahya Rahmat-Samii. “The field equivalence principle: Illustration of the establishment of the non-intuitive null fields”. In: *IEEE Antennas and Propagation Magazine* 42.4 (2000), pp. 122–128.

[85] Jacob Robinson, Seelig Sinton, and Yahya Rahmat-Samii. “Particle swarm, genetic algorithm, and their hybrids: optimization of a profiled corrugated horn antenna”. In: *Antennas and Propagation Society International Symposium, 2002. IEEE*. Vol. 1. IEEE. 2002, pp. 314–317.

[86] Wesley D Sacher et al. “Polarization rotator-splitters in standard active silicon photonics platforms”. In: *Optics express* 22.4 (2014), pp. 3777–3786.

[87] S.A. Schelkunoff. *Advanced antenna theory*. Applied mathematics series. Wiley, 1952. URL: https://books.google.com/books?id=AiVTAAAAMAAJ.

[88] SA Schelkunoff. “Some equivalence theorems of electromagnetics and their application to radiation problems”. In: *Bell System Technical Journal* 15.1 (1936), pp. 92–112.

[89] Bing Shen et al. “An integrated-nanophotonics polarization beamsplitter with 2.4× 2.4 μm2 footprint”. In: *Nature Photonics* 9.6 (2015), pp. 378–382.

[90] Osamu Shimomura, Frank H Johnson, and Yo Saiga. “Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea”. In: *Journal of cellular and comparative physiology* 59.3 (1962), pp. 223–239.

[91] Constantine Sideris, Elad Alon, and Ali Hajimiri. “An Ultra-Fast Simulation Technique for Optimizing Electromagnetic Structures with Arbitrary Metallic or Dielectric Materials”. preprint. Nov. 2016.

[92] Constantine Sideris, Elad Alon, and Ali Hajimiri. “Globally Optimal Current Distributions for Maximizing Performance Metrics of Electromagnetic Structures”. preprint. Nov. 2016.

[93] Constantine Sideris and Ali Hajimiri. “An integrated magnetic spectrometer for multiplexed biosensing”. In: *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*. IEEE. 2013, pp. 300–301.

[94] Constantine Sideris and Ali Hajimiri. “Design and implementation of an integrated magnetic spectrometer for multiplexed biosensing”. In: *IEEE transactions on biomedical circuits and systems* 7.6 (2013), pp. 773–784.

[95] Constantine Sideris et al. “Automated design of a 3D printed waveguide surface coupler”. In: *2015 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*. IEEE. 2015, pp. 318–319.

[96] Ole Sigmund. “A 99 line topology optimization code written in Matlab”. In: *Structural and multidisciplinary optimization* 21.2 (2001), pp. 120–127.

[97] Ole Sigmund. “Morphology-based black and white filters for topology optimization”. In: *Structural and Multidisciplinary Optimization* 33.4-5 (2007), pp. 401–424.

[98] Ole Sigmund and Jakob Søndergaard Jensen. “Systematic design of phononic band–gap materials and structures by topology optimization”. In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 361.1806 (2003), pp. 1001–1019.

[99] Richard Soref. “The past, present, and future of silicon photonics”. In: *IEEE Journal of selected topics in quantum electronics* 12.6 (2006), pp. 1678–1687.

[100] Dov Stekel. *Microarray bioinformatics*. Cambridge University Press, 2003.

[101] Dennis M Sullivan. *Electromagnetic simulation using the FDTD method*. John Wiley & Sons, 2013.

[102] Nan Sun et al. “CMOS RF biosensor utilizing nuclear magnetic resonance”. In: *IEEE Journal of Solid-State Circuits* 44.5 (2009), pp. 1629–1643.

[103] Allen Taflove and Susan C Hagness. *Computational electrodynamics*. Artech house publishers, 2000.

[104] Yusuke Tanaka et al. “A versatile multi-modality serial link”. In: *2012 IEEE International Solid-State Circuits Conference*. IEEE. 2012, pp. 332–334.

[105] TT Taylor. *A discussion of the maximum directivity of an antenna*. 1948.

[106] Nam K Tran and Gerald J Kost. “Worldwide point-of-care testing: compendiums of POCT for mobile, emergency, critical, and primary care and of infectious diseases tests”. In: *Point of Care* 5.2 (2006), pp. 84–92.

[107] Anna J. Tudos, Geert A. J. Besselink, and Richard B. M. Schasfoort. “Trends in miniaturized total analysis systems for point-of-care testing in clinical chemistry”. In: *Lab Chip* 1 (2 2001), pp. 83–95. DOI: 10.1039/B106958F. URL: http://dx.doi.org/10.1039/B106958F.

[108] J.L. Volakis, A. Chatterjee, and L.C. Kempel. *Finite Element Method Electromagnetics: Antennas, Microwave Circuits, and Scattering Applications*. IEEE Press Series on Electroma. Wiley, 1998. ISBN: 9780780334250. URL: https://books.google.com/books?id=55q7HqnMZCsC.

[109] J.L. Volakis and K. Sertel. *Integral Equation Methods for Electromagnetics*. Electromagnetics and Radar. Institution of Engineering and Technology, 2012. ISBN: 9781891121937. URL: https://books.google.com/books?id=hfHvKyrKfJIC.

[110] Hua Wang. “Magnetic sensors for diagnostic medicine: CMOS-based magnetic particle detectors for medical diagnosis applications”. In: *IEEE Microwave Magazine* 14.5 (2013), pp. 110–130.

[111] Hua Wang and Ali Hajimiri. *Design of a frequency-shift-based CMOS magnetic sensor array for point-of-care (PoC) biomolecular diagnosis applications*. Boca Raton, FL, USA: CRC Press, 2011.

[112] Hua Wang, Constantine Sideris, and Ali Hajimiri. “A frequency-shift based CMOS magnetic biosensor with spatially uniform sensor transducer gain”. In: *Custom Integrated Circuits Conference (CICC), 2010 IEEE*. IEEE. 2010, pp. 1–4.

[113] Hua Wang et al. “A frequency-shift CMOS magnetic biosensor array with single-bead sensitivity and no external magnet”. In: *2009 IEEE International Solid-State Circuits Conference-Digest of Technical Papers*. IEEE. 2009, pp. 438–439.

[114] Hua Wang et al. “A magnetic cell-based sensor”. In: *Lab on a Chip* 12.21 (2012), pp. 4465–4471.

[115] Hua Wang et al. “An ultrasensitive CMOS magnetic biosensor array with correlated double counting noise suppression”. In: *Microwave Symposium Digest (MTT), 2010 IEEE MTT-S International*. IEEE. 2010, pp. 616–619.

[116] Joseph Wang. “Survey and summary from DNA biosensors to gene chips”. In: *Nucleic Acids Research* 28.16 (2000), pp. 3011–3016.

[117] Xu Wang et al. “Lithography simulation for the fabrication of silicon photonic devices with deep-ultraviolet lithography”. In: *IEEE Group IV Photonics Conference (San Diego, CA, USA 2012), paper ThP17*. 2012.

[118] Chen Yang et al. “3D printed RF passive components by liquid metal filling”. In: *2015 28th IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*. IEEE. 2015, pp. 261–264.

INDEX