Circuit Design and Technological Limitations of Silicon RFICs for Wireless Applications

by

Donald A. Hitko

Bachelor of Science in Electrical Engineering
GMI Engineering & Management Institute, 1994

Master of Science in Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 1997

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

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2002

Author _____

Department of Electrical Engineering and Computer Science
May 24, 2002

Certified by _____

Charles G. Sodini
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by _____

Arthur C. Smith
Chairman, Committee on Graduate Students
Department of Electrical Engineering and Computer Science

Circuit Design and Technological Limitations of Silicon RFICs for Wireless Applications

by

Donald A. Hitko

Submitted to the Department of Electrical Engineering and Computer Science
on May 24, 2002, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Semiconductor technologies have been a key to the growth in wireless communication over the past decade, bringing added convenience and accessibility through advantages in cost, size, and power dissipation. A better understanding of how an IC technology affects critical RF signal chain components will greatly aid the design of wireless systems and the development of process technologies for the increasingly complex applications that lie on the horizon. Many of the evolving applications will embody the concept of *adaptive* performance to extract the maximum capability from the RF link in terms of bandwidth, dynamic range, and power consumption—further engaging the interplay of circuits and devices is this design space and making it even more difficult to discern a clear guide upon which to base technology decisions. Rooted in these observations, this research focuses on two key themes: 1) devising methods of implementing RF circuits which allow the performance to be dynamically tuned to match real-time conditions in a power-efficient manner, and 2) refining approaches for thinking about the optimization of RF circuits at the device level. Working toward a 5.8 GHz receiver consistent with 1 GBit/s operation, signal path topologies and adjustable biasing circuits are developed for low-noise amplifiers (LNAs) and voltage-controlled oscillators (VCOs) to provide a facility by which power can be conserved when the demand for sensitivity is low. As an integral component in this effort, tools for exploring device level issues are illustrated with both circuit types, helping to identify physical limitations and design techniques through which they can be mitigated. The design of two LNAs and four VCOs is described, each realized to provide a fully-integrated solution in a 0.5 μm SiGe BiCMOS process, and each incorporating all biasing and impedance matching on chip. Measured results for these 5-6 GHz circuits allow a number of poignant technology issues to be enlightened, including an exhibition of the importance of terminal resistances and capacitances, a demonstration of where the transistor f_T is relevant and where it is not, and the most direct comparison of bipolar and CMOS solutions offered to date in this frequency range. In addition to covering a number of new circuit techniques, this work concludes with some new views regarding IC technologies for RF applications.

Thesis Supervisor: Charles G. Sodini
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

Wow! Maybe that says it all. I can remember when it took me by surprise to hear a colleague mention taking five years to complete the Ph.D.—five years after the Master's, that is. How little I knew then what deciding to continue beyond the Master's would entail for me. But I recall another colleague attributing his decision to stay for the Ph.D. program to thinking: "It seemed like a pretty neat place to spend a few years." Albeit at a different location, and although "a few years" may have proven to been a euphemism, I still found his statement to hold throughout my time here.

Attending graduate school at a university like MIT is mostly about the people that it brings together, for this is what makes the experience so rewarding. But rather than implicating anyone, I would instead like to express my gratitude to each of you by mentioning some of your personal contributions which have meant so much to me. So here goes, in no discernible order. First, thank you for assuring me that I was born to be a graduate student. How right you were. Thank you for the Tosci's runs, and for being around the group almost as long as I have been. Thanks to those who collectively surmounted the phenomenon of always playing one bad inning. Thank you for the time spent crushing tennis balls, and for waiting in perpetuum for a window seat. Thanks for getting together to work on calendar pages, and to the many predecessors here who were featured in those pages. Thanks to those with whom hallway hockey was anointed. Perhaps we can have a reunion in 30 years to see if the puck is still there. Thank you for the many helpful job-related discussions, for sustaining some semblance of organization about the office, for always being upbeat, and for infusing new enthusiasm into the group. Thanks to those who provided helpful advice on FFTs, replica biasing techniques, device noise, pointy-haired pursuits, computers, the art of Re-use, RF circuits, and when to shoot the puck. My apologies if I did not do particularly well in taking the advice. Thank you for the die photos, which dressed up numerous presentations as well as this thesis. Thanks to those who made playing (and coaching!) hockey so much fun, and to those who put up with me taking their P.E. classes six times. Of course, this would not have happened without the initial encouragement to get started on the hockey "hobby", for which I am very thankful. Thank you for always putting on a good show whenever I came to watch, and for coming to watch my feeble efforts. Thank you for the Esplanade skates, and especially those at 2AM. Thank you for getting up at 4AM to observe re-enactments of historical events. Finally—although still in no discernible order—thank you for the 1AM dinners (of which Mom would not approve) and for ice cream in (on?) fire trucks; I'm glad I could be there for your 48th.

In their support of this work, we are indebted to SRC (through contract number 2000-HJ-794), IBM Microelectronics in Burlington, Vermont, and HRL Laboratories in

Malibu, California. With a big assist from Jeff Gross, the folks at IBM supplied some excellent focused-ion beam rework to enable experimentation with the oscillators. HRL lent their support through a fellowship, making space for me on two of their IC fabrication runs, providing access to noise measurement equipment for amplifiers and oscillators, and by continually demonstrating their remarkable patience.

The readers of this thesis have proven nothing short of amazing, a task made even more daunting by my manner of writing and lack of a knack for timing. A big thank you to Charles Sodini, Jesús del Alamo, Hae-Seung Lee, and Dan McMahill, each of whom are undoubtedly grateful that FrameMaker® has no provision for footnoting a footnote. In particular, Charlie deserves a commendation for being dragged through this procedure on my behalf with two theses *and* an area examination. Discussions with Jesús regarding devices and noise have always been very enlightening. I am also very appreciative of the time Dan has spent on this, as he has offered a great deal of insight and many suggestions which are reflected throughout this work.

Of course, none of this would have occurred without the unwavering support of my family. To my parents, may you both enjoy a relaxed forthcoming Father's Day (Mother's Day wishes were passed along in my Bachelor's thesis) and a happy (upcoming) 36th anniversary. I'd also like to congratulate Trisha and Eric for their perseverance, and wish you continued good tidings with the restaurant. Finally, many thanks are owed to the rest of the family for keeping up with me during my extended disappearance to Massachusetts.

As for what comes next, well, we will have to see. But as a start, the late "Badger" Bob Johnson might have said it best in noting, "It's a great day for hockey." Sounds good to me.

Donald A. Hitko
Cambridge, Massachusetts
May 29, 2002

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the largest growth areas in electronics over the past decade has undoubtedly been in applications of wireless communication. Semiconductor technologies have been a key to this growth, bringing added convenience and accessibility through advantages in cost, size, and power dissipation. Wireless products and systems thrive on this increased utility, the commercial momentum of which has been fueling further investment in integrated circuit designs and technology. The resulting advancement in system capabilities has developed greater interest and receptiveness on the part of the consumer, leading to more applications being envisioned, and necessitating more available spectrum to support the wireless infrastructure. To obtain this added bandwidth and alleviate interference, the frequencies of the communication channels are necessarily edging upward, placing yet more demands on the technologies used to implement the wireless systems.

Several recently opened ISM bands in the 5-6GHz range have been allocated for unlicensed operation of broadband wireless links between portable devices, computers, and the Internet. While always subject to change, the spirit of these National Information Infrastructure (NII) systems has been described by the FCC in a 1996 ruling [1]:

> NII/SUPERNet devices [can] provide short-range, high-speed wireless digital information transfer and could support the creation of new wireless local area networks (LANs) as well as facilitate access to the National Information Infrastructure without the expense of wiring. These devices may further the universal service goals of the Telecommunications Act by offering schools, libraries, health care providers, and other users inexpensive networking alternatives which may access advanced telecommunications services.

Three unlicensed bands are set aside by the ruling: 5.15-5.25GHz, 5.25-5.35GHz, and 5.725-5.875GHz, of which the band centered at 5.8GHz allows the highest transmit power levels. With 150MHz of allotted bandwidth and up to 1W of transmit power, data rates of 1GBit/s are conceivable over links covering short haul distances.

While the products and applications in this field are still evolving, the one constraint is that the consumer market will determine the acceptable end product price based on convenience, functionality, and a comparison with substitutes. Along with a consideration of form factor issues, the price the market will bear effectively sets a bound on the technologies which can be used in realizing the product. In conjunction with circuit design techniques, the limitations of these technologies determine the performance that can be achieved in the required system components, a set of capabilities around which the wireless link specifications must be drawn. These specifications in turn define the features and functionality, creating a reinforcing loop in the product design space as illustrated in Figure 1-1.

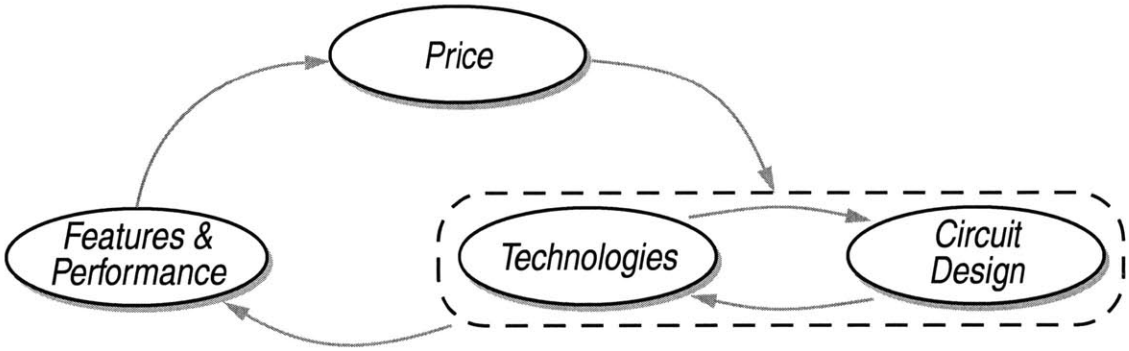

Figure 1-1. System dynamics model for IC technology-based markets.

Size and cost constraints may, for example, make the use of a discrete dielectric resonator oscillator impractical for the RF upconversion stage in the transmitter of a portable communications device. Integrating the RF link within a GaAs or InP MMIC may yield good performance while addressing the size issues, but may represent too costly a solution for the application. A circuit realized in silicon might be a less expensive alternative; however the transistor and passive device parasitics are more significant in a silicon IC technology.[1] For an oscillator, the increased transistor terminal parasitics result in a higher phase noise, potentially interfering with weaker signals in adjacent channels or degrading the sensitivity of a receiver.[2] In a transmitter, oscillator phase noise can limit

[1] The semiconducting nature of silicon results in appreciable terminal capacitances to the substrate (e.g., collector-substrate capacitance in a bipolar transistor) which are not significant when semi-insulating materials are used for the substrate material. Device designs in III-V semiconductors also typically employ mesa structures that lower access resistances and junction areas when compared to more planar device topologies.

[2] Oscillators and phase noise are covered in Chapter 3.

how closely channels may be set which, for a fixed bandwidth, reduces the achievable data rate. Conversely, by reducing the signal to noise ratio available to the demodulator, phase noise in a receiver increases the bit error rate. Hence the noise of an oscillator in a wireless communications system perceptibly impacts the allowable number of users, realizable data rates, and the quality of the link, exemplifying the importance of the device technology. Making the right technology choice plays a crucial role in determining the commercial success of a product.

1.1 Thesis Contributions

Among the realizations from considering the nature of product development is that a better understanding of how an IC technology affects critical RF signal chain components would greatly aid the design of both wireless systems and future process technologies for the increasingly complex applications that lie on the horizon. The fundamental interplay between devices and circuits in this design space confounds the issue, making it difficult to discern a clear guide upon which to base technology decisions. In designing key RF components—such as oscillators, mixers, and amplifiers—circuit techniques and topologies can be developed to mitigate device limitations and to break traditional design trade-offs by extending the concept of scalable circuit performance in radio transceivers. Furthermore, as the properties of a wireless communication channel tend to be highly uncertain, significant power savings can be realized by using circuits designed to allow a dynamic adaptation to changing operating conditions. In light of these observations, the focus of this thesis centers upon two key themes:

- The optimization of RF circuits at the device level. Two broad classes of circuits are considered in this research: linear time-invariant (LTI) and linear time-variant (LTV). The LTI class of circuits is studied by designing and characterizing a pair of 5.8 GHz low-noise amplifiers (LNAs), and the LTV side is investigated by constructing and measuring a set of four voltage-controlled oscillators (VCOs). Approaches for exploring device level issues are developed with both circuit types, helping to identify physical limitations and design techniques through which they can be mitigated. Other RF circuits, such as mixers, can then be considered through these approaches as a combination of LTI and LTV elements. By carefully crafting an experiment around a set of designs in a BiCMOS process and tracing measured circuit performance back to device level issues, some new views are offered on directions that should be taken in IC technologies for RF applications.

- Devising methods of implementing RF circuits which allow the performance to be dynamically tuned to match real-time conditions in a power-efficient manner. Once the physical limitations are accommodated in the design, extracting the opti-

mum performance from a technology becomes a matter of dissipating the required power in the circuit. It may not, however, be either necessary or feasible to operate the transceiver circuits under this condition at all times. Signal path topologies and adjustable biasing circuits are developed to provide a facility by which power can be conserved in RF circuits when the demand for performance is low, providing flexibility without compromising the operation when optimum performance is required. Incorporation of adaptability at the circuit and system levels is paramount in expanding the capabilities and increasing the utilization of wireless communication links, and yet remains a largely untapped resource in this field.

These themes represent two areas in which innovation will hold significant implications for the future of RF/microwave integrated circuits and the wireless applications in which they are used. An introduction to these ideas and other related concepts is provided in the following sections.

1.1.1 Optimization of RF Circuits at the Device Level

One of the salient characteristics of RF circuit design is that the active devices are often pushed near their physical limits of operation, resulting in a high degree of correlation between the performance of an individual transistor and that of the circuit. As the signal frequency increases toward the rate where small-signal gains fall to unity, the ability to compensate device shortcomings through feedback mechanisms becomes constrained. Thus, an important area of investigation for this regime of operation is to identify the device features that are limiting circuit performance in key RF transceiver components. Most of the components can be represented by models of either the linear time-invariant or linear time-variant variety. Through linearity, both models assume that the RF signals being processed are small enough not to appreciably impact the operation of the circuits that are processing them.[3] Additionally, in LTI circuits, the parameters of the elements remain constant—a condition observed in many amplifiers and filters. Time-variant circuits include oscillators, mixers, and prescalers, and are characterized by possessing gains, impedances, noise power, etc., which vary with time—often in a pattern that is periodic. For both classes of circuits, models are discussed as a means of eliciting the mechanisms responsible for constraining the circuit performance. However, this determination cannot be made in a vacuum, as circuit techniques and topologies can be developed to work around limitations (to an extent).[4] Pushing the performance envelope in RF can only happen through a concurrent optimization of circuits and technology.

---

[3] A mixer processes either the RF or IF signal and produces the other; the applied oscillator drive can be considered as creating a time-varying operating point for the signal and the noise sources in the circuit. More will be said about mixers in Chapter 3.

[4] Note the earlier point about the limited ability to implement compensation via feedback at RF.

Knowledge gleaned about the performance-limiting factors identified through these approaches can be applied at a number of levels. First, even within the confines of a chosen process, a designer has considerable latitude with transistor selection, sizing, and 2-D layout geometry. These degrees of freedom may be used to better optimize RF circuits in pushing the limits of a technology, and to make more evident where those limits lie. Similarly, the tools described in this work can be used to guide the selection of a technology for a chosen application. A bipolar or BiCMOS technology may involve more masks than a comparable CMOS process, but the higher transconductance and lower noise per unit current of a bipolar transistor may provide advantages which can improve the quality of service to cost ratio of the overall system. Knowledge of the technological limitations yields insight into the extent to which performance can be expected to benefit. Finally, it is important to understand where device enhancements result in better circuits, and where changes in the device needed to realize the enhancements may hurt performance more than it helps. A thinner base and higher collector doping concentration can be employed to increase the transistor f_T,[5] but whether a faster switching response and a higher current gain translate into improved RF circuits can be determined through the models that are proposed in the chapters which follow. Answers to such questions may be surprising, and are vital in setting directions for the continued development of IC process technologies.

As a direct illustration of the impact a transistor technology can have upon RF circuit performance, comparative LNA and VCO designs are implemented in the CMOS and bipolar halves of a BiCMOS process. Different physical limitations are encountered based on the devices being used, and thus different solutions are reflected in the circuits. Each of the designs is geared toward a receiver in a 1GBit/s wireless network operating at 5.8GHz, for which a high level of sensitivity may be required to support the data rate. But rather than choosing and designing to a given specification, the emphasis here is upon finding the peak performance that can be extracted from a 0.5μm SiGe BiCMOS process, and then determining how these limits change as a function of the technology and power consumption.

1.1.2 Trade-off Between Quality of Service and Power Consumption

To meet consumer expectations in an increasingly sophisticated market, wireless communication systems need to provide an acceptable level of performance under defined

[5] The f_T of a transistor is the frequency at which its current gain (in the common-emitter or common-source configurations) falls to unity, and is often used as a figure of merit in comparing devices and technologies.

worst-case conditions. The interpretation of acceptability in these networks is one of supporting low latency, high data rate, ubiquitous access—necessitating a highly capable link. However, it should be recognized that the data rates will not always be 1GBit/s and the channel may not always demand high sensitivity; designing to operate around a set of worst-case conditions drains power without always buying performance. When less demanding scenarios are common, the utility of the system may be increased by acting upon real-time information about the link and the data being transmitted. Methods of reducing power consumption by dynamically trading off quality of service levels that are not required may be feasible. By incorporating adaptability into RF circuits, the operation of transceiver components can be adjusted so that power is never burned to support performance levels—measured in terms of gain, linearity, noise, impedance matching, delay, etc.—that are unnecessary for a given transmission. When the information being demanded is only of moderate data rates, bandwidth efficiency is less of a concern, allowing a given bit error rate to be achieved for rather modest signal to noise ratios. In this case, the power consumed by the receiver could be reduced from the peak sensitivity settings. Similarly, when data rates are low and the network is lightly loaded, phase noise requirements in the transmitter can be relaxed as 1) a greater RMS phase error can be tolerated in the modulator, and 2) adjacent channel interference is not as big an issue when the neighboring channels are unused.

The challenge is to implement the adaptability without compromising the peak performance that can be attained by the circuits, and to provide for reliable system operation under all conditions. Power consumption can be adjusted by changing the bias current, the supply voltage, or both together as appropriate. For the LNAs presented herein, the voltage is seen to have little effect, so the bias current is the control by which the noise figure and gain can be tuned to meet the instantaneous demand. However, the input and output impedances of a transistor stage also change with the bias, leading to an undesirable change in the matching characteristics of an amplifier built around it. This consideration leads to the development of a "switchless" switched-stage bipolar LNA, and an accompanying base current source biasing circuit to provide adjustable performance while maintaining 50Ω impedance matches.[6] In an oscillator, the optimum bias current is tied to the supply voltage and thus the two should be adjusted together for maximum efficiency. For the bipolar and CMOS VCO topologies discussed in Chapter 6, biasing circuits are developed to allow control of the bias current about a "default" setting, extending the range of operation while minimizing the cost to the oscillator phase noise. Together, this

[6] Source and load impedances of 50Ω are assumed to be presented to the LNAs and VCOs.

collection of designs illustrates another underlying theme: the features and performance of transceivers for wireless applications can be enhanced as much through innovations in the biasing and buffering circuitry as it can through developments in the RF signal path itself. Both of these aspects are emphasized in the later chapters of this manuscript.

1.2 A Preview of the Thesis

Taking form around the characteristics expected of evolving high data rate, bandwidth on demand, wireless networking applications, the essence of this work is embodied in a set of designs targeted at the U-NII 5.8GHz band.[7] Pushing the performance of RF circuits requires optimization across the circuit and device levels, a concurrency reflected in the presentation of this thesis where circuit and device considerations have been intertwined. Crossing between the fields makes for a technical and instructional challenge; an understanding of both circuits and devices is presumed in an attempt to convey the message in a minimal amount of time. In an effort to cater to readers with differing backgrounds, supporting information is included through footnotes to hopefully answer the questions that crop up for some readers without cluttering the text for others. An ample bibliography can be found at the end, sorted into topics for ease of reference.

Almost by definition, inductors play a pivotal role in many RF circuits; a modeling paradigm that provides accuracy and flexibility in the design of inductors is essential. The approach adopted for this work is described in Chapter 2, and will become a poignant issue later as differences between the expected performance and measured results are investigated. Chapter 3 then follows with a framework for thinking about noise and device limitations in linear time-variant circuits. Although discussed within the context of oscillators, consideration is also given to extending the model to mixers. The effects of cyclostationarity in the sources of noise are scrutinized, as is the importance of properly discerning the components which comprise the terminal currents of a transistor. From there, a topological and technological foray into the issues of designing low-noise amplifiers is detailed in Chapter 4, providing a background for concepts such as impedance matching, noise matching, and transistor stability. Following on the heels of this discourse is the presentation of two fully integrated LNA designs, where the technology considerations of the preceding chapter are translated into bipolar and CMOS implementations which are subsequently compared through measurements. Concluding this work in Chapter 6 is the discussion of an experimental set of four voltage-controlled oscillators.

[7] U-NII is the chosen acronym for the Unlicensed National Information Infrastructure.

The bipolar versus CMOS angle returns, but is cast next to additional examples which demonstrate some of the conclusions mined by studying the time-variant nature of oscillators. As the journey through this material covers a great deal of ground in both circuit design and technology, Chapter 7 recaps the highlights. Hopefully the trip will be a rewarding one.

Chapter 2

An Approach to Spiral Inductor Modeling

Having all but disappeared from the many disciplines of integrated circuit design which sought their circumvention where at all possible, inductors are now in the midst of a renaissance. Within the context of RF and microwave applications, the renewed prevalence of inductors represents a confluence of usefulness and feasibility. Inductances of a few tenths to a few tens of nanohenries—with reasonable associated qualities—prove to be both beneficial and quite realizable in ICs designed to process signals at frequencies of around 1 GHz or higher. Values outside of this range generally contribute little to the circuit response and can be difficult to achieve in structures having inductance as the dominant trait.

Inductors can be commonly found as dissipationless conduits for DC bias levels, as elements in impedance transformation and filtering networks, in the determination of time constants and characteristic frequencies, and also in providing local feedback for either stabilization or degeneration purposes. A quick calculation reveals that 1.35 nH yields a 50 Ω impedance at 5.8 GHz; a range of useful impedances at RF frequencies is thus fairly easily covered by available inductances. Beyond this, however, a different set of desirable inductor characteristics may exist for each of the applications mentioned in the preceding list. When functioning as an isolating element in a DC bias path[1], design constraints for an inductor often favor realizing the largest possible inductance in a given available die area. Conversely, in a resonant tank circuit for an oscillator, it is the loss in the tank that typically needs minimizing at some chosen frequency. As with the scalable resistors and capacitors at the disposal of IC designers in most semiconductor processes, it is vital that the designer of an RF circuit be able to optimize inductors in terms of the value (inductance), die area, and the parasitics associated with the components. An approach to enabling these design-oriented optimizations is presented in this chapter.

[1] This usage is often referred to as an "RF choke" in radio parlance.

2.1 Integrated Inductances

Due to the inherent ease of adding transistors that it affords, the integrated circuit medium naturally lends itself to active incarnations of inductors. While a variety of inductorless circuit techniques have been considered [2], the more feasible themes for operation at microwave frequencies include single transistor terminal impedances and simple transconductor/gyrator-capacitor constructs. With appropriate terminations, a bipolar transistor[2] can exhibit inductive behavior over some frequency ranges at either the collector or the emitter [3][4]. Alternatively, the frequency response of an inductor can be synthesized using transconductors or gyrators along with capacitors in feedback circuits [5][6]. Regardless of the approach, the essential inductor property being replicated by any of the active inductance-simulating circuits is the impedance; all of these transistor-based "substitutes" unfortunately fail to yield many of the salient qualities which real passive inductors quite closely approximate (no power consumption, no noise, a wide dynamic range, and an insensitivity to supply and temperature variations). As a result, in wireless applications where performance and low power consumption are crucial, active inductors have found little usage.

Conversely, the requisite constituent of any passive inductor is simply a metal, and integrated circuit interconnect levels, bond wires, and package leads all certainly qualify.[3] Bond wires have self inductances of approximately 1 nH per millimeter of length and can be stitched between two pads on the same die [7][8] or used as inductors in the more conventional connection between a pad and the lead frame [9]. To obtain larger values, the package leads can also be incorporated with the wire bonds [10]; this can make an additional 2-5 nH available to on-chip circuitry.[4] By virtue of offering better and thicker metal that is further removed from the IC substrate and its associated losses, bond wire and lead frame inductances can provide significantly higher quality factors than can be achieved with planar spirals and three dimensional solenoids [11] built from interconnect metallization. An argument may also be made that less die area is consumed by the bond wire and lead frame inductors than by the monolithic forms. But these gains come at a price, a price that is manifest as design time and uncertainty.

[2] While implementations with bipolar transistors seem to be more commonly found, many of the same approaches also work with FET devices.

[3] Polysilicon has been used as a ground shield material, but is generally not considered appropriate for the windings of an inductor due to its high resistance (even when silicided) and proximity to the substrate relative to the metal levels in an IC process.

[4] This is a typical range of inductances associated with pins in small outline plastic packages. Metal packages optimized for microwave applications can have inductances smaller than this, and larger DIP or QFP style packages might have 10 nH or more riding along with each pin.

Planar spiral inductors are defined lithographically, resulting in component parameters that are subject only to variation in deposition and etch processes.[5] Inductances involving bond wires may be additionally affected by die placement tolerances (except when die to die bonds are used), bonding height variation, and being pushed around during plastic encapsulation (for packaging). While all of these problems are most likely solvable, they do represent a set of manufacturing issues that need to be addressed within a product flow should bond wire inductors be used. Furthermore, the need to model bond wires and/or package leads accurately for circuit design can result in complex electromagnetic simulations and perhaps even costly design iterations. Again, it should be appreciated that these challenges are tractable. However, the approach taken in this work is to recognize that while bond wires and lead frames may feature lower loss, planar spirals are often good enough[6] in many applications, and they have come to be reasonably well understood in terms of modeling and optimization. Some of the research which has yielded this wealth of understanding is reviewed in the section which follows.

2.2 Selected Recent Efforts in Spiral Inductor Modeling

One approach to modeling a complex geometry is by partitioning it into segments which are more readily analyzed and then linking together the solutions. Though not the first to apply this technique to spiral inductors, a recent effort by Long and Copeland [12] represents perhaps the most complete such treatment to date. Coupled microstrip line sections and microstrip bends (corners) were chosen as the units of analysis in this work, both of which are well covered in the extant literature. Each line segment within a rectangular planar spiral is represented by the lumped-element π-network shown in Figure 2-1. A model for an N-turn spiral inductor consists of $4N$ such sections, joined together by a series inductance and shunt capacitance which represent the current crowding effects at each corner of the winding. The self inductance of segment n within the spiral (denoted L_n) and the mutual inductances between this conductor and every other segment of the spiral parallel to it (represented by the dependent current sources) are calculated using

[5] This is, of course, also true for the 3-D solenoid inductors mentioned earlier. Unfortunately, these structures largely remain as curiosities due to having a higher loss per unit inductance compared to planar spirals and being burdened with a lack of any other particularly redeeming qualities.

[6] Good enough? This is a "purposely vague" statement if there ever was one. In this context, good enough might mean that the circuit performance has become limited by something other than the inductor (e.g., varactor Q), or perhaps that the inductor loss is already low enough for the application. Examples of the latter case include DC biasing, emitter degeneration, and even matching networks when some operational bandwidth is desired.

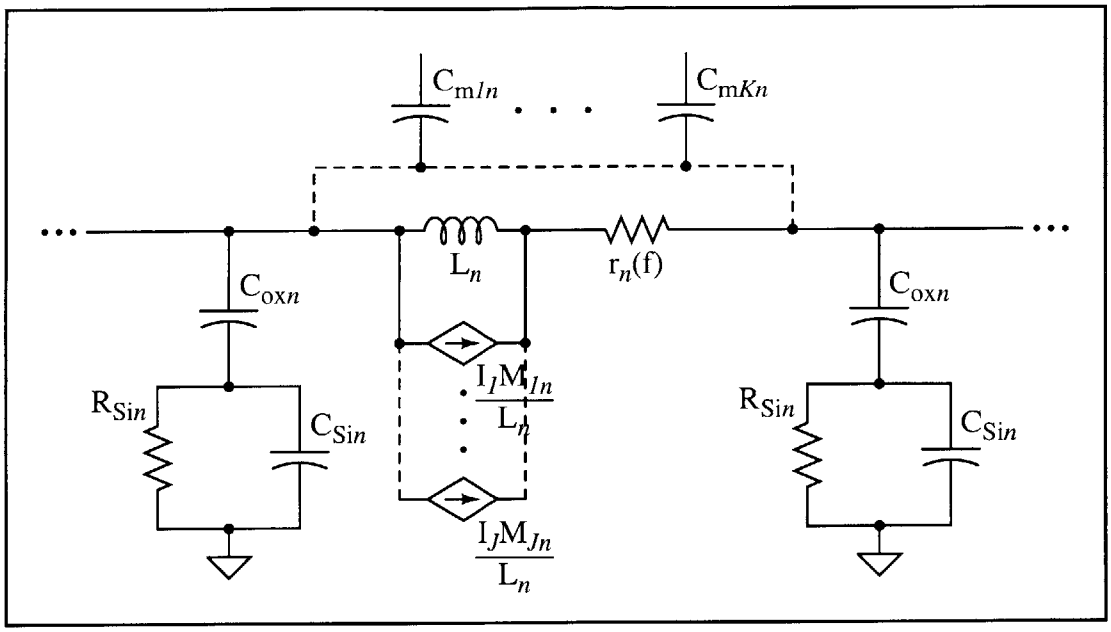

Figure 2-1. Coupled line segment model proposed by Long and Copeland [12].

closed-form expressions derived by Grover [13]. The mutual capacitances between traces (C_m) and the total self capacitance from a given trace to the substrate are computed with techniques borrowed from coupled microstrip lines. A representation for the substrate in this microstrip analysis and in the model above (R_{Si}, C_{Si}, and C_{ox}) is provided by the work of Hasegawa, et al. [14], in which microstrip transmission lines over a compound dielectric of SiO_2 on Si were considered. This model for a spiral inductor segment is then completed by adding a frequency dependent resistance, $r_n(f)$, estimated from a set of formulas that have been empirically fit to losses measured in rectangular conductors [15].

Following this procedure for each section of a winding results in a decidedly complex circuit model for a spiral inductor. Some degree of simplification may be pursued by neglecting terms that are generally small. As a starting point, the series inductance representing the higher order magnetic storage modes induced at the corners[7] is usually of little consequence for the frequencies and spiral geometries of interest (for RF circuits) and can usually be ignored. Another possibility is that the electric field interaction among segments may be dominated by immediately adjacent lines or perhaps by the underpass connection to the center of the winding. Either case allows the number of mutual capacitance elements to be abridged. But even with these reductions, a model is left that—despite possessing simple geometric inputs—remains unwieldy in the circuit design process.

[7] Depending on one's personal view of Maxwell's equations, these higher order modes of energy storage are either the result or the cause of the "current crowding" effect observed as a filament of current navigates a corner.

Irrespective of whether these simplifying assumptions can be made, the authors suggest collapsing the resulting concatenation of line segments and corners into a compact circuit model for the purposes of simulation and optimization. A typical compact representation for an inductor integrated on a silicon substrate[8] is exhibited in Figure 2-2, and is similar to that used by Long and Copeland for each segment of a spiral except that all of the intersegment coupling terms are folded into one inductance (lumped together with the self inductance) and one capacitance (appearing as the shunt element C_s). An approximate equivalent circuit based on this network can be fit to the complete model, taking the sums of the individual inductances, resistances, and substrate capacitances as initial estimates in the procedure. While this numerical fit does produce a representation more amenable to rapid calculation, any subsequent change in the inductor design necessitates another modeling iteration beginning with the parameter calculations for each segment. This is an unfortunate circumstance in that the optimization of circuits with inductors is itself frequently iterative in nature; trading away some thoroughness in modeling for a simpler procedure may thus be considered a potentially worthwhile exchange.

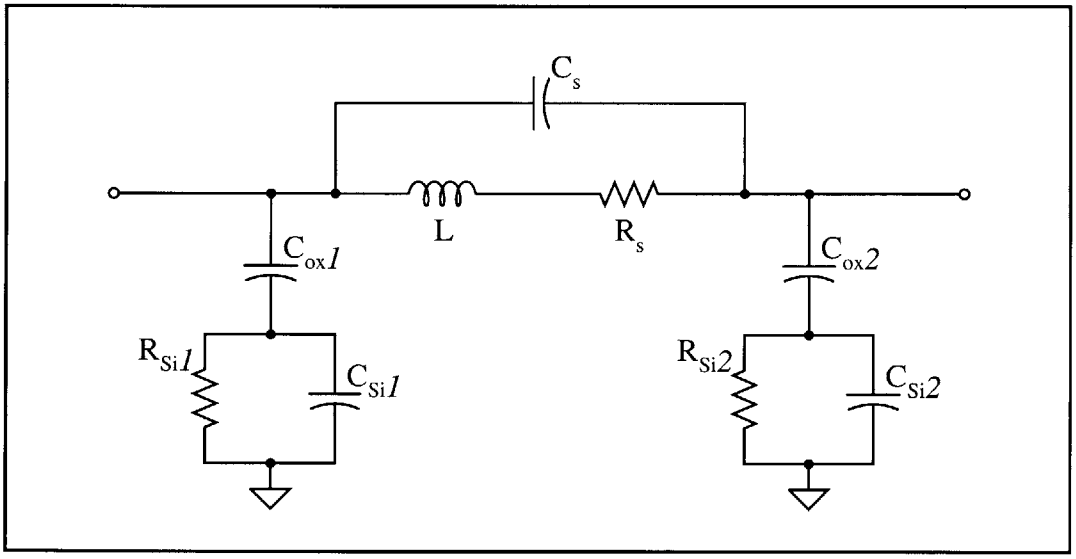

Figure 2-2. Compact lumped-element model for an inductor on a silicon substrate.

An alternative modeling paradigm is to consider the entire spiral inductor structure and attempt to reason the first-order dependencies that capture its essential characteristics.

[8] The shunt legs of the π-network representing the substrate make this model specific to the silicon medium. A model for an inductor on a semi-insulating substrate would drop the oxide capacitance term. In addition, the high resistivity of materials such as GaAs and InP makes for negligibly large equivalent substrate resistors, leaving only a capacitance to ground at either end of the network.

The hope is that this "big picture" approach can yield sufficient accuracy while eliminating computational steps between the spiral geometry data and a compact model representation by focusing on the relevant parasitic effects. A physical model proposed by Yue and Wong [16] adheres to this methodology. In this work the inductance[9] is considered laden with three dominant parasitics: series resistance, underpass capacitance, and the effect of the substrate. With the one added assumption that the substrate parasitics are equally distributed (e.g., C_{ox1}=C_{ox2}=C_{ox}), the corresponding circuit network is that shown previously in Figure 2-2.

Achieving scalability through physics-based formulations for each of the mentioned effects is a key element in this work. Taken as geometric parameters in describing the spiral are the width of the conductor that forms the winding (w), the total length of this conductor (l), the number of turns in the winding (N), and the width of the underpass conductor (w_{up}) employed to reach the inner terminal of the spiral. The first parasitic, series resistance, is assumed to originate from the familiar *resistivity*×*length*÷*area* characteristic of imperfect conductors:

$$R_s = \frac{\rho l}{w t_{eff}}, \tag{2.1}$$

where t_{eff} represents the effective conductor thickness and is used to model the effects of eddy currents within the conductor. Two sources of these currents—which oppose the applied RF signal—are discussed by Yue and Wong: self induction (from currents within the same trace, also known as the skin effect) and induction via currents flowing in adjacent traces (proximity effect).[10] Enlisting the aid of an electromagnetic field solver to analyze some representative cases, the authors propose that, at 1GHz and given typical dimensions[11] for interconnect metallization, proximity effects are not significant for uniplanar spirals. Considering then the current distribution in a microstrip line, an equivalent conductor thickness can be derived for use in Equation 2.1:

$$t_{eff}(f) = \delta(f)[1 - e^{-t/\delta(f)}]. \tag{2.2}$$

---

[9] Most of the modeling effort by Yue and Wong, and all of the discussion pertaining to it which follows, concentrates on the parasitic elements in the spiral model. For the inductance itself, the authors relied upon the Greenhouse method [18], about which more will be said within the context of alternative inductance formulations in Section 2.2.1.

[10] Ah, the joys Faraday has brought to light. Time-varying magnetic fields induce electric fields in any material (notwithstanding idealized conductors) through which they pass. Eddy currents thus also flow in the substrate, another mechanism by which R_s can increase.

[11] Metal traces with a 20μm width and 2μm spacing were simulated. To represent the materials involved, 1μm Al metal layers were chosen with an interlevel dielectric consisting of 1μm SiO_2.

In this expression, t represents the physical conductor thickness and $\delta(f)$ the skin depth of the conductive material ($\delta(f) \propto f^{-1/2}$). At 1 GHz, the skin depth of a deposited Al layer is 2.8 μm; given a 2 μm metal deposition thickness, t_{eff} is reduced by nearly 30% (to approximately 1.4 μm). This example illustrates that significant increases in the effective series resistance can indeed accrue at frequencies where the skin depth is still appreciably larger than the thickness of the metal.

The next parasitic effect incorporated into this simplified model is the capacitive coupling that shunts portions (or all) of the spiral inductance. As embodied in the approach espoused by Long and Copeland, these coupling terms exist between each pair of segments in a winding. Yue and Wong, however, suggest that these intersegment terms will be small compared to the overlap capacitance between a spiral and its underpass, and that this coupling mechanism can be adequately represented by a single parallel plate capacitor of area Nww_{up} connected across the series components of the winding:

$$C_s = \frac{\varepsilon_{ox} N w w_{up}}{t_{ox(m-m)}}. \tag{2.3}$$

The permittivity and thickness of the interlevel dielectric layer separating the "plates"—usually some variant of SiO_2 in Si-based integrated circuits—are denoted ε_{ox} and $t_{ox(m\text{-}m)}$, respectively. If an inductor were to be constructed with a wide metal winding in a process featuring five or six metal levels, it may be desirable to reduce this capacitance by using an intermediate metal layer (e.g., metal 3 in a five level metal back-end technology) for the underpass rather than the level immediately beneath the spiral. Trade-offs such as this, and the various loss mechanisms in spiral inductors, will be discussed further in Section 2.3.

Each of the remaining parasitic mechanisms instituted by Yue and Wong are owed to the substrate on which the inductor sits. Considering the now-standard SiO_2 on Si circuit model [14], the authors submit that the oxide capacitance, substrate capacitance, and substrate conductance should all scale with the area occupied by the metallization (lw). Somewhat arbitrarily assuming that the components of this substrate parasitic can be split evenly across the inductor, the final pieces of the simplified model fall into place as:

$$C_{ox} = \frac{\varepsilon_{ox} l w}{2 t_{ox(m-Si)}}, \tag{2.4}$$

$$C_{Si} = (l w C_{sub})/2, \text{ and} \tag{2.5}$$

$$G_{Si} = 1/R_{Si} = (l w G_{sub})/2. \tag{2.6}$$

Analogous to the formulation for the spiral to underpass capacitance, ε_{ox} and $t_{ox(m\text{-}Si)}$ characterize the dielectric between the winding and the surface of the Si substrate. For the substrate itself, C_{sub} and G_{sub} represent (per unit) area capacitance and conductance terms. Though physical in nature, these terms may essentially be treated as fitting parameters, established empirically through data on measured spirals. This collection of substrate elements completes a simplified yet promising picture of spiral inductors consisting of the most significant parasitic effects and constructed with models that relate the effects to geometry. While this model forms an excellent starting point, a couple of additional developments may offer improvements; these are discussed in Section 2.2.1.

A third tact toward addressing the problem of modeling spiral inductors that has been investigated is the construction of an electromagnetic field solver simplified to the extent possible for handling this one specific case. One such effort by Niknejad and Meyer has resulted in ASITIC[12] [17], a tool that combines a set of more computationally-efficient techniques that has been developed for solving EM fields around a metal winding on a multi-layered substrate, with a graphical interface for conveniently generating 2-D spiral layouts of interest. Although the simplifications do appreciably reduce analysis times, the simulations are still too lengthy to provide for efficient inductor optimization. In fact, the procedure used in ASITIC for optimizing a geometry to meet design constraints still relies upon a collection of approximate lumped-element solutions. This results in a point along the simplicity-accuracy design trade-off that is similar to the other modeling approaches that have been presented.

2.2.1 Alternative Inductance and Series Resistance Formulations

The "divide and conquer" inductor models just described segment the spiral into a conjoined set of straight-line sections for which the self and mutual inductance terms can be calculated (for each section) using Grover's closed-form expressions for rectangular[13] conductors [13]. The total inductance realized by the spiral can then be computed by summing over all the sections comprising it, a technique generally referred to as the Greenhouse method [18]. While known to yield usable results over a reasonably broad range of geometries, this method is somewhat labor intensive, and becomes difficult to apply to non-rectangular spirals; an expression that could directly provide the inductance with comparable accuracy from simple geometric parameters would be a preferable solution. Several such expressions have been proffered by Mohan, et al. [19], wherein

[12] Named more for what it's used rather than for what it is, ASITIC originates from Analysis and Simulation of Inductors and Transformers for Integrated Circuits [17].

[13] Here, "rectangular" refers to both the cross-sectional shape of the conductor and the layout.

some quasi-physical formulas have been fit to the EM simulation results for a representative library of planar spiral inductors. A comparison against measured data for 60 spiral structures demonstrates that the reported expressions perform as well as the field solver offered in ASITIC,[14] typically yielding inductances within 5% of those extracted from measurements.[15]

Defining the outermost dimension of a spiral geometry as d_{out} and the innermost as d_{in}, the inductance of a spiral can be characterized by the number of turns (N), the average diameter $d_{avg} = (d_{out} + d_{in})/2$, and the fill ratio of the spiral, where the latter is defined as $\rho = (d_{out} - d_{in})/(d_{out} + d_{in}) = W_{sh}/d_{avg}$. The second form of the fill ratio expression recalls an approximation where each side of a spiral is represented as a current sheet of width W_{sh}, while the length of the sheets and the separation between opposing sheets are both characterized by the distance d_{avg}. This gives rise to a geometric mean distance formulation for the total inductance of the spiral:

$$L = \frac{\mu_0 N^2 d_{avg} c_1}{2}\left(\ln\left(\frac{c_2}{\rho}\right) + c_3\rho + c_4\rho^2\right), \tag{2.7}$$

which is at least mildly satisfying in that the inductance is proportional to: the length of the conductor, the square of the number of turns, and a multiplicative factor that accounts for the mutual coupling from opposite sides of the spiral. The constants c_i are fitting parameters dependent upon the shape of the spiral, and are listed below in Table 2.1.

Table 2.1: Fitting Parameters for GMD Inductance Formulation, from Mohan, et al. [19]

Spiral shape	c_1	c_2	c_3	c_4
Square	1.27	2.07	0.18	0.13
Hexagon	1.09	2.23	0.00	0.17
Octagon	1.07	2.29	0.00	0.19
Circle	1.00	2.46	0.00	0.20

[14] The correspondence here is not terribly surprising as the data points to which the expressions were fit were generated by the same EM field solver. What is of note, however, is that a relatively low-order model (represented by the inductance expressions) can represent with reasonable accuracy a range of spiral inductors likely to be useful in integrated circuits.

[15] One observation on the error analysis offered by the authors is that the *smallest* inductance among the measured test set is 2.5 nH, and that the three inductors of less than 3 nH availed as data points are associated with the largest relative errors listed in the experiment. In designs intended to operate at frequencies much above a few gigahertz, the inductances will generally fall below the range spanned by the comparison set.

Another component of spiral inductors that warrants further consideration is the series resistance imposed by the winding and the manner in which it may vary with frequency. The models discussed earlier in this chapter consider only the skin effect, making the assumption that the various coupling mechanisms within the inductor structure have a negligible effect on the resistance. Yue and Wong explored this assumption through electromagnetic simulations of coupled line segments [16], but their representation of three adjacent traces does not necessarily correspond to that of a spiral inductor where the magnetic flux passing through the center of the spiral can have a much higher density than witnessed around the periphery of the spiral where the fields are not so tightly constrained. As a result, current crowding will first be observed in the inner-most turn of a spiral winding, an asymmetry not captured in the preceding analyses.

By examining the electromagnetic simulation results for a range of spiral geometries, Kuhn and Ibrahim [20] propose that a more representative distribution of the magnetic fields impinging upon the winding is that the intensity is at a maximum at the inner turn and then decreases linearly turn-by-turn. Considering the magnetic field normal to the winding and averaged (in magnitude) over the inner turn, the magnitude of the induced eddy current in this turn becomes equal to that of the applied RF excitation current at a frequency characterized by:

$$\omega_{crit} = \frac{3.1p}{\mu_0 w^2} R_{sheet}, \tag{2.8}$$

where the spiral geometry is captured by the turn pitch (p) and the width of the metal winding (w), while R_{sheet} denotes the sheet resistivity of the metal in Ohms per square.[16] Noting that, for most spirals, the metal spacing is much smaller than the width, it becomes evident that proximity effects gain significance at frequencies inversely proportional to the width of the trace used to form the inductor.

Next, formulating the power dissipated by the eddy currents within the winding, the authors use the characteristic frequency in Equation 2.8 to provide a rough estimate of the effective series resistance of the spiral in terms of its DC resistance (R_{DC}):

$$R_{eff} \approx R_{DC}\left[1 + \frac{1}{10}\left(\frac{\omega}{\omega_{crit}}\right)^2\right]. \tag{2.9}$$

[16] While the factor of 3.1 in the numerator of Equation 2.8 may look tantalizingly close to π, this in fact is little more than a fitting parameter resulting from an observation made by the authors (in the EM simulations) that the magnetic flux density normal to the inner-most turn of the spiral can be approximated as $B_N = \mu_0 I/(1.54p)$ over a fairly wide range of geometries.

This quadratic increase with frequency eventually stalls, however, as the presence of the eddy currents counters the magnetic field that was initially assumed, resulting in a more linear relationship at higher frequencies. Kuhn and Ibrahim submit that the transition between the quadratic and linear regimes of current crowding will occur at a frequency roughly a factor of 4 to 6 beyond ω_{crit}. Though this limits its range of validity, the approximate expression for effective resistance (Equation 2.9) does catch some interesting frequencies in the low gigahertz RFIC design space. Calculating for the cases of 1μm thick aluminum in widths of 10μm and 20μm, the corresponding critical frequencies are found to be 1.2GHz and 600MHz, respectively. Hence a region of resistive behavior quadratic in frequency up to 2-3GHz might be expected in typical spiral inductor applications.

It can be witnessed in these examples and the preceding relationships that the proximity effect does not depend on either the size of the inductance or the number of turns in the spiral, but rather just the width of the conductor and its resistivity. Being induced by the same magnetic field-based mechanism, the skin effect is similarly independent of the spiral geometry, and also begins to play a role in inductor loss over the same frequency ranges discussed in the context of current crowding. Perhaps the best lesson to be taken from this work is that the frequency-dependent nature of conductor loss is a difficult modeling challenge and, as a result, any simplified model is likely to only work well in a restricted range of frequencies and geometries. Some first-order considerations are captured by the models presented here which allow for optimization within limitations, but a final-stage spiral design should still be submitted to an electromagnetic simulation tool when inductor loss is of critical importance.

2.3 Spiral Inductor Loss Mechanisms

If there is anything that can be considered an absolute certainty with integrated inductors, it is that the devices will fall short of their lossless ideal. A cross-sectional diagram of a spiral inductor sitting on an insulating layer covering a silicon substrate is shown in Figure 2-3, where an external signal has been applied to the metal winding which in turn carries the signal current I. Some of the signal energy will be dissipated by resistance encountered in the winding, and is generally the dominant spiral inductor loss mode at low frequencies. While the resistance will effectively increase with frequency (as discussed in the preceding section), other loss mechanisms will begin to play even more significant roles at higher frequencies. First, the potential associated with the applied signal creates an electric field for which the substrate is essentially an RF termination. At

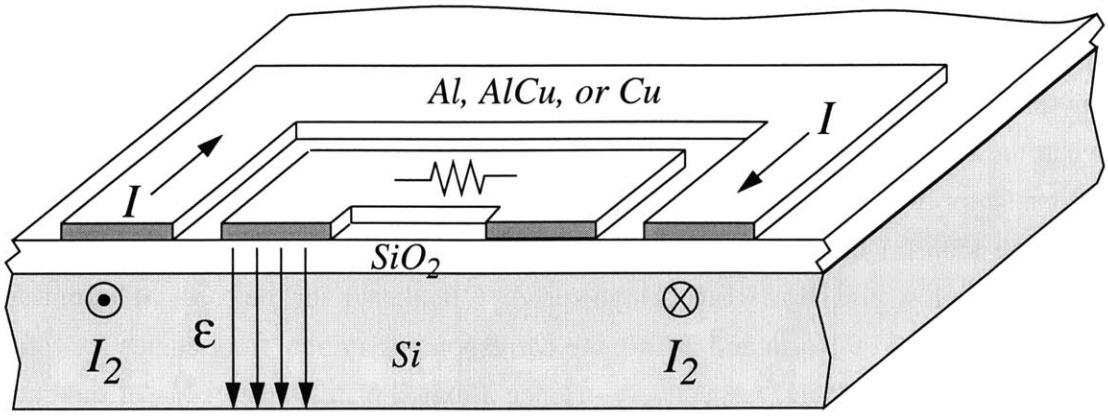

Figure 2-3. Loss mechanisms in a spiral inductor.

lower frequencies—in typical silicon material—charge carriers in the substrate are able to respond and "follow" the electric field, and the field thus terminates at the interface between the oxide and the silicon.[17] As the signal rates increase, however, the carriers can no longer keep pace, and the electric field subsequently penetrates into the substrate as sketched above in Figure 2-3. When the charges move in response to, but lag behind the applied field, the resultant (capacitive) current contains a real (i.e. in-phase) component that can be modeled by an insulator having a dielectric constant with an imaginary term that is frequency dependent. A capacitor constructed from such an insulator is lossy and can be represented in circuit form by the addition of a parallel substrate resistance [14]. While potentially addressable through the incorporation of patterned ground shields [21], this capacitive loss mechanism begins to limit inductor quality at moderate RF/microwave frequencies.[18]

A third effect is that the magnetic field generated by the signal current flowing in the spiral also penetrates into the substrate. Analogous to a transformer where the field generated by one winding passes through another, a current is induced in the substrate (shown in Figure 2-3 as I_2) flowing in the direction opposite that of the intended signal. Continuing the transformer analogy, this eddy current flows in the secondary (substrate) side as a result of signal energy applied to the primary winding and, just as with a trans-

[17] This behavior is why the lumped oxide capacitance approximation, typically accepted without question by IC designers, works at all. Without an explicit conductor or other source of carriers (e.g., a MOSFET source/drain region) present beneath the oxide, this approximation becomes dubious at gigahertz frequencies in lightly doped (e.g., 10-20Ω-cm) silicon, and even earlier when more heavily doped wafers are used.

[18] And yes, this is another purposely vague description. Generally this effect becomes prominent in the neighborhood of a few gigahertz, although the location of the "neighborhood" depends on the area occupied by the spiral, oxide thickness, and substrate resistivity. More regarding this topic can be found in the paper by Hasegawa, et al., which analyzed SiO_2 on Si microstrip [14].

former, loss encountered in the "secondary" due to finite substrate resistivity can be observed through the primary side. At higher RF and microwave design frequencies, this becomes a very significant loss term in silicon-based spiral inductors.

2.3.1 Inductor Quality Factor

Enough has been written in this chapter alone to be suggestive of the potential for inductor loss, but some means of quantifying the loss is needed. The concept of a quality factor, a ratio between stored and dissipated energies within an element, proves not only enlightening but also quite useful for design and optimization. Despite the apparent fundamental energy-based origins, however, there remains considerable debate surrounding the definition of quality factor (often called simply just "Q"). Invariably, some of the questions being debated pertain to application-specific contexts, although a similarly enthusiastic discussion can be found regarding *which* measure of stored energy should be used in the assessment of quality factors.

For the case of an inductor, the most conventionally-used definition of Q involves the energy that an external source is able to store in the inductance divided by the energy dissipated in the device. Under this interpretation, the electrical energy storage elements (capacitors) associated with the inductor effectively compete against the external source; at the self-resonant frequency, the only energy taken from the source will be that dissipated by the loss mechanisms in the inductor, and hence the Q will be zero. Stated mathematically, this involves a subtraction of the peak energy stored in electrical form from that in the magnetic fields of the inductor [22]:

$$Q \equiv 2\pi \frac{\left|\text{Peak Magnetic Energy} - \text{Peak Electrical Energy}\right|}{\text{Energy Loss in One Oscillation Period}}. \tag{2.10}$$

Given this formulation and the lumped-element model shown in Figure 2-2, the quality factor of a spiral inductor can be calculated as:[19]

$$Q = \frac{\omega\left[\frac{L}{Z_s^2} - C_s - C_{ox}(1 - s\tau_{sub})^2 - C_{Si}(s\tau_{sub})^2\right]}{\frac{R_s}{Z_s^2} + \frac{(s\tau_{sub})^2}{R_{Si}}}, \tag{2.11}$$

where the series inductor components are grouped together as $Z_s = R_s + j\omega L$ and the effective time constant associated with the substrate parasitic elements is denoted:

[19] For the truly curious, an element-by-element derivation is provided in the Appendix.

$$\tau_{sub} = \frac{R_{Si}C_{ox}}{1 + sR_{Si}(C_{ox} + C_{Si})}. \tag{2.12}$$

Equation 2.11 can be observed to collapse to $\omega L/R_s$ at low frequencies, the relationship expected in a simple model of an inductor having only a series resistor loss element. But more generally, this definition is also consistent with the approach of directly calculating quality factors from measured small-signal parameters, being equivalent to the oft-used forms $-Im\{y_{11}\}/Re\{y_{11}\}$ and $Im\{z_{11}\}/Re\{z_{11}\}$ [23].[20] This correspondence, and the convenience it affords, undoubtedly is a significant contributor to the continued popularity of this manifestation of Q.

Unfortunately, a shade of murkiness begins to be cast over this picture of quality factor in noting the many circuit applications which introduce capacitance in conjunction with an inductor. It can be argued, for example, that the Q associated with a resonator is not accurately reflected when the above definition is used to describe a constituent inductance. Consider first the extreme case of an inductor used at self-resonance, or equivalently, lowering the resonant frequency of the inductor by adding an ideal capacitor. At resonance, the electrical and magnetic energy storage forms will balance, resulting in $Q = 0$ for the inductor via Equation 2.11. It is evident, however, that energy is indeed being stored in the resonator, which should therefore have $Q > 0$; this would appear to violate the intuitive notion that the Q of the whole should not exceed the Q of any of its parts.[21]

One way out of this quagmire is to use bandwidth-based definitions for quality factor, provided that the circuit networks under consideration are approximately second-order and that the loss is not prohibitively high. For a resonator, Q can be observed directly from the frequency response as the ratio of the center frequency to the 3dB bandwidth. A consistent approach for an inductor makes use of the same definition, and can be accommodated by the addition of (lossless) capacitance as required to resonate the inductor at the frequency of interest—a manipulation that can be performed mathematically given a two-port characterization of the inductor. For further discussion regarding the practice and pitfalls in quality factors, the reader is referred to a paper by O [24] in which a number of Q metrics are analyzed and compared across some measured inductors.

---

[20] For the equivalence to hold, the inductor must effectively be connected in a one-port configuration (i.e. having one of the inductor terminals grounded). Regardless of the physical connection, using the short-circuit admittance (y) parameters enforces this. However, $z_{11} \neq 1/y_{11}$ unless the second terminal is physically a small-signal ground.

[21] The most satisfying relationship for an LC tank would be $1/Q = 1/Q_L + 1/Q_C$, where Q_L and Q_C denote the individual quality factors for the inductor and capacitor.

It would seem then that there is no easy answer to be found in the debate over quality factor definitions, in that the most appropriate description will always be application-dependent. For an oscillator circuit, Q values based on bandwidth considerations will generally be more relevant, although the calculations should incorporate transistor parasitics associated with the oscillator itself in addition to any output loading.[22] Conversely, when inductive behavior is desired for tuning or stabilizing a gain stage, the metric defined in Equation 2.10 is likely to be of greater consequence. Pragmatically though, for inductances smaller than a few nanohenries, differences among the various definitions of Q are not large—in either the value or the frequency at which the value peaks. While these differences should not be dismissed as insignificant, design decisions are not likely to be appreciably impacted by the working definition chosen for quality factor.

2.4 Quick Turn Models for Spiral Inductors

Earlier sections in this chapter have discussed typical applications of inductors, investigated a few efforts geared toward providing models of them, and explored loss as one of the predominant concerns encountered in using integrated inductances. As intimated earlier, however, a means of rapidly generating circuit models from geometric descriptions is required to enable the design of circuits using inductors. The models are used for both circuit simulation and also for balancing the various loss mechanisms within the inductor itself—by scaling the geometry—to optimize a chosen metric (e.g., Q for a given inductance) over a desired range of frequencies. The purpose of what follows is not to propose any ground-breaking modeling advances, but rather to describe a collection of approximate tools which work well together and provide a convenient rapid modeling capability.

For the spiral inductors used in conjunction with the circuits presented throughout this thesis, a first-pass calculation of the inductance is obtained from the current sheet formulation tendered by Mohan, et al., and presented in Equation 2.7. The associated parasitic elements are computed following the approach of Yue and Wong as described by Equations 2.1 through 2.6. Unit capacitance and conductance terms (i.e. per unit area) are required in this approach for modeling the substrate, and the microstrip on silicon work of Hasegawa, et al. [14] is tapped to derive these components. Lightly-doped[23] substrates

[22] This is sometimes called the "loaded" Q of the resonator, even when the only loading is from transistor parasitic elements. Maximizing loaded Q for an oscillator can be an important consideration in sizing the inductive portion of a resonant tank.

[23] Silicon substrates used for RFIC applications are typically p- material in the 10-20Ω-cm range.

support a mode of propagation at RF frequencies that is transitional in nature between a slow-wave mode and a more familiar quasi-TEM arrangement,[24] a behavior that is modeled by an oxide capacitance (ε_{ox}/t_{ox}), a substrate capacitance (ε_{Si}/t_{Si}), and a substrate conductance (σ_{Si}/t_{Si}).[25] Finally, to allow comparisons across candidate spiral inductor structures, the loss corresponding to each spiral is quantified using the relationship for Q established in Equation 2.11; this particular quality factor metric was chosen largely for its ease of calculation. These equations have been coded into a MATLAB [25] routine so that a suite of lumped-element model component values (L, R_s, C_s, C_{ox}, C_{Si}, and R_{Si}) can at once be generated for a list of inductors specified by a set of geometric parameters. Once the optimal geometry has been identified from the model data, FastHenry [26] analyses are performed on the chosen spiral structure to more accurately determine the series resistance and inductance at the frequency of interest.

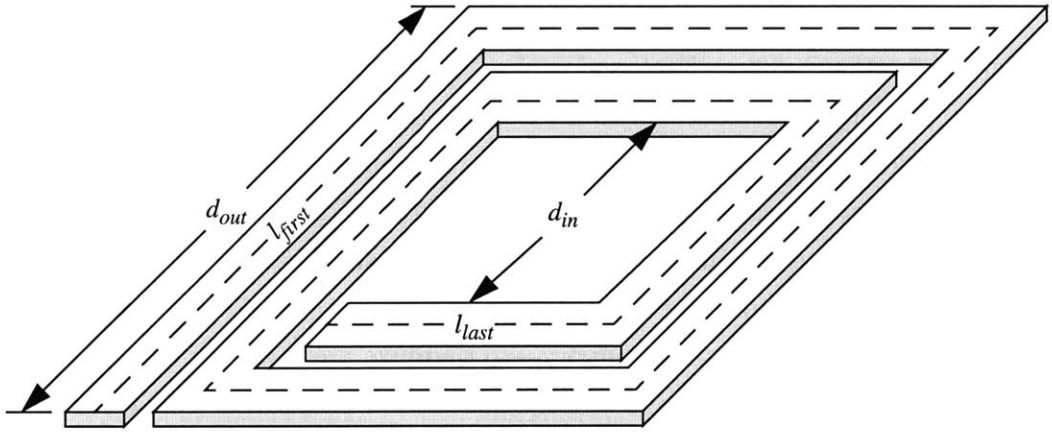

Figure 2-4. Geometry parameters describing a square spiral inductor.

While the implementation of this approach is relatively straight-forward, one non-trivial aspect is determining the total length (l) of the metal traces comprising the spiral (required in the expressions for R_s, C_{ox}, C_{Si}, and R_{Si}). For the purposes of this model, and as indicated by the dashed line in Figure 2-4, the length of the spiral should be measured

[24] The difference in the propagation modes pertains to the behavior of the carriers in the substrate in response to the electric field impressed by the applied signal. A quasi-TEM mode predominates at higher frequencies, when the carriers in the substrate are no longer able to respond quickly enough to charge and discharge the oxide capacitance. Conversely, the slow-wave mode is characterized by a movement of charge at the substrate-oxide interface in response to the time-variant electromagnetic fields. The latter mode results in a higher effective capacitance per unit length of (microstrip) trace, resulting in a lower propagation velocity.

[25] Each of these per unit area terms can be modified to account for the fringing effects (i.e. width dependence) prevalent when the trace width is much smaller than the substrate thickness. The classical microstrip treatments can also be applied to this multiple dielectric case.

along its center line. Considering the geometry of a square spiral having N turns, an outer dimension d_{out}, metal width w and spacing s, the length can be expressed as:

$$l = l_{first} + 2\left(\sum_{k=0}^{2(N-1)} [d_{out} - w - k(w+s)] \right) + l_{last}, \tag{2.13}$$

where the lengths of the first segment of the spiral $l_{first} = d_{out} - w/2$ and the last segment $l_{last} = d_{out} - w/2 - (2N-1)(w+s)$ are indicated separately. Inserting the expressions for the first and last segments, evaluating the summation, and simplifying the result, the length of metal comprising a square spiral can then be quickly determined for an integer number of turns:

$$l = d_{out} + (4N-1)(d_{out} - w) - (2N-1)^2(w+s). \tag{2.14}$$

The quick-turn spiral inductor model described above can essentially be used for arbitrary shapes, although inductance fitting coefficients are only available for the more commonly encountered squares, hexagons, octagons, and circles. Lengths associated with hexagonal and octagonal inductors can be addressed through variants of Equation 2.13, or by approximating the shapes as being circular for the purposes of calculating the series resistance and substrate parameters. True circular spirals are characterized by a radius that decreases toward the center at a fixed angular rate; the length of a circular spiral may be posed in terms of its initial radius (r_i) and final radius (r_f) as:

$$l = \int_0^{2\pi N} \left[r_i - \left(\frac{r_i - r_f}{2\pi N} \right) \theta \right] d\theta, \tag{2.15}$$

assuming $r_i > r_f$. Noting that $r_f = r_i - N(w+s)$, and performing the integration, the length of trace comprising a circular spiral is also seen to be readily determined:

$$l = \pi N[2r_i - N(w+s)]. \tag{2.16}$$

This completes a study of spiral inductors that has been geared toward—but not limited to—use in silicon-based integrated circuits. The silicon medium introduces loss and additional modeling complexity into a device already rife with coupling mechanisms. But a number of recently-advanced treatments handle many of the most significant effects in terms of geometric dependencies, yielding some hope that inductors need not be relegated to the design by trial-and-error scrap heap.

Appendix: Calculation of Inductor Quality Factor

While quality factors can be derived from measured data, it is also satisfying to know that the energy-based perspective of Q provided by Equation 2.10 can be directly applied to a circuit model of an inductor. Grounding one port of the spiral inductor model and driving the other with a sinusoidal voltage source (y-parameter configuration) of amplitude V_1, the circuit to analyze appears as shown below.

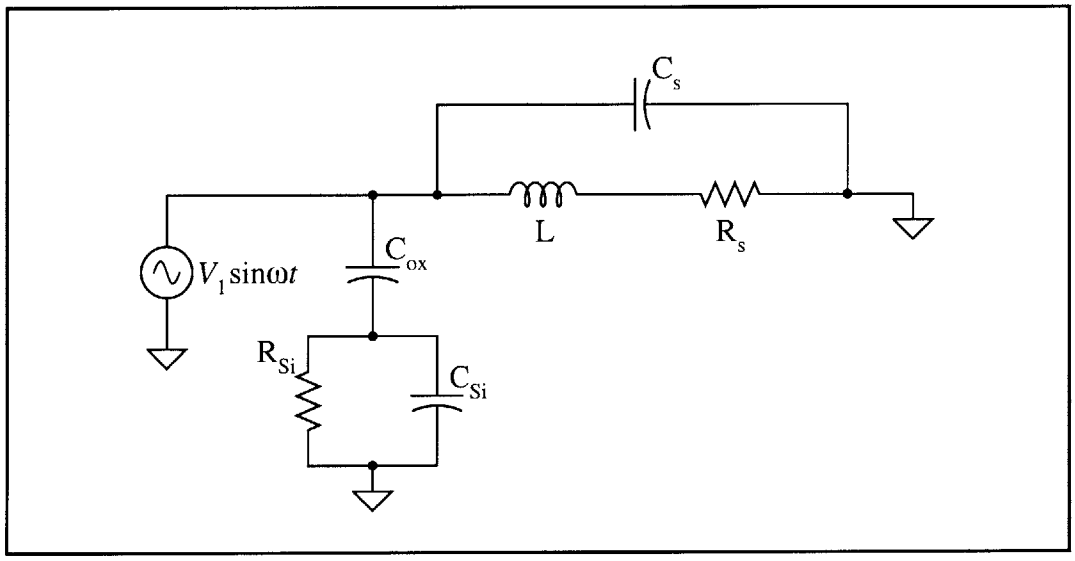

Figure 2-5. Circuit model for calculation of inductor quality factor.

In response to the applied voltage, the peak current flowing in the inductor will be:

$$i = \frac{V_1}{R_s + j\omega L}, \tag{2.17}$$

and thus the maximum energy stored in the inductance is:

$$W_L = \frac{1}{2}Li^2 = \frac{L}{2}\frac{V_1^2}{(R_s + j\omega L)^2}. \tag{2.18}$$

The same current flows in the resistance R_S, prompting an energy loss per oscillation period ($\omega/2\pi$) as given by:

$$W_R = R_s\frac{i^2}{2}\left(\frac{2\pi}{\omega}\right) = \frac{R_s}{2}\frac{V_1^2}{(R_s + j\omega L)^2}\left(\frac{2\pi}{\omega}\right), \tag{2.19}$$

where the factor of one-half is needed to get the RMS value of the current. If all of the capacitances in the circuit were negligible, the inductor quality factor would then be:

$$Q = 2\pi\frac{W_L}{W_R} = \frac{\omega L}{R_s}, \tag{2.20}$$

as expected and as observed at low frequencies.

Unfortunately, the capacitances all too often cannot be neglected, and so the analysis and subsequent expressions for Q become more complex. The voltage applied to the circuit also falls across the shunt capacitance C_s, imparting a peak energy:

$$W_{C_s} = \frac{1}{2}C_s V_1^2. \tag{2.21}$$

The substrate network divides the applied signal, with the voltage across the substrate RC being given by:

$$V_{Si} = \frac{sR_{Si}C_{ox}}{1 + sR_{Si}(C_{ox} + C_{Si})}V_1 = s\tau_{sub}V_1, \tag{2.22}$$

with τ_{sub} being introduced to simplify later expressions.[26] Summing the peak energies stored in the oxide and substrate capacitances yields:

$$W_{C_{sub}} = \frac{1}{2}C_{ox}V_1^2(1 - s\tau_{sub})^2 + \frac{1}{2}C_{Si}V_1^2(s\tau_{sub})^2, \tag{2.23}$$

completing the litany of energy storage terms. Finally, V_{Si} is also impressed across the substrate resistance R_{Si}, dissipating an energy per period:

$$W_{R_{Si}} = \frac{V_1^2}{2R_{Si}}(s\tau_{sub})^2\left(\frac{2\pi}{\omega}\right), \tag{2.24}$$

making the total loss per period in the inductor:

$$W_{loss} = W_{R_s} + W_{R_{Si}} = \frac{2\pi}{\omega}\frac{V_1^2}{2}\left(\frac{R_s}{Z_s^2} + \frac{(s\tau_{sub})^2}{R_{Si}}\right), \tag{2.25}$$

where $Z_s = R_s + j\omega L$ is used for further simplification. Quality factor is then constructed with the capacitive energy terms subtracting from W_L, matching the expression shown earlier in Equation 2.11:

$$Q = 2\pi\frac{W_L - (W_{C_s} + W_{C_{sub}})}{W_{loss}} = \frac{\omega\left[\frac{L}{Z_s^2} - C_s - C_{ox}(1 - s\tau_{sub})^2 - C_{Si}(s\tau_{sub})^2\right]}{\frac{R_s}{Z_s^2} + \frac{(s\tau_{sub})^2}{R_{Si}}}. \tag{2.26}$$

[26] It will be well worth the extra notational confusion!

Chapter 3

Time-Variant System Modeling

Together with a collection of passive elements, oscillators, mixers, and amplifiers form the canonical infrastructure of RF and microwave transceivers. Oscillators and mixers are central to the frequency translation and synthesis processes, and amplifiers can be targeted for low noise, power, intermediate frequency, and signal-limiting applications. Understanding the manifestation of noise in these circuits is essential for extending RF system performance while reducing the costs of development. Where the front-end components were once considered largely within the province of design-by-experience heuristic approaches, greater sophistication in the modeling is becoming increasingly vital as technological limitations are pushed by demands for increased bandwidth and functionality in wireless systems.

Two important distinctions that can be drawn with respect to modeling any circuit pertain to linearity and time-invariance. Linearity is defined by the principle of superposition—perhaps one of the greatest friends of engineers everywhere—and implies that a system yields the same behavior regardless of the size of the signals or the number of them that are applied. Stated more formally, if a system responds with an output $y_1(t)$ when subjected to an input $x_1(t)$, and similarly generates $y_2(t)$ given a different input $x_2(t)$, then a composite and scaled input $ax_1(t) + bx_2(t)$ will result in an output $ay_1(t) + by_2(t)$ when applied to the same system if it is linear. Time-invariance is another property that lends itself to simplification, implying that there is no inherent context of absolute time within the system. A given input signal generates the same output irrespective of when that input is applied. Framed within a fixed time reference, if the input $x_1(t)$ is delayed by a time τ: $x_1(t\text{-}\tau)$, then the corresponding output is simply $y_1(t)$ delayed by the same interval: $y_1(t\text{-}\tau)$. The combination of these two qualities results in a linear time-invariant (LTI) system that can be completely characterized by a single impulse response, a greatly beneficial trait that allows the response to an arbitrary input to be determined from a convolution integral.[1]

[1] This property also opens the door to all the usual frequency-domain analysis tools.

Among the active elements in RF transceivers, amplifiers can generally be considered LTI over a useful range of inputs, with some measure of the limits to the linearity being conveyed through distortion or compression levels. Linear time-invariant models have also been used to help cast light upon design trade-offs for oscillators [27]-[29] and mixers [30]. Though often yielding intuitively-appealing analogies, the LTI-based models of oscillators and mixers fail to catch some fundamental mechanisms of their operation, potentially leading designers astray. This chapter investigates these limitations and presents an approach for using time-variant models to analyze noise in oscillators and mixers.

3.1 A Plausibility Argument for Time-Variant Models

The first widely-accepted[2] framework for considering oscillator noise was tendered in the mid-1960s by Leeson [28], wherein a plausible argument was crafted to explain the spectrum observed around the carrier generated by an oscillator circuit. Instilling slightly more rigor into the framework, Lindenmeier subsequently used an LTI description to demonstrate that the spectrum primarily consists of phase noise, and that the spectral density arising from a resonant oscillator (excluding flicker noise sources) could be expected to behave as [29]:

$$\mathcal{L}\{\Delta f\} = 10\log\left[\left(\frac{kTF}{P_o}\right)\frac{1}{Q^2}\frac{1}{(\Delta f/f_o)^2}\right]. \tag{3.1}$$

Although not introduced until well after the initial theories were advanced, $\mathcal{L}\{\Delta f\}$ has become the commonly-used notation to indicate the noise power in a 1 Hz bandwidth at a frequency displaced Δf from the carrier, expressed in dB relative to the carrier power (P_o), and with the units usually expressed as dBc/Hz. In Equation 3.1, the oscillator is characterized by the quality factor of the resonator (Q) and a multiplicative excess noise parameter (F) that is used to fit the noise contributions of the circuit.[3] This relationship also lends itself to a satisfying energy-based interpretation: as the energy that can be held in storing the state of the oscillator becomes larger, the susceptibility of the (state-space) trajectory to perturbations from noise decreases, subsequently lowering phase noise. Increases in the carrier power are a reflection of having more energy stored in the resonator, and a

[2] This may be an indication that it was the first widely-*understood* treatise concerning oscillator noise. Expositions concerning variations in the frequency of oscillation can be traced back to the 1930s [31].

[3] The excess noise parameter was originally conceived to be related to the noise factor of the amplifier stage around which the oscillator was formed, giving rise to its designation as F.

higher Q means that less of it is dissipated during an oscillation cycle. Conversely, higher carrier frequencies necessitate smaller-valued resonant tank elements; the corresponding reduction in energy available to store the oscillator state leads to increased phase noise.[4]

Despite the lack of a clear idea as to what the excess noise factor is or on what it depends, Leeson's model was the predominant view of oscillator noise for about 30 years. More recent developments, however, have illustrated the importance of incorporating the time-varying behavior exhibited by oscillators. As a demonstration, consider an ideal LC tank in isolation. Once energy has been introduced into the system, the energy will iterate between being stored in the inductor and the capacitor, indefinitely producing a sinusoidal signal as depicted by the solid traces in Figure 3-1. The dashed lines overlay the perturbation response of the system to two impulsive noise events equal in amplitude but occurring at different locations within the oscillation period. In the first case, an impulse of energy is injected into the LC circuit when the signal is at an extremum. The amplitude of the oscillation responds immediately, but the phase of the signal is unchanged. The dynamics of amplitude restoration essential to oscillators cause the disturbance to decay with time, leaving behind no lasting effect to the state of the system.[5] However, when the same noise event is introduced near a zero-crossing in the waveform, a decidedly different picture emerges as the phase of the oscillation is offset by the disturbance. While no effect

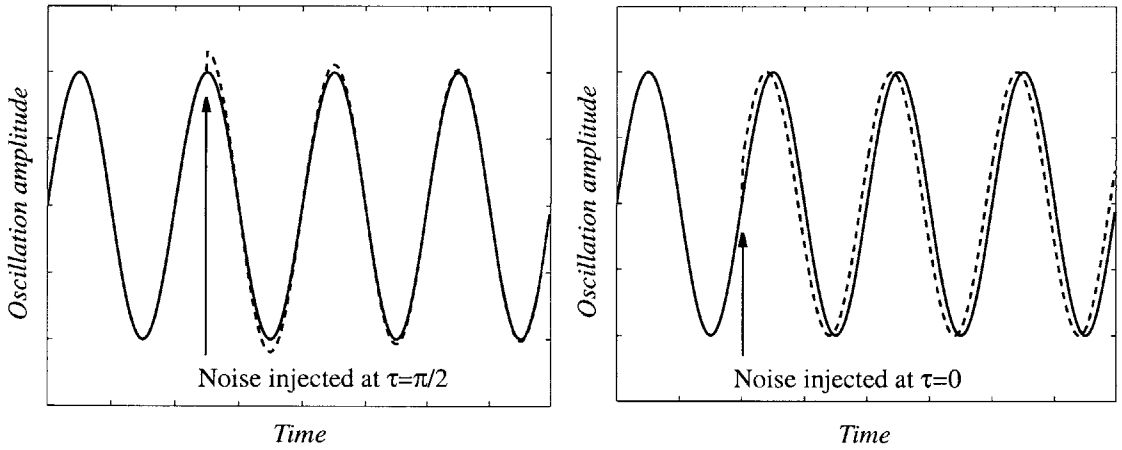

Figure 3-1. Pictorial argument for a time-variant oscillator model, adapted from [34].

[4] An alternative view is to consider the resonator as filtering the spectrum of the noise, limiting the effective noise bandwidth impinging upon the oscillator to $f_{BW} = f_o/Q$. The phase noise of the oscillator is then proportional to the power spectral density of the noise and the square of the noise bandwidth (i.e. $S_\phi \propto kTFf_{BW}^2$) [32].

[5] Some type of amplitude-limiting mechanism is required for a steady-state oscillation to exist [27]. Fortunately, there is such a mechanism inherent to the behavior of many transistor circuits, as gain usually falls with increasing signal swing.

upon the signal amplitude is witnessed, oscillator dynamics lack a complementary mechanism to restore phase, and so the disturbance persists.[6]

As a second example, a single-balanced mixer is shown in Figure 3-2, where a differential local oscillator (LO) drives the emitter-coupled pair and the incoming RF signal is coupled to the lower transistor. From the perspective of signal transmission to the IF outputs, three different modes of operation are exhibited on a periodic basis as determined by the applied LO. When v_{LO^+} is more than a few kT/q above v_{LO^-}, all of the current established by the tail transistor passes through the left side of the differential pair, enabling a cascode amplifier connection between v_{RF} and v_{IF^+}. Under the opposite condition ($v_{LO^-} - v_{LO^+} \gtrsim 3kT$), v_{IF^+} is pulled toward the supply voltage through the collector load and the transmission from the RF input to the positive IF output is very small. Conversely, the switching core operates in a high differential gain mode during the interval that the LO is between these extremes, with approximately half the RF signal energy being directed toward either output.

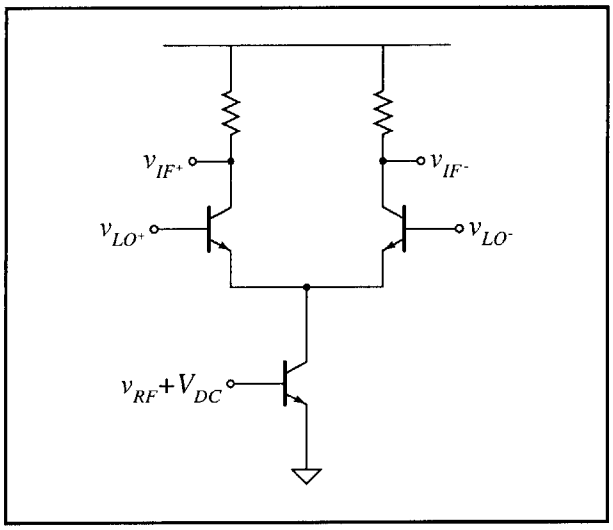

Figure 3-2. Simplified schematic of a single-balanced mixer.

Given the three distinct modes evidenced in the signal gain that varies periodically with the LO, it should come as no surprise that an LTI-based model would fall short of capturing essential mixer behavior. The time-varying nature of the circuit not only modulates the magnitude and effect of each noise source (as observed in the oscillator example), but also fundamentally underpins the frequency conversion process. While it may be tempting to try using time-averaged values for gains in the calculation of the mixer noise

[6] A phase-locked loop (PLL) provides a phase-restorative function for an oscillator, which is why the phase noise spectrum of an oscillator is attenuated by a PLL.

figure, even an ideal switching core aliases noise in the input current from frequencies around each LO harmonic to the same IF band, appreciably increasing the noise figure above that of the input transconductor. A method of capturing the effects of time-variant behavior is clearly needed to help guide the mixer and oscillator designs that are ubiquitous in RF circuits, and to provide greater insight into the selection and optimization of semiconductor technologies for these applications. This chapter develops an approach to using time-variant models in the analysis and design of oscillators, and concludes with a discussion of how the same method can be extended to mixers.

3.2 Oscillator Analysis Using a Linear Time-Variant Model

Once the assumption of time-invariance in oscillators and mixers has been observed to be precarious, the question becomes where to turn next. Fortunately, the problem is not completely unconstrained, as both types of circuits operate under a periodic steady-state condition; mixers and oscillators may not be time-invariant, but periodicity limits the scope of the analysis and leaves open the possibility of using Fourier-based techniques. Furthermore, linearity in these RF circuits can be seen to hold for the small amplitudes generated by noise sources and exhibited by most RF and IF signals [33]-[35], a property that simplifies the task considerably. The use of a periodic time domain approach to analyzing the effects of noise in oscillators was first illustrated by Kärtner in 1990 [36]. Kärtner introduced the application of perturbation methods for solving the differential state-space equations which represent the behavior of an oscillator and noted at this time that linearity should persist. However, the presentation in this section is based on an interpretation first offered by Hajimiri and Lee [38].

A linear time-variant (LTV) system can be characterized by an impulse response, albeit one that is time-varying. While its calculation is often bypassed for LTI circuits in favor of frequency-domain approaches, an impulse response can be directly computed with the aid of any circuit simulator. Following the development by Hajimiri and Lee [34], an oscillator is first subjected to a transient analysis run until the steady-state condition is reached. An impulse of energy is then introduced into the oscillator by applying a narrow pulse[7] from a current source, and the output signal that results is compared

[7] When an input signal is much shorter in duration than the response time of the circuit to which it is applied, the circuit does not respond to the shape of the input but rather just the energy contained within it. A discrete Fourier transform (DFT) of the signal approximating an impulse can be used to validate this criterion. The transform of an impulse is a constant, and the DFT of the pulse used in simulations should closely approximate this through the frequencies of interest.

against the undisturbed oscillation to gauge the impulse response of the circuit. This procedure is depicted schematically in Figure 3-3(a-c).

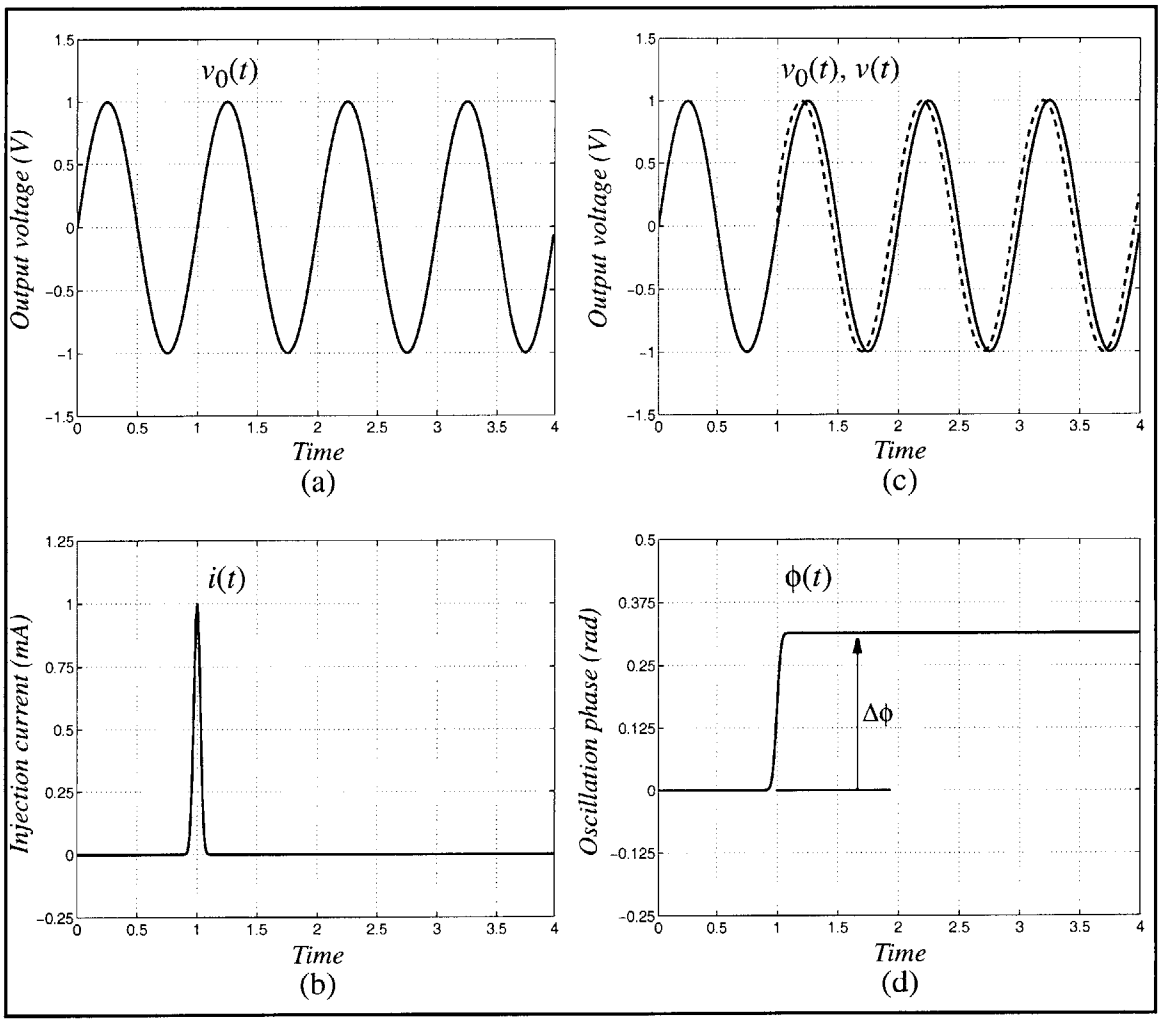

Figure 3-3. Calculation of the impulse sensitivity of an oscillator; (a) steady-state oscillation voltage, (b) current injected to approximate an impulse function input, (c) impulse perturbation response compared with the steady-state oscillation, (d) phase of the output signal in the perturbation response.

The impulse of charge introduced by the current source deposits onto circuit capacitances that store the state of the oscillator, instantaneously changing the value of the voltage state variable. Continuing surreptitiously from this new state, the oscillation thus withstands a step in phase; the change in phase can be calculated by measuring the extent to which the zero-crossing points in the output waveform have shifted relative to the corresponding points in the unperturbed steady state. Normalizing the time shift to an oscillation period of 2π radians and plotting the phase difference at the time of each zero-

crossing, a step function of height $\Delta\phi$ is observed as shown in Figure 3-3(d). Using this definition of phase, the output of the oscillator—incorporating the disturbance to the system shown in Figure 3-3(b)—can then be described as $v(t) = V_0 \sin(\omega_0 t + \phi(t))$, with $\phi(t) = 0$ taken as the steady-state condition.

For an LTV system, the output can be determined for an arbitrary input $i(t)$ by superposition once the impulse response $h(t,\tau)$ has been found. Assuming the step function impulse response that is characteristic of every oscillator, the output phase can be related to an input current by:

$$\phi(t) = \int_{-\infty}^{\infty} h_{\phi}(t,\tau) i(\tau) d\tau = \frac{1}{q_{max}} \int_{-\infty}^{t} \Gamma(\omega_0 \tau) i(\tau) d\tau, \tag{3.2}$$

where the amplitude of the impulse response—a periodic function relative to the oscillation period—is represented as $\Gamma(\omega_0\tau)/q_{max}$. This integral evaluates trivially when the input is an impulse function, yielding the value of the integrand at the moment the impulse occurs ($t = \tau_i$):

$$\phi(t)\big|_{t=\tau_i} = \frac{\Gamma(\omega_0 \tau_i)}{q_{max}} q_i. \tag{3.3}$$

Again harkening to the dynamics of oscillators where the step in phase resulting from an impulse of current into a capacitor is instantaneous, the sensitivity of the oscillator phase at an offset τ_i into the oscillation period can then be evaluated as:

$$\Gamma(\omega_0 \tau_i) = \frac{q_{max}}{q_i} (\Delta\phi\big|_{t=\tau_i}), \tag{3.4}$$

where q_i is the charge delivered by the impulse, q_{max} is a normalizing constant representing the amount of charge that is displaced in storing the state of the oscillator (the product of the total capacitance holding the state and the maximum voltage swing developed across this capacitance),[8] and $\Delta\phi$ is measured from the transient simulation waveforms as illustrated in Figure 3-3(c,d).

Repeating the analysis just described for injection times spanning one oscillation period ($0 \le \omega_0\tau < 2\pi$), an impulse sensitivity function (ISF) is mapped out which relates how susceptible an oscillator is at any given point to having the phase of its state bumped

[8] Expressing q_{max} separately in this relationship is not strictly necessary as the charge swing could just as easily be folded into the impulse sensitivity Γ. However, the appearance of q_{max} in the phase noise calculations derived using this LTV approach lends comfort in that it will help yield a form more closely resembling the historical LTI models.

off the steady-state trajectory. A typical ISF is shown in Figure 3-4 along with the corresponding oscillator output signal; it precisely quantifies the phase change that is observed in the oscillation for a fractional charge disturbance q_i/q_{max}. The ISF is also periodic in time, making available a Fourier series representation where each term in the series is multiplied with the input current $i(\tau)$ inside the superposition integral of Equation 3.2. Viewed in this light, the nth term in the Fourier series—$c_n \cos(n\omega_0\tau + \theta_n)$—multiplies noise around frequency $n\omega_0$ in the spectrum of $i(\tau)$ by the coefficient c_n, and then down-converts the integrated noise at this harmonic to the baseband spectrum of $\phi(t)$ via the cosine function.

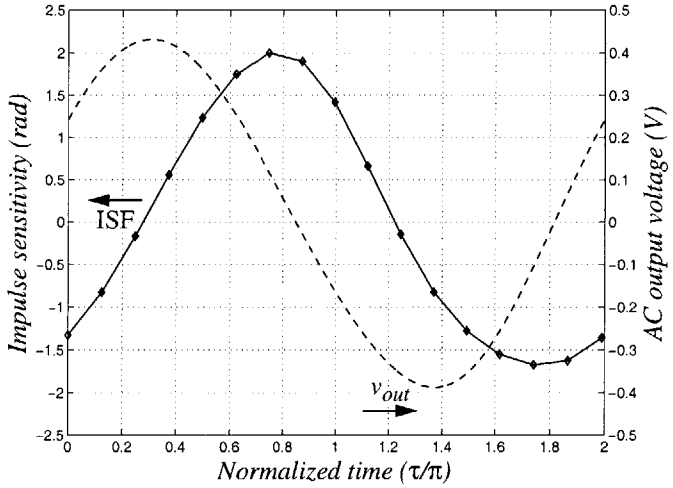

Figure 3-4. Typical oscillator impulse sensitivity function.

Phase noise refers to the sidebands accompanying a voltage signal as would be measured on a spectrum analyzer and not the phase. However, the spectrum of the oscillator output will follow that of $\phi(t)$ except at frequencies very close to the carrier—small offset frequencies for which the narrowband PM approximation no longer holds.[9] Thus, the single-sideband output phase noise spectrum of the oscillator (in the f^{-2} region) for a white noise source having a uniform spectral density $\overline{i_{nd}^2} \equiv \overline{i_n^2}/\Delta f$ is shown to be [34]:

$$\mathcal{L}\{\Delta f\} = 10\log\left[\left(\frac{\Gamma_{RMS}}{q_{max}}\right)^2 \frac{\overline{i_{nd}^2}}{8\pi^2 \Delta f^2}\right]. \tag{3.5}$$

The RMS value of the ISF (Γ_{RMS}) arises from the summation of the Fourier series coefficients acting upon a constant-valued noise spectral density, and q_{max} is indicative of the

[9] A more analytical approach to demonstrating the congruence between the spectrum of the phase disturbance and that of the oscillator voltage is offered by Hajimiri in [37].

output signal swing (i.e. the carrier power). Over some range of frequencies close to the carrier, the phase noise spectrum exhibits a sharper f^{-3} characteristic as a result of flicker noise; the phase noise spectral density in this region can be calculated by [34]:

$$\mathcal{L}\{\Delta f\} = 10\log\left[\left(\frac{\Gamma_{DC}}{q_{max}}\right)^2 \frac{\overline{i_{nd}^2}}{8\pi^2\Delta f^2}\frac{f_{1/f}}{\Delta f}\right], \tag{3.6}$$

where $f_{1/f}$ used to denote the transistor flicker noise corner frequency.[10] Equating these two expressions and solving for the offset frequency Δf, the transition between the f^{-3} and f^{-2} regions in the phase noise spectrum is witnessed to occur at:

$$\Delta f = \left(\frac{\Gamma_{DC}}{\Gamma_{RMS}}\right)^2 f_{1/f}. \tag{3.7}$$

With equations 3.5 and 3.6 in hand along with a noise source representation, the phase noise of an oscillator can be precisely determined using only the standard time-domain transient analysis capability of a circuit simulator to calculate impulse sensitivity functions. As the accuracy afforded by any analysis is ultimately constrained by modeling approximations, this technique allows full device models to be incorporated, implicitly assimilating many second-order considerations that affect oscillator noise. The price paid for this acuteness is increased computational complexity. As with any theory, however, the real utility is in learning how to improve designs by applying results observed through the theory, and not just in using it for numerical iteration; the remainder of this chapter investigates some of the lessons learned from studying an LTV model.

3.3 Optimizing Devices and Technology for Oscillators

Within the context of oscillators, veritable tomes can be found covering integrated passive circuit elements.[11] For a chosen circuit application, be it to minimize loss or to realize a desired component value, the passive device properties to optimize are generally clear and techniques have evolved to guide the designs. But the challenge is heightened significantly when it comes to the transistors in these circuits. First, there are no singular measures such as inductance or quality factor that describe a transistor, and hence it is not evident what should be optimized. Even if there was some defining metric to use, it is then also not clear what might be done to optimize it. Variations in the layout geometry can be

[10] Other components can generate flicker noise as well, but transistors are usually the dominant source in integrated circuits.

[11] See, for example, Chapter 2 for spiral inductors, or the many other references discussing inductors, capacitors, and varactors that are highlighted in Appendix B.

used to trade off device parasitics, and process parameters in next generation technologies could be adjusted to improve oscillator performance. Thus far, however, widespread acceptance of a method for quantifying the impact of a semiconductor process technology in RF circuits has not yet materialized.

The preceding section described an LTV model for oscillators which, owing to its linearity and generality, can be harnessed as a tool for exploring issues in transistor optimization and design. Each noise source in an oscillator, and within the transistors themselves, can be analyzed independently in this fashion and their resulting contributions to phase noise summed. Whereas the time-variant models discussed in the literature [38]-[41] have previously only considered transistors in terms of a single equivalent noise source, it is important to recognize the limitations involved in making this approximation. The equivalent source model assigns the same time-variant behavior to each source of thermal and shot noise in the device being represented; this tact may work well at low frequencies or when there is a single dominant noise source, but the accuracy begins to suffer at higher design frequencies as the device parasitic capacitances store an increasingly larger fraction of the energy in the resonant tank.[12]

For illustrating the application of the LTV model through a representative case study, a typical LC oscillator configuration[13] is presented below in Figure 3-5. V_{BASE} is an RF ground, allowing the common-base transistor to provide the requisite gain from the

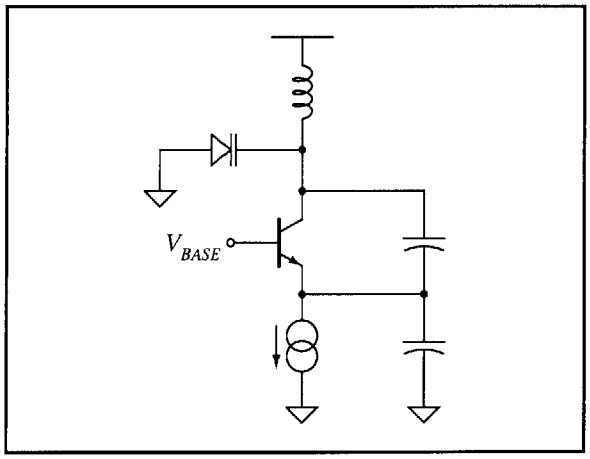

Figure 3-5. Schematic of single-ended oscillator used in noise analysis.

[12] The transistor parasitic capacitances—as determined by the size and layout of the devices—are relatively insensitive to the oscillation frequency, but the explicit capacitors used in the resonator have to shrink in order to realize higher frequencies.

[13] This particular configuration is known as a Colpitts oscillator, and is commonly found in both discrete and integrated manifestations.

emitter to the collector node where the resonator is connected. The capacitors indicated in the schematic couple a fraction of the voltage signal in the resonator back to the input of the gain stage as a positive feedback mechanism, and also combine with the inductor and varactor in determining the frequency of oscillation. Although the average bias current is set by the current source, the base-emitter voltage follows the oscillation, and thus the transistor actually remains off for much of the period. To be able to analyze each source of noise independently, the terminal parasitics are removed from the model for the common-base transistor and are reconstructed as depicted in Figure 3-6, where the nodes of the intrinsic device are marked B′, C′, and E′. The terminal resistances are separated apart from the transistor, and the base-collector capacitance is split across the base resistance in the usual fashion. Given this implementation, the impulse sensitivity function associated with each noise source can be calculated by placing the current source injecting the impulse functions (as discussed in Section 3.2) across the component that generates the noise being considered.

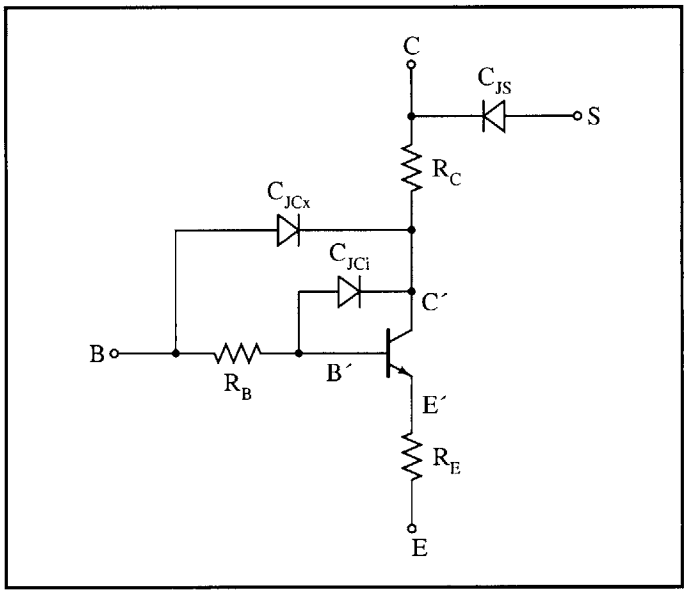

Figure 3-6. Bipolar transistor with terminal parasitic elements.

Choosing the inductance (with a typical integrated Q~6) and capacitors to resonate at 5.8GHz,[14] and biasing the oscillator with 2.8mA from a 3V supply, the ISFs (denoted by Γ) determined using the LTV approach for each of the resistances are plotted in Figure 3-7. A critical observation is that the sensitivities are non-zero during the positive

[14] The component values are chosen from an oscillator design presented in Chapter 6; the inductor is a spiral realizing 0.6nH, while the upper and lower feedback capacitors are 0.75pF and 2pF, respectively.

half-cycle of the oscillator (for $\tau \lesssim 0.8\pi$), during which time the transistor is switched off. *Thermal noise from resistances can still reach the resonator and affect the oscillator phase, even in the absence of shot noise.* Hence it is seen that attributing the same time-variant sensitivity to each source of noise within a transistor can be problematic; the importance of treating each noise component individually—with its own ISF—is apparent in this 5.8GHz example, and will only become more important as design frequencies for RF applications increase.

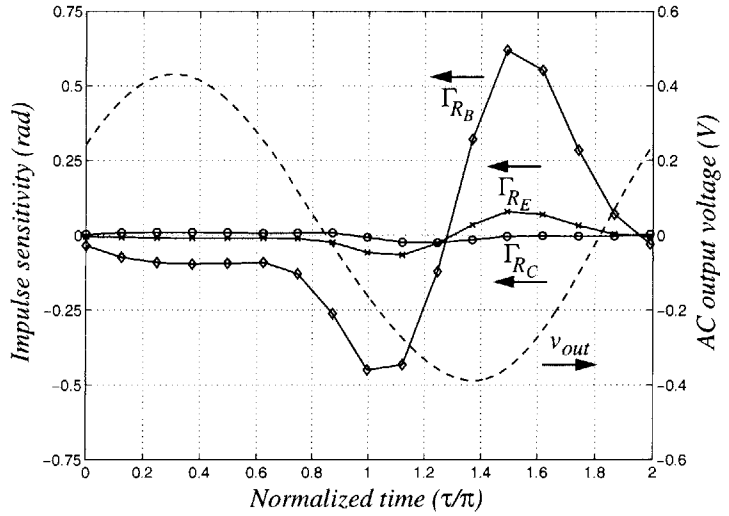

Figure 3-7. ISFs associated with transistor terminal resistances.

Another observation is that the ISFs associated with the base and emitter resistances are similar in shape, as might be expected since both touch the base-emitter junction and thus should have a roughly comparable effect on the oscillator.[15] In fact, the ISFs for these resistances are essentially scaled versions of each other, where the scale factor is the ratio of the resistances: $\Gamma_{R_B}/\Gamma_{R_E} \cong R_B/R_E$. Though it represents the behavior for a hypothetical unit current impulse, the ISF does effectively encode information about the source of the noise (i.e. $\Gamma \propto R$). This should not be too surprising, however, in that the same consequence is witnessed in LTI noise calculations where the resistance not only determines the noise power but also plays a role in any transfer function involving the noise source. Revisiting Equation 3.5, the expression for phase noise presented earlier demonstrated a dependence upon the product $\Gamma_{RMS}^2 \overline{i_{nd}^2}$; while the current noise term is inversely proportional to resistance ($\overline{i_{nd}^2} \propto 1/R$), the sensitivity to this noise increases

[15] For low-noise amplifiers (an LTI example), the base and emitter resistances do have the same effect, with the sum of the two (divided by the source resistance) contributing to the noise figure.

with resistance ($\Gamma_{RMS}^2 \propto R^2$). Lower phase noise would thus be expected as transistor terminal resistances decrease, a rather satisfying conclusion relayed through the ISFs.[16]

The base resistance in this example has been modeled as a constant, although this is an approximation and should not necessarily be assumed. Base sheet resistivity is a function of the integrated majority carrier concentration in the base region [43]; when high injection levels are present in the transistor, the intrinsic base resistance can be appreciably lowered as collector current increases. But for typical bias conditions found in oscillators and with the heavily-doped bases enabled by heterojunction and graded-base SiGe transistors, deviations from the DC bias value of R_B are commonly less than 10% over the entire oscillation swing. Neglecting the bias-dependency of the base resistance may not be possible with all oscillators and technologies, but the LTV framework can also accommodate parameters that vary in tune with the oscillation. The next section picks up the discussion of this topic and continues the example begun here.

3.4 Cyclostationarity in Oscillators

Through plausibility arguments and direct computation, the sensitivity of an oscillator to a given source of noise has been demonstrated to be time-variant. Adding to this complexity is that the power of a noise source itself can vary with time. When the properties of a random variable follow a periodic pattern, the source represented by that random variable is described as being cyclostationary. The base and collector shot noise effects in bipolar transistors, for example, are modulated by the instantaneous value of the currents flowing as a result of the oscillation and are therefore cyclostationary.[17] While this property has been observed in the literature (e.g., [38],[40]-[42]), it will be shown that calculating the time-varying power spectral density of the noise from the terminal currents is an approximation that can lead to significant errors in microwave frequency designs.

However, since the power exhibited by a shot noise source is proportional to the current, determining the base and collector currents as a function of time would seem a good first step. For the oscillator example introduced in the previous section, the currents are calculated with a transient analysis and plotted in Figure 3-8(a,b) versus time normal-

[16] Caution should be exercised in carrying this result too far: the relationship between a resistance and the oscillator phase noise is not necessarily a linear one. The amplitude of the oscillation (q_{max}) might be affected, and ISFs for other noise sources can also change. More about this intricacy will be said in Section 3.4.

[17] Similarly, in MOSFETs, the thermally-induced gate and drain current noise sources are modulated by the oscillation.

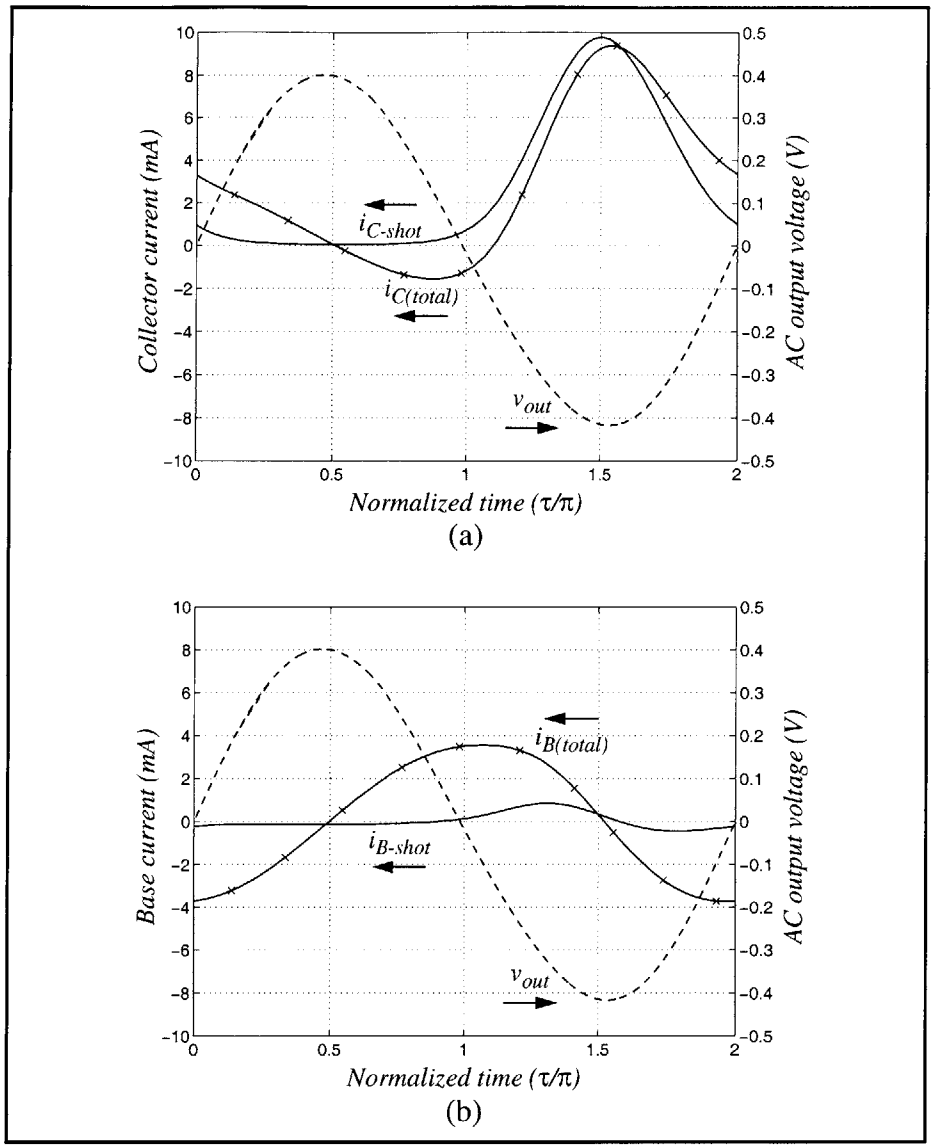

Figure 3-8. Transistor collector (a) and base (b) currents calculated for the single-ended oscillator example; the terminal currents are marked with x's and the components that generate shot noise are shown with the solid traces.

ized to the oscillation period. The currents flowing into the collector (i_C) and base (i_B) terminals are marked with x's, and are seen to be relatively large—on the order of milliamps—even when the transistor is expected to be off (for $0 \lesssim \tau \lesssim \pi$). A clue as to the nature of these currents is observed in the behavior of i_B during the same time interval; if the output voltage for this duration is represented as $v_{out}(\tau) = V_0 \sin(\tau)$, then the base current appears to be of the form $i_B(\tau) = i_b(\tau) = -I_{B0}\cos(\tau)$, and looks suspiciously as if a capacitance might be involved. This is exactly the case. Noting that the base is tied to an AC ground, the potentials across the transistor junctions are related to the output signal:

$v_{bc} = -v_{out}$ and $v_{be} = -v_{out}/n$, where n is set by the capacitive feedback ratio and the indication of time has been dropped for simplicity. Considering the depletion capacitances of an off-state transistor:

$$i_b = C_{jE}\frac{dv_{be}}{d\tau} + C_{jC}\frac{dv_{bc}}{d\tau} = -\left(\frac{C_{jE}}{n} + C_{jC}\right)\frac{dv_{out}}{d\tau}, \tag{3.8}$$

lending credence to the form guessed earlier.[18] From this argument, the realization is that the *shot noise is not determined by the total transistor terminal currents flowing into the base (i_B) and collector (i_C); significant portions of the terminal currents are associated with charging and discharging the base-collector, base-emitter, and collector-substrate depletion regions, capacitive components which do not generate shot noise.* Subtracting the displacement terms from the total terminal currents yields the components of i_B and i_C that generate shot noise and are indicated by the solid traces in Figure 3-8(a,b). It is notable that transistor displacement current components are clearly evident in this 5.8GHz example, and will become increasingly significant in designs at higher frequencies due to the larger *dV/dt*'s and relative junction capacitances that are encountered.

One method for calculating the displacement currents that magically disappeared via subtraction in the plots above would be to solve relations such as Equation 3.8, including the variations in the junction capacitances over forward and reverse bias. However, circuit simulators are highly optimized for solving differential equations like this, and already contain the models. A simulator can be used to solve the problem directly with the schematic shown in Figure 3-9. The oscillator, built around Q0, is the same as before, with the additional base bias circuitry being represented here by a voltage source and an output resistance. Transistors Q1 and Q2 are identical to Q0, with the exception that the forward junction currents are made arbitrarily small by dividing the saturation current model parameters (IS, ISC, and ISE) by a very large number (e.g., 10^{10}).[19] Zero-valued current sources implement open circuits, allowing the displacement currents for the base-collector and collector-substrate depletion regions to be measured independently in Q1 and for the base-emitter junction capacitance via Q2.[20] The junctions of Q1 and Q2

[18] The junction capacitances are also a function of v_{out} but will not undergo too dramatic a change over the range of voltages being subjected to them in this example, and so this dependency has been dropped for the purposes of illustration.

[19] The goal is to zero out the forward-bias currents which would primarily flow in the base-emitter junction of Q2. However, most SPICE-like simulators encounter convergence difficulties when the saturation current parameters are set to zero. Reducing each parameter by the same factor is probably gratuitous in this case, but serves to keep the V_{CE} offset (for $I_C = 0$) unchanged.

[20] The fourth terminal of Q2 indicates the substrate node. When not explicitly shown, the p-type substrate in which the transistors are formed is tied to the lowest circuit potential (ground).

are then exercised by a set of ideal unity-gain buffers (voltage-controlled voltage sources) which copy the terminal potentials from the oscillator, completing the schematic solution.

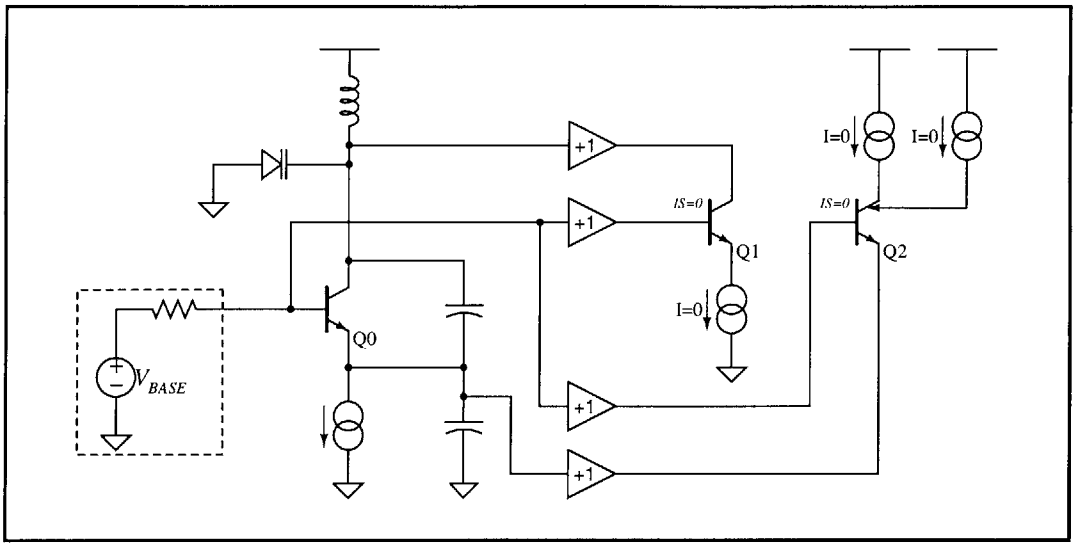

Figure 3-9. Schematic used to solve for the transistor displacement currents in the single-ended oscillator example.

Returning from the brief explanatory interlude, the transistor currents contributing shot noise to the oscillator can be expressed using a constant current value and a multiplicative scale factor $\alpha_i(\tau)$ that varies periodically with the oscillation:

$$i_{C\text{-}shot}(\tau) = \alpha_{i_C}(\tau) i_{C\text{-}shot,\,max}, \tag{3.9a}$$

$$i_{B\text{-}shot}(\tau) = \alpha_{i_B}(\tau) i_{B\text{-}shot,\,max}. \tag{3.9b}$$

Somewhat arbitrarily, the scale factors are chosen such that $0 \le \alpha_i(\tau) \le 1$, and then the maximum instantaneous shot noise current values can be used to establish the noise spectral density (e.g., $\overline{i_{nd}^2} = 2qi_{C\text{-}shot,\,max}$). The time-domain response of a cyclostationary shot noise process can thus be represented as:

$$i_n(\tau) = \alpha_i(\omega_0\tau) i_{max}(\tau), \tag{3.10}$$

where the periodic nature is again made explicit in the scale factor and $i_{max}(\tau)$ is the noise current associated with $i_{C\text{-}shot,\,max}$ or $i_{B\text{-}shot,\,max}$ as appropriate. Given $i_n(\tau)$ as an input, the oscillator phase can be calculated by superposition; inserting the previous expression into Equation 3.2 gives:

$$\phi(t) = \frac{1}{q_{max}}\int_{-\infty}^{t}\Gamma(\omega_0\tau)i_n(\tau)d\tau = \frac{1}{q_{max}}\int_{-\infty}^{t}\Gamma(\omega_0\tau)\alpha_i(\omega_0\tau)i_{max}(\tau)d\tau, \tag{3.11}$$

from which it becomes evident that cyclostationary noise behavior can be handled as a white noise source of constant power when an effective ISF is defined by:

$$\Gamma_{eff}(\omega_0\tau) = \alpha_i(\omega_0\tau)\Gamma(\omega_0\tau). \tag{3.12}$$

Applying this definition in the ongoing oscillator example, the effective ISFs are illustrated in Figure 3-10. The sensitivities to impulses of current applied from the collector to emitter and the base to emitter are shown with dashed lines. Weighting these ISFs by the shot noise components (α_i) from Figure 3-8, the RMS and DC values of the resulting Γ_{eff} curves can then be used in the phase noise calculations described in Section 3.2.

Having in hand the ISFs for each of the terminal resistances (Figure 3-7) and the effective ISFs for the base and collector shot noise currents, contributions to the phase spectrum can be determined for each source of noise originating in the oscillator gain transistor. But the analysis is not yet complete; noise from the bias circuitry and resonator loss elements remains to be considered. Though not explicitly shown in the schematics, loss components in the varactor, inductor, and MIM capacitors produce thermal noise. The transfer functions for power in an LTI network hold regardless of whether the power comes from a signal or from noise, so any equivalent circuit representing loss in the resonator will also be valid for the consideration of noise.[21] Over at least a modest range of frequencies, transformations between series and parallel RL and RC sections can be used to represent any RLC network with a single resistance alongside a group of ideal inductors and capacitors [44]. For a parallel resonant structure as used in the oscillator example of this chapter, a shunt resistance (from the collector node to an RF ground) is generally the most convenient form. The value of the equivalent resistance (R_{eq}) can be evaluated analytically at resonance by applying the series/parallel transformations to each loss element in the resonator, or by $R_{eq} = 1/Re\{y_{11}\}$, where y_{11} is determined for the one-port network consisting of the inductor, varactor, and capacitors with the transistor and current source removed. The ISF for the resonator noise is calculated by sourcing the impulses of current between an RF ground and the collector node (the location of the equivalent resistance), and the result from this analysis was shown previously in Figure 3-4 under the guise of being a "typical example". The noise spectral density associated with the resona-

[21] Although oscillators and mixers are time-variant circuits, portions of them can still be considered as LTI. Identifying sections that are LTI is part of the art that is oscillator and mixer design.

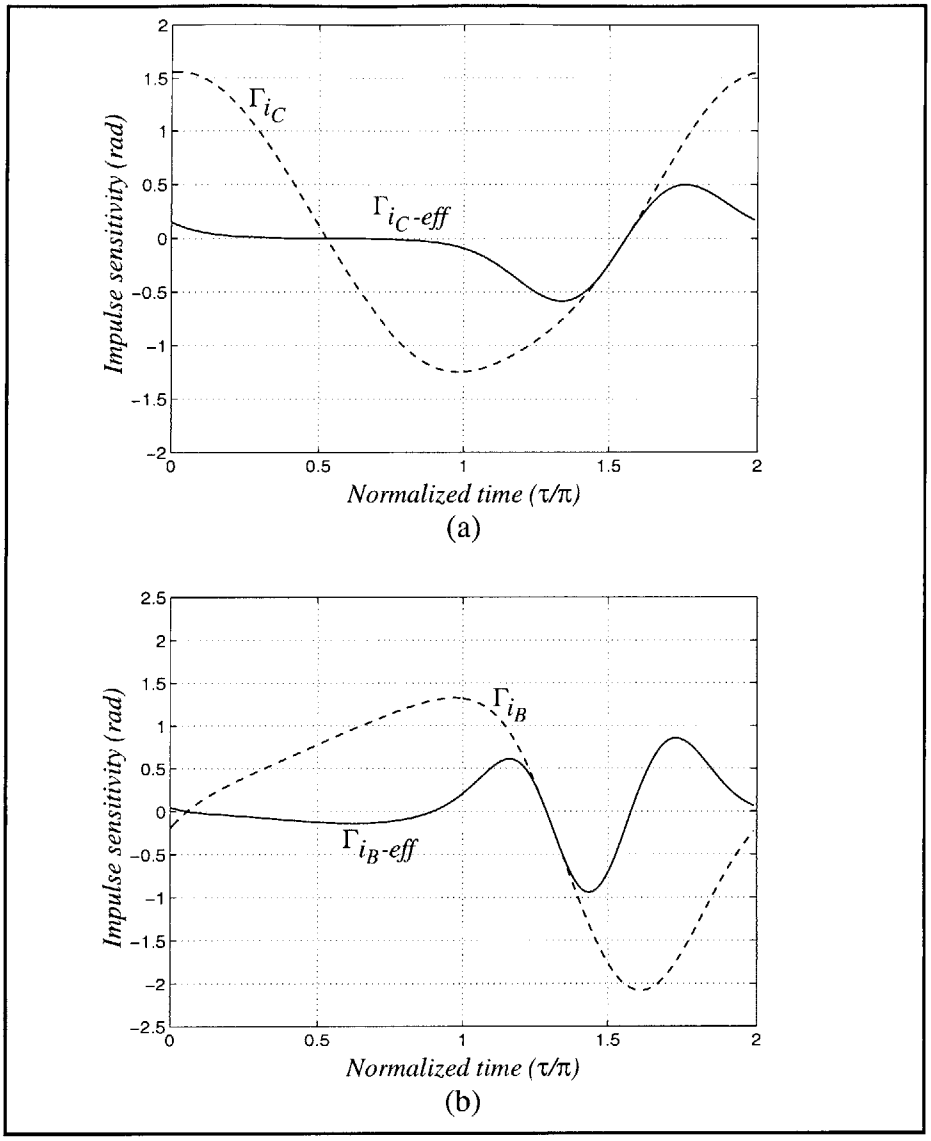

Figure 3-10. Calculation of the effective ISFs for the collector (a) and base (b) shot noise sources for the single-ended oscillator example.

tor loss is $\overline{i_{nd}^2} = 4kTR_{eq}$, allowing direct contributions to phase noise on the part of the resonator to be calculated in a manner consistent with the other oscillator noise sources.

Bias circuitry can also play a significant role in oscillator performance, although careful design should be able to minimize its impact on phase noise.[22] Noise in the tail current can clearly affect the phase, although the parameters of the current source itself do not change appreciably over the oscillation period, and thus a single output equivalent cur-

[22] Careful design along with added die area and power consumption might be a more accurate statement. Noise considerations are what prohibit making bias circuits with arbitrarily small devices and currents.

rent noise source can be derived (following the usual LTI methods) to represent the current source transistor and associated reference bias circuitry. The ISF waveform for this equivalent noise source in the single-ended oscillator example indicates a prevailing fundamental Fourier series component, so a reasonable approximation to use in the phase noise computation for the tail source is the current noise spectral density exhibited near the frequency of oscillation.[23] Similarly, the effect of noise on the base bias voltage can be analyzed with an output equivalent noise source and an ISF determined for the location of the base bias. But the impact of this source on the phase of the oscillator should readily be engineered to be smaller than the effect of the base resistance, and so it is neglected here.

By now, and in perhaps the most delayed punch line in modern technical literature, the phase noise of the oscillator example begun in the preceding section can finally be calculated. Using the impulse sensitivities and noise spectral densities that have been discussed and illustrated throughout this 5.8GHz case study, the contributions of each noise source to the f^{-2} region in the oscillator spectrum are plotted in Figure 3-11. The trace shown for each component represents the phase noise that would result if that component

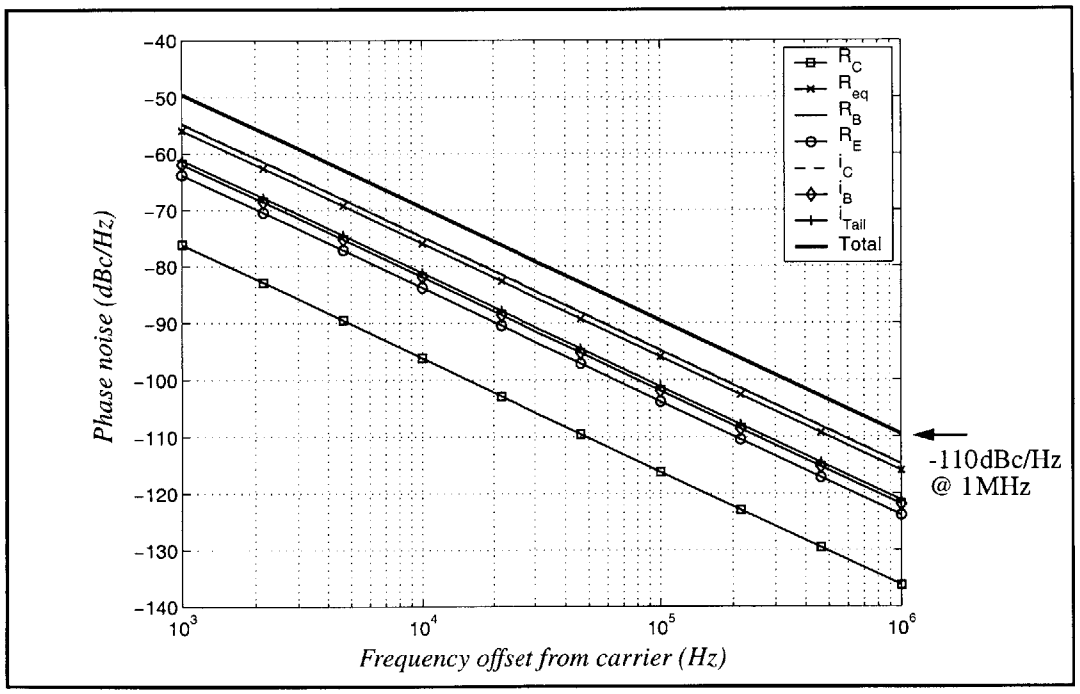

Figure 3-11. Calculation of phase noise for single-ended oscillator example.

[23] More precisely, the current noise spectral density at the output of the tail source should be determined at DC and near every harmonic of the oscillation, with the frequency-dependent noise densities being used in the Fourier series expansion of the superposition integral as described in Section 3.2. This approximation does, however, get the most significant term right, and will provide an indication of whether noise in the tail source remains a concern.

were the only source of noise in the oscillator. For this example, nearly equal contributions from the collector shot noise and the base resistance are observed—the collector shot noise is almost indistinguishably higher—with noise from the resonator loss (R_{eq}) mechanisms trailing not far behind. This balance between thermal and shot noise represents an optimal trade-off, and can be achieved through proper transistor sizing. Following behind at distances of about 5-7dB are the noise sources associated with the tail current, base current, and the emitter resistance. Finishing at the bottom, because of its small size and location at the output, the collector resistance is rendered inconsequential. Summing the noise powers for each of the seven contributors:

$$\mathcal{L}\{\Delta f\} = 10\log\left[\frac{1}{8\pi^2\Delta f^2 q_{max}^2}\sum_k \Gamma_{RMS_k}^2 \overline{i_{nd_k}^2}\right], \tag{3.13}$$

the oscillator phase noise spectral density is witnessed to be -109.6dBc/Hz at a frequency offset 1MHz from the carrier.[24]

To underscore the importance of properly handling the transistor displacement currents in an analysis of oscillator noise, the phase noise is "calculated" again, but this time (mistakenly) using the full base and collector currents to determine the effective ISFs associated with the shot noise sources. As shown in Figure 3-12, this approximation overestimates the contributions from i_B and i_C, leading to the erroneous conclusion that the oscillator is dominated by shot noise; the resultant phase noise spectrum is 3dB too high (-106.7dBc/Hz at a 1MHz offset)—an amount which could, for example, significantly impact receiver sensitivity were it to be real.[25] That the components of the terminal currents associated with junction capacitances do not contribute shot noise is demonstrated later in Section 6.3, where the phase noise of the example oscillator is measured to be notably *lower* than the calculation presented in Figure 3-12 (and equal to the calculation in Figure 3-11). Perhaps of even greater salience, however, is that improperly accounting for the displacement currents can lead to sub-optimal designs. Based on using the full terminal currents in the analysis, a designer may strive to either reduce the bias current in the

[24] A similar comparison could also be performed for the phase noise in the f^{-3} region, although the only two components carrying appreciable flicker noise energy will be the collector shot noise and the tail source. The transition to this frequency characteristic is expected to occur at very small carrier offsets for the SiGe bipolar technology used as the basis for this example, so flicker noise is not considered further here.

[25] One particularly treacherous example is reciprocal mixing in a receiver. Sideband energy in the LO downconverts adjacent channels and interferers to the desired IF. The in-band noise from the adjacent channels can translate into an SNR degradation that is dB for dB with the phase noise power (i.e. 3dB higher phase noise results in a 3dB decrease in the SNR).

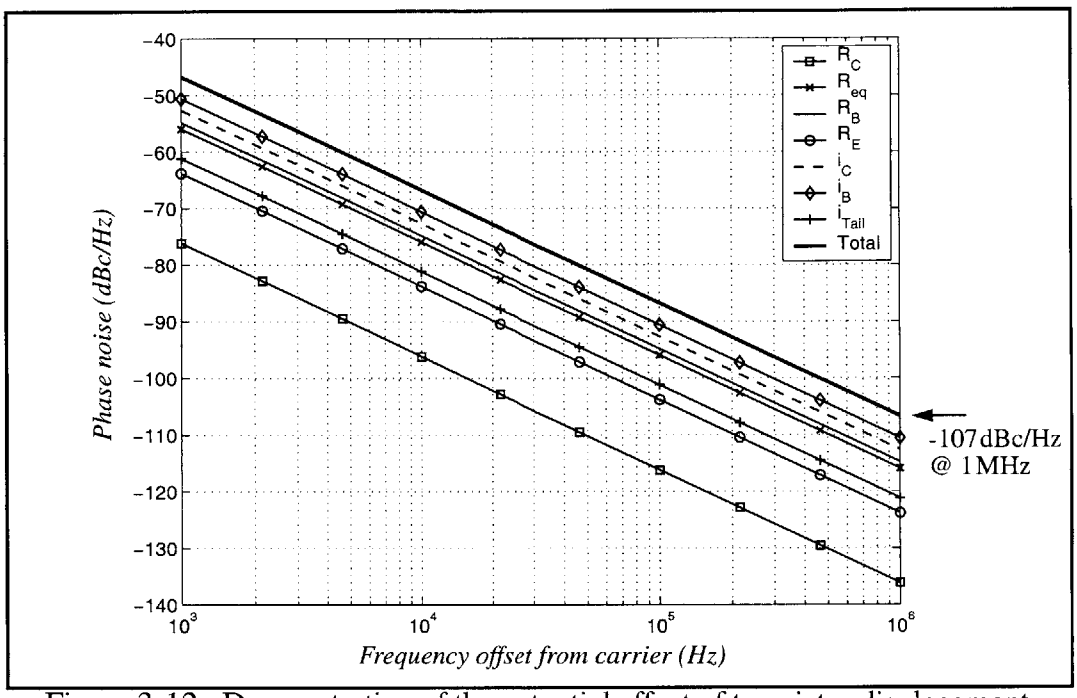

Figure 3-12. Demonstration of the potential effect of transistor displacement currents in the calculation of oscillator phase noise.

oscillator or to increase the transistor current gain[26]—unfortunate decisions which would actually be likely to worsen the phase noise performance in this 5.8GHz example. Being able to evaluate design judgments like these is where the merit of the LTV model really begins to shine.

Another interesting observation stemming from the discussion of cyclostationarity is that the effects of transistor shot noise can be reduced through the design of the oscillator circuit. Perusing again the plots of Figures 3-8 and 3-10, it is evident that shot noise is generated only during a fraction of the period, and that this interval can be timed to occur when the sensitivity to the noise is near a minimum. In this respect, the Colpitts construction performs admirably in that the transistor only turns on to restore energy to the resonant tank as the oscillation approaches its minimum value—the very moment when the ISFs associated with the transistor currents tend toward zero.[27] An important sidelight demonstrated in Figure 3-8 is that the duration for which the transistor conducts, and hence the degree to which the shot noise is unweighted, is primarily controlled by the

[26] The transistors in oscillators often operate at current densities well below that corresponding to the peak f_T, and thus the RF current gain—sometimes denoted $\beta(f)$ —can be impacted at the circuit design level by using smaller geometry devices. In the device/process design, there are any of a number of approaches available to improve f_T and/or the DC current gain.

[27] This technique also applies to the drain and gate noise currents in a MOSFET-based oscillator.

dynamics of the circuit and not the device. *Once the transistor f_T is high enough relative to the frequency of oscillation, further increases in f_T will not appreciably improve the oscillator in any regard.*[28] In fact, re-engineering the transistor stack for a higher f_T (e.g., thinning the base) might entail trade-offs that could actually hinder phase noise performance. These types of technology issues can be explored for RF circuit applications with the aid of LTV models, and will be pursued further in subsequent chapters.

At various points throughout this presentation of the time-variant approach to oscillators, the reader might reflexively ask a very innocuous-sounding question: "What happened to Q?" From Leeson's plausibility argument as well as general intuition, it would seem that the resonator quality factor should play a readily apparent role. But not only was there no mention of Q to be found in the entire development of the LTV model, Figure 3-11 indicates that noise originating from within the resonator ranked behind both collector current and base resistance in generating phase noise for the oscillator analyzed in this chapter. Indeed, at first glance, something would appear to be uncomfortably amiss.

This first impression would be wrong, however, as part of the story is told by q_{max} and the rest through the impulse sensitivity function. For a chosen oscillator circuit and bias point, the amplitude of the oscillation decreases with increasing resonator loss, and as Q decreases, so too does the charge displaced by the oscillation (q_{max}). The other incarnation of quality factor is hiding in the ISF, as it should be expected that lower Q resonators will not preserve the oscillator state as well as those with higher quality factors. Returning to Equations 3.3 and 3.4, for a given fractional charge disturbance injected by an impulse of current, the amount by which the phase is perturbed will be higher when lower Q resonant elements are used. Hence the RMS value of the ISF will increase as Q decreases. So, in effect, Q is alive and well—it just does not explicitly appear in the phase noise relationships. Furthermore, a valid argument may be made *against* an explicit reference to Q; noting that the loading on the resonator in the Colpitts oscillator example (Figure 3-5) changes as the transistor switches on and off, it becomes evident that the effective (loaded) Q is also cyclostationary. Quality factor alone cannot include this time-varying sensitivity, but an ISF directly calculated from applied impulses does capture the behavior, providing a more complete representation of oscillator dynamics than a mere mention of Q can offer.

[28] In some circuits, higher transistor f_T can be traded for lower power consumption. But in an oscillator, the optimal bias points are more acutely determined by the passive devices than by the active. This will be seen more clearly in Chapter 6.

The voyage into time-variant oscillator models begun in this chapter is now complete. A graphical plausibility argument provided some initial motivation, and then gave way to the formalism introduced by Hajimiri and Lee where an oscillator is represented as a linear time-variant system characterized by an impulse sensitivity function and having a phase output and a noise input. This approach was then harnessed as a means of exploring trade-offs in semiconductor technology, working at the device level and examining the impact that various transistor considerations can have on oscillator performance. In so doing, the importance of treating each transistor noise source with its own ISF was illustrated, as was the necessity of subtracting displacement components of the terminal currents for the purposes of calculating shot noise. By using a representative 5.8GHz oscillator as an example to illustrate many key ideas, a number of the technological directions being sought for RF circuits began to be unearthed. The work here does not end with oscillators, however, as many of the concepts that have been discussed can be extended to other time-varying circuits commonly found in RF applications. An overview of mixers is presented in the next section as a generalization of the LTV methodology.

3.5 Extending the Linear Time-Variant Model to Mixers

For oscillator circuits, a linear time-variant system view has been seen to be an effective approach for considering how the sensitivity of the phase changes in response to the large-signal oscillation conditions that are present. Like oscillators, mixers also exhibit a periodic variation with respect to sources of noise, although the periodicity in mixers is tied to an applied LO signal rather than the intrinsic circuit behavior. The technique of characterizing a circuit with a set of impulse responses directly computed using a simulator also applies, although in this case the phase is arbitrary; of greater concern in mixers is the magnitude of the noise competing with the signals being processed. Most crucially, however, is that linearity still holds[29] for noise and small inputs in the time-variant representation of mixers. As utilized in the analysis of oscillators, linearity allows noise sources at the device level to be considered independently, paving the way for a more informed repartee on technology-related optimizations, directions, and trade-offs.

Illustrating again through an example, a single-balanced mixer is fleshed out in slightly more detail in Figure 3-13. As if it were a low-noise amplifier (LNA), the input stage is designed for simultaneous noise and impedance match to the RF source at 5.8GHz, and biasing is shown through a bias tee for the purposes of this analysis.[30] The

[29] Useful as it may be, the tag line "linearity still holds" probably won't sell many books.

transistors comprising the switching pair have their terminal parasitics represented by the same external network used in the oscillator example (depicted in Figure 3-6), and the collector loads—which will generally be tuned to the desired IF band—have been replaced in favor of large-valued inductors that provide DC current paths but are RF opens at all frequencies of interest. Loss in (what would be) the tuned collector network is represented by the resistors R_{eq}, yielding the output impedance to which the IF load ($2R_{eq}$) is matched. An anti-aliasing low-pass filter inserted at the output completes the circuit and prepares the simulation data for a subsequent DFT computation. Implemented as a third-order Butterworth, the filter provides a maximally-flat response and a corner frequency chosen to be 15GHz. Provided that the corner frequency is sufficiently above the desired IF range and far enough below the simulation sampling frequency (4THz in this example) to minimize aliasing, the noise analysis should not be affected by this filter.[31]

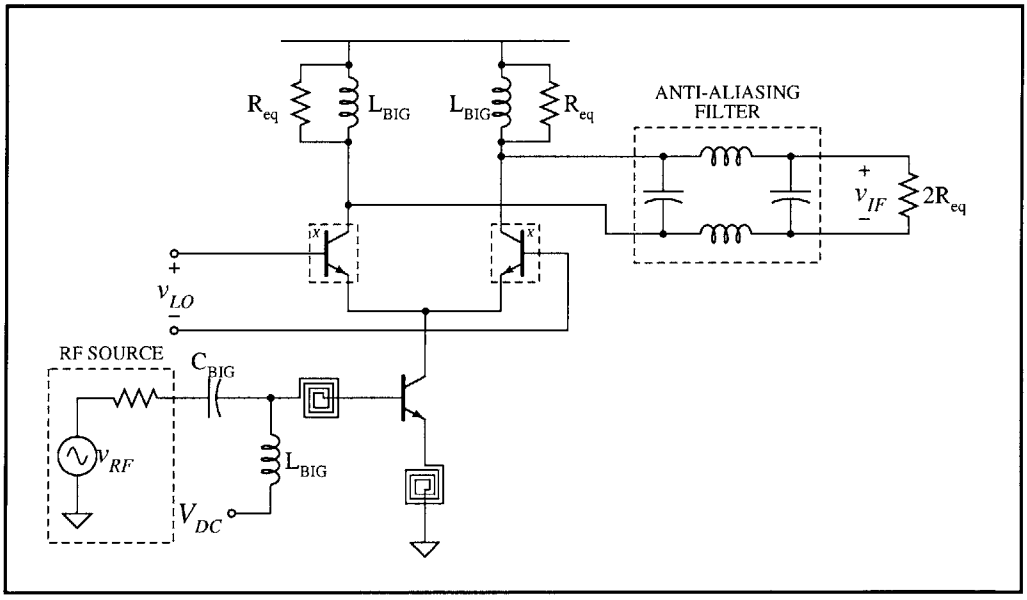

Figure 3-13. Single-balanced mixer schematic for noise simulation.

In a mixer, the steady-state operating condition is defined with an LO applied and all other signal sources set to zero. The balanced IF output in this mode appears as the solid trace in Figure 3-14(a); at some point after the circuit has reached its steady state, an impulse of current $i(t)$ is injected at the location of one of the mixer noise sources, pro-

[30] Conveniently, more will be said about the design of low-noise amplifiers in the next chapter.

[31] Although it may be tempting to set the corner frequency low enough to filter the LO component from the transient response, an accurate calculation of the conversion gain is required for frequencies around the LO and its harmonics. As always, it is best to place the filter corner out beyond the band of interest.

ducing a perturbation response $v(t)$ measured at the output (and indicated with the dashed line). Whereas considerations with oscillators limit the discussion to output phase and the associated dynamics insure a step function impulse response, there is no such luxury in the analysis of mixers. The full impulse response must be used, and can be calculated as the difference between the perturbation response and the steady state: $h(t) = v(t) - v_0(t)$, achieving the result plotted in Figure 3-14(b).[32] The basic shape of the impulse response is contributed by the anti-aliasing filter, although the amplitude, delay, and amount of dispersion all vary with the timing of the impulse within the LO period.

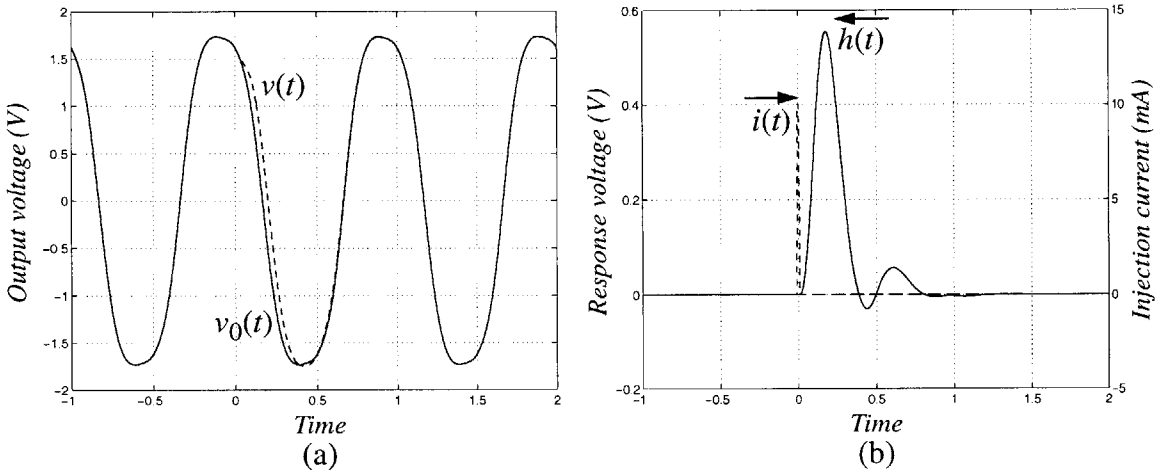

Figure 3-14. Calculation of the impulse response of a mixer; (a) impulse perturbation response compared with the steady-state condition, (b) applied impulse and calculated impulse response.

This operation of injecting the current pulse and measuring the impulse response is then repeated at different time offsets so that exactly one LO period is spanned by an integer number of samples (injections). A non-integer number of samples would create a periodic discontinuity in the time-domain and introduce errors in the spectrum. This effect is generally mitigated in spectral estimation problems using discrete Fourier transforms by windowing the time-domain data so that the values at both ends of the window fall to zero. But windowing would necessitate the sampling of multiple periods, a step which can easily be avoided by choosing the appropriate times to inject the current impulses. Furthermore, being that there is no retention of state[33] in a mixer and that the period is precisely

[32] The actual unit impulse response is $h(t)$ scaled by the weight of the impulse $i(t)$. However, this step will be picked up in the frequency domain when the ratio of $H[k]$ to $I[k]$ is taken to yield the conversion gain.

[33] Unlike an oscillator, there is no circuit state being stored in a mixer. As such, each impulse response will both start and end at zero.

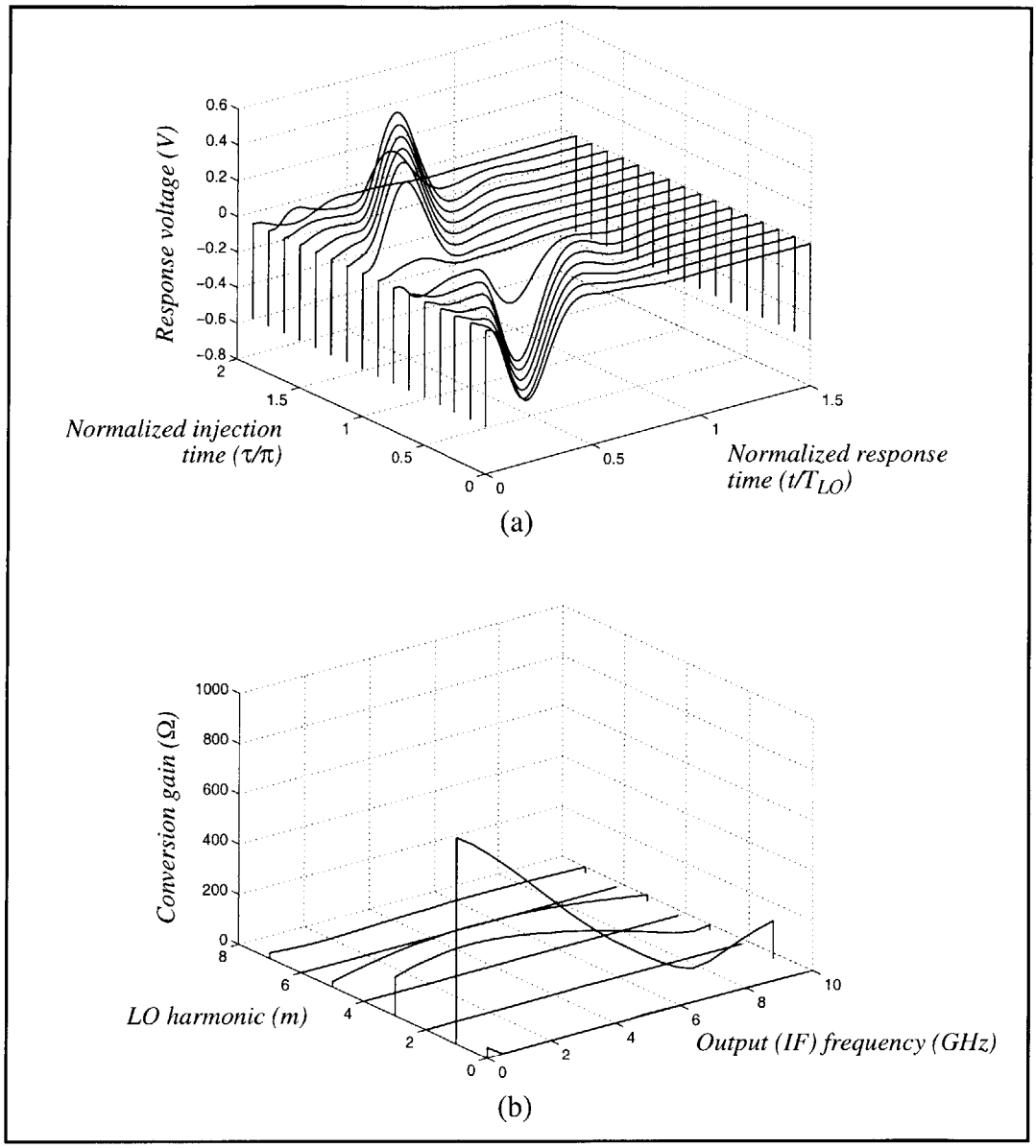

Figure 3-15. Mixer response to noise injected at the input of the switching pair; (a) response to impulses injected at moments throughout one LO period, (b) 2-D DFT of impulse response data giving conversion gain for noise about each LO harmonic as a function of the chosen IF.

determined by the applied LO, all of the injection times and impulse responses can be captured with a single transient analysis—provided that enough time is given between applied impulses for the response to settle. The injection current source can be specified with a SPICE pulse statement; an example for a 5GHz LO is shown in Figure 3-16, and gives the circuit an initial 5ns to reach steady-state. Each pulse has a duration of 4ps at an amplitude of 10mA, and uses 3ps for the rise and fall times. Ten periods are allotted for settling

```
Iinj <+node> <-node> pulse(0 10m 5n 3p 3p 4p 2012.5p),
```

Initial value *Final value* *Delay* *Rise time* *Fall time* *Pulse width* *Period*

Figure 3-16. SPICE statement for a periodic current impulse.

between injection events, and the impulse is advanced 1/16th of an LO period (12.5ps) with each occurrence. Placing this source at the output of the transconductor (the equivalent output noise current generator for the input stage), the 16 impulse responses subsequently calculated and illustrated in Figure 3-15(a)—arrayed along the left planar axis—cover exactly one period, or $\tau = 2\pi(n/N)$ with $N = 16$ and $n \in [0, N-1]$. While τ denotes the position within the LO period where the impulse is injected, the units of time (t) on the right horizontal (x) axis delineate the impulse response, with $t = 0$ for each response set at the center of the applied impulse that generated it. A continuous variation in the transmission gain from the noise source to the output is evident along the LO axis in this example,[34] reinforcing the need for a time-variant model.

Now that the circuit response has been characterized and aligned into its two independent time dimensions, a 2-D DFT can be used to determine the frequency response presented by the mixer to the transconductor equivalent output source. The result of the transform, where each row in the DFT, $H[k,m]$, has been normalized to the scaled input signal $NI[k]$, appears in Figure 3-15(b). Having two measures of time as an input, the magnitude of the DFT will be gauged along two measures of frequency. For the y-axis, in absolute (rather than normalized) units, the DFT samples occur at:

$$f_y[m] = \frac{m}{N} f_{s_y} = \frac{m}{N} N f_{LO} = m f_{LO}, \tag{3.14}$$

where N is the number of points along the y-dimension of the DFT, and f_{s_y} is the sampling frequency in this direction. With N points sampled on precisely one LO period, the y-axis of the DFT is observed to give information about the harmonics of the applied LO signal. The x-axis is the more conventional time-frequency transform, and represents the frequency at the output (IF) node of the mixer. Thus, the plot in Figure 3-15(b) describes the conversion gain to any given IF frequency for signals and noise appearing that distance away from any harmonic of the LO. Put another way, if an IF is chosen, then reading up the y-axis at this value of f_{IF} gives the gain for each frequency that converts to that IF.

[34] A sinusoidal LO drive has been used in this example. This tends to be the case in practice, as generating square waves becomes difficult at frequencies well into the gigahertz range.

Given the 5GHz oscillator drive used in this example, the magnitude of the DFT at the point located by $m = 1$ and an IF of 800MHz yields the conversion gain for inputs at frequencies of 4.2GHz and 5.8GHz. This value, and the rest of the gain terms associated with an 800MHz IF, are plotted in Figure 3-17(a); the even-order rejection demonstrated in the data is inherent to the balanced output of the mixer.

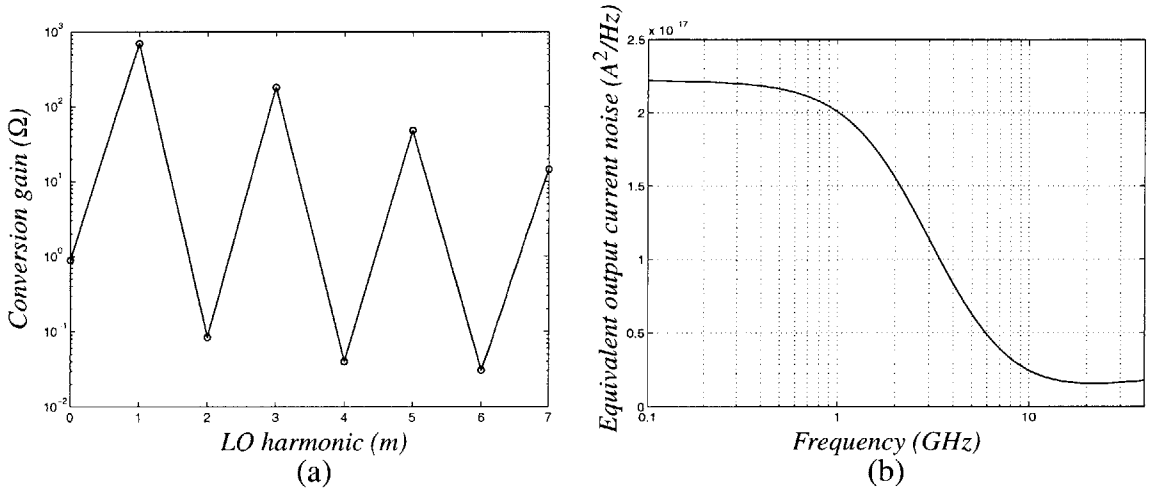

Figure 3-17. Calculation of the noise in a mixer resulting from the input transconductor; (a) conversion gains from input of switching pair to an IF output at 800MHz, (b) equivalent output current noise from RF transconductor stage.

Now that the conversion gains have been determined, the noise at the mixer output can be calculated once the spectrum of the noise source has been ascertained. As the input transconductor stage is not appreciably impacted by the switching core, a conventional LTI small-signal analysis suffices to produce the equivalent output current noise spectrum shown in Figure 3-17(b). The roll-off in the transconductor output noise at 1GHz is owed to the inductive emitter degeneration used for matching the RF input. Taking the values of the current noise spectral density ($\overline{i_{nd}^2}$) at each frequency that downconverts to the chosen IF, applying the corresponding gain factors, and summing the contributions gives:

$$\overline{v_{ond}^2} = H_0^2 \overline{i_{nd}^2}\Big|_{f_{IF}} + \sum_{m=1}^{\infty} H_m^2 \left[\overline{i_{nd}^2}\Big|_{mf_{LO}-f_{IF}} + \overline{i_{nd}^2}\Big|_{mf_{LO}+f_{IF}} \right]. \tag{3.15}$$

Although the summation of the mixer output noise in Equation 3.15 is over all LO harmonics, the data in Figure 3-17(a) indicate that the conversion gain for $m = 9$ would be approximately a factor of 100 smaller than the largest contributor ($m = 1$), and so higher order terms can be reasonably neglected. The equivalent input noise can be obtained by

finding the conversion gain of the mixer at the desired RF input frequency. One factor in the signal conversion gain resides in the switching core and is represented by H_1; the other component is the transconductance of the input stage at the RF frequency ($G_{m\text{-}RF}$). Dividing the mixer output noise spectral density by the product of the gain terms yields:

$$\overline{v_{ind}^2} = \frac{\overline{v_{ond}^2}}{H_1^2 G_{m\text{-}RF}^2}. \tag{3.16}$$

The input-referred noise can be directly compared with the noise of the source itself, commonly expressed in dB as the noise figure:[35]

$$NF = 10\log\left(\frac{\overline{v_{ind}^2}}{4kTR_s}\right), \tag{3.17}$$

where R_s represents the source resistance of the RF input. This procedure can then be repeated independently for each source of noise in the mixer, and the sum of each of the input-referred noise powers can be used in Equation 3.17 to calculate the overall noise figure. Continuing the example of this section, the output noise spectral density resulting from noise in the transconductor stage, as computed from the data in Figure 3-17, is:

$$\overline{v_{ond}^2} = 4.07 \times 10^{-15}\ \mathrm{V^2/Hz}. \tag{3.18}$$

The switching core conversion gain is seen to be $H_1 = 879.9\Omega$ from the DFT,[36] and a small-signal AC analysis performed on the low-noise transconductance input stage gives $G_{m\text{-}RF} = 36.9\mathrm{mS}$, translating into a noise figure of 6.9dB. This would be the noise figure of the mixer were the switches ideal. Other non-ideal effects in the transistors can then be included and evaluated with regards to the impact they have on this "baseline" number.

For more detail into the treatment of mixers via LTV methods, the reader is referred to a paper by Hull and Meyer [35]. The key idea, however, is that in separately handling the sources of shot and thermal noise within the circuit, time-variant models allow issues of technological limitations to be more fully illuminated within the context of RF/microwave applications. To what extent do the transistor terminal resistances impact

[35] The noise from the source resistance has been included in the calculation of the equivalent output noise current of the input stage, which is why the noise factor argument in Equation 3.17 does not appear in the more commonly found "1 + ..." form.

[36] As might be guessed from looking at the conversion gain data, the equivalent output resistance (R_{eq}) used in these simulations is $1\mathrm{k}\Omega$; most of the rest of the input signal energy (i.e. the reason that the transresistance $H_1 \neq 1\mathrm{k}\Omega$) goes into the higher harmonics.

mixer noise figures? Would having a higher f_T device help? Or is performance more a function of the junction capacitances? If the limitations are stemming from parasitic resistances and capacitances, are bipolar transistors providing any benefit over what could be achieved in a potentially simpler CMOS technology? Questions like these are important to ask at all levels of design to improve the performance of RF front-end circuitry. A thorough circuit design should consider whether it is better to size the current steering transistors for peak f_T operation or if larger devices could reduce the noise figure by cutting the thermal noise contributed by the switching core. Should multiple base contacts be used to cut the base resistance, or does the concomitant increase in base-collector and collector-substrate capacitances hurt mixer performance? A similar analysis can be studied at a system design level to aid in choosing which process to use for a desired circuit application. Does having access to, say, 0.18μm SiGe bipolar transistors improve the performance or power consumption characteristics of transceiver circuits relative to a cheaper 0.6μm homojunction technology?

Moving one design level higher, the techniques discussed in this chapter, applied to a class of evolving wireless circuit applications, can finally begin to provide guidance for process development. Would doping the collector more heavily to support higher peak f_T values make for better RF performance? Or would penalties associated with higher junction capacitances and lower output resistances outstrip the benefits? Is it more important to address problems inherent to processing thinner bases or to focus on issues pertaining to lateral dimensions? At an even higher level, questions such as whether BiCMOS should continued to be pursued for tightly integrated RF systems can be addressed. Linear time-variant circuit modeling fills an important role in carrying a systematic analysis through all levels of a design, a process by which a better determination can be made in the directions for advancing future generations of semiconductor technologies, and one which will be explored through some real designs later in this thesis.

Chapter 4

Low-Noise Amplifiers

As a fixture in RF and microwave receivers, the low-noise amplifier (LNA): provides an impedance match to the antenna to enhance signal reception, amplifies the incident RF signal before it encounters higher noise receiver circuitry, and helps isolate the antenna from local oscillator tones and unwanted mixing products generated in the receiver chain. Being commonly-employed and fulfilling requirements in wireless systems that are difficult to circumvent with architectural innovations, LNAs make for an interesting study from a technology perspective. The realization of an LNA pushes fundamental limits in device speed, noise, and power consumption in ways that can be addressed in only a very limited fashion by circuit techniques, and thus the design of these amplifiers yields an interesting point of comparison between semiconductor processes as well as valuable insight into the continued development of RF technologies.

The specific goals for an LNA will always be application dependent, but the design focus in future wireless systems will shift toward extracting the maximum capability from the RF link in terms of bandwidth, dynamic range, and power consumption. A key concept in next-generation RF links will be *adaptive* performance; in the context of an LNA, the question is whether power can be traded off for sensitivity and gain in a manner that does not unduly compromise system operation. Performance will be measured in terms of the best possible noise figure that can be achieved while reasonable gain, linearity, isolation, power dissipation, and impedance matching are all maintained. The discussion in this chapter focuses primarily on noise rather than linearity because 1) the linearity levels that accompany integrated LNAs optimized for noise can suffice for many wireless applications, and 2) noise is the more fundamental problem in that it cannot be circumvented with circuit techniques or added power consumption.[1]

[1] Linearity can often be addressed either by reducing the LNA gain or by lowering the impedance levels and increasing the power consumption to make up for the lost gain. The latter approach can accommodate the case of having a weak desired RF signal with a strong interferer in an adjacent channel. For further discussions of linearity, the reader is referred to [45][46].

By the metrics of noise figure and power dissipation at microwave frequencies, bipolar transistors generally outperform CMOS devices as a result of possessing a higher transconductance per unit current, and hence the development geared toward obtaining "optimal" LNA performance is presented here in terms of bipolar devices. Details at the implementation level are covered in Chapter 5, packaged with a comparison to CMOS that illustrates the differences in the design approach, physical limitations, and achievable circuit performance associated with the two device technologies. The exploration into LNAs begins with a look at impedance matching; although seemingly foreign to the realm of IC design, the realization of appropriate source and load impedances represents a key factor in extracting performance from transistors at high frequencies. Performance at high frequencies is the game that is RF circuit design, and any consideration of technology issues in this arena stands to benefit from a thorough understanding of matching.

4.1 Impedance Matching

In lower frequency ranges where transistor gain is plentiful, signal power is easy to synthesize, and signal to noise ratios are large, circuit designs are often conducted in the voltage domain with low output impedance, high input impedance stages continually regenerating the signal energy. Power gain is a much less available commodity at microwave frequencies, and transferring as much of the signal as possible from one stage to the next—while incurring minimal noise—becomes a significant design challenge. This is the problem of impedance matching: aligning the impedances presented by the source and load to a transistor stage with that required to extract the desired performance from the stage. Toward this end, discrete circuit designs typically employ passive matching networks to transform the source and load impedances to something more conducive to getting a chosen active device to function productively. An equivalent, and sometimes more convenient, viewpoint is that the matching elements transform the input and output impedances of the transistor to match that of the applied source and load. The same story is true when the medium switches to integrated circuits, although the ability to realize nearly arbitrary device sizes opens an additional degree of freedom that can be exploited to reduce the burden on the frequently irksome matching networks.

To illustrate the concept of impedance matching and the effects of device sizing, consider first a lone common-emitter bipolar transistor. Making a set of reasonable first-order approximations,[2] the impedance looking into the base of the transistor can be expressed as:

$$Z_{in} = R_B + \frac{r_\pi}{1 + s r_\pi C_\pi}, \tag{4.1}$$

where R_B denotes the base resistance (intrinsic and extrinsic components) and r_π and C_π represent the base-emitter resistance and capacitance terms in the hybrid-π model. Base resistance scales with the emitter periphery and (the junction capacitance portion of) C_π is a function of emitter area, allowing both elements to be tailored by the transistor geometry. Furthermore, the diffusion capacitance term in C_π is proportional to the collector current while r_π has the inverse relationship, and so every component of the input impedance is subject to the device design. The bias current and transistor size are usually intertwined, and it is at times useful to think of the collector current *density* as being fixed while the transistor is scaled; in this case both r_π and C_π can also be viewed as relating to device size and hence residing directly under the discretion of the designer.

Matching is not always about maximum power transfer, however, as source and load impedances might instead be chosen for realizing minimum noise, maximum output power, maximum linearity, or for reasons of stability [47]. Noise matching relates to the equivalent input noise sources of a device rather than the impedances; returning to the example of a common-emitter transistor stage, the equivalent input voltage noise is seen to be:

$$\overline{v_n^2} = \overline{v_{R_B}^2} + R_B^2 \overline{i_B^2} + \left(\frac{R_B}{\beta(f)} + \frac{1}{g_m} \right)^2 \overline{i_C^2}, \tag{4.2}$$

when the base ($\overline{i_B^2}$) and collector ($\overline{i_C^2}$) shot noise sources are considered along with the thermal noise in the base resistance (denoted $\overline{v_{R_B}^2}$). The small-signal current gain, appearing here as $\beta(f)$ but also known as h_{21} and h_{FE}, rolls off with frequency due to the presence of C_π. Even when the input is short circuited, voltage noise appears across the base-emitter junction from the shot noise sources flowing through resistances in the base-emitter loop, and thus the noise currents appear in Equation 4.2 as well as in the equivalent input current noise:

$$\overline{i_n^2} = \overline{i_B^2} + \frac{1}{\beta^2(f)} \overline{i_C^2}. \tag{4.3}$$

Neglecting the otherwise obvious correlation between $\overline{v_n^2}$ and $\overline{i_n^2}$, the source resistance that minimizes the overall circuit noise involves the ratio of the two:

---

[2] For the purposes of this analysis, the base-collector and collector-substrate capacitances (C_μ and C_{CS}, respectively) are ignored, as are the collector and emitter terminal resistances and the small-signal output resistance resulting from base-width modulation.

$$R_{s_{opt}} = \sqrt{\overline{v_n^2}/\overline{i_n^2}}. \tag{4.4}$$

Considering the collector current density to be fixed,[3] the transistor current gain remains nearly constant as the device geometry is scaled, and so it is clear that equivalent input current noise ($\overline{i_n^2}$) is directly proportional to the emitter area in both the base and collector shot noise terms under this condition. The relationship between the voltage noise source and transistor geometry is initially a little hazier, but can be simplified by weighing the first two terms in Equation 4.2. This comparison highlights that the base shot noise component can be dismissed for the case $I_B R_B \ll 2kT/q$, a relationship that will generally hold for the large devices and low currents desired in LNAs.[4] Since the smallest emitter widths will almost always be used to minimize base resistance in low noise applications, the voltage noise can be viewed as being inversely proportional to emitter area when the current density is held constant.[5] The optimum source resistance for driving a common-emitter stage thus also decreases linearly with emitter area, allowing any desired source impedance to be accommodated by appropriately scaling the device size.

This flexibility indeed proves very useful; unfortunately however, the real part of the input impedance is very different from $R_{s_{opt}}$, and both move in the same direction with transistor sizing, making a simultaneous noise and impedance match difficult to achieve with a single common-emitter device. Furthermore, the input resistance approaches R_B for frequencies beyond f_β,[6] an impedance which is too small to afford an easy match in most RF systems. An additional degree of freedom is needed to alleviate this quandary, and a frequently-employed technique in the narrowband amplifiers which typify wireless applications is to add some inductive degeneration in the emitter, resulting in an LNA

---

[3] This is not as blatantly restrictive as it might seem, as holding the current density constant will often be the approach taken in designing an LNA. For a given bipolar transistor technology and operating frequency, there is a single current density that will result in the minimum noise. More is said about this consideration in the following section.

[4] A typical base current in a silicon-based bipolar transistor used in an LNA might be 30μA, and the base resistance—which had better be much less than 50Ω—is generally in the neighborhood of 10Ω. The product of these values is safely less than twice the thermal voltage (kT/q) at room temperature.

[5] To be more explicit, the base resistance will decrease linearly with increasing emitter area while the required collector current will increase. Neglecting the term involving the base shot noise current, each of the voltage noise components are inversely proportional to device size.

[6] The corner frequency in the small-signal common-emitter current gain is often denoted f_β. Given the single-pole model that nicely approximates the input impedance of most bipolar transistors, f_β can be expressed as f_T/β_0, where β_0 is the low frequency value of the current gain. While the high frequency limit of the input impedance might not be the most expedient property to explore in a circuit design, this model is sometimes used as a method of extracting the base resistance from measured s-parameter data.

input stage as shown below in Figure 4-1. The arrows in the diagram indicate the directions generally accepted by convention: Γ_{in} is the reflection coefficient looking into the device and Γ_{opt} is the reflection coefficient that the source should present to the device in order to minimize system noise.

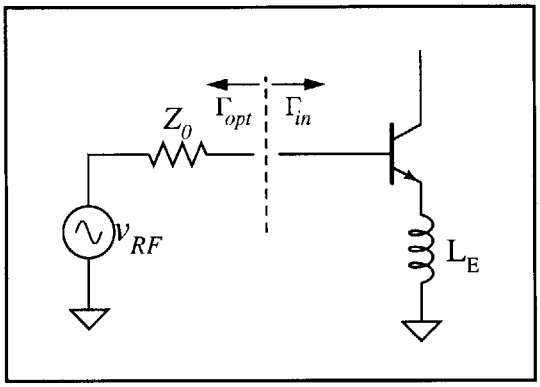

Figure 4-1. Input impedance-matching using an emitter inductance in a common-emitter amplifier stage.

The presence of the inductor creates a potential at the emitter terminal that follows the input signal but leads the intrinsic base-emitter voltage by 90°. This feedback mechanism interacts with C_π to lend an additional real component to the input impedance as given in:

$$Z_{in} = R_B + \frac{r_\pi}{1 + sr_\pi C_\pi} + j\omega L_E(1 + \beta(f)) \cong R_B + \frac{g_m}{C_\pi}L_E + j\left(\omega L_E - \frac{1}{\omega C_\pi}\right), \qquad (4.5)$$

where the approximation has assumed frequencies far enough above f_β for C_π to dominate the base-emitter junction impedance. Shot noise in the base and collector currents also flows in L_E, appending two new terms to the input equivalent voltage noise source:

$$\overline{v_n^2} = \overline{v_{R_B}^2} + (R_B + j\omega L_E)^2\overline{i_B^2} + \left(\frac{R_B}{\beta(f)} + \frac{1}{g_m} + j\frac{\omega L_E}{\beta(f)}\right)^2\overline{i_C^2}, \qquad (4.6)$$

Conversely, when the input is an open circuit, the inductive feedback loop implemented by L_E is not completed, and thus the current noise is unchanged from Equation 4.3.

Again neglecting correlation between the noise sources, Γ_{in} and Γ_{opt} (at a single frequency) are plotted in Figure 4-2 for increasing emitter inductances.[7] The real part of the input impedance moves from a value near R_B toward the 50Ω circle.[8] Moving slightly

[7] The data shown are representative of having a typical bipolar transistor sized and biased for a low-noise amplifier application.

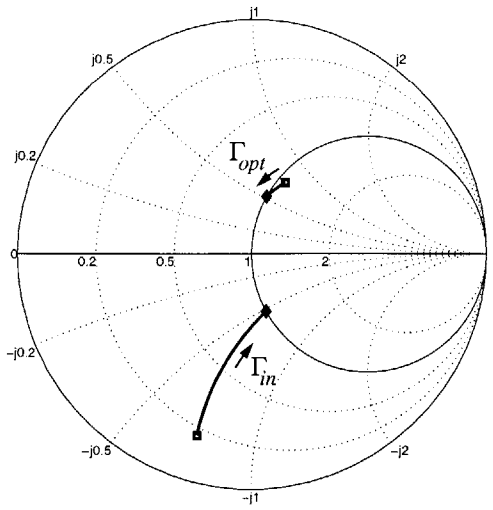

Figure 4-2. Effect of emitter inductance on the input impedance of a common-emitter amplifier stage where the arrows indicate the direction of increasing L_E; correlation in the equivalent input noise sources has been neglected.

in the direction of lower resistances is Γ_{opt}, although the absence of correlation in this example has exaggerated the variation in $R_{s_{opt}}$; to a first approximation, and hence for the purposes of design, $R_{s_{opt}}$ is essentially unchanged by the addition of L_E. Another effect rendered by assuming uncorrelated equivalent input noise sources is that the imaginary part of the optimum source impedance equals that in the conjugated input impedance (i.e. $X_{opt} = -X_{in}$). This would be a desirable trait, as Γ_{in} and Γ_{opt} could then both be closed to the center of the Smith chart (a 50Ω source impedance) by prefixing a base inductance to the stage depicted in Figure 4-1. In practice, however, correlation between $\overline{v_n^2}$ and $\overline{i_n^2}$ contributes a component to the optimum source reactance $X_{opt} = -(X_{in} + X_{corr})$,[9] where the latter term is proportional to frequency and the base-collector capacitance [48]:

$$X_{corr} = \frac{\omega C_\mu}{g_m} \sqrt{\frac{\overline{v_n'^2}}{\overline{i_n'^2}}}. \tag{4.7}$$

The ratio under the radical in Equation 4.7 consists of the fully-correlated portions of the equivalent input noise sources (calculated with C_μ considered in the device model). The correlation reactance prevents a simultaneous noise and power match from precisely

[8] More generally, the "50Ω circle" should be called the Z_0 circle. Impedances plotted on a Smith chart can be normalized to any characteristic impedance; it need not be 50Ω.

[9] At first, the concept of a correlation impedance might seem odd. Imagine, however, if the input noise consisted of a single frequency characterized by a fixed phase offset between the perfectly correlated voltage and current components. This portion of the noise could then be cancelled by a judicious selection of the source impedance.

occurring, and effectively represents an error term in the design of LNAs. It is usually not a large error at frequencies reaching into the lower end of the microwave spectrum, but should be remembered within the context of technology development. Advanced silicon-based bipolar technologies often rely upon heavily-doped collectors to support the higher current densities needed for achieving faster (higher peak f_T) transistors, a trend which may begin to undermine continued improvement in LNA performance due to the concomitant increase in C_μ. There are, of course, many other intricacies involved with LNAs, and the following sections take a deeper look at a handful of narrowband amplifier topologies and some of the shadows they cast in the semiconductor technology space.

4.2 Technology Considerations for LNAs

In a quest to realize gain and controlled impedances in a low noise and low power framework, the design of an LNA brings together a number of poignant technology issues. Investigating a range of circuit topologies and possibilities of developing techniques to circumvent limitations are important steps in determining the extent to which these issues represent fundamental barriers in performance. But before embarking on such an investigation, a means of considering comparative circuit performance is needed, and noise figure provides a convenient first-order description for LNAs in RF and microwave systems. Indicating the extent to which the signal to noise ratio (SNR) is degraded by a component, noise figure proves to be a salient measure of wireless receiver sensitivity.[10]

Given an equivalent input noise source representation of an element, and assuming that the current ($\overline{i_n^2}$) and voltage ($\overline{v_n^2}$) noise components are uncorrelated, the noise figure (expressed in dB) for that element can be calculated by:

$$NF = 10\log\left(\frac{\overline{v_{R_s}^2} + \overline{v_n^2} + Z_s^2\overline{i_n^2}}{\overline{v_{R_s}^2}}\right) = 10\log\left(1 + \frac{\overline{v_n^2} + Z_s^2\overline{i_n^2}}{\overline{v_{R_s}^2}}\right), \tag{4.8}$$

where Z_S is the source impedance and $\overline{v_{R_s}^2}$ is the thermal noise associated with the source resistance. Referring to the noise expressions derived for the common-emitter stage in

[10] There are two caveats that should be remembered in using noise figure. First, noise figure provides no indication of gain; a wire may have a 0dB noise figure, but its decided lack of gain fails to protect the SNR from noise in subsequent stages. In this sense, noise measure could be a better metric from a system perspective. A second criticism leveled at noise figure is that it is a function of the source resistance, a property which can lead to confusion when the resistance is not fixed. In absolute terms, an amplifier having a 3dB noise figure and meant for a 75Ω system contributes more noise than an amplifier with a similar 3dB noise figure but intended for 50Ω applications. The ratio of the noise in the two cases is 75Ω/50Ω.

Section 4.1, noise figure is seen to depend upon the magnitude of the transistor noise sources—and thus upon R_B, I_C, and I_B—with a corner frequency in the behavior determined by the f_T of the device through $\beta(f)$.[11] With the prevalence of shot noise in both $\overline{v_n^2}$ and $\overline{i_n^2}$, it comes as no surprise that bipolar transistors generate less noise when operated at lower current levels. The mitigating factors in this trend are that the current gain begins to fall and the intrinsic portion of the base resistance increases as I_C is reduced. Finding the optimum bias current (density) is thus a vital first step in designing low-noise amplification stages.

Denoting the argument to the logarithm function in Equation 4.8 as F, varying the source impedance to find its minimum value, and expressing the approximate result in terms of explicit transistor parameters gives [49]:

$$F_{min} \cong 1 + \frac{1}{\beta_0} + \sqrt{\frac{1}{\beta_0} + 2\frac{qI_C}{kT}(R_B + R_E)\left(\frac{1}{\beta_0} + \frac{f^2}{f_T^2}\right)}. \tag{4.9}$$

This result has been simplified by restricting the range of frequencies to be above f_β ($\omega \gg [r_\pi C_\pi]^{-1}$), further assuming that the low frequency small-signal current gain ($\beta_0 = \partial I_C / \partial I_B$) equals the DC (large-signal) current gain ($\beta_{DC} = I_C / I_B$), and by neglecting the base resistance in comparison to r_π.[12] Writing the collector current as a product of the current density (J_C) and the emitter area (A_E), and then normalizing the base and emitter resistances to this area yields:

$$F_{min} \cong 1 + \frac{1}{\beta_0} + \sqrt{\frac{1}{\beta_0} + 2\frac{qJ_C}{kT}(R'_B + R'_E)\left(\frac{1}{\beta_0} + \frac{f^2}{f_T^2}\right)}, \tag{4.10}$$

with R'_B and R'_E representing the base and emitter resistances associated with a device having an emitter geometry of the minimum width and $1\,\mu m^2$ in area. A subtle change indeed, but this form demonstrates that the minimum noise figure achievable in a chosen

[11] Though not explicitly shown earlier, the effect of the emitter resistance in the equivalent input noise sources is additive with the base resistance, and can be incorporated by replacing R_B with $R_B + R_E$ in Equations 4.2 and 4.6. While the situation may change in highly-scaled devices, the emitter resistance has routinely been made quite small (relative to the base resistance) in modern polysilicon emitter technologies, to the point where R_E does not make a strong contribution to the noise.

[12] The statement $r_\pi \gg R_B$ is very similar to an assumption made previously in the noise analysis of Section 4.1, namely that $(kT/q) \gg I_B R_B$. One further approximation made in Equation 4.9 is that the collector current ideality factor is dropped from the expression for the transconductance (i.e. n_F is assumed to be unity). This is a very good approximation for nearly all silicon-based bipolar transistors in use today.

technology is a function of the current density (β_0, R'_B, and f_T all depend on J_C) and not the transistor size; the optimum source impedance associated with the minimum noise figure decreases as the device grows, but F_{min} remains essentially unchanged.[13]

By way of completing the story, the unity current-gain frequency for a bipolar transistor can be related to the current density—for bias currents well below the onset of base pushout—as:

$$f_T = \frac{1}{2\pi[\tau_F + (kT/q)(C'_{jE} + C'_\mu)/J_C]}, \tag{4.11}$$

where τ_F denotes the forward base transit time and the base-emitter C'_{jE} and base-collector C'_μ junction capacitances are defined per unit area analogously to the resistances in Equation 4.10. The current density at which the optimum (lowest) F_{min} occurs can then be determined by inserting this expression for f_T, approximating β_0 and R'_B as constants, and solving $\partial F_{min}/\partial J_C = 0$; this algebraic handiwork provides [49]:

$$J_{C_{opt}} \cong 2\pi(C'_{jE} + C'_\mu)\frac{kT}{q}\sqrt{\frac{\beta_0 f^2}{1 + 4\pi^2\beta_0\tau_F^2 f^2}}. \tag{4.12}$$

While subject to the same limitations as the noise figure expression from which it is derived (restricted, most notably, to frequencies above f_β), Equation 4.12 does convey the often-overlooked significance of the role played by transistor junction capacitances in the performance of RF/microwave LNAs. That the optimum current density is proportional to $C'_{jE} + C'_\mu$ indicates that *the attainable noise figure which can be extracted from a technology, and the power consumption required to realize this minimum noise condition, will both decrease as the junction capacitance is reduced.* When the source resistance is fixed, LNA power dissipation and noise figure will approximately follow the square root of $J_{C_{opt}}$; the power saved by lowering the optimum bias current density through a reduction in capacitance is partially offset by the necessity of using a larger transistor to provide a noise match to the source.

Surprisingly mute in all of this behavior is the ubiquitous current gain $\beta(f)$. Primarily bound by junction capacitances in the LNA regime of RF frequencies and low noise bias currents, $\beta(f)$ goes largely unaffected as $J_{C_{opt}}$ and $C'_{jE} + C'_\mu$ scale together. As

[13] This is more true for the purposes of an LNA, for which noise constraints mandate the minimum emitter width, than it is in general. Until current crowding starts to become an issue, increasing the emitter width for a given emitter area and bias current yields a higher transistor f_T due to reductions in both C_μ and τ_F. The changes in τ_F result from a second-order effect involving current spreading in the collector [50].

might be expected, the thermal noise contributors in the transistor are independent of capacitance while the shot noise sources follow in tune with the bias current, yielding an increase in the optimum source resistance as $J_{C_{opt}}$ decreases. For frequencies not too close to f_T, the largest term in the voltage noise expression (Equation 4.2) is the base resistance thermal noise; ignoring the somewhat smaller $\overline{i_B^2}$ and $\overline{i_C^2}$ terms for the moment, and recognizing that a current density of $J_{C_{opt}}$ corresponds to having $\beta(f) = \sqrt{\beta_0}$, a rather crude approximation for the optimum source resistance can be proposed:

$$R_{s_{opt}} \cong \frac{1}{A_E}\sqrt{\beta_0 R'_B \frac{kT/q}{J_{C_{opt}}}}. \tag{4.13}$$

An appendix could be filled with the litany of assumptions that have gone into Equation 4.13, but this relationship does capture the prevailing behavior and demonstrates that a reduction in the optimal collector current density ($J_{C_{opt}}$) necessitates an increase in transistor emitter area (A_E) by $\sqrt{J_{C_{opt}}}$ to maintain a desired source resistance. As before, the emitter stripes are assumed to be drawn at the minimum width allowed by the technology, R'_B is the base resistance for a $1\,\mu m^2$ device,[14] and the low frequency current gain is taken to be equal to the DC current gain. Furthermore, the operating frequency has been constrained to a band of frequencies far enough above f_β for C_π to dominate the base-emitter impedance, but yet enough below f_T so that $\beta(f)$ is still appreciably greater than unity. And finally, it should not be forgotten that correlation between the equivalent input noise sources has been, well, forgotten in the derivation of $R_{s_{opt}}$.

Couched within this discussion of technology considerations is a framework that has been established for designing low-noise amplifier stages based upon common-emitter topologies. The approach begins by identifying the collector current density that minimizes noise at the frequency of interest (via Equation 4.12).[15] It is important to note that *this current density, denoted $J_{C_{opt}}$, is a technology parameter and is subject only to the design frequency*; there are no "knobs" to be turned at the circuit or system design levels. Next on the agenda is to size the device to place Γ_{opt} on the 50Ω circle by scaling the emitter area (Equation 4.13 or, more accurately, using Equation 4.4). The choice of a 50Ω characteristic impedance is more tradition than physical law; while an LNA is typically

[14] Also as before, the emitter resistance can be incorporated by summing it with the base resistance (i.e. $R'_B \rightarrow R'_B + R'_E$).

[15] Being tacitly practiced here is a fairly typical microwave design ploy: in adjusting the bias current and looking at F_{min}, the source impedance is allowed to go wherever it needs in order to minimize the noise at each bias condition. Having the impedances "track" the swept parameter proves exceedingly useful, and is the reason why so many definitions of gain are used in the design and analysis of microwave circuits.

constrained by interfacing with an input from an antenna and perhaps a low-loss filter, moving toward higher impedances might result in lower power consumption, provided that the corresponding loss in gain does not become too severe.[16] With the optimum source resistance having been chosen, the design is completed by setting first the emitter inductance and then the base inductance to provide an input impedance of 50Ω as reflected in Equation 4.5. Tuning L_E allows a 50Ω resistance to be established looking into the base, and L_B is used to resonate the residual input capacitance. By appropriately sizing the transistors and inductors, this approach—summarized pictorially in Figure 4-3—yields the optimum noise figure afforded by a technology while transforming the input impedance to provide a power match. Considering that only one active device has been used, it should not be surprising that this topology will prove difficult to beat.

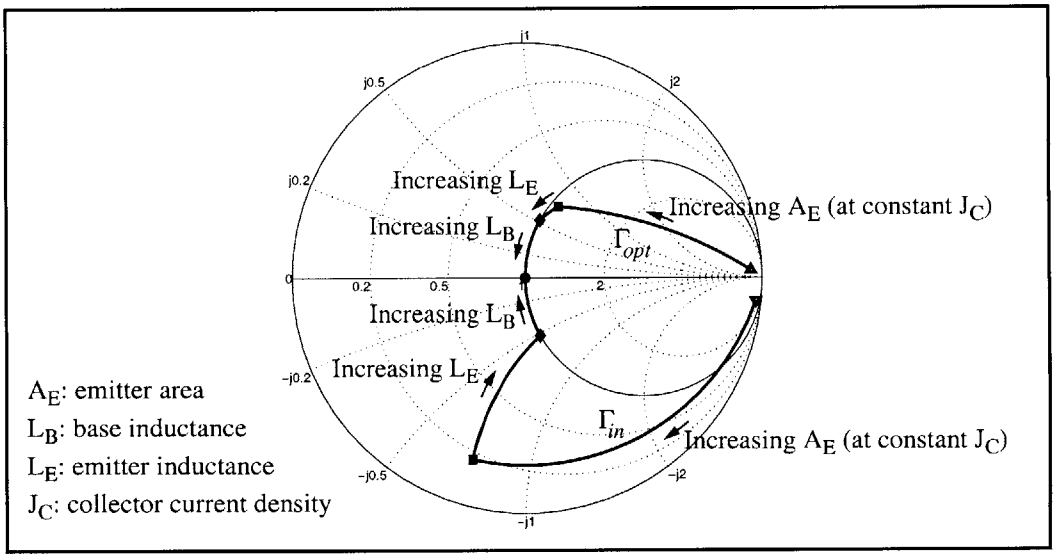

Figure 4-3. Illustration of achieving a simultaneous noise and power match in a common-emitter transistor stage; correlation in the equivalent input noise sources has been neglected.

4.3 Alternative LNA Topologies

Noting that design simplicity often carries the day in RF and microwave circuits, a good place to begin a search into alternative LNA topologies is with an examination of the

[16] Regardless of the frequencies involved and whether voltage or power is being considered as the signal, transistors fundamentally remain transconductance devices and thus need to drive large impedances to realize gain. If all of the circuit impedances could be scaled inversely with the transistor size, then the available gain (to first order) would remain the same as the bias current is reduced. Unfortunately, in an integrated circuit, the maximum load impedance is limited by the inductor loss, and thus the decrease in transconductance resulting from the lower bias current associated with a higher source resistance will reduce the gain.

remaining single-transistor amplifier connections. With all of the emphasis in the preceding sections upon analyzing a common-emitter stage, it should be no small consolation that the same equivalent input noise sources represent a common-base transistor [51], and so all of the expressions derived for the optimum current density, source impedance, and noise figure apply equally well to the common-base configuration [52]. Some second-order differences in gain and noise figure exist due to subtleties in the base-collector capacitance feedback mechanism [53], but are generally not too significant for LNA applications operating at frequencies below f_T.[17] One difference that is significant, however, is the input impedance; for a common-base transistor:

$$Z_{in} = R_E + \frac{1}{g_m}\left(\frac{\beta(f)}{1+\beta(f)}\right) + \frac{R_B}{1+\beta(f)}, \tag{4.14}$$

with the emitter resistance (R_E) now being explicitly included. This impedance is Z_{in} for the common-emitter stage divided by the factor $1+\beta(f)$, and will typically be small in magnitude and predominantly resistive. Except for low bias currents, the input resistance will be too small to yield a good power match to most RF sources—a problem that unfortunately seems to keep recurring.

A convenient solution for the common-emitter design was afforded by the addition of some series-feedback inductance, a technique for which the analogue in common-base stages is to attach a capacitive impedance at the base terminal. When the base is shunted (to an RF ground) through a capacitance C_B, the input resistance of the common-base device increases by $C_\pi/(g_m C_B)$. For this capacitor to have its intended effect, however, the base bias will need to be applied through some sort of reasonably large RF impedance. Coupling the bias through an inductor or a resistor would certainly qualify; the catch is that for the transistor to function properly (i.e. to provide gain and low noise), the impedance presented by C_B must be small over the range of frequencies for which the LNA is being designed. Nothing is new here either, as much the same can be said for the inductance in the common-emitter stage. Grounding the emitter through a large L_E degenerates the gain and increases the equivalent input voltage noise (Equation 4.6), the latter taking

[17] The feedback admittance y_{12}—which reduces the transistor gain, increases the noise figure, creates stability issues, and may possibly contribute to global warming—can be expressed as $y_{12} = -sC_\mu/[1 + s(C_\mu + C_\pi)R_B]$ for a common-emitter device, and is a factor of $g_m R_B$ larger when the same transistor is used in the common-base configuration (this holds up to about f_T, and neglects the transistor output resistance that contributes a $1/r_o$ term). This does not tell the whole story, however, as $g_m R_B$ will be on order of unity for typical silicon bipolar LNAs. As frequencies approach f_T, the available gain from emitter to collector falls more rapidly than that from base to collector so, for similar y_{12} values, the common-base connection exhibits lower gains and higher noise figures than the common-emitter variant [53].

$R_{s_{opt}}$ along with it. The distinction between these two configurations, and one which proves to be of consequence, is that f_T works with the input resistance in the common-emitter case and against it when using a common-base transistor. Approximating g_m/C_π as ω_T, the resistance added by an emitter inductance to the input of a common-emitter device is $R_{in} \cong \omega_T L_E$, allowing the impedance of this inductor to be expressed as $Z_{L_E} \cong \omega R_{in}/\omega_T$. The extra input resistance induced by C_B in a common-base stage has the reciprocal relationship: $R_{in} \cong 1/(\omega_T C_B)$, with the capacitor imposing an impedance $Z_{C_B} \cong \omega_T R_{in}/\omega$ between the base and (an RF) ground. The weaker ground that results in the common-base stage—the impedance is higher by a factor of ω_T^2/ω^2—makes simultaneously obtaining gain with both noise and power input matches a very difficult (if not impossible) task indeed. This story may change for designs targeting frequencies very near, or beyond, the transistor f_T. Furthermore, higher frequencies may allow a convenient transmission line implementation of the base circuitry; the DC bias could be applied through a high characteristic impedance $\lambda/4$ trace, and an open-ended stub (or radial stub) could be used to provide a narrow-band capacitance at the base. But except for broadband LNA applications or those that push the very limits of the available transistor technology, a common-base input stage yields no fundamental advantage over a common-emitter topology, and mostly just results in a more contentious design.

The last of the three single-transistor configurations is the emitter-follower. So commonly seen as a buffering stage due to its high input impedance, low output impedance, and nearly unity voltage gain, that a common-collector device can yield power gain through its impedance transformation properties might easily be overlooked. The maximum available (power) gain attainable from a transistor used in this fashion is fairly well approximated as $G_{ma} \cong 10\log(A_v^2|Z_{in}|/|Z_{out}|)$,[18] where A_v is the voltage gain, Z_{in} is the same as that looking into a common-emitter stage when $j\omega L_E$ (in Equation 4.5) is replaced by the load impedance, and Z_{out} corresponds to the input impedance of a common-base stage (as given in Equation 4.14) when the resistance of the source driving the emitter-follower is included with the transistor base resistance (i.e. $R_B \rightarrow R_B + R_s$). Although $|Z_{in}|/|Z_{out}|$ follows $\beta(f)$ for lower frequencies—extending approximately to the point where $\beta(f) = g_m(R_B + R_s)$—this ratio of impedances can be shown to be greater than unity for $\beta(f) = 1$, indicating that power gain is in fact available beyond f_T.[19] The

[18] The maximum available gain of a transistor is the power gain that can be obtained when simultaneous conjugate impedance matches are presented at both the input and the output. It is only defined over the range of frequencies for which a transistor is unconditionally stable, which is to say that the device will not oscillate for any passive termination that could be affixed to either the output or the input. Emitter-followers specifically tend not to cooperate with capacitive loads, a topic that Section 4.4 will cover in some detail.

gain at such frequencies will not be large, yet this observation should at least lend some sense of the potential for realizing a common-collector RF amplifier.

Given the apparent feasibility of amplification, the next question pertains to how the noise performance of an emitter-follower compares, and the answer is that the equivalent input noise sources for this connection are also very similar to those for a common-emitter stage. Taking the output from the emitter instead of the collector includes the base current in the numerator of the current gain expression, and so all of the terms involving shot noise are multiplied by the factor $[\beta(f)/(1+\beta(f))]^2$, resulting in:

$$\overline{v_n^2} = \overline{v_{R_B}^2} + \left(R_B + \frac{1}{g_m}\right)^2 \left[\frac{\beta(f)}{1+\beta(f)}\right]^2 \overline{i_B^2} + \left(\frac{R_B}{\beta(f)} + \frac{1}{g_m}\right)^2 \left[\frac{\beta(f)}{1+\beta(f)}\right]^2 \overline{i_C^2}, \tag{4.15}$$

for the voltage noise source and a current noise represented by:

$$\overline{i_n^2} = \left[\frac{\beta(f)}{1+\beta(f)}\right]^2 \overline{i_B^2} + \left[\frac{1}{1+\beta(f)}\right]^2 \overline{i_C^2}. \tag{4.16}$$

The appearance of the dynamic emitter resistance ($1/g_m$) in the $\overline{i_B^2}$ component of the voltage noise stems from the emitter not being at ground. However—as this component will generally remain only a minor contributor to the voltage noise—the expressions shown above will be essentially equivalent to those for a common-emitter transistor until $\beta(f)$ approaches unity, allowing the previously-derived relationships for optimum noise figure, current density, and source resistance to be applied with reasonable accuracy.

There is one significant drawback to the common-collector configuration, and this is in having only a base-emitter junction sitting between the output and input terminals. Despite its reputation as a buffer amplifier, the isolation afforded by an emitter-follower decreases in conjunction with the current gain, and so very little decoupling is provided as frequencies begin to push toward f_T. Possessing a reverse transconductance given by $y_{12} \cong -1/[R_B + \beta(f)/g_m]$ for any usable frequency, this conductance will be roughly a factor of C_π/C_μ higher throughout the RF spectrum for a common-collector device than that exhibited by the other two incarnations.[20] The increased interaction between input and output loading that results at higher frequencies hinders the performance of a follower as a voltage buffer and limits its usefulness as an RF amplifier when isolation is required between an antenna and internal circuitry. Furthermore, poor isolation makes for a difficult design as the matching networks and impedances on one side of the amplifier influ-

[19] At first this may seem to run afoul of fundamental physics, being that the power gain cannot be greater than one when both the voltage and currents gains are less than unity. However, $\beta(f)$ refers to the *common-emitter* current gain; the current gain from base to emitter is $1+\beta(f)$.

ence those on the other side.[21] Compounding the gravity of this consequence is that the collector of a forward-active bipolar transistor, when used as a common terminal, offers little opportunity to affect the transistor impedances through local feedback—a trait conveniently exploited in both the common-base and common-emitter forms. As a result, when the goal is to achieve a fully-integrated solution, it appears quite likely that the small improvement in intrinsic noise figure offered by an emitter-follower will be swamped in a sea of matching network issues.

It is seen from the preceding discussion that the common-base and common-collector connections are very similar to the common-emitter device in terms of noise performance throughout most of the RF frequency range. Small differences do exist among the transistor configurations in how they are inhibited by parasitic resistances and capacitances; bipolar transistors feature a little more (power) gain when used as common-emitter devices, exhibit slightly lower noise in the emitter-follower form, and yield some potentially useful impedance properties along with intermediate gain and noise performance when the base is taken as the common terminal.[22] That the common-emitter connection has become the predominant choice for LNA input stages is owed largely to the convenience that it affords; the alternative uses of a single transistor entail a greater design challenge and fail to offer improvements of any real consequence to RF performance. While stages can be cascaded to increase gain or bandwidth at the cost of power consumption and noise figure, a couple of multiple transistor incarnations—the cascode and the f_T doubler—may also be worth investigating as alternatives for LNA applications.

4.3.1 Unity Current Gain Frequency (f_T) Doubler

A recurrent theme in this section has been to examine the effects of falling current gain on noise, gain, and impedances as the frequencies being handled approach the tran-

[20] A more commonly-encountered measure of isolation for RF purposes is phrased in scattering parameters—a set of small-signal parameters that characterize a network in terms of unitless gains. Expressed in dB, s_{12} conveys the rejection an amplifier provides (measured at its input) of a signal driven onto its output port. For a common-emitter or common-collector transistor, to a first-order degree of accuracy, and assuming that the source and load resistances are equal to a characteristic impedance Z_0, $s_{12} \cong -2y_{12}Z_0(R_B + Z_\pi)/(Z_0 + R_B + Z_\pi)$, where the base-emitter impedance is denoted $Z_\pi = \beta(f)/g_m$. This isolation term can be considered as the reverse transconductance multiplied by the resistance in the RF source and a factor that accounts for loss in the base-emitter circuit.

[21] A device offering perfect isolation is also known as being unilateral, a description rooted in the notion that signals can only propagate through the device in the forward direction (i.e. the reverse gain $s_{12} = 0$).

[22] The higher power gain of the common-emitter device refers to the maximum stable gain (or maximum available gain, if the transistor is unconditionally stable) at low frequencies. For higher (RF) frequencies, the common-base transistor can provide slightly higher gains.

sistor f_T. Observing the deleterious effects of low current gain at RF frequencies, one might begin to ponder whether f_T is fundamental, or rather if there is some means of increasing the unity current gain frequency beyond the device f_T. As an affirmative answer, enter a class of circuits loosely known as f_T doublers. By way of first-order thinking, if the input signal is impressed across two base-emitter junctions placed in series, the effective input capacitance—and thus the input current—is halved. By itself, this series connection accomplishes little; the input voltage splits across the two base-emitter junctions and thus each transistor produces half the output current that a single device would have generated. However, if the collectors of the two series transistors can be connected such that the output currents add in phase, then the overall output current would be the same as that sourced through a single transistor. With the associated input current having been lowered by a factor of two, the current gain—and hence the unity current gain frequency—is effectively doubled.

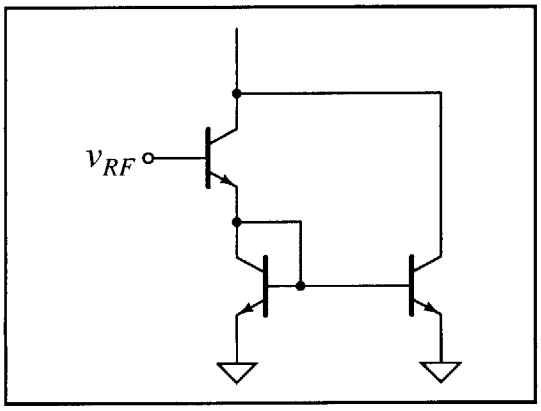

Figure 4-4. Battjes' f_T doubler circuit.

One particularly elegant single-ended embodiment of the f_T doubler concept was implemented by Battjes in the late 1970s and is depicted in Figure 4-4 [54][55]. The current mirror in the emitter of the input transistor ensures that the same current flows in both devices driving the output, effectively applying the transconductance to the full input voltage. Meanwhile, the $2C_\pi$ loading of the mirror increases the input impedance, lowering the input current associated with a given output signal level. Assuming matched transistors each having current gain characteristics described by $\beta(f)$, the overall current gain (h_{21}) realized by this circuit can be expressed as:

$$h_{21} = 2\beta(f)\left[\frac{1+\beta(f)}{2+\beta(f)}\right]. \tag{4.17}$$

The current gain of the Battjes doubler thus follows $2\beta(f)$ up to a pole frequency located where $\beta(f) = 2$ and a subsequent zero at $\beta(f) = 1$. A doubling is effected for frequencies below $f_T/2$, beyond which the pole-zero doublet indicated in Equation 4.17 causes the current gain to asymptotically approach $\beta(f)$.

While this increase in f_T is real, and this technique has proven useful for wideband amplifiers [56][57],[23] it is important to realize that there is no corresponding increase observed in the power gain (i.e. this is an f_T doubler, not an f_{MAX} doubler). Concomitant with the doubling of the current gain is a halving of the output impedance driven by the transconductance, leaving the available (power) gain unchanged. Similarly, the shot and thermal noise contributions from each transistor are also unaffected by splitting and recombining the signal path, resulting in equivalent input noise levels that are actually elevated above the single-transistor common-emitter stage due to the cascaded devices. Hence, while their simplicity of connection and biasing allows f_T doubler stages to easily replace common-emitter transistors in a variety of applications, the higher noise and equivalent power gain of the doubler fail to make a compelling argument in the realization of low-noise amplifiers.

4.3.2 Cascode

Another connection that is often invoked in the quest for improved frequency response is the cascode: a DC-coupled common-emitter, common-base cascade that reuses the same bias current for each stage. Shown with impedance matching inductors in Figure 4-5, the conventional thinking behind the cascode is that the input loading of a Miller-multiplied C_μ is suppressed by keeping the voltage gain across the common-emitter device near unity and instead obtaining the gain via a common-base transistor. However, as capacitive loading tends not to be an issue for reactively-matched narrowband amplifiers, this suppression property is not significant in most wireless applications.[24]

---

[23] One modification of the f_T doubler circuit shown in Figure 4-4 is to balance the collector-emitter voltages by adding a diode-tied device in the collector of the mirror output transistor. This can be important when using III-V HBTs not only for self-heating considerations, but also because the transistor f_T generally falls as V_{CE} is increased beyond a certain level. GaAs, InGaAs, and InP feature a pronounced velocity overshoot effect; the electrons initially fill a low effective mass conduction band and then reach a much heavier band at higher energies, causing the collector transit time to reach a minimum at low V_{CE} values [58]. Secondly, III-V transistors fabricated via MBE often have thin (and lightly-doped) collectors that are fully depleted for V_{CE}~0. In this case C_μ is no longer reduced by increases in V_{CE}.

[24] This property of a cascode can be more of a hindrance than a help, as the impedance-reducing quality of shunt feedback can be usefully employed to aid the realization of an input match. One published cascode amplifier has featured an impedance inserted in series between the two transistors, implemented for the explicit purpose of re-instantiating the Miller effect [59].

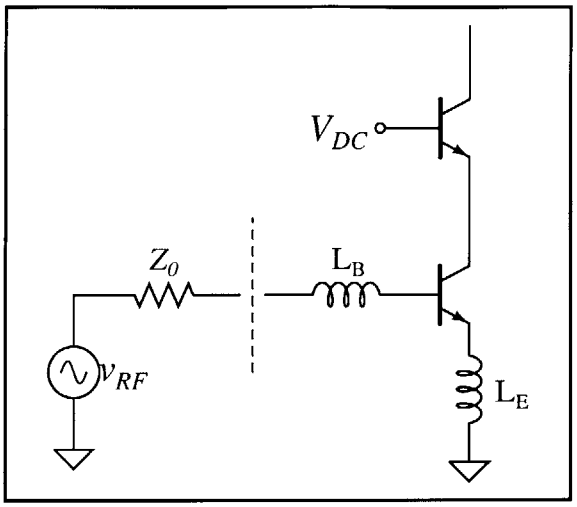

Figure 4-5. Cascode amplifier stage with input impedance matching.

Of greater consequence is the manner in which the feedback path represented by C_μ limits gain and stability in a common-emitter stage. As an illustration, consider the inverting amplifier configuration shown in Figure 4-6. The op-amp represents the open-loop gain afforded by the transistor ($A_v = -g_m Z_{out}$), and the base-collector capacitance connects around this gain element to a node defined by the intrinsic (R_{Bi}) and extrinsic (R_{Bx}) portions of the base resistance. Driven through a source having an impedance R_S, the magnitude of the voltage gain is limited to f_μ/f, where $f_\mu = [2\pi C_\mu(R_s + R_{Bx})]^{-1}$, independent of the intrinsic transistor gain and assuming perfectly-resonated (infinite) input and output impedances.[25] The cascode amplifier can provide a win by sidestepping this feedback issue and the associated stability problem of having appreciable loop gains

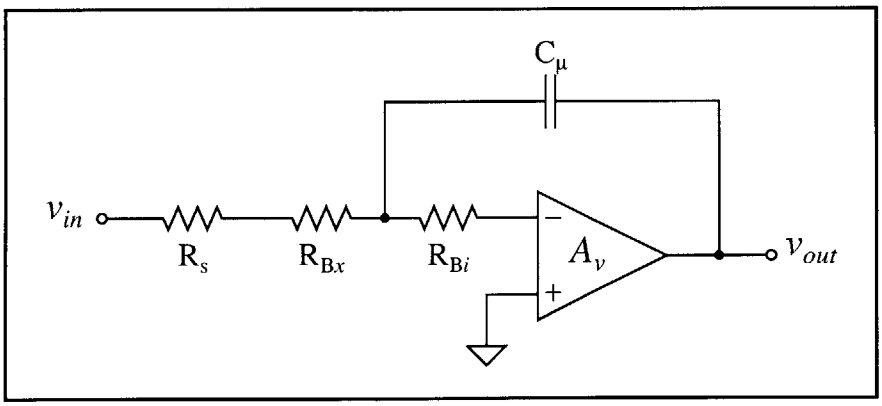

Figure 4-6. Intrinsic feedback within a common-emitter stage.

[25] Particularly at low frequencies, the realizable transistor gain will generally be restricted to be well below this value, as the inductive source impedance required to resonate the input capacitance can lead to stability problems. See Section 4.4 for a more detailed analysis.

at high frequencies. But tempering any newfound enthusiasm is that the common-base transistor is also limited by a feedback mechanism as the base resistance degenerates the gain. An additional limitation with the cascode connection is that it largely only trades one stability problem for another; both sides of this trade-off will be examined in Section 4.4.

Regardless of the competing set of issues that the cascode introduces, its increased output impedance does yield a palpable improvement in power gain, where the maximum available gain from the cascode is on the order of $10\log(g_m r_o)$ higher than is achievable from a common-emitter device at the same bias current.[26] Also improved through cascoding is the isolation; the attenuation from output to input is much higher than can be achieved from other single-stage designs. In fact, the low internal impedance of the cascode helps to isolate the input from the output to a greater extent than some two stage designs manage to achieve. But these benefits do not come without a price; the stacked devices eat into supply headroom,[27] and the interstage noise mismatch leads to a higher noise figure than could be achieved with a cascade of common-emitter stages.[28] The second of these costs summarizes a general lesson to be taken from studying the f_T doubler and cascode topologies: adding more devices in the signal path, even in these single-stage designs, can only hurt the amplifier noise figure. Unfortunately, there are no circuit techniques which can be found to overcome this fundamental technological hurdle.

4.3.3 Effects of Inductor Loss

As exemplified in this chapter, inductors play a prevalent role in the design of LNAs for RF/microwave receivers. The gain and noise characteristics of an amplifier stage can be dramatically improved over a finite bandwidth by incorporating inductors as impedance matching elements. Along with inductance, however, inductors bring with them a number of loss mechanisms that can significantly impact the amplifier. Noting that the series resistance terms associated with the base and emitter inductors add to the tran-

---

[26] Sadly, much of this gain is thrown away in the inductor loss, but the improvement can be more appreciable with MOSFETs and III-V transistors due to their lower output resistances (r_o).

[27] When the supply voltage has been fixed due to other constraints, the stacked devices might actually be a benefit. Splitting the supply voltage across two transistors can mitigate V_{CE} breakdown and self-heating concerns, and could potentially lead to a higher f_T (see Footnote 23).

[28] The higher noise figure of the cascode is sometimes explained as resulting from the suppressed gain of the input transistor. Even in the cascode structure, however, the noise contributions of the common-base device are still reduced by the available gain of the common-emitter transistor which precedes it. A more correct way of looking at the higher noise figure is to recognize that the output impedance of the common-emitter device is generally not particularly close to Γ_{opt} for the cascode transistor, and it is this mismatch that increases the noise.

sistor terminal resistances R_B and R_E, it can be seen (through Equation 4.9) that the LNA noise figure might sustain an appreciable hit from inductor loss—particularly when on-chip spirals are being used.[29] To help convey a sense of the magnitude, it is useful to look at a representative example. A typical silicon-based bipolar transistor, sized to provide an optimum noise match to a 50Ω source when biased at $J_{C_{opt}}$, might have a base resistance of 8Ω and a 1Ω resistance in the emitter terminal. Assuming a current gain of 80 at DC and an f_T/f ratio of 4 at the operating frequency (associated with $J_{C_{opt}}$), the minimum achievable transistor noise figure (NF_{min}) is determined to be 1.50dB. Borrowing inductor sizes of 0.7nH and 0.4nH for L_B and L_E from the 5.8GHz design example carried out in Chapter 5, taking $Q = 5$ as a representative value and attributing all of the loss to a series resistance, an additional 8Ω needs to be considered with the base and emitter resistances in Equation 4.9 to account for the inductors. From this calculation, the loss in the matching inductors is seen to bump the minimum realizable LNA noise figure up by about 0.4dB (to 1.91dB)—a rather significant effect indeed!

Continuing with the model of combining inductor loss with the terminal parasitic elements, it becomes evident that the source resistance needed to minimize the amplifier noise is also affected. The inductors contribute both thermal noise and additional impedances to the equivalent input noise voltage (as expressed in Equation 4.6) while leaving the corresponding noise current (Equation 4.3) unchanged. As the ratio of these equivalent sources relates to the optimum source resistance, $R_{s_{opt}}$ increases due to the loss in L_B and L_E. To compensate, the emitter area of the input transistor can be increased—keeping the same current density—to reestablish an optimum noise condition to match the input source. Finally, though not surprisingly, the reactive portion of Γ_{opt} also fails to emerge unscathed; the parasitic capacitances associated with the inductors supplement the reactances of C_π and C_μ. The matching inductance (mostly in L_B) thus needs to be reduced to return Γ_{opt} to the $X = 0$ line along the center of the Smith chart.[30] An iterative design loop results, where inductor loss necessitates a larger input transistor, but the extra capacitance associated with the larger transistor and the inductor allow a smaller—and hence lower loss—inductor to be used. Fortunately, this loop generally reaches a satisfactory conclusion after just two or three iterations.[31]

---

[29] These inductor series resistance terms impact every measure where R_B and R_E appear, encroaching upon many aspects of amplifier performance (stability and isolation included).

[30] Or, to phrase this in a manner less reliant upon Smith charts, the reduction in the matching inductance is needed to return the amplifier input resonance to the desired center frequency in the face of the additional parasitic capacitances brought on by the inductors.

[31] As is often true of high frequency amplifier design: no pain, no gain.

The preceding example brings to light another lesson regarding integrated implementations of matching and tuning networks: what might seem like a win when conceived with lossless components may well be washed out in the loss of real passive devices, especially when the attendant area penalty is considered. As one example, consider the previously-discussed base-collector capacitance and its deleterious effects upon the gain and isolation of a common-emitter transistor. In principle, it should be possible to neutralize this capacitance over a narrow band of frequencies by resonating it with a shunt inductor.[32] However, the loss associated with an integrated inductor is likely to actually *lower* the base-collector impedance below that of the intrinsic junction, in turn hurting the amplifier performance more than it helps. Similarly, using an inductor to resonate the capacitance sitting on the inner node of a cascode would seem to be an opportunity to improve upon the stability, gain, and noise figure characteristics of an LNA. But here again, inductor loss will generally erase any benefits in both gain and noise figure. Furthermore, as will be shown in the following section, stability in the cascode can be better handled as a consideration in sizing the common-base transistor. This is not to dismiss the general utility of inductors in many settings, but merely to indicate that some caution should be exercised in evaluating each new application.

4.4 Transistor and Amplifier Stability

To the conventional analog IC designer, analyses of stability are rooted in descriptions of loop gain and phase carrying names like Routh, Nyquist, and Bode. Venturing into a culture of microwave circuit design that has grown up around two-port representations, stability parameters derived from impedances or (equivalently) reflection coefficients, and Smith chart-based stability circles tends to leave one's mouth agape.[33] An improved scenario would be to establish a comfort level in believing that these different considerations of stability are at least consistent. Rather than attempting to demonstrate mathematical equivalency, the goal here is to interpret some typical stability circles for a transistor stage and then to try to understand why these conditions of potential instability might exist.

Traditional microwave definitions of stability parameters and circles can be completely established from a set of small-signal (e.g., s-parameter) measurements on a two-

[32] A DC-blocking capacitor will be needed in series with the neutralizing inductor unless the preferred bias condition is $V_{BC} \cong 0$.

[33] The reverse situation is also true: most microwave designers are no more comfortable with phase margins than IC designers are with stability circles.

port element. Stability circles, for which definitions have been supplied in the Appendix, delineate a region of impedances that can lead to instability. Not surprisingly, these circles are drawn on a Smith chart, and are defined for both source and load impedances. Source stability circles for a common-emitter transistor at two frequencies have been plotted in Figure 4-7.[34] In this case, the Smith chart represents the set of all possible (passive)

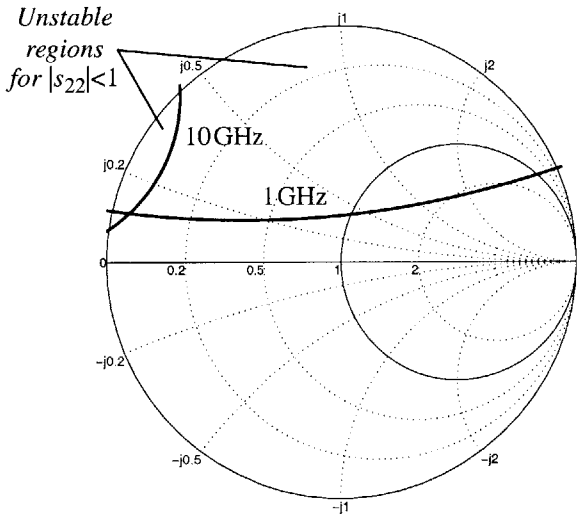

Figure 4-7. Source stability circles for a common-emitter transistor.

source impedances, and the circles indicate the collection of those impedances for which $|\Gamma_{out}| = 1$. Having a reflection coefficient greater than unity indicates that the signal being reflected from a port is larger than the signal incident upon that port, meaning that the element is supplying power to the terminal being measured via the applied test signal. If a low-loss impedance (i.e. a load) is connected to this port, an oscillation could be sustained as a significant portion of the signal power reflected from the transistor would be returned by (and not transferred to) the load. Hence, the region corresponding to $|\Gamma_{out}| > 1$ is the set of source impedances that can lead to instability, and to which side of each circle this lies can be most easily determined by checking the reflection at one point. Usually, the easiest impedance to check is the characteristic impedance Z_0 (e.g., 50Ω):

$$\Gamma_{out}\big|_{Z_s = Z_0} \equiv s_{22}, \tag{4.18}$$

[34] Note that these circles are for the transistor itself; lower frequency analog designs with gain- and bandwidth-limiting resistive and capacitive impedances generally don't consider that the transistor alone may be unstable.

as this is the definition of s_{22}. If $|s_{22}| < 1$, then stable operation will result when the source impedance is equal to the characteristic impedance,[35] and thus the stable region is the one that contains the center of the Smith chart.

Returning to the common-emitter example plotted in Figure 4-7, checking to ensure that $|s_{22}|$ is less than unity at 1GHz and 10GHz verifies that the unstable regions are located inside the circles. Both of these regions correspond to inductive impedances that are not overly lossy, although the loss needed to avoid potential oscillations is lower at 10GHz as is seen from the narrower swath cut by the circle calculated at that frequency. To get a better feel for why an inductive source impedance can create instability with a common-emitter device, it is instructive to separate the base-collector capacitance and (extrinsic) base resistance from the remainder of the transistor as shown below in Figure 4-8. An inductive source termination has been affixed at Γ_s and the loop has been

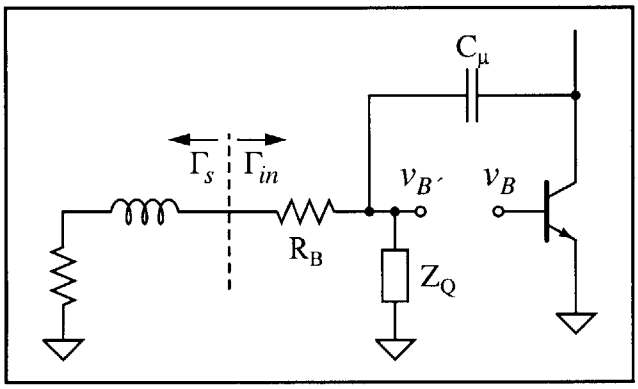

Figure 4-8. Loop gain analysis of a common-emitter transistor.

broken at the base terminal, where Z_Q represents the loading of the base-emitter junction on the feedback loop. Taking v_B as the input, the transistor provides a voltage gain $-g_m Z_{out}$, a quantity that should exceed unity (magnitude) if the transistor is to be useful as a gain stage. Then, ignoring Z_Q for a moment, the feedback network is observed to form a series LC resonator having a finite quality factor (Q) limited by the base and source resistances. At resonance, the voltage swing on the internal node of a series LC circuit is a factor of Q larger than the signal applied across the resonator, and so voltage gain in the feedback path is also available. Furthermore, provided that $Q > 1$, the resonant behavior of C_μ and the inductive source impedance will exhibit 180° of phase shift for a range of frequencies below resonance. Given these conditions, and even incorporating the loading

[35] Of course, the output impedance must also be in its region of stability for the transistor stage to be deemed unconditionally stable.

of Z_Q, it is not exceedingly difficult for there to be substantial loop gain from v_B to $v_{B'}$ while having 0° net phase shift—precisely the requirements for oscillation.[36]

Another interesting example comes from the emitter-follower and the load impedances which might result in instability with this stage. Defined analogously to the source stability circles mentioned earlier, load stability circles calculated at 1GHz and 10GHz for a common-collector transistor are illustrated in Figure 4-9. Verifying that $|s_{11}| < 1$ determines that the stable regions are those inside the circles, leaving most of the capacitive half of the Smith chart as potentially unstable territory. While a loop gain analysis could

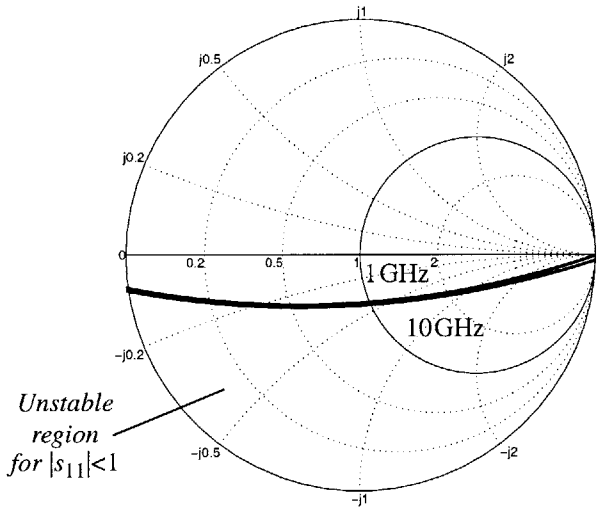

Figure 4-9. Load stability circles for an emitter-follower transistor.

also be constructed for this case, it is similarly insightful to look at the input impedance of an emitter-follower stage when the load being driven is capacitive. Revisiting the expression for the input impedance of a common-emitter device having some impedance from the emitter to ground (Equation 4.5), and denoting the load capacitor as C_L, the impedance observed looking into the base terminal of the follower is:

$$Z_{in} = R_B + Z_\pi + \frac{1}{j\omega C_L}(1 + \beta(f)) \cong R_B + \frac{1}{j\omega(C_\pi + C_L)} + \frac{1}{j\omega C_L}(-j|\beta(f)|), \qquad (4.19)$$

where frequencies have been assumed to be far enough above f_β in the approximation for the base-emitter impedance to be dominated by C_π. Further simplification provides:

$$Z_{in} \cong R_B - |\beta(f)|\left(\frac{1}{\omega C_L}\right) - j\frac{1}{\omega(C_\pi + C_L)}, \qquad (4.20)$$

[36] Pardoning the plurality having stated this as two conditions to be met, this condition for oscillation is sometimes referred to as the Barkhausen criterion.

for which the real part is negative when $|\beta(f)| > \omega R_B C_L$. The presence of a negative resistance looking into a terminal is equivalent to the condition $|\Gamma| > 1$; power can be supplied at the input by the transistor being used as a one-port element, creating the possibility for an oscillation to be sustained.[37]

If there is a moral to this story, it is that capacitance in the emitter of a transistor can result in unstable operation, particularly when the loading presents a fairly high impedance. This type of loading is exactly what occurs inside a cascode, where the common-base transistor sits atop the collector-substrate and base-collector capacitances of the common-emitter device.[38] Consequently, negative resistance is generated at the (intrinsic) base of the cascoding device wherein it "touches" the output through a base-collector capacitance. Instability with the cascode can thus result when an inductive (resonant) load is appended. While there may be enough loss in the load inductor to mitigate any stability issues, another possibility for addressing stability is availed in pondering the layout geometry of the common-base transistor. Keeping in mind that the collector current has already been established in relation to the emitter area of the input transistor, the design affords the freedom to avoid negative resistance by sizing the common-base transistor as suggested in Equation 4.20. Stability in the cascode stage can be improved by increasing the ratio $R_B/|\beta(f)|$, although the direction in which to move the device size depends on where along the f_T versus J_C (collector current density) curve the transistor currently resides. A larger transistor tends to reduce both base resistance and f_T for a given bias current—but not necessarily by the same factor. The emitter width could also be increased for a chosen emitter area; base resistance will increase while f_T would be largely unchanged. Designing with these layout parameters tends to be an exercise in balancing among a number of competing effects, but a reasonable improvement in stability can generally be achieved at a small expenditure in noise figure and available gain.

This glimpse at stability in transistor and amplifier stages has been intended as an introduction to some microwave circuit design concepts and to show how these measures are consistent with notions of stability in other realms. Feedback paths within a device alone suffice for creating potential instability when conspiring impedances are affixed at the input or output. The presence of these destabilizing feedback paths is owed to transis-

[37] Though it results more from the design approach than anything topological, there is a class of microwave circuits possessing this characteristic that often goes by the moniker "one-port negative resistance oscillators".

[38] The output resistance of the common-emitter transistor also loads the cascode device, but as this resistance is generally high-valued in bipolar transistors, it is easily overwhelmed at microwave frequencies by capacitances loading the collector.

tor limitations and, as such, can be influenced by device sizing and process design. Capacitance in the emitter circuit of a bipolar transistor would not lead to difficulties except for a phase shift in the input signal through C_π; similarly, it is due to resonant behavior with C_μ that stability concerns arise when the base terminal is driven through an inductance.[39] Hence, even in designs where impedance matching is employed, minimization of intrinsic and extrinsic transistor capacitances becomes important in the extraction of innate device performance for high frequency operation.

4.5 Lessons in Low Noise

While there may not be a single electronic component that is ever required to "do it all", the circuit that comes the closest might be the low-noise amplifier in portable RF receivers. These amplifiers are required to provide gain and a reasonable degree of linearity, to deliver low noise with minimal power consumption, and to operate at high frequencies while still rendering a significant degree of isolation. In perusing these competing demands and delving into some LNA design concepts, the common-emitter stage with base and emitter input-matching inductors is found to perform well in a surprisingly large number of areas. Coupled with its relative ease of design, this broad appeal has made the common-emitter device a difficult champion to unseat as the LNA input stage in narrowband RF applications.

As the primary concern in LNAs, noise is minimized in bipolar transistors at low bias current densities, before shot noise contributions begin to outweigh thermal noise with increasing current levels. Operating at these characteristically low current densities, the base transit time plays only a secondary role in determining the minimum noise figure (NF_{min}), the associated optimum current density ($J_{C_{opt}}$), and the resulting transistor f_T; it is the base-collector and base-emitter junction capacitances that prove crucial in LNAs. Noise performance becomes squeezed—and in which an optimum is found—between thermal noise (largely) from the base resistance and a collector shot noise component that increases with frequency as the transistor current gain falls. As a result of heavy doping in the emitter (i.e. $N_{D_E} \gg N_{D_C}$), C_{jE} is usually much larger than C_μ, and $\beta(f)$ is primarily limited by the base-emitter junction capacitance at low-noise current densities. Given the predominance of C_{jE}, true HBTs can provide an advantage over homojunction and graded-base transistors for LNA applications.[40] A valence band discontinuity can be engineered

[39] Field-effect transistors are not immune either; the same effects act upon the gate-to-source and gate-to-drain capacitances of FETs.

into an NPN HBT at the base-emitter junction, allowing a reduction in C_{jE} via lower emitter doping levels.[41] But all of this emphasis on the role played by C_{jE} should not relegate C_μ to wallflower status; the base-collector capacitance also imposes a number of important limitations in high frequency amplifier design—even in those intended for narrowband application. Issues with stability, isolation, and reduced gain have all been traced to C_μ, not to mention a correlation reactance that prevents a simultaneous noise and power match at the input of an LNA device. The capacitance of both intrinsic junctions hence proves to be quite important indeed.

Unfortunately, the trend in modern bipolar devices is toward thinner bases to reduce the base transit time, a move which actually begins to work *against* LNA noise performance and power consumption. After scaling the base thickness (downward), the doping concentration is generally increased to compensate for the higher base sheet resistivity resulting from the thinner layer. Then, an increase in the collector doping usually follows, lest pronounced base pushout (i.e. the Kirk effect) weigh in and spoil the higher peak unity current-gain frequencies being sought through the reduction in τ_F. Thinner bases thus result in higher per unit area junction capacitances, imposing quite a tax in a regime where base resistance and junction capacitances are the limiting factors. While perhaps contrary to conventional thinking, the real improvement in noise figure that has been observed in advanced silicon-based bipolar technologies is owed to the continued shrinking in lithographic (i.e. lateral) dimensions. Narrower emitter stripes reduce both base resistance and junction area, effectively overcoming some of the penalties associated with recent directions in vertical stack design.

This chapter has focused on realizing high-performance LNAs with bipolar transistors. As technologies scale, CMOS designs are demonstrating good results at increasingly high frequencies. Intrinsically, however, bipolar devices feature a higher transconductance for a given bias current [61], and therefore deliver more gain and lower noise while consuming less power than CMOS—provided that the same generation of lithographic tools is available for both technologies. Despite this caveat, more will be said about CMOS implementations in the following chapter.

---

[40] Many of the SiGe bipolar transistors found today are actually of the graded-base variety in which there is no heterojunction at the base-emitter interface.

[41] With the valence band discontinuity creating a barrier inhibiting the flow of holes from the base to the emitter, the emitter doping can be reduced to concentrations even below that found in the base.

Appendix: Two-port Stability Circles

Stability circles prove to be a handy concept in the design of transistor stages used as amplifiers and oscillators. Grounded in two-port network theory, stability circles consist of the collection of impedances which, when presented as a termination on one port, result in a unity reflection coefficient at the other port. Many good texts on microwave amplifier design provide derivations of stability circles through the algebraic manipulation of complex numbers;[42] though these derivations are not repeated here, the results are summarized in this Appendix.

First solving for the output reflection coefficient of a two-port network having an arbitrary termination at the source port, and then equating the magnitudes of the numerator and denominator, a circle is found in the source reflection coefficient plane centered at:

$$\Gamma_{ssb} = \frac{s_{22}\Delta^* - s_{11}^*}{|\Delta|^2 - |s_{11}|^2}, \tag{4.21}$$

and described by a radius:

$$\rho_{ssb} = \frac{|s_{12}s_{21}|}{|\Delta|^2 - |s_{11}|^2}. \tag{4.22}$$

In both expressions, Δ represents the determinant of the two-port matrix representation:

$$\Delta = s_{11}s_{22} - s_{12}s_{21}. \tag{4.23}$$

Plotted on a Smith chart, this circle defines the range of source impedances which may lead to instability in the transistor stage. Load stability circles are similarly defined, and can be calculated by interchanging s_{11} and s_{22} in Equations 4.21 and 4.22.

There are seemingly many other measures of stability, often scalar metrics, encountered in microwave circles. Some have names (Linville and Stern), and others are simply known by designations such as "k", "B", and more recently "μ". These figures of merit have been formulated to provide a simple numerical test for stability and to indicate a degree of stability in much the same way that a phase margin is used. However, as these alternatives are also rooted in two-port representations, they fundamentally offer no information beyond that exhibited in a graphical analysis via stability circles.

---

[42] An excellent work by Carson [60] carries out derivations using both generalized scattering parameters and admittance parameters. Admittances (y-parameters) might present a more comfortable notation for some readers.

Chapter 5

An Exercise in Designing Low-Noise Amplifiers

While the preceding chapter investigated the fundamental limitations confounding the realization of high frequency, low-noise amplifiers, this chapter offers a look at how closely a couple of LNA designs can approach these technological limitations within the purview of some added constraints. For receivers characteristic of high data rate wireless networks, the LNA must provide a very low noise figure and reasonable gain to render the sensitivity required to support higher-order digital modulation formats.[1] A 3dB increase in the signal to noise ratio reaching the detector following the RF front end allows the data rate to be doubled for a given available bandwidth; used another way, the same 3dB enhancement could also lower the bit error rate by a factor of 100 to 1000 for a chosen data rate [62].[2] Hence, even seemingly incremental improvements in receiver sensitivity can dramatically enrich the performance of a wireless link when transmit power is limited.

Presented in this chapter are two LNAs which target a 150MHz band, centered at 5.8GHz, that has been set aside under the auspices of the unlicensed (wireless) national information infrastructure (U-NII) [1]. The principal focus of these designs is to achieve the lowest possible noise figure attainable with a 0.5μm BiCMOS technology possessing SiGe graded-base NPNs, and to compare the trade-offs between bipolar and CMOS LNA implementations when the transistors are afforded the same lithographic dimensions and accompanying passive devices. Another key feature is to architect some means of reducing the power dissipated in the LNA when maximum receiver sensitivity is not required, and to do so in such a fashion that does not detract from the attainable sensitivity when it is required. The challenge is to allow gain and noise figure to be dynamically traded for lower power consumption as the demand for bandwidth and the operating environment change, but to maintain reasonable impedance matches and isolation under all conditions. A fully-integrated solution is the goal, using on-chip spiral inductors for sim-

[1] "Higher-order" in this case refers to a signalling scheme that encodes multiple bits per symbol.

[2] The bit error rate is a nonlinear function of the signal to noise ratio (SNR). Bit errors are reduced more abruptly at higher SNRs.

plicity and compactness, and incorporating the effect of pads (for wirebonds or RF probes) at the input and output.[3] Biasing circuitry will prove to be a critical component of the design, both in enabling adaptability and to the extent that it may hinder the RF performance of the amplifier. Each of these considerations will be explored and, where possible, the impact upon LNA parameters such as gain, noise figure, and linearity will be assessed. Results from both simulations and measurements will be presented, providing a basis for commentary on devices and technology for wireless applications.

5.1 A Bipolar 5.8GHz LNA Design

As illustrated in the previous chapter, a fairly well-known process exists for tuning a common-emitter stage to extract the optimum noise performance available from a technology while providing an input (impedance) match over a narrow frequency band. Following this procedure, and using the spiral inductor modeling techniques discussed in Chapter 2, the stage shown in Figure 5-1 forms the starting point for this 5.8GHz LNA design. A cascode was chosen to improve the isolation, to achieve a higher gain (albeit

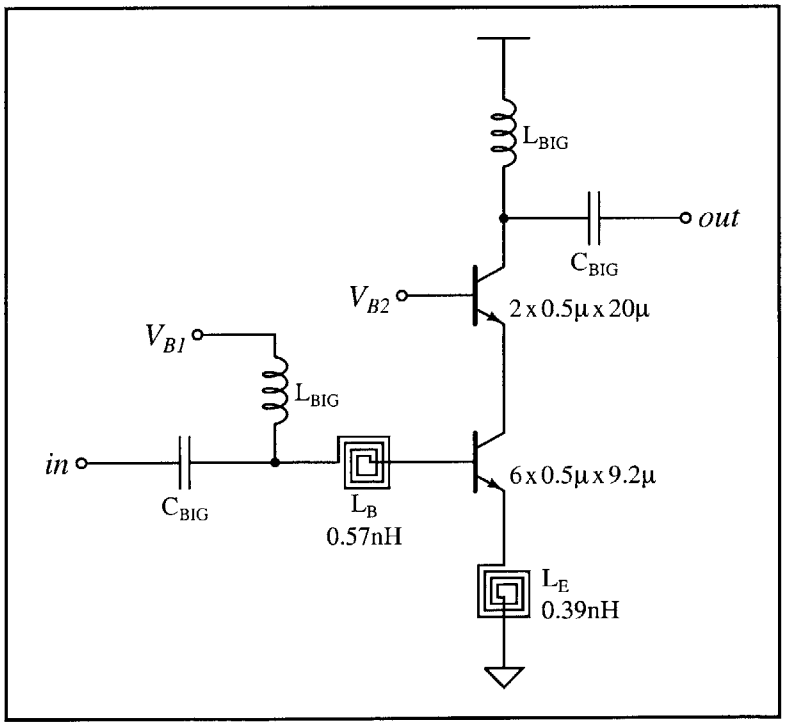

Figure 5-1. Cascode stage tuned for minimum noise at 5.8GHz.

[3] Although some LNA designs (and most transistor test cells) that are characterized on wafer have the loading of the pads de-embedded from the measured response, this luxury can not be afforded to an amplifier that is designed to interface with off-chip signals and components.

only marginally higher), and—for reasons that will become clear later—to increase the output resistance. The input transistor is realized as a six emitter device so that the collector region can be made approximately square in layout, a desirable trait since a square geometry minimizes the collector-substrate capacitance (C_{jS}) for a given emitter area.[4] The optimum source resistance ($R_{s_{opt}}$) for this stage is made equal to 50Ω by choosing 27.6μm^2 for the emitter area, the last 8% of which has been included to compensate for the loss expected in the base and emitter inductors.[5] With the collector current density needed to minimize noise at 5.8GHz coming in at a shade over 0.1 mA/μm^2 for this 0.5μm technology, a bias current of 2.8mA establishes $J_{C_{opt}}$. Input matching is completed by adding spiral inductors sized to provide 0.57nH and 0.39nH, respectively, in the base and emitter, for which the scalable models (Chapter 2) suggest Qs of 19 should be possible.[6] Topping off this stage is a cascode device, where the size reflects a balance of noise, gain, and stability. Minimum-width emitters are used to keep base resistance low, and the slight negative resistance that results at the collector helps cancel loss in the output matching network. While this two emitter device will have a rather non-square geometry, its C_{jS} can be tuned out. Rather than trying to reduce this capacitance, a geometry covered by the scalable transistor model is chosen instead.[7]

5.1.1 Output Matching Network Design

A cascode stage with an input impedance tuned for low noise at 5.8GHz has been realized; the next piece to be added is a network that transforms the output impedance of the cascode to match whatever loading will be imposed on the LNA. Here—and although this need not be the case in an integrated receiver where the load might be the input of an on-chip mixer—a 50Ω load is the target.[8] The basic idea embodied by the matching network is to allow power delivery to the load while maximizing the impedance presented to

---

[4] A square geometry seeks to balance the area (vertical) and perimeter (sidewall) contributions to an implanted region. While the optimal point resides along a fairly broad minimum, the savings can be rather substantial. In some technologies, the same approach can also be taken to reducing the base-collector capacitance for a given emitter area. However, when the bipolar devices employ base windows isolated by shallow trench or LOCOS oxidation, or when mesa structures (e.g., selective epitaxial growth) form the transistors, the perimeter capacitance is negligible.

[5] Thermal noise from loss mechanisms within the spirals contributes to the equivalent input voltage noise of the LNA stage, increasing the optimum source resistance above the desired 50Ω value. A larger input transistor can be used to compensate (Section 4.3.3).

[6] These quality factors are estimated at 5.8GHz. The spiral inductor models will, unfortunately, prove to be decidedly optimistic. However, the results shown including these models will yield a good demonstration of what should be possible with best-in-class inductors.

[7] This choice reflects a design philosophy where it is deemed best to reduce uncertainty in places where there is little to be gained by doing otherwise. The scalable model set being used has only been verified for one and two emitter devices of up to 20μm in (emitter) length.

the collector of the cascode device so as to maximize the gain of the LNA. Conceptually, given ideal inductors to resonate the base-collector (C_μ) and collector-substrate capacitances, a voltage gain—measured from the intrinsic base of the common-emitter transistor—of $(g_m r_o)^2/2$ could be achieved. This analysis ignores many non-idealities: losses in terminal resistances, the voltage dropped across the emitter inductor, a current gain between the two transistors that is not quite unity, and issues with stability; what is clear, however, is that adding shunt inductance to create a parallel resonance at the collector is desirable. Further amenities include incorporating a DC-blocking capacitor into the matching network (always a nice touch), and being able to accommodate the capacitance of the output pad in the impedance match.

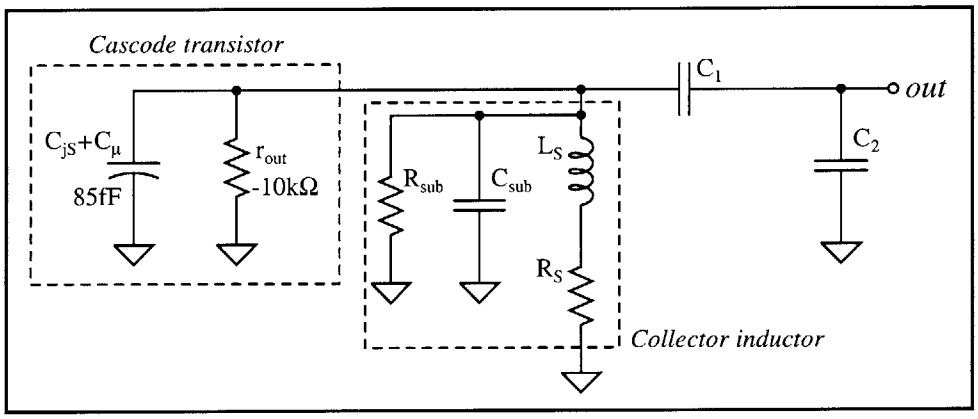

Figure 5-2. Design of the output matching network for the cascode LNA stage.

From these considerations, the topology that naturally falls out is the tapped-capacitor resonator as shown above in Figure 5-2 [63]. The impedance looking into the cascode stage is modeled as a parallel RC, where the capacitance is $C_{jS} + C_\mu$ and the output resistance is negative due to the capacitive loading in the emitter of the common-base transistor. Sizing the collector inductor appropriately yields the desired resonance, although the inductance is accompanied by substrate parasitics (indicated as R_{sub} and C_{sub}) and winding losses (R_s) which will limit the maximum impedance (and hence the LNA gain) that can be realized. Capacitors C_1 and C_2 implement a voltage divider which transforms the output impedance by the factor $(1+C_2/C_1)^2$; C_1 also serves as a DC block, while C_2 may be composed in part by the capacitance of the output pad. As often seems to be the case when spiral inductors are involved, an iterative design procedure

[8] In fact, if the mixer input essentially serves as nothing more than a transconductor, an explicit output load may not be needed. However, it may still be beneficial to resonate any capacitive loading on the collector so that all of the output current from the LNA reaches the mixer switching core.

results where the chosen inductor affects the output resistance and capacitance being matched to the load. The best approach is usually to begin with an estimate of the equivalent resistance associated with loss in the inductor, allowing a target inductance to be chosen on the basis of the desired center frequency and Q (bandwidth) for the impedance match. A spiral geometry can be then designed and modeled, enabling further refinements in the matching network.

In terms of the inductor model parameters from Chapter 2 discussed in association with the substrate (C_{ox}, C_{Si}, and R_{Si}) and the interwinding capacitance (C_s), when frequencies are at least reasonably high ($\omega \gg [R_{Si}(C_{ox}+C_{Si})]^{-1}$), an equivalent parallel RC model suffices to represent the loading:

$$R_{sub} \cong R_{Si}\left(1+\frac{C_{Si}}{C_{ox}}\right)^2, \tag{5.1a}$$

$$C_{sub} \cong \left(C_s + \frac{C_{Si}C_{ox}}{C_{Si}+C_{ox}}\right). \tag{5.1b}$$

The series components of the inductor can also be reworked into a parallel equivalent circuit (which holds for a narrow range of frequencies about a chosen ω_0). As long as the inductor Q is not too low (i.e. $\omega_0 L_s \gg R_s$), the inductance is unchanged in the series-to-parallel transformation, while the equivalent parallel resistance associated with the winding loss is:

$$R_p \cong R_s\left(\frac{\omega_0 L_s}{R_s}\right)^2. \tag{5.2}$$

Collectively, these terms render an equivalent output resistance of the cascode stage given by the parallel combination of three resistances:

$$R_{EQ} = \left(\frac{1}{R_p}+\frac{1}{R_{sub}}+\frac{1}{r_{out}}\right)^{-1}. \tag{5.3}$$

Now availed of this output resistance, and adapting from Lee [63], the fractional bandwidth of the output matching network can be expressed as $Q = R_{EQ}/(\omega_0 L_s)$. As an intermediate result, the Q of the load-side RC network (the parallel combination of C_2 and the 50Ω load designated as R_L) can be determined:

$$Q_L \equiv \omega_0 R_L C_2 = \sqrt{\frac{R_L}{R_{EQ}}(Q^2+1)-1}, \tag{5.4}$$

allowing the capacitances in the matching network to be calculated:

$$C_2 = \frac{Q_L}{\omega_0 R_L}, \tag{5.5a}$$

$$C_1 = \frac{C_2(Q_L^2+1)}{QQ_L - Q_L^2}. \tag{5.5b}$$

Notably absent among these matching element calculations is the capacitance being tuned out (i.e. $C_{jS}+C_{\mu}+C_{sub}$), the presence of which was one of the reasons behind having a matching network in the first place. A good initial cut at including this capacitance is simply to increase the center frequency used in Equations 5.4 and 5.5 by the ratio:

$$\frac{\omega_0'}{\omega_0} = \sqrt{1+\frac{C_{jS}+C_{\mu}+C_{sub}}{C_1C_2/(C_1+C_2)}}, \tag{5.6}$$

reflecting the fraction of the total network capacitance that lies at the collector node.[9] Here again, the solution is achieved iteratively as C_{sub} is a component of the inductor; an initial pass reveals that inductances in the 2-4nH range yield reasonable capacitance values, where the only unwavering constraint is that C_2 must be large enough to absorb the output pad. Choosing and modeling a spiral inductor of 2.7nH gives a quality factor of 17, numbers which furnish capacitances of 0.18pF and 0.54pF to complete the matching network sketched in Figure 5-3. Some adjustments will be made in the indicated values as further additions and refinements are made, but at last the design is beginning to resemble an impedance-matched low-noise amplifier suitable for usage at 5.8GHz.

In getting to this point, a common-emitter device was first sized so that the minimum transistor noise figure is obtained with a 50Ω source resistance. Base and emitter inductors were then added for matching, followed by a cascode device for improved isolation and gain. At this stage, a model of inductor loss was introduced and an impedance transformation network was affixed at the output. Each of these steps imposes an additional limitation on the ultimate sensitivity within reach of a technology, and may also exact a penalty in the power consumption associated with this (or any given) level of sensitivity. A survey of the performance implications is arrayed in Figure 5-4, where the dashed lines represent the baseline metrics of a common-emitter device.[10] At 5.8GHz, the NF_{min} posted by the 0.5μm SiGe bipolar technology chosen for this work is 1.5dB, a number which is a function of the current density and not the emitter area.[11] Setting the transistor size such that Γ_{opt} sits on the 50Ω circle, the 50Ω noise figure (i.e. without

[9] Often, this first-order approximation will be good enough at this point in the design. The final optimization is best performed through simulation so that models of the parasitics associated with the capacitors and the pads can easily be incorporated.

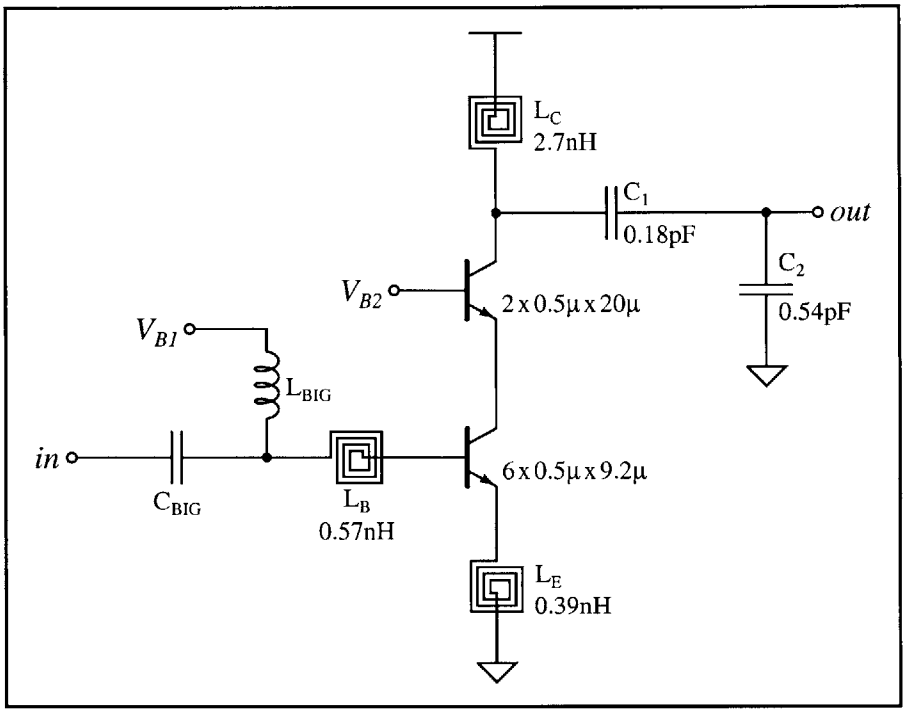

Figure 5-3. Cascode LNA gain stage tuned for 5.8GHz.

using an input matching network) is observed to be several tenths of a dB higher at 1.85dB. The other half of the picture—the "A" part of LNA—can be examined through the available gain (G_a) and the maximum available gain (G_{ma}) obtainable from the amplifier.[12] Much like NF_{min}, the maximum available gain is independent of transistor impedance levels and therefore provides a point of comparison that does not depend on the size of the device. Available gain then is to G_{ma} as (the 50Ω) noise figure is to NF_{min}—the load impedance is the only variable, taking on the value required to maximize the gain when the source impedance is taken as being fixed. For the same common-emitter transistor having Γ_{opt} on the 50Ω circle at 5.8GHz, the maximum available gain is 15.5dB, with 13.5dB being available when a 50Ω source impedance is provided.

[10] These baseline measurements of a common-emitter transistor represent figures-of-merit that are often cited in the description of a technology, and particularly for those technologies with RF aspirations. NF_{min} is a useful point of comparison between devices, as is the frequency where G_{ma} falls to unity—a frequency commonly known as f_{MAX}. A little caution with this latter "metric" should be heeded, however, as it is sometimes defined by either extrapolating the (maximum) unilateral gain or the maximum stable gain, both of which can lead to higher "f_{MAX}" numbers.

[11] This is not to say that NF_{min} is entirely independent of transistor geometry. The base resistance and the base transit time (to an extent) both vary with the emitter width.

[12] Rigorously, the maximum available gain is only defined when the transistor or amplifier being characterized is unconditionally stable. However, it is customary to instead show the maximum stable gain (G_{ms}) for frequencies where G_{ma} is undefined.

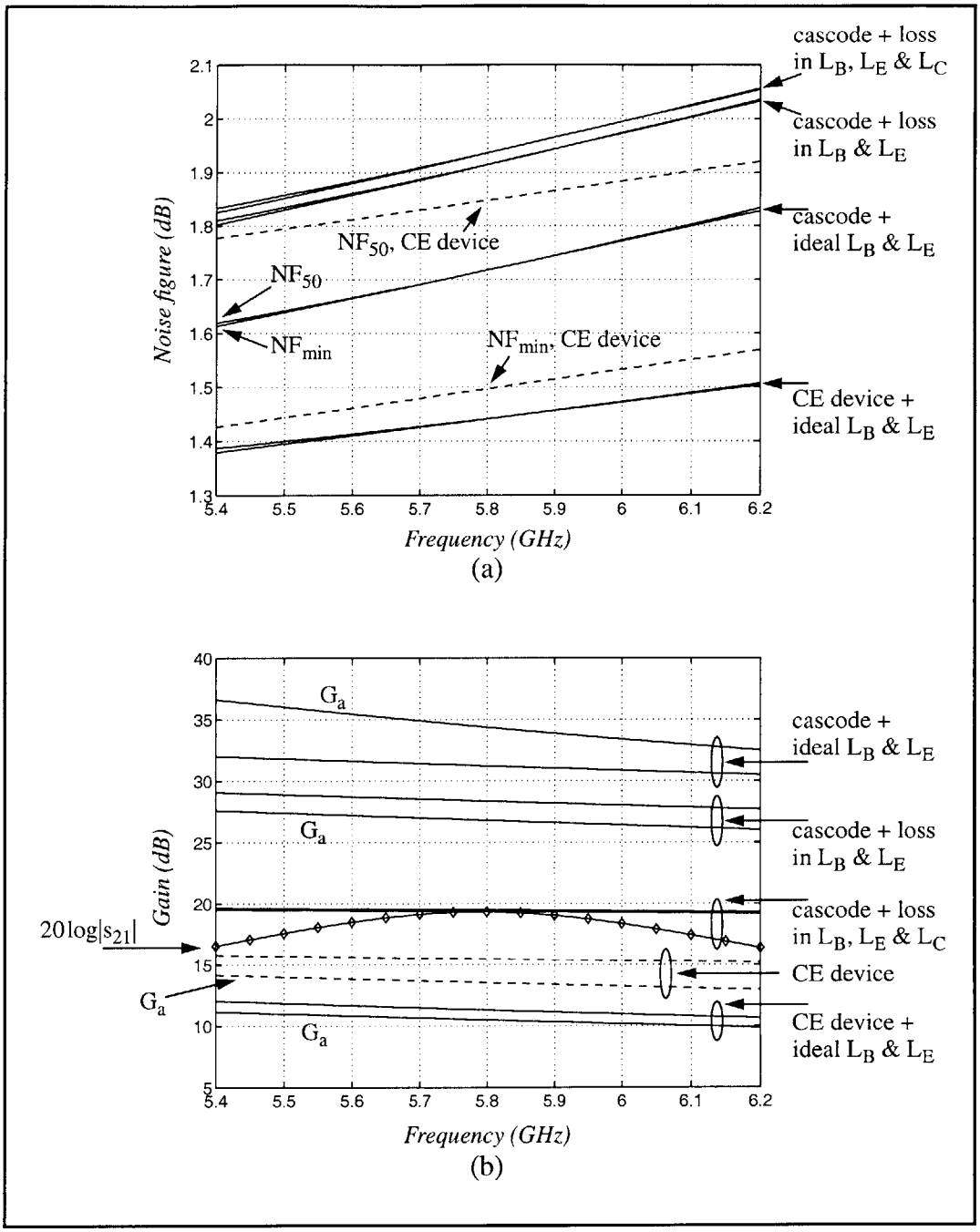

Figure 5-4. Evolution of noise figure (a) and gain (b) in a cascode LNA stage as inductor loss (Qs of 17-19) is included. In (a), the upper trace of each pair is the 50Ω noise figure (NF_{50}) and the lower is the minimum noise figure (NF_{min}). In (b), each set of traces represents the maximum available gain (G_{ma}) and the available gain (G_a) for the condition listed to the right, where G_a is labeled when it differs appreciably from G_{ma}. By convention, the maximum stable gain (G_{ms}) is shown when G_{ma} is undefined. Marked with the diamonds is $20\log|s_{21}|$ for the matched stage when a 50Ω impedance is provided by both the source and load.

Next, adding in ideal base and emitter tuning inductors closes the noise figure to NF_{min}, reflecting the fact that the optimum source impedance (for minimum noise) is now 50Ω. As this is, of course, a narrowband match, the noise figure can be seen to deviate from its minimum possible value for frequencies away from the 5.8GHz design target. The small *reduction* in noise figure (by 0.06dB) is owed to the degenerative effect of the emitter inductance, an effect further evinced by the concomitant 4dB drop in gain (G_{ma}). Notably lower still is the available gain, falling 0.8dB below G_{ma}. A difference like this between G_a and G_{ma} indicates that some signal is being lost at the input, which is to say that the input impedance is not precisely matched to the source. That a simultaneous match for minimum noise and optimum power transfer would not be achieved was surmised in Section 4.1, as correlation between the equivalent input noise generators gives rise to an additional reactive component in the noise-minimizing source impedance. For this LNA, the decision has been made to realize the best possible noise figure, a choice which will limit the input return loss and give up approximately 1dB in gain due to the resulting mismatch.

With the base and emitter inductors—ideal versions at that—leaving only 10.5dB of available gain, a cascode structure would seem to be an imperative. The addition of a common-base transistor to the stack, after recalibrating the inductances slightly to return Γ_{opt} to the center of the Smith chart, boosts the (maximum stable) gain by 20dB, albeit at a cost of almost 0.3dB in noise figure. In this case, the available gain is greater than G_{ms}, implying that the load impedance needed for optimum gain from the cascode falls outside the stable region. This aside, having an LNA with 31dB of gain at 5.8GHz would be nice; unfortunately, such numbers will merely be distant memories once even moderate inductor loss is considered. The effects of the series resistance and substrate parasitics associated with the spirals are demonstrated in two steps: first looking at just the input side and then adding in loss from the output matching inductance. Sustaining an additional 0.2dB hit is the noise figure, where most of the depreciation is observed to result from the base and emitter inductors with Qs of 19. Precipitous drops in gain are seen at both steps, with G_a falling to 27dB due to loss in the input network alone, then plummeting another 7.5dB once the expected Q of 17 is instilled in the collector inductor. Sadly, while the extra 20dB of gain made available by cascoding seems impressive, more than 10dB is given right back by way of inductor loss.[13] Concluding the picture, and shown with the diamond-marked trace in Figure 5-4(b), is the simulated s_{21} (in dB) for the LNA gain

[13] Keep in mind that these inductor quality factors are about as high as can be realized in an integrated technology; the actual inductors in this work will fall far short of these lofty expectations.

stage when the output is drawn from the tapped capacitor (as indicated in Figure 5-3). Although the available gain has been used throughout this example, the actual insertion gain of the LNA placed in a 50Ω network should be able to approach G_a at the design center frequency.[14] As shown, with the output matching network transforming the collector impedance to 50Ω, the realized gain of the amplifier touches G_a at the band center.

Among the sea of data just presented, the benchmarks to keep in mind are the 1.9dB noise figure along with a gain of about 19dB, accomplished with 2.8mA of bias current and on-chip spiral inductors having quality factors in the high teens (17-19). This furnishes a slightly more pragmatic view of what should be possible at 5.8GHz with a fully-integrated solution in a 0.5μm silicon-based bipolar technology. Of course, the LNA is not yet complete. First, these numbers have been gauged in simulation using ideal bias sources. Then, given the significance of inductor loss, what will be the effect of loss in the pads and the other passive components? And, perhaps of greater consequence is that this design implies a fixed bias point; how do we now think about saving power when the peak levels of performance are not required? At the least, however, the noise figure, gain, and current consumption numbers provide useful reference marks for evaluating the gravity of each additional issue being faced as well as the efficacy of proposed work-arounds.

5.1.2 Operating at Reduced Power Consumption Levels

If there is one defining challenge of wideband, wireless networks, it may be in accommodating a wide dynamic range in data rate and signal power. The highest levels of sensitivity may not always be demanded of the receiver, but robust system operation requires that a high level of isolation and reasonable input and output matches be maintained by the front-end LNA regardless of the incoming SNR. At the same time, however, being able to trade away some sensitivity for reduced power consumption will become an essential feature in wireless connectivity, and finding ways to implement this adaptability will be a focal point of design innovation. Of all the ideas that may be envisioned for providing a low power "mode" of operation, the simplest concept would be to just turn off the LNA and bypass it with a switch when the received SNR is high. Unfortunately, a switch would also provide unfettered access to the antenna for the many tones and frequency products generated in the subsequent mixer stage, a prospect which dismisses the bypass switch idea rather quickly. But even were isolation not an issue, another repercussion of using a switch is that such an "all-or-nothing" solution would probably be far from opti-

[14] Insertion gain is simply the measured s_{21}^2, but will equal the transducer gain (G_t) when the source and load impedances are Z_0—the characteristic impedance of the system.

mal. Instead, offering some power savings for conditions of moderate SNR—where some LNA gain is still needed—will likely be a more useful scenario.

Very little has been said to this point regarding the supply voltage for low-noise amplifiers. The reason for this treatment is that there is very little to say. As long as the supply voltage affords sufficient headroom for RF signal swing given the impedances at the collector node, it does not have a material effect on the performance of the LNA; with no improvement to be had for higher voltages, the amplifier should already be operating at the minimum which allows the required swing.[15] While the supply voltage may only wield limited leverage, the bias current in the cascode remains a possible handle by which to trade power and performance. Cutting the current clearly reduces power at the expense of gain and noise figure, but doing so also rather dramatically affects the input impedance.[16] Both the resistive portion (dominated by $\omega_T L_E$ as discussed in Section 4.1) and the imaginary component (largely C_π) change significantly with collector current, taking the desired input match along with them.[17] With such a high degree of sensitivity in the impedance, it becomes evident that trying to save power merely by lowering the bias current may not be the best of options.

Having seemingly ruled out all other possibilities, designing a parallel LNA stage—optimized for lower power operation—suddenly seems a much more attractive alternative. This stage can sacrifice an optimum noise match for the sake of operating at a reduced bias current, but still must provide 50Ω input and output impedances, and should minimally impact the "high performance" mode evolved in Figure 5-4.[18] While using an LNA with parallel stages necessitates choosing between them based upon the received SNR and the required data rate, a topology that avoids passing the RF signal through a switch would be a preferable solution, as the presence of an explicit switch is a harbinger of sub-optimal performance. Placed at the input, an RF switch adds loss and increases noise figure; when found at the output—as in many adjustable gain LNAs—power is

[15] Increasing V_{CE} may lead to slightly larger gains, that is, unless the collector regions have already been fully depleted.

[16] The transistor output resistances also change, but variation here will be less of a concern than the base-emitter impedance because the output impedance of the LNA is primarily limited (in magnitude) by loss in the collector inductor.

[17] It is at least mildly conceivable that some type of impedance stabilization circuit could be implemented to maintain an input match as the bias current is adjusted. However, any such compensation scheme is likely to directly impact the noise figure (through loss) and may also involve a fairly hefty power consumption overhead.

[18] Some latitude is begged for the "high performance" moniker in that this is meant as a relative description; to avoid passing judgment in comparison with discrete implementations or other technologies, it might more accurately be called the "peak performance" mode.

being dissipated to generate signal that is only being shunted away. Conversely, having a "switched LNA" without a switch may sound like a conundrum, but will prove quite feasible by using the signal path transistors to perform the switching function.

In the investigation of alternative LNA topologies (Section 4.3), it was argued that a common-base input offered no fundamental performance advantage over the common-emitter design. But, instead of noise and gain, if the question is to find the smallest device that can yield an input match, then the grounded-base transistor has something to offer. Granted, an emitter inductance was needed with the common-emitter device to increase the impedance looking into the first stage. However, this was for a relatively large transistor biased at several milliamps of current; for smaller geometries, base resistance alone can exceed 50Ω, and the fact that the input impedance is a factor of $1+\beta(f)$ lower when the same device is used in the common-base configuration is amenable to the plight of needing a reduced power mode. If the second stage input transistor can be made small enough, it can be wired directly to the base of the common-emitter device and bleed off little of the incoming RF signal power when the high performance mode is desired.

Recalling from the previous chapter, the input impedance of a common-base transistor includes the terminal resistances as well as the small-signal base-emitter impedance represented by the transconductance:

$$Z_{in} = R_E + \frac{1}{g_m}\left(\frac{\beta(f)}{1+\beta(f)}\right) + \frac{R_B}{1+\beta(f)}. \tag{5.7}$$

Conversely, when the device is switched off, the loading imposed at the RF input by this transistor is related to its base-emitter junction capacitance (C_{jE}). This sets up a trade-off in sizing the common-base input device: a larger emitter area will drag down the gain and noise performance in the high sensitivity setting, but lowers the bias current required to achieve a 50Ω input impedance in the reduced power mode. As a first step in examining this quandary, model data shows an almost minimum-sized 0.5μm by 2μm emitter geometry to present a zero-bias C_{jE} of 7.5fF and 15Ω of emitter resistance. When biased at a 0.75mA collector current level, the nominal values for base resistance and unity current-gain frequency are observed to be 176Ω and 56GHz, respectively.[19] With such large terminal resistances leaving only 20Ω for $1/g_m$, a bias current of 1.3mA is needed to achieve an input match. While this would allow power consumption to be cut by more than a factor of two, an even larger reduction may be within reach. Bumping the emitter length to

[19] The 0.75mA was chosen as a starting point for this example as being greater than the 0.5mA required for the $1/g_m$ term to yield 50Ω, and enough smaller than 2.8mA (as used in the high performance stage) to make this effort worthwhile.

5μm scales R_B in proportion to the emitter periphery (down to 82Ω) and R_E inversely with the emitter area (now 6Ω), doing so at the cost of a junction capacitance that increases with emitter area (up to 19fF).[20] Possessing a still considerable f_T running in the neighborhood of 43GHz, a 50Ω input resistance can be had with this 0.5μm by 5μm transistor for a mere 0.7mA of collector current. To go further in diminishing power dissipation would entail disproportionately larger transistor sizes as the terminal resistances asymptotically approach zero.[21] Thus, the 19fF C_{jE} for the 5μm emitter (length), a number which represents less than 4% of the total LNA input capacitance in the high performance mode, is taken as being a suitable compromise between power and performance.

The next step in the design is to cascode the common-base stage, doing so for all the same reasons expressed earlier: improved isolation, increased available gain, and a higher output resistance.[22] Finally, connecting the collectors of the cascode devices and coupling the inputs of the stages together through the capacitor C_C results in the LNA topology shown in Figure 5-5. Inductor L_{E2} resonates the input capacitance of the grounded-base transistor to deliver a 50Ω match, and L_{DC} provides a DC path for the bias current in this stage. Becoming more apparent with the paralleling of the stages is why the increased output resistance obtained by cascoding is so important: for precisely the reason of being able to ignore it. As long as it is high (in magnitude) relative to the equivalent resistance representing loss in the collector inductor, the output impedance of the LNA will be the same in both the high performance and low power modes—a tenet which greatly simplifies the output matching network that will be required.[23] Furthermore, the importance of isolation is magnified in this architecture as the inactive stage provides a feedback path around the operative stage. In the high performance mode, the feedback through the common-base stage is negative and tends to inhibit the gain. For the low power stage the feedback is positive; while this may tend to augment the gain, stability

[20] Base resistance may be a fairly strong function of the collector current density, a detail that is currently being ignored in this first-order scaling analysis.

[21] Another side-effect of continuing to increase the common-base input transistor size is that its C_π begins to be noticed in the input impedance. This visibility requires more series inductance to tune it out, thereby inflating the spiral inductor loss (and hence noise figure) in *both* modes.

[22] The 50Ω source impedance shunts the emitter inductance from the perspective of the output impedance, which is why the second common-base transistor still provides a measure of improvement.

[23] As noted in Figure 5-2, capacitive loading from the cascode transistors is also an important component of the output matching network. However, the collector-substrate capacitance (C_{jS}) will not change as the LNA mode is toggled, and C_μ will not deviate too significantly as the base-collector junctions of the cascode transistors will see a fairly strong reverse-bias condition regardless of whether the stage is "on" or "off". If the change in C_μ was to present a problem in the output impedance, the two cascode devices could always be designed to match (perhaps at the cost of a little gain).

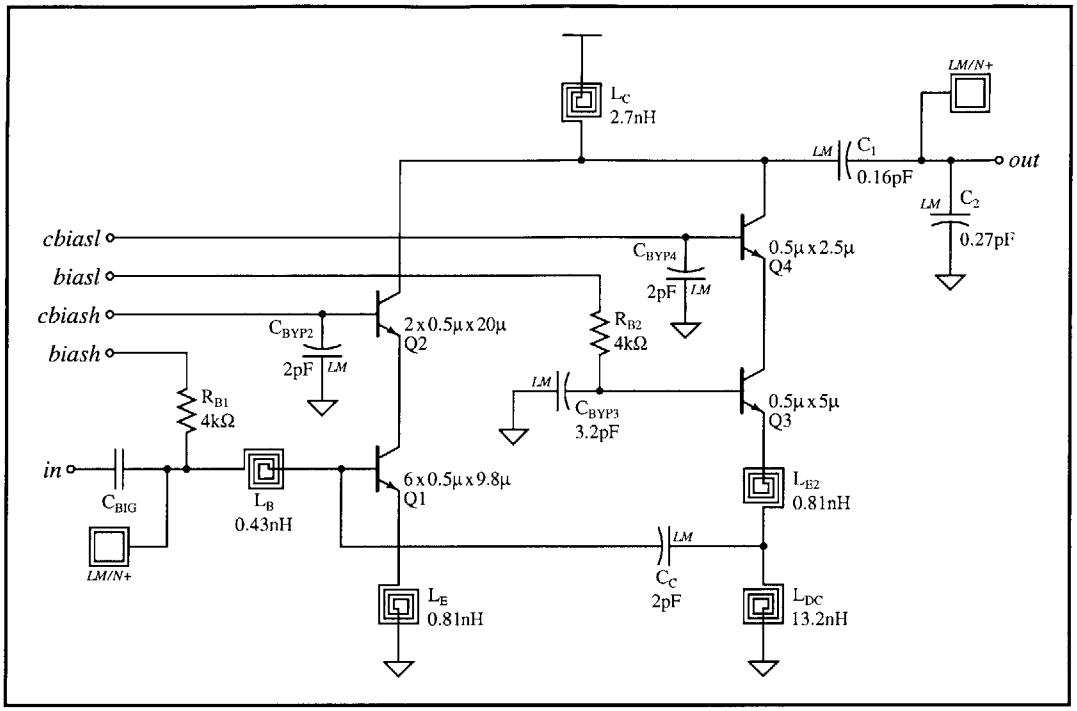

Figure 5-5. RF signal path for the 5.8GHz switched-stage LNA.

may also be compromised. Thus, having two transistors in series to turn "off" in each path is a noteworthy attribute of this parallel stage LNA.

Another important consideration for this design is to use probe pads which are shielded from the substrate by a low impedance ground layer [64][65]. Formed by a rectangle of the 2µm thick uppermost metal over a heavily-doped n+ region tied to an RF ground (the LNA supply voltage), the shielded pads help minimize loss in and coupling through the substrate.[24] As an added bonus, these pads are easily modeled by an oxide capacitance and a small series resistance to ground, where the resistance is that of the n+ ground plane; the predictability resulting from the recognizable behavior is a boon to designing for the desired impedances at the input and output. Tweaking the matching networks and sprinkling in bypass capacitors on the base bias voltages are the final touches, completing the LNA signal path with the exception of a DC block on the input.[25] The capacitance needed for coupling the RF input at 5.8GHz is small enough to easily inte-

[24] Assuming a p-type silicon substrate, the ground shield can be formed in a bipolar technology by using a combination of the n+ subcollector and n+ collector sinker (or reach-through) implants. CMOS technologies can make use of an n+ source/drain diffusion region for the shield. Such trickery becomes unnecessary in processes that make use of semi-insulating substrates.

[25] The "LM" designation in the schematic is an abbreviation for "last metal" and indicates the top plate of the MIM capacitors. In this context, "last" is meant in the as fabricated, "bottom-up" sense.

grate with a high quality MIM structure, but was left off chip in this design so that the on-chip biasing circuitry would have an externally visible node, affording the capability of overriding the bias point during evaluation using a bias tee should the need arise.

With these latest additions to the LNA, the question beckons: "What price adaptability?" Answering this query via simulation, and as illustrated in Figure 5-6, the achievable noise figure suffers a 0.25dB hit and the available gain withers considerably when the low power stage is merged into the design. Much of the drop in gain is owed to an increase in the emitter inductance L_E, which has grown by nearly a factor of 2.5 to return the input impedance in the high performance mode to 50Ω. Next, models for the probe pads are added along with subcircuit descriptions of the parasitics associated with each resistor and capacitor.[26] Sitting 10μm above their n+ diffusion ground planes in this five level metal technology, the pads are barely noticeable with 50Ω source and load impedances. Of slightly greater consequence are the capacitors; although the noise figure is only marginally affected (an increase of 0.04dB), around 0.25dB of gain is lost due to non-idealities inherent to C_C, and another 0.25dB evaporates in the output matching capacitors. Hence, the realities of integrated components and implementing a reduced power

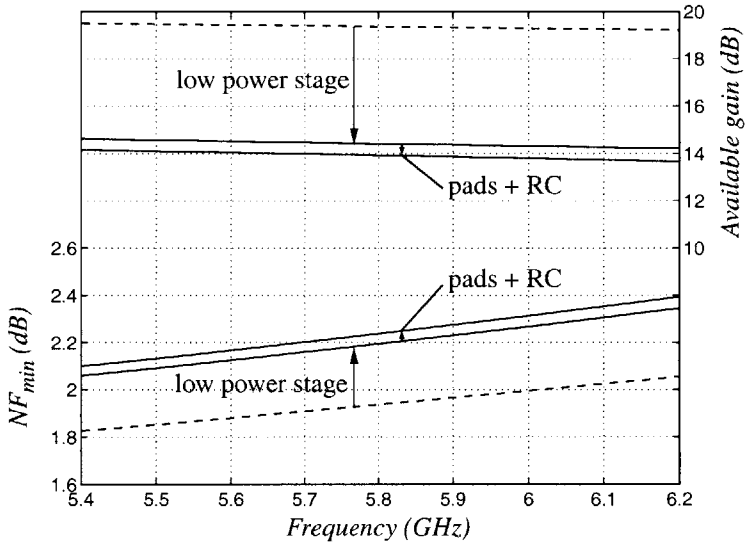

Figure 5-6. Effects observed in the 5.8GHz LNA of adding the low power stage and including parasitics associated with the pads, resistors, and capacitors.

[26] The resistors are implemented with base (handle) poly on which silicide formation has been masked. The capacitors are realized as MIM devices using a metal plate layer inserted (in the process flow) between the top two interconnect levels. To reduce capacitance to the substrate, the resistors and capacitors are formed over a grid of deep trench isolation pockets. The shielded probe pads are modeled as having 31fF of oxide capacitance in series with 5Ω of ground plane resistance.

mode have lowered the gain to 14dB and elevated the noise figure to 2.25dB. But this amplifier now very closely reflects the performance that can be realized in a 5.8GHz wireless networking application. The only piece—albeit an important one—missing from the design at the moment pertains to the four nodes hanging off the left side of the schematic in Figure 5-5, nodes from which the bias conditions for each of the signal path transistors need to be established.

5.1.3 A Switched Current Source Bias Circuit

If the writing to this point in the chapter has been at all effective, it should be evident that a good deal of effort has been invested to keep noise in the RF signal path at a minimum; to trample over this with noise from biasing circuitry would render this effort inert. One approach to minimizing the impact is to reduce the generation of noise in the bias network itself, a demanding task that usually entails high power consumption and may only prove mildly effective. Alternatively, the bias circuit can be decoupled from the signal path over the RF band of interest, thereby alleviating concerns about the extent to which noise is produced in this circuit. Bias decoupling is easy enough to accomplish with the grounded-based transistors where the base voltage can be bypassed by an RF short (e.g., a capacitor). However, the high performance stage of the switched LNA makes use of a common-emitter device wherein the bias and RF input are applied to the same terminal. The ideal scenario here would be to supply the base bias through a large-valued inductor. An inductance effectively detaches the bias circuit at higher frequencies, but yet preserves a low impedance path by which to set the bias point—seemingly the perfect solution.

Unfortunately, and as starkly noted in the design of the output matching network, integrated inductors are laden with parasitics that curtail their impedance at RF; a relatively low capacitance polysilicon resistor can yield a higher impedance and thus more effectively shield the input match and isolate the LNA from bias noise.[27] Coupling the bias through a resistor, however, comes with an issue of its own: higher resistances are needed for isolation but also increase the output impedance of the base bias supply at DC, providing a mechanism through which the bias current can be modulated by low frequency distortion products. The trade-off is depicted in Figure 5-7, where a diode reference with equivalent voltage and current noise generators is considered together with the thermal noise of the bias coupling resistance (R_{B1}) as being the only sources of noise in an

[27] Using a resistor instead of an inductor will also save a sizable chunk of die area. As only the base current will need to be supplied through the resistor, and since matching will not be a concern, the polysilicon width can be kept rather small.

otherwise ideal LNA. As seen in part (b) of this figure, the resistance needs to be in the range of several kilohms to adequately insulate the LNA from noise in the bias circuit—values which are large enough to inhibit the linearity of the amplifier. Of particular concern will be intermodulation products created as second-order distortion terms act upon adjacent signals in the LNA pass band. These signals may be received from the antenna or by feeding through a subsequent mixing stage, upon which non-linearities can act to produce a component in the base voltage located at the difference between the frequencies of the competing signals. Effectively, a time-varying bias current results, modulating the gain of the LNA and thereby creating in-band distortion.

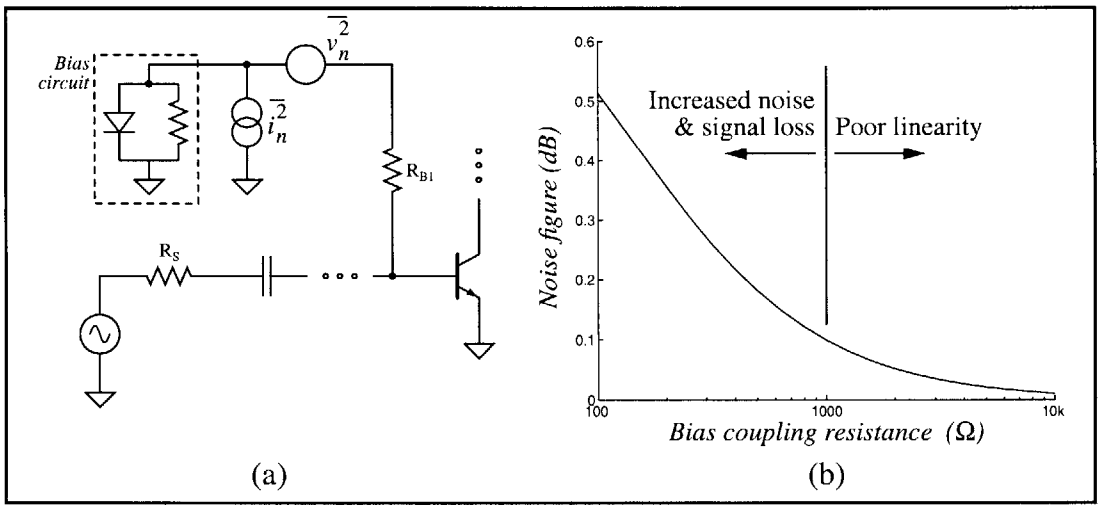

Figure 5-7. Effect of the bias coupling impedance on noise and linearity.

Rather than trying to strike a balance between noise and linearity, the solution to this dilemma is to go to the extreme of having a very large output impedance. If a current source is used to drive the base instead of a voltage source, then there will be no means by which a time-varying bias current can be supplied because the base will be incrementally open circuited for low frequencies. With the RF coupling capacitor also providing a DC block, any low frequency distortion components in the base voltage will fall across the "deaf ears" of the current source instead of the resistor R_{B1}, immunizing the bias current and LNA gain against these non-linearities.[28] Of course, the challenge now is to accurately establish the collector current via a base current, and to do so in the face of a transistor (DC) beta that can vary all over the process map. One concept for accommodating this

[28] This may set a pragmatic limit on how big C_{BIG} (the RF coupling capacitor) should be. While this device should be sized to present a very small impedance in the LNA pass band, it should preferably have a reasonably large impedance for frequencies from DC out to roughly the bandwidth that the LNA will see.

variation is portrayed in Figure 5-8. Transistors Q1 and Q2 form a ΔV_{BE} pair through which the currents are forced to match by the mirror MP1-2. Under this condition, the base-emitter voltages of Q1-2 differ by an amount $\Delta V_{BE} = (kT/q)\ln(I_{S2}/I_{S1})$, where the ratio I_{S2}/I_{S1} is equal to the device layout multiple $m = 4$. This potential difference falls across the resistor R_{BIAS}, creating a supply-independent reference current that is proportional to absolute temperature (PTAT)—characteristics that are desirable for steadying the gain and input match of the LNA over operational changes in temperature and supply voltage. The next addition to the circuit is Q3, which provides a measure of current gain to reduce the base current induced error between I_{C1} and I_{C2}. But instead of tapping directly into the supply, the base drive being furnished by Q3 is first passed through a current sensing device—the diode-tied MP3. Assuming that the NPN base current is dominated by the reverse injection of holes from the base into the emitter, Q1 and Q2 will flow equal base currents despite the difference in emitter area, giving $I_{DP3} = 2I_{B1}$.[29] Then, when matched to Q1, the LNA common-emitter input transistor will have the same base drive requirements, allowing I_{C1} to be replicated in the LNA stage by appropriately ratioing a current source device slaved to MP3. By way of saving power in the biasing circuitry, the reference device Q1 is scaled down from the LNA transistor by a factor of

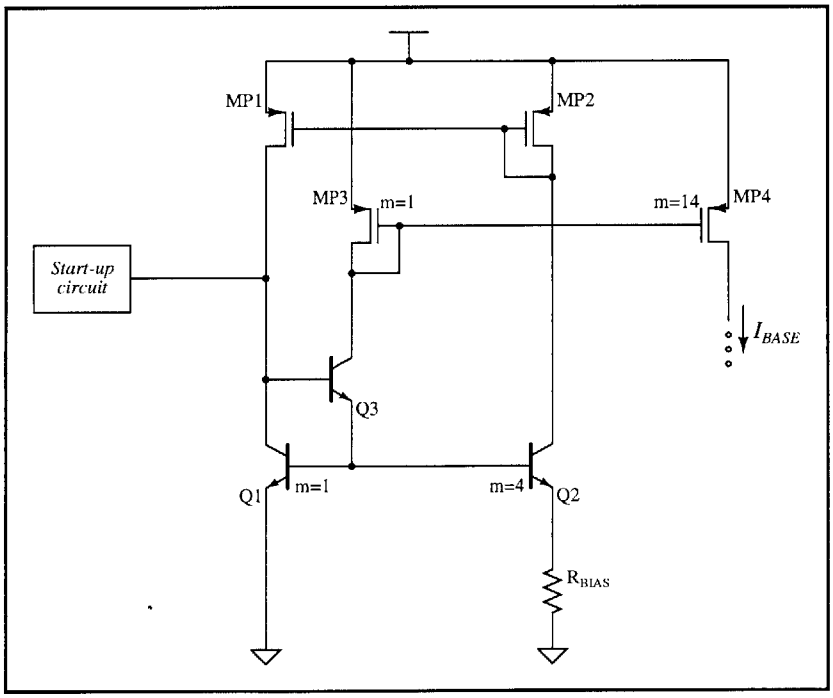

Figure 5-8. Concept for the base current source bias scheme.

[29] Reverse injection is the predominant source of base current in silicon homojunction and SiGe graded-base transistors. This will generally not be true for HBTs, with which some beta-compensation resistors in the base leads of Q1 and Q2 may be required.

$m = 28$; the current I_{BASE} thus needs to be similarly larger than I_{C1}, and can be delivered by sizing the current source MP4 relative to MP3 as a device multiple of $m = 14$.

Fleshed out in slightly more detail, the complete solution is presented in Figure 5-9. At room temperature (taken to be 300K), the base-emitter voltage of Q1 will be 36mV larger than that of Q2, so 370Ω for R_{BIAS} sets the current in each side of the ΔV_{BE} cell to about 100μA. The emitter of Q1 is drawn at 0.5μm by 1.8μm to be approximately a factor of 28 smaller than the common-emitter LNA transistor, a geometry which will not provide the best possible transistor matching.[30] Unfortunately, constructing the LNA device as 28 unit copies of Q1 would compromise its RF performance. However, being that the absolute value of the LNA bias current is already subject to the fabrication tolerances in the resistance of R_{BIAS}, trying to match transistors to better than one part in ten would be a futile effort anyway. The same thinking runs behind the relaxed layout in sizing the p-channel current sources relative to the sensing device MP3. Then, to allow compensation for process variations, an external trim is incorporated by way of a bias adjustment pin. Applying a voltage to this pin effectively alters the value of R_{BIAS} by substituting for some of the current flowing into the resistor or pulling more through Q2. As one example, shorting the *bias_adj* pin to ground lowers R_{BIAS} from 370Ω to 270Ω (370Ω in parallel with the 1kΩ of R_{ADJ}). To use the "default" (nominal) bias condition in the LNA, *bias_adj* may simply be left floating.

Support for the low power mode is easily accomplished by appending another current source device (MP5) to MP3 along with a set of switches (MN1-4). The digital input *lna_ctl* is provided as a means of selection, where applying a logic high activates the high performance mode and a logic low chooses the low power stage using the pass transistors MN3-4. Serving to pull the base of the off-state input device (weakly) toward ground are MN1-2, where a high resistance pull-down is actually preferable so that the base-emitter junction capacitance of the off transistor does not load the input. For the switches themselves, the choice was to minimize the capacitance on the *biash* and *biasl* nodes so as to not shunt the current sources at 5.8GHz. Although p-channel devices would likely contribute less noise and would also offer the possibility of further increasing the output resistance by biasing them as cascode transistors, these effects were deemed to provide little additional benefit in this design and thus smaller n-channel switches with simplistic on/off biasing were used. The remaining need in the LNA signal path is to bias the cascode transistors, for which a $2V_{BE}$ reference is produced by supplying some current (via MP6)

[30] The emitter area ratio used in the design is actually a little more than 28. This helps to compensate for the resistance of the emitter inductor used in the LNA stage.

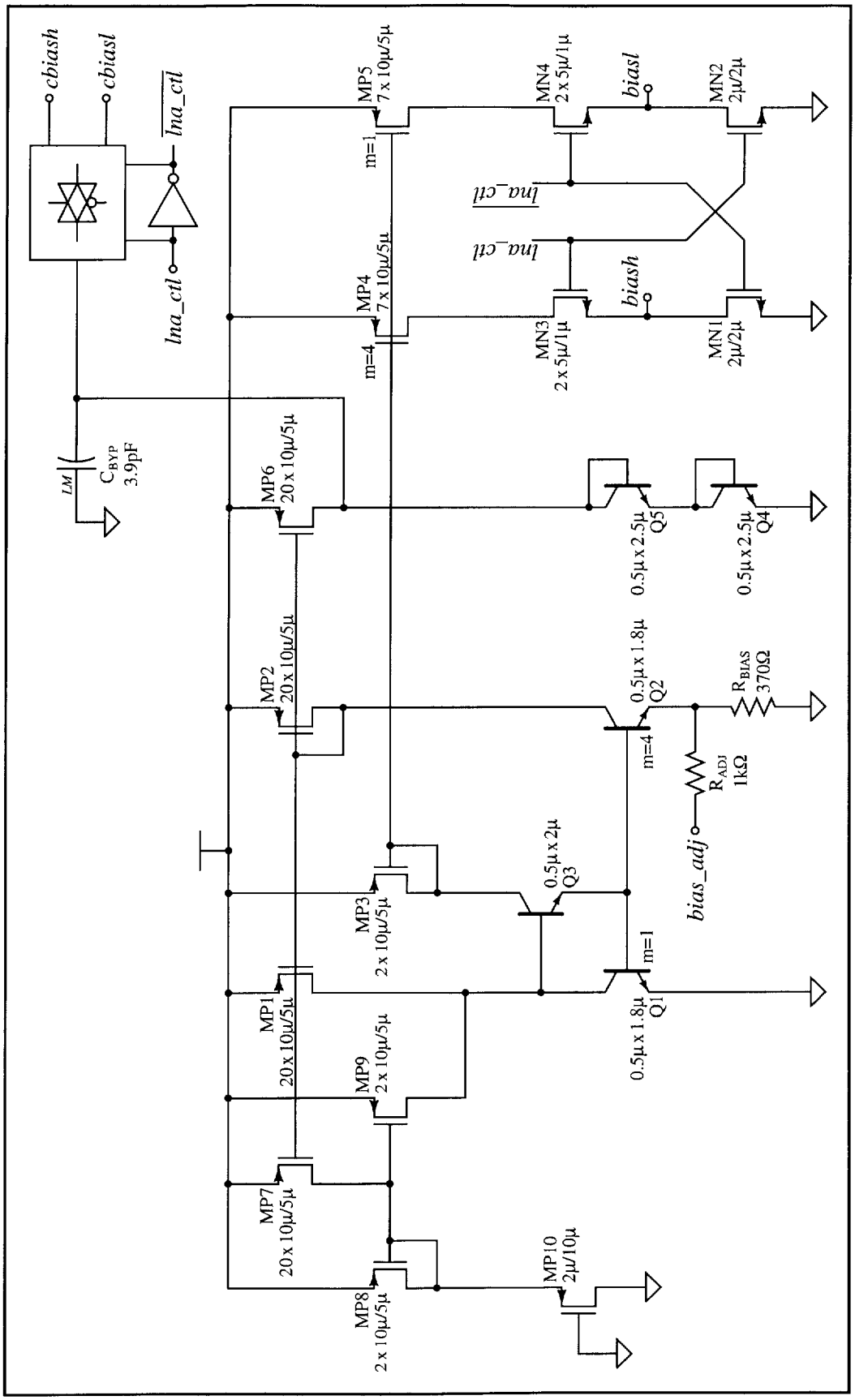

Figure 5-9. Biasing circuitry for the switched-stage 5.8GHz LNA.

through Q4 and Q5. A switch arrangement similar to MN1-4 selects the active stage and provides a pull-down for the other. Having already noted the relative independence of LNA noise performance and gain to V_{CE}, the $2V_{BE}$ common-base reference level is geared toward maximizing the supply headroom available at the collector node for signal swing. Given the high impedance at this node, available swing range will be the key conspirator in gain compression.[31]

Together with the current mirror MP1-2, the ΔV_{BE} cell (Q1-2) forms a regenerative bias circuit for which a zero current condition is also a stable state. When the base of Q3 is low and the gate bias on MP1-2 is at supply, then no current will be drawn and thus no gain resides in the circuit from which to escape the zero state. For the purposes of facilitating the transition out of this state at power-up, MP7-10 comprise a start-up circuit that initializes some current into the regenerative quad. As soon as the supply voltage exceeds two p-channel device thresholds, MP10 will begin pulling current through MP8; this current in turn is mirrored into the regenerative bias cell by MP9. Once current flows in the ΔV_{BE} pair, MP7 will tie off the MP8-9 current mirror as the 100μA it sources will easily overpower the (approximately) 10μA availed by MP10.[32] The factor of 10 in margin should ensure reliable shutdown of the start-up circuit over all variations in temperature, supply, and process.[33]

Any way it is sliced, this bias circuit entails a lot of transistors; that this circuit is going to be used to power a *low-noise* amplifier may appear to have been forgotten in the design process. However, the high impedance furnished by the current source bias scheme should isolate the signal path from this gadgetry, a promise borne out in the results shown in Figure 5-10 of simulating the entire LNA design. By comparison with the numbers obtained using the ideal sources (last seen in Figure 5-6), it is observed that the noise fig-

[31] The 1dB (gain) compression point is a common metric of linearity and relays the signal power level at which the gain supplied by the device to a single input tone has fallen by 1dB from its small-signal limit. Care should be taken in reading this measure as to whether it is referring to the input or output signal level at which the 1dB of gain compression is observed.

[32] While the start-up injection into the ΔV_{BE} pair is turned off, the current drawn through MP10 is not. Under worst-case conditions (e.g., -40C, 5 V supply, fast p-channel devices), the quiescent current in the start-up circuit can rise to 50μA. Although this is a fairly small overhead compared to the milliamps of current flowing in the LNA, this may not be case in all circuits for which the use of a regenerative biasing circuit is considered.

[33] For the technology-conscious, this base current source biasing circuit makes use full use of the trappings of a BiCMOS process. In fact, it could be implemented using only bipolar devices; many bipolar processes offer lateral PNPs which can replace each p-channel MOSFET used in this design with the exception of the "resistor" MP10. Switching between modes becomes a little more complex, and may best be accomplished by providing the base drive for the base current source transistors through switched emitter-followers rather than trying to use bipolar devices as a pass transistors.

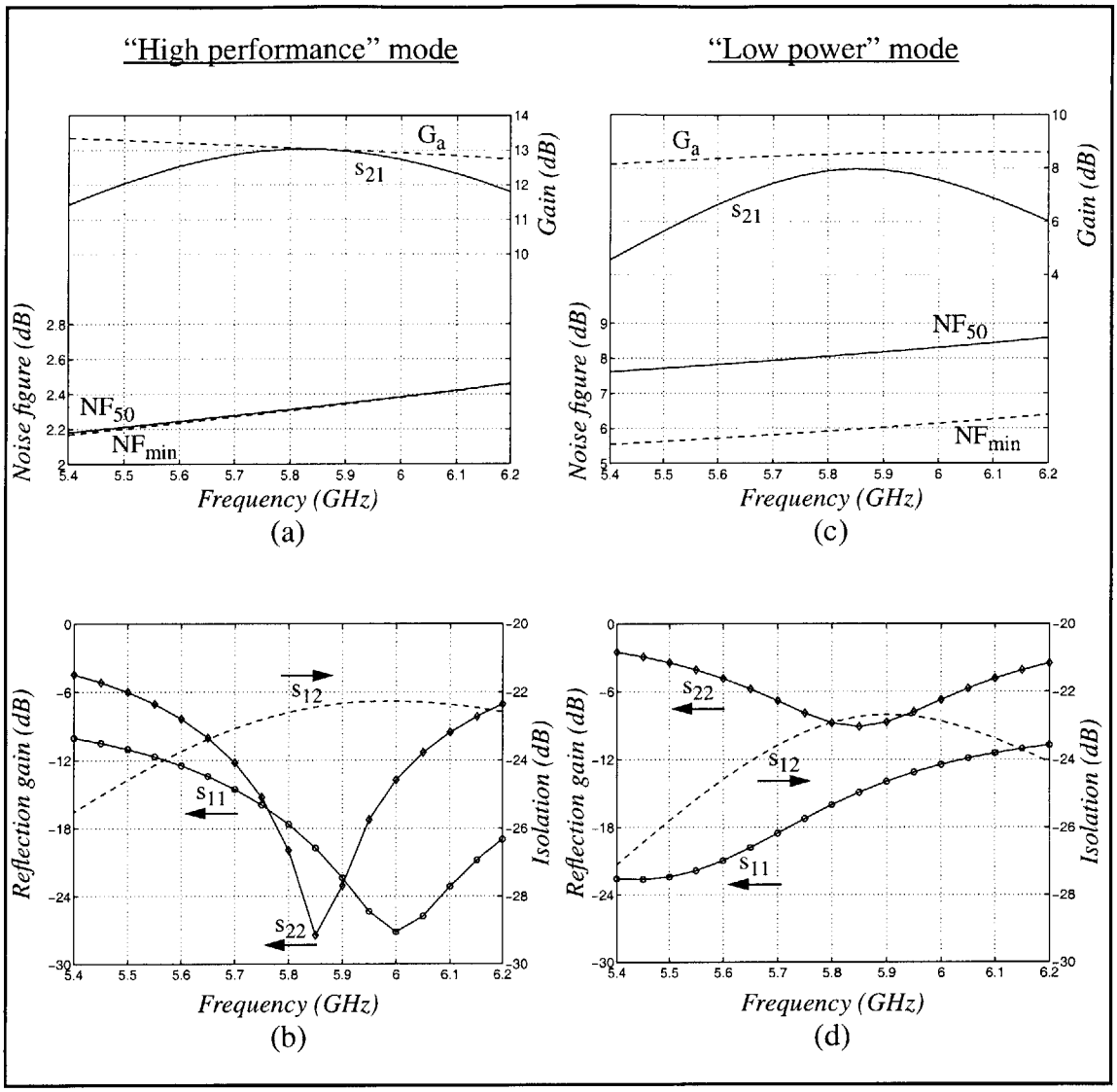

Figure 5-10. Simulated performance of the 5.8GHz switched-stage LNA in both the (a,b) "high performance" and (c,d) "low power" modes. Note the scale change between the modes in representing the gain and noise figure.

ure (in the high performance mode) has crept up a mere 0.06dB and that the gain has been taxed less than 1dB. The final numbers are 13dB of gain and a 2.3dB noise figure in the 8mW high performance setting, while the 2mW low power mode provides 8dB of gain and an 8dB noise figure. Input return loss is better than 14dB for a 200MHz bandwidth about 5.8GHz; looking at the output, the return loss is at least 7dB in the low power mode and 12dB when high sensitivity is being demanded. In either case, isolation is better than 22dB for frequencies anywhere near the band of interest. These numbers reflect full device-level simulations including all parasitics associated with the passive components and pads, and have been gauged with real-world 50Ω source and load impedances.

While the aforementioned set of simulations have rested on a 3V supply and operation at room temperature, the ΔV_{BE} bias cell should provide a high degree of robustness in the operation of the LNA should these conditions change. The essence of gain is proportional to g_m, as are terms in the real and imaginary components of the input impedance; keeping the transconductance constant, the insertion gain and input return loss remain relatively unaffected by temperature as reflected in Figure 5-11(a). Calculated at 5.8GHz for a temperature range of -40 to 100°C, gain droops only slightly at elevated temperatures and $|s_{11}|$ curls up marginally toward the extremes. Following suit in Figure 5-11(b) with supplies ranging from 2V to 5V, the LNA gain witnesses only a slender increase with higher voltages and the return loss remains high despite the presence of voltage-dependent junction capacitances.[34] Such independence, from both supply and temperature variations, will go a long way in preserving the integrity of the RF link in a portable platform.

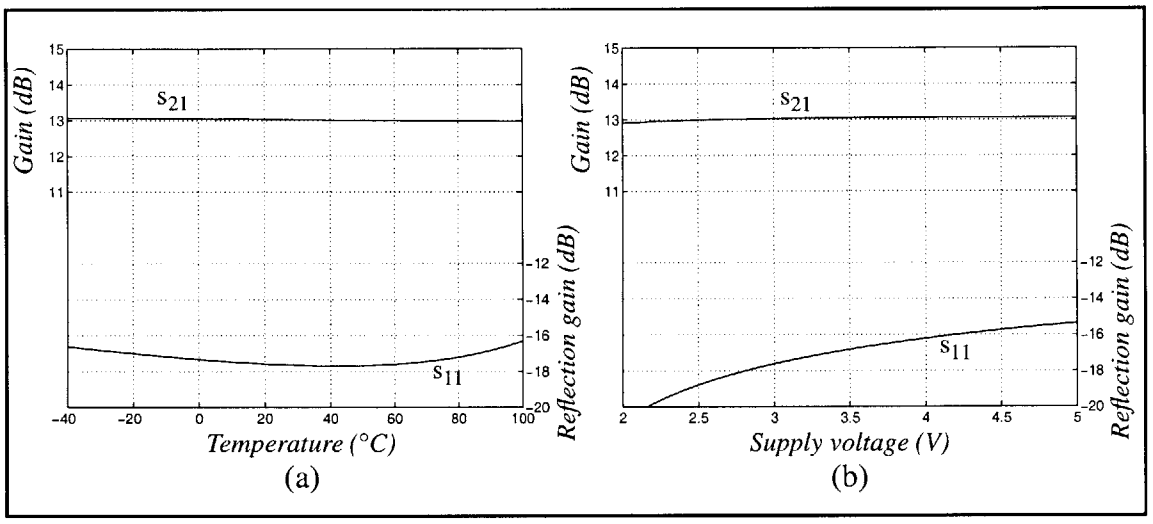

Figure 5-11. Simulated behavior of the 5.8GHz switched-stage LNA over (a) temperature and (b) supply variations.

5.1.4 Layout Design of the 5.8GHz Switched-Stage LNA

Having been convinced that the LNA circuits are fulfilling the obligations originally set forth, attention now turns toward the layout. A die photo of the finished chip, measuring 1.34mm by 1.18mm,[35] is furnished in Figure 5-12. Placement of the five spiral

[34] Though measured at the 5.8GHz band center for the purposes of this illustration, this is at a frequency where $\partial|s_{11}|/\partial f$ is quite high, not at the minimum in $|s_{11}|$. This selection of frequencies exacerbates the extent to which the return loss is actually changing; temperature and voltage coefficients—particularly in the junction capacitances—actually result in a slight shift in the center frequency.

[35] The chip is this size, the die photo is somewhat larger.

inductors, each sitting above a grid of deep trench isolation pockets, largely dictates the floorplan as an exercise in minimizing inter-spiral coupling. From the left, and moving counterclockwise, the inductors can be identified in the die photo as L_B, L_E, L_{DC}, L_{E2}, and L_C. With the exception of the DC bias path inductor L_{DC}, each spiral consists of the top two metal levels strapped together using copious amounts of vias.[36] For L_{DC}, as its primary goal is to be invisible from an RF perspective, a multi-level winding is designed instead. Composed of a last metal trace that spirals inward in a clockwise fashion, an array of vias reaching down two levels, and then a metal 3 winding that mirrors the upper spiral but continues the clockwise direction, this multi-level arrangement increases inductance by creating (positive) mutual coupling vertically as well as laterally. Each of the input matching inductors (L_B, L_E, and L_{E2}) are optimized for quality factor using wide metal traces, while narrower line widths are adopted in L_C and L_{DC} where Q is not as important. Were it not for these differing considerations, inductors L_{E2} and L_{DC} might conceivably be jointly realized as one tapped spiral.

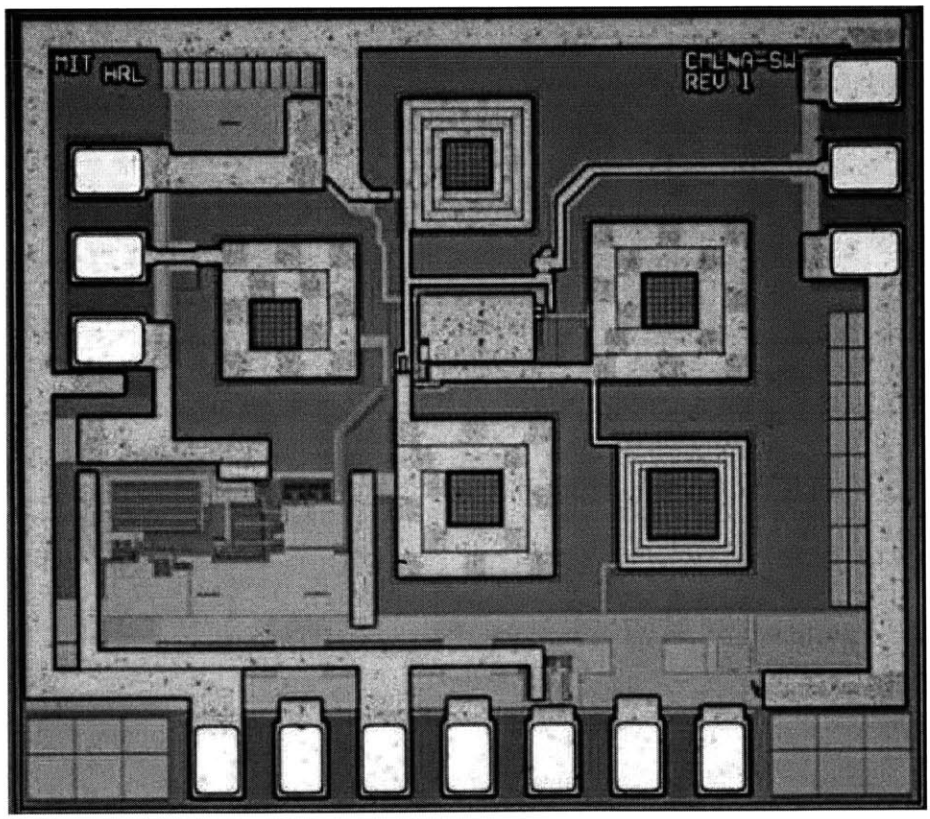

Figure 5-12. Die photo of 5.8GHz switched-stage LNA IC (CMLNA-SW).

[36] The sets of vias can actually be seen in the photo as the darkened "blotches" in the spirals.

Next in order of conspicuousness are the pads, with the RF input (on the left) and output (on the right) designed for contact by coplanar waveguide GSG probes. The RF pads transition to a 50Ω microstrip arrangement with a last metal signal trace over a lower metal ground plane. Along the bottom is a set of pads for a bypassed supply probe having four power pins interspersed with three grounds. The power pins, from left to right, are: the LNA supply voltage, the bias circuit supply voltage, the LNA mode select input (*lna_ctl*), and the *bias_adj* pin. The LNA signal path and the bias circuit operate from the same voltage and could share a supply, but have been given separate pins in this design so that the currents can be individually measured. ESD protection is provided near the pads in the form of clamping FET devices between the supply pins and ground, plus a classical double diode, series resistor scheme for the digital mode select input.[37] The remaining pin, *bias_adj*, already feeds into resistors and the emitter of an NPN, and so no additional precautions were taken.

Sitting in the lower-left corner is the bias circuit, where it occupies approximately the same amount of area as one of the spiral inductors. The big chunk of metal in the middle of the chip forms the shared top (ground) plate of the capacitors that bypass the common-base bias voltages; other supply-filtering capacitances are added in at various locations throughout the chip where space and convenience allow. Finally, as evident in the upper-right corner, the name CMLNA-SW has been affixed to this design. Though having no feasible pronunciation, the first part of the acronym identifies this chip as one of the C-band Monolithic Low-Noise Amplifiers. The SW suffix specifically calls out the switched (bipolar) version to differentiate it from the CMOS implementation.

In the final tally, the addition of a low power mode that protects the impedance matches, sustains isolation, and provides some gain has come at a cost of two inductors, 0.25dB in noise figure, and about 5dB of gain. In exchange, however, this allows a real-time option of reducing the power consumption by a factor of four when peak sensitivity is not required of the receiver. Beginning with the 1.5dB NF_{min} and the 15.5dB G_{ma} exhibited by this 0.5μm graded-base bipolar technology at 5.8GHz, this design finished with an LNA featuring a noise figure 0.8dB higher and an insertion gain about 2.5dB lower (even after cascoding) than the intrinsic device measures. It should be remembered in citing these results that inductor quality factors in the high teens (17-19) have been assumed—numbers which are about the best that could be expected in a silicon technology at these

[37] The FETs intended for clamping ESD transients are some fairly wide n-channel devices that have been ruggedized with a larger drain region and by adding some resistance between the gate terminal and ground. Negative transients are clamped by the drain to substrate diode, while positive transients turn on the transistor by coupling energy from the drain to the gate via C_{GD} [66].

frequencies. Nonetheless, these metrics will be useful to keep in mind during the exploration of using CMOS as an alternative.

5.2 CMOS LNA Design

Notwithstanding the dramatic appearance of CMOS transistors in the preceding section, a rather bipolar-centric view into the world of LNAs have prevailed throughout this presentation. Lest CMOS be relegated to the vagaries of behind-the-scenes biasing duties, it should be emphasized that the small-signal models of bipolar and CMOS transistors are essentially the same—particularly at RF frequencies where C_π dominates the base-emitter junction impedance. The differences really only reside with the values and dependencies of the model parameters; as such, the topological and technological arguments made in the context of bipolar designs also apply here. To be sure, the transconductance of CMOS transistors is lower for any given bias current, and the same can be said for the output resistance.[38] But so too are most of the device capacitances, perhaps with the notable exception of the gate-drain capacitance (C_{GD}), which tends to be of the same order as C_μ in RF-sized transistors. By way of terminal resistances, that encountered in the extrinsic source-drain regions is usually around 500Ω-μm for n-channel devices and about a factor of three higher for their p-channel brethren [67]. In symmetric MOSFETs, this resistance can be assumed to split equally between the source and drain terminals (e.g., a 100μm wide n-channel transistor will be encumbered by 2.5Ω of extrinsic source resistance, a number in the same ballpark as the emitter resistance of a reasonably-sized NPN). Resistance in the gate terminal, although distributed across the width of the channel, can be modeled for noise, delay, and f_{MAX} purposes by a single lumped resistor taking on the value [68]:

$$R_G = \frac{1}{3} R_\square \left(\frac{W}{nL} \right), \tag{5.8}$$

where $R_\square$ represents the sheet resistivity of the gate material (polysilicon or silicide) and n is the number of gate fingers. The gate resistance can readily be made fairly small by using multiple finger layouts; revisiting the example of the 100μm wide device, assuming a 0.5μm (drawn) channel length and a silicided poly layer having a resistivity of 4.5Ω/□, a layout with 10 gate fingers would exhibit 3Ω of gate resistance. So, at least in the com-

[38] Transistor output resistance is probably not too compelling an issue in RF applications as impedances tend to be limited by inductor loss anyway. As further evidence, there are certainly plenty of MESFET and HEMT devices to be found in microwave and mm-wave designs.

parison of terminal parasitics, CMOS devices stack up quite well against bipolar transistors. As will be seen next, however, noise is another matter altogether.[39]

5.2.1 Noise in CMOS Transistors

Regardless of the chosen device technology, when noise is the substantive issue, having a solid understanding of its various origins is the cornerstone of the design process. Fortunately, the similarities observed earlier between the small-signal models for bipolar and CMOS transistors also extend to the presence of noise sources. The thermal noise from the terminal resistances is treated exactly as discussed in the previous chapter for bipolar devices, even though the effects of geometry and process scaling upon the resistances may differ. There are also analogues to the base and collector current noise sources observed in bipolar transistors, although the mechanism is considerably different as shot noise processes—classically speaking—do not occur in MOSFETs.[40] Instead, thermal agitations vary the velocity of the carriers in the channel, lending a noise component possessing the power spectral density:

$$\overline{i_D^2} = 4kT\gamma g_{d0} \tag{5.9}$$

to the drain current which is proportional to the zero-bias drain conductance (g_{d0}) of the device. In the long-channel limit, g_{d0} simply equals the transconductance, but can appreciably exceed g_m in shorter channel transistors. Similarly, as measured drain noise seems to outpacing g_{d0} in modern technologies, the "excess noise" factor γ can be significantly inflated beyond the value of 2/3 observed at long channel dimensions.[41]

Charge in the channel, as mirrored across the oxide capacitance, also induces a noise current in the gate. The presence of an induced gate current was recognized in FETs by van der Ziel in the 1960s [69][70], but that it is of consequence for the design of RF amplifiers in CMOS has only fairly recently been remembered [71]. Resulting from the finite transit time carriers require in getting from the source into and through the channel, the phase of the input current undergoes a shift relative to the gate-to-source voltage

---

[39] With all apologies to Jim Abrahams, the Zucker brothers, and the rest of the creative crew responsible for foisting the *Airplane* movies into our collective conscience: "As will be seen next, however, noise is another matter."

[40] This may not be entirely true, particularly for highly scaled CMOS transistors. Ballistic transport in the channel [72] and gate leakage current [73] both lead to the presence of shot noise. Furthermore, some argue that channel noise in subthreshold might have shot noise origins [61].

[41] Clear explanations behind this increase are still being sought. Much like an effective mass, many of these noise "factors" are essentially the right answer divided by the wrong answer, where the "right" answer is the measured result and the "wrong" answer is that calculated from conventional physics models.

applied to a MOSFET. This phase shift becomes noticeable at higher frequencies, conferring upon the FET a real component in the input impedance that can be modeled by placing a gate conductance:

$$g_g = \frac{\omega^2 C_{GS}^2}{5 g_{d0}} \tag{5.10}$$

in parallel with C_{GS}. Fluctuations in the channel charge, and thus in the delay which gives rise to this conductance, are also responsible for the noise current flowing into the gate. The gate noise has an associated power spectral density:

$$\overline{i_G^2} = 4kT\delta g_g, \tag{5.11}$$

where δ is the excess gate noise factor. Applying the classical long-channel approximations renders a value of 4/3 for δ, although short channel transistors have played witness to an exacerbation of this effect as well. While both the gate and drain noise currents are rooted in the same charge flow mechanism, these terms are only partially correlated in that they have different dependencies upon the location of a given noise disturbance within the potential distribution along the length of the channel.[42] Conventional models peg the correlation coefficient at $c = j0.395$, but as the lengths within the devices shrink, the channel position dependencies converge and the degree of correlation increases.

Equivalent input noise generators can be derived for MOSFETs in much the same fashion as performed for bipolar transistors. Some care must be taken in the CMOS case to manage the correlation between the drain and induced gate noise sources, although a reasonable first-order cut at noise behavior can be obtained merely by substituting the corresponding terminal identifiers into the relationships expressed in the previous chapter.[43] The real danger in omitting $\overline{i_G^2}$ from the design process is that doing so leads the noise optimization astray. Just as would be the case with bipolar transistors were there no base shot noise (or if the current gain was infinite), the goal of minimum noise would steer toward the use of vanishingly small devices biased at correspondingly small currents, and then compensating for the gain using a high Q input matching network. But in practice, this matching network would present a large source impedance to the FET through which the gate noise current would flow, saddling the amplifier with a sub-optimal noise figure.

---

[42] A nice graphically-reinforced derivation of the differences between the induced gate and drain noise currents can be found in a thesis by Sepke [74].

[43] This should not come across as being completely heretical as the base and collector current shot noise sources in bipolar devices are also partially correlated [75]. However, the correlation in bipolar transistors is routinely ignored for frequencies below f_T.

Unfortunately, the absence of gate noise is exactly the scenario professed by most incarnations of CMOS circuit simulation models to date.

After determining that induced gate noise was responsible for the departure of the measured noise figure from their expectations for a 1.5GHz LNA design (in a 0.35μm process), Shaeffer and Lee have suggested a revised noise analysis of tuned MOSFET amplifiers, offering [71][76]:

$$F = 1 + \frac{R_{L_G} + R_G}{R_s} + \frac{\gamma}{\alpha}\frac{\chi}{Q_{in}}\frac{f}{f_T} \tag{5.12}$$

as a more complete description of the noise factor for a common-source transistor. This definition provides a value for F intermediate between F_{50} and F_{min}, as it assumes that the RF source (having an output resistance R_s) is reactively matched to the device through an inductance L_G added before the gate terminal. This inductor may come attached to some series resistance (R_{L_G}), which will directly add to the gate terminal resistance (R_G) component defined in Equation 5.8. The ratio γ/α refers to the drain noise current, where α indicates the extent to which the transconductance falls short of drain conductance: $g_m = \alpha g_{d0}$. Noise and loss in the input circuit is reflected through its quality factor:

$$Q_{in} = \frac{1}{2\pi f C_{gs}(R_s + R_G + R_{L_G})}, \tag{5.13}$$

which also equals the voltage transformation ratio (voltage gain) provided by the input matching network at resonance. Finally, the χ parameter brings the induced gate noise into the noise factor expression as:

$$\chi = 1 + 2|c|Q_{in}\sqrt{\frac{\delta\alpha^2}{5\gamma} + \frac{\delta\alpha^2}{5\gamma}(1 + Q_{in}^2)}, \tag{5.14}$$

where δ and c respectively are the excess noise factor and correlation coefficient associated with the gate noise current. The one remaining piece that should be included in the analysis is the extrinsic source terminal resistance, which can easily be accommodated by lumping it together with R_G in Equations 5.12 and 5.13.

Now, what remains is to determine F_{min} by finding the source resistance that minimizes the noise factor expression provided above. For a chosen device size (and hence a given C_{GS}), this task is equivalent to locating the optimum Q_{in}. As a representative case study, a minimum-length n-channel MOSFET in a 0.5μm technology will be featured, assuming the gate and extrinsic source resistivities are as exemplified in Section 5.2 and that the long-channel approximations apply reasonably well to the noise parameters.

Selecting a transistor gate width of 134.2μm—split into 22 gate fingers of 6.1μm each—and a drain current of 5mA, the optimum source resistance for 5.8GHz operation is determined to be 91Ω (Q_{in} of 1.8), at which point NF_{min} is 1.82dB.[44] Reducing the drain current in this device to 1.67mA increases the minimum noise figure to 2.52dB, found for an $R_{s_{opt}}$ of 89Ω. By way of providing some insight into how well the BSIM3v3 models available for this CMOS technology might compare, the same two bias points are simulated using the commercial tool Spectre®, the results of which are plotted below in Figure 5-13. Interestingly, the simulator model returns higher noise figure numbers despite the absence of an induced gate current mechanism. Reasons for this behavior are not clear, but it is believed that some of the additional transistor parasitics (e.g., C_{GD}) incorporated into BSIM3v3—and left out of the analytical Shaeffer work—might be playing a significant role at this frequency.[45] Also surprising, and difficult to fathom as conforming to the actual device characteristics, is that the BSIM model predicts an increase in the minimum noise figure of only 0.1dB as the bias current is cut by a factor of three. So, uncertainty persists, but this is the best information available for purposes of this design. The "guess" here is that the BSIM model is probably not terribly far removed from reality for higher bias points at these device sizes, but then the Shaeffer development better captures the dependencies of the noise performance on transistor width and drain current. Effectively, a linear mental construct will be employed where the y-offset is supplied by BSIM and the slope is taken from the analytical analysis. Neither model seems to account for all of the pertinent effects, but such is the quandary of IC design.

5.2.2 A 5.8GHz CMOS LNA

Despite the differences in the manifestation of noise within CMOS devices as compared to bipolar, the process for designing an LNA in CMOS is the same. The basic approach was outlined in Chapter 4, where the first step is to size the transistor to give an $R_{s_{opt}}$ of 50Ω.[46] One of the key differences between the technologies, however, is that reaching this goal may not be feasible with CMOS given the bound of (reasonably) finite power consumption. Recalling the example in the previous section wherein the optimum source resistances were in the 90Ω range, it is apparent that a considerably larger device—

[44] Establishing both the device size and the drain current sets the transistor f_T as applied to Equation 5.12. This is a slightly over-constrained version of the problem Shaeffer and Lee formulate as the fixed g_m optimization [76].

[45] Gate terminal resistance is not included in BSIM3v3 either, but is added into the simulation by using a subcircuit representation for the MOSFETs.

[46] More generally, it should be said that the idea is to achieve a noise match by sizing the transistor to give an optimum source resistance that matches the output resistance of the source; this need not be 50Ω.

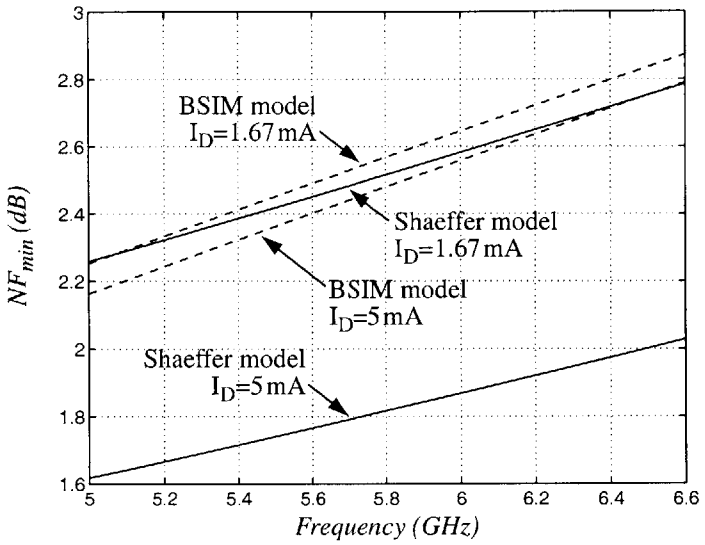

Figure 5-13. Comparison of models for the minimum noise figure of a common-source transistor at two drain current levels. Solid lines represent results of the Shaeffer model [76] and dashed lines are simulation results using full BSIM3v3 models.

operated at correspondingly higher drain currents—would be required to obtain a noise match to a 50Ω source. In light of this limitation, a better design approach is to solve for the minimum noise figure (via Equation 5.12) that can be achieved given a fixed power constraint. Optimizing under this condition finds the transistor width that furnishes the lowest noise while keeping constant the product of the drain current and the gate voltage required to support this current.[47] As long as the channel is not driven too deeply into inversion (i.e. the amount by which the gate-to-source voltage exceeds the threshold is relatively modest), a fairly broad minima exists around an input matching network having $Q_{in} \cong 3.8$—a solution that covers a reasonably wide range in the power dissipation being tolerated of the design [76]. Using this result, and ignoring parasitic resistances for the purposes of this calculation, Equation 5.13 recommends a C_{GS} of 144fF for a 5.8GHz amplifier. Then, given a 0.5μm technology having a gate oxide capacitance (C_{ox}) of 3.9fF/μm^2, the usual $C_{GS} = (2/3)WLC_{ox}$ relationship suggests an optimum transistor width of 111μm. This is close to the width of 134.2μm chosen for this design, a number arrived at through a slightly different set of approximations.[48] Reapplying the Shaeffer

[47] Increasing the device width lowers the required V_{on} (gate overdrive potential) for a given drain current (I_D), but in turn necessitates a larger I_D to maintain the transistor f_T. Somewhere in this relationship the optimum balance is struck.

[48] Lee [77] provides a 750μm-GHz rule of thumb for the optimum transistor width, then observes that variations of 20% about this number typically lead to degradations of only 0.1-0.2dB.

model to this device with a 50Ω source resistance, the reactively-matched noise figures are 2.01dB for the higher sensitivity (5mA) mode and 2.74dB for the reduced power setting (where I_D is 1.67mA). Judging from these numbers, it appears that the power constraint has cost between 0.2dB and 0.25dB in noise performance.

Forming the LNA around this device, a cascode transistor is appended for the same arguments made previously in light of the switched-stage implementation. As can be seen in Figure 5-14, the geometry of the cascode was selected so that the interconnection of the two transistors could be accomplished by merging the appropriate source and drain regions. An inductance of 0.43nH is used in the source of the lower device to move the input impedance to the 50Ω circle, from which point the match is completed by adding another 2.4nH before the gate. The output matching network follows directly from that used in the bipolar design, where the same inductor (designated L_D in this case) has been used to facilitate comparisons in the performance of bipolar and CMOS technologies.[49] As the input capacitance of a MOSFET is only a weak function of the transistor bias cur-

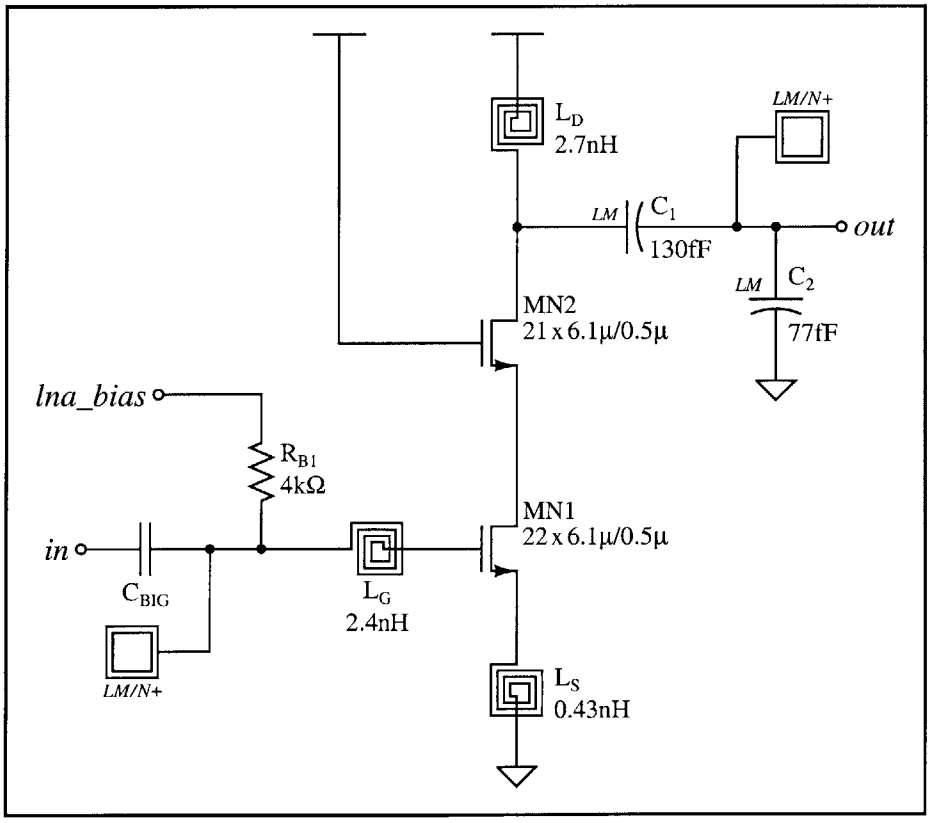

Figure 5-14. RF signal path for the 5.8GHz CMOS LNA.

[49] This design choice allows a more direct comparison of the gains exhibited by the CMOS and bipolar LNAs, without having to worry about differences in the signal power dissipated in the output matching network.

rent, a single stage suffices to meet the goal of maintaining an impedance match while allowing sensitivity and power consumption in the LNA to be traded for one another. The remaining piece is the 4kΩ bias coupling resistor; because the insulating gate of a MOSFET already provides an incremental open in the bias path at low frequencies, this resistance will not lead to the linearity issues noted in the switched-stage design.[50] This is not to say that the concerns regarding noise contributions from the bias circuit have been alleviated in that this part of the picture painted in Figure 5-7 still holds; the impedance presented by the FET will be matched to the RF source in the band of interest, and so having a large bias coupling resistance remains important. As an additional feature, developing a gate bias that compensates for temperature-induced variations in the transconductance of the input transistor will prove handy, as is ensuring that changes in the supply voltage do not steer the LNA away from its intended bias current. The approach to addressing these challenges will provide another interesting set of parallels between designing with field-effect devices in comparison to bipolar transistors.

5.2.3 Stabilized Biasing for the CMOS LNA

One of the most salient characteristics of bipolar transistors is the exponential relationship between the base-emitter voltage and the collector current, a relationship which holds remarkably well over the transistor's useful operating range. While the details of establishing a current in CMOS are not as logarithmically clean, the ΔV_{BE} concept used in biasing the bipolar LNA can similarly be applied to obtain a constant g_m in an MOS transistor. The gate-source voltage is used instead of V_{BE}, and the difference exhibited by two MOSFETs flowing the same current can be expressed as:

$$\Delta V_{GS} = c_m \sqrt{\frac{2I_D}{\mu C_{ox}} \frac{L}{W}}, \tag{5.15}$$

where the larger of the two devices is realized as m copies of the W/L-sized smaller transistor—a proportionality captured in the coefficient $c_m = 1 - m^{-1/2}$. When the currents in each side of the ΔV_{GS} pair are equalized by a current mirror, a resistor (denoted R_{BIAS}) soaking up the difference in the gate-source potentials establishes a current:

$$I_D = \frac{2}{\mu C_{ox}} \frac{L}{W} \frac{c_m^2}{R_{BIAS}^2}, \tag{5.16}$$

provided that an initialization mechanism ejects the circuit from the $I_D = 0$ state.[51] By employing this current as a reference, the μC_{ox} dependence of the transconductance can

[50] The linearity problems arise because of the base current in the bipolar transistors.

be cancelled in any MOSFET it serves to bias, taking the first-order source of temperature variation along with it.

This is the technique adopted for stabilizing the gain of the CMOS LNA, for which the gate bias is provided by the circuit drawn in Figure 5-15. Resistors R_{BIAS} and R_{ADJ}

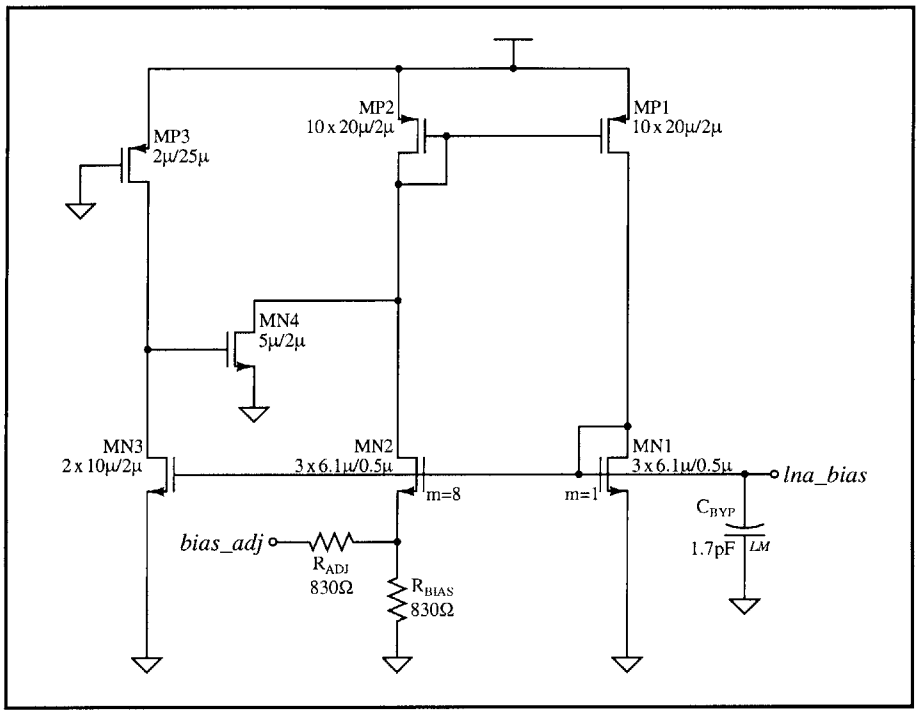

Figure 5-15. Biasing circuitry for the 5.8GHz CMOS LNA.

are sized in conjunction with the width of MN1 to conveniently establish the desired 1.67mA and 5mA bias current levels in the LNA; the low power mode is enacted by leaving the *bias_adj* pin floating, and the high performance setting is chosen by simply tying *bias_adj* to ground. The selection of two discrete operating modes is somewhat arbitrary and is largely done as a matter of convenience; in practice, the sensitivity afforded by this amplifier can be traded for lower power consumption along a continuum of points via the *bias_adj* pin. Completing the design—and ensuring that the voltage applied to *bias_adj* does not merely select among different values of zero for the bias current—is the start-up circuit formed by MP3 along with MN3-4. When power is first applied to the LNA, the gate of MN4 will be pulled high by the "resistor" MP3, in turn drawing some current through the mirror MP1-2. Once the ΔV_{GS} cell reaches its destina-

[51] Although the analysis provided here rests on the conventional square-law expressions for MOSFETs, the ΔV concept should be applicable to virtually any transconductance device. The resistance term appears squared because $g_m \propto \sqrt{I_D}$.

tion current, MN3 will easily out drive MP3, thereby shutting off the kick-start injection supplied by MN4.

Although topologically different from the 5.8GHz design implemented with bipolar devices, this entirely CMOS version has strived to achieve the same set of goals. Borrowing the same 3V supply, the performance expected of the CMOS LNA at 15mW and 5mW power consumption levels has been cast forth in Figure 5-16. Models for the RF-shielded pads and all of the passive devices have been included, while BSIM3v3 representations are called upon for the MOSFETs.[52] At 10dB, the gain observed in the high performance setting is 3dB lower than that expected from the switched-stage design, despite

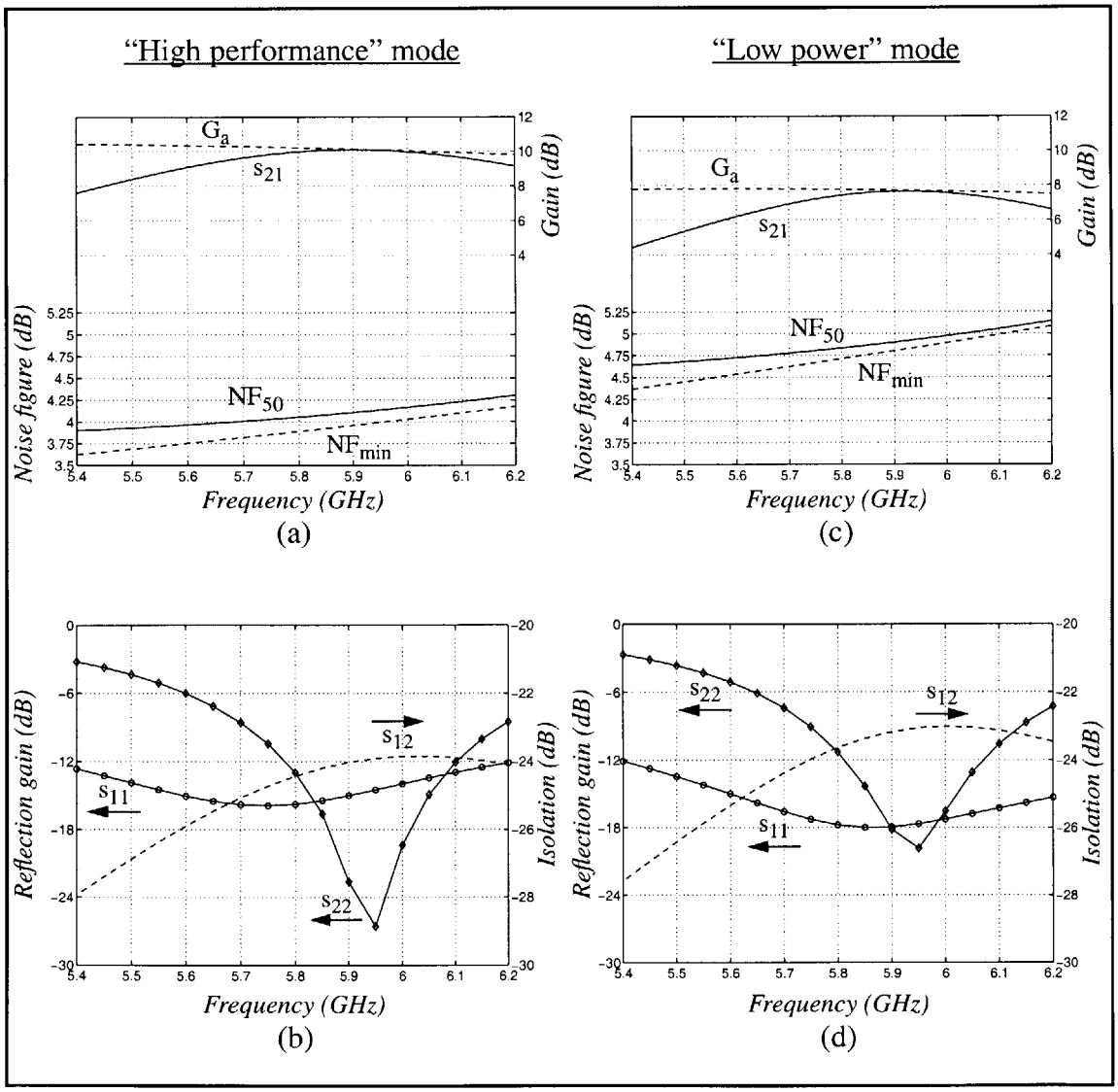

Figure 5-16. Simulated performance of the 5.8GHz CMOS LNA in both the "high performance" (a,b) and "low power" (c,d) modes. Noise figure calculations are based upon BSIM3v3 circuit models which have not implemented induced gate noise.

the highest (drain) impedance an on-chip inductance can buy and a bias current that is higher by nearly a factor of two—a testament to the higher transconductance of bipolar transistors.[53] With a 50Ω source resistance, the noise figure at 5.8GHz comes in at a shade over 4dB, a number which is expected to be approximately correct—despite the absence of induced gate noise—as per the comparison in Section 5.2.1. While the CMOS LNA began with a 1dB poorer NF_{min} for the intrinsic device, another sizable portion of the 1.7dB increase over the noise figure of the bipolar design occurs from the higher loss associated with the larger total inductance needed to realize an input match (2.8nH for L_G+L_S versus 1.2nH for L_B+L_E).[54,55] From this it may be concluded that if power consumption were not a consideration, the noise figure would improve on two fronts: the larger transistor width would provide a Γ_{opt} closer to the 50Ω source resistance, and would also necessitate less inductance to achieve an impedance-matched input. Of course, power is a concern, which is why a 5mW setting has also been characterized. In reducing the bias current by a factor of three, the gain would not be expected to fall too dramatically since—at best—$g_m \propto I_D^{1/2}$; as seen in the plots of gain, simulations predict a drop of about 2.5dB. Noise figure creeps up by about 0.75dB when the bias current is reduced to 1.67mA, although the actual number may be another 0.5dB (or so) higher considering that the induced gate noise has been neglected.

As the only device impedance that undergoes much of a change over bias (i.e. the output resistance) is effectively washed out by inductor loss, little is affected in the return losses and isolation as the bias current is adjusted—a fact borne out in comparing parts (b) and (d) of Figure 5-16. Furthermore, all three measures should be similar to that obtained with the bipolar design, albeit each for different reasons. Isolation largely rests with the feedback capacitance and, as noted earlier, the gate-drain capacitance in an RF-sized CMOS device is about the same as the base-collector capacitance expected in a bipolar transistor. The transformation ratio, and hence the Q, of the output matching network for each amplifier is predominantly set by the inductor loss, and the same spiral inductor has been used in both designs. On the input side, were the matching inductances lossless, the

[52] Resistance in the gate terminal, as described by Equation 5.8, is added extrinsically.

[53] Recalling from Figure 5-6, it is worth noting again that the parallel stage design employed to provide an input impedance match at two bias current levels cost the bipolar implementation about 4.5dB in available gain.

[54] Inductor quality factors in the high teens (at 5.8GHz) are being assumed in this design as well.

[55] The lower gain available from the CMOS transistors compared to the bipolar devices also takes a bite, increasing the noise penalty associated with adding the cascode. Contrary to what is sometimes written, the noise of the cascode is reduced by G_a of the input transistor; however, the noise of the cascode device is generally magnified by the high (and highly non-optimal) source impedance it sees from the lower device.

CMOS LNA would have a higher Q impedance match due to C_{GS} of the input device being smaller than the corresponding C_{π} term in the bipolar implementation. With on-chip spirals, however, the loss in the inductors tends to scale faster than the inductance, and the higher loss experienced in the CMOS design acts to widen its input bandwidth. Regardless, if the numbers seen in these simulations are any indication, a 5.8GHz LNA in a 0.5μm CMOS technology appears quite usable for the types of wideband wireless applications being considered throughout this work.

5.2.4 Layout Design of the 5.8GHz CMOS LNA

Although a number of different technology-based limitations have been encountered along the way, the approach to designing an LNA in CMOS has mirrored the switched-stage effort in many ways. The layout of the IC is no exception; designated CMLNA-CM (the suffix indicates CMOS), the die photo that appears in Figure 5-17 illustrates the striking resemblance. As the input impedance characteristics of MOSFETs have lent themselves to a single stage amplifier, there are two fewer inductors and a correspondingly smaller 1.03mm by 0.96mm die area.[56] Residing on the left side is the RF input while the output is taken from the right, both of which make use of RF-shielded GSG pads and on-chip 50Ω microstrip traces leading to and from the amplifier. Noting the 4-turn spiral—in place of the 1.5-turn L_B—encountered immediately after the input pads, the larger matching inductance needed for L_G is clearly evident. In helping to reduce coupling into the substrate, each of the inductors—as well as the poly resistors and MIM capacitors—is formed over a grid of the deep trench isolation pockets available in this technology.

Looming along the bottom of the photo is a set of pads that allows the chip to be powered from the same 4 power, 3 ground, bypassed supply probe which will be used for the switched-stage LNA. The far left and far right pads in this row furnish the supply voltage, where the pin farthest to the right provides a separate connection for the biasing circuitry. Adjacent to this pin is the *bias_adj* control, leaving the extra pad—freed by not needing a separate mode select input—to be used as an off chip connection to the gate voltage of the p-channel current mirror. Should the start-up circuit fail to initialize the regenerative bias cell, this pin can be stimulated as an alternative means of escaping the zero condition. The bias circuit itself is quite compact, fitting entirely inside a 170μm by 100μm rectangle. Finally, protection from ESD transients is included in the same manner

[56] For the curious: yes, this die photo is shown (approximately) to scale relative to the switched-stage LNA.

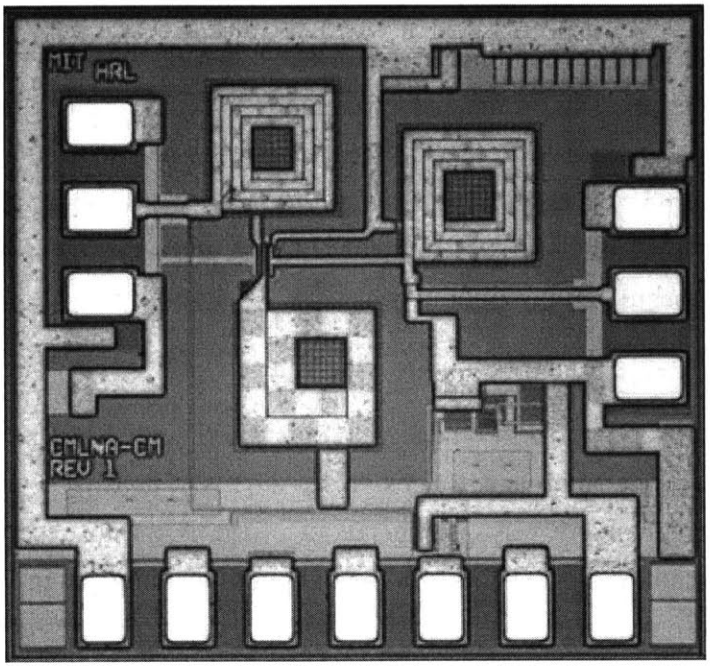

Figure 5-17. Die photo of 5.8GHz CMOS LNA IC (CMLNA-CM).

prescribed for the bipolar design, with extended drain MOSFETs placed between the supplies and ground and double diode clamps on the analog input.

Looking back, this LNA design began with a CMOS transistor having a G_{ma} of 10.6dB and a 50Ω noise figure of 2.7dB within its reach, then finished with 10dB of gain and a noise figure anticipated to be in the 4dB neighborhood.[57] At least in terms of noise, this would seem to suggest that the CMOS technology has been burdened to a greater extent (than the bipolar) in supporting the requisite degree of return loss at the input and output along with reasonable amounts of gain and isolation. The 1.3dB difference between the "starting" and "ending" noise figures for the CMOS design is the "reality tax"; in comparison, the bipolar version finished within 0.8dB of the intrinsic device NF_{min}. Hence, the additions required to turn a raw device into a useful LNA have seemed to exacerbate the noise disparity between the two device technologies. While this conclusion has been based on the most complete information available at the time of the design, the question of salience for the design of wireless systems is: "Given bipolar and CMOS technologies at the same lithographic node and providing the same accompaniment of passive devices, what is the real difference in performance?" The answer comes next.

[57] The achievable noise figure of 2.7dB is derived from the "linear mental construct" described in Section 5.2.1. The y-offset in this construct is the 2.5dB NF_{min} calculated from the BSIM model and shown in Figure 5-13; the slope portion is the 0.2dB penalty for failing to provide an optimum noise match.

5.3 Measured Results for the 5.8GHz LNAs

Two 5.8GHz LNA designs have been completed, one entirely in CMOS and the other relying upon bipolar transistors. Both designs provide fully-integrated 50Ω in, 50Ω out solutions with on-chip biasing and a means of saving power when reduced sensitivity is permissible. To provide the most direct comparison possible, the designs have been fabricated together, sitting beside one another on a set of wafers produced in a 0.5μm BiCMOS technology. All characterization has been performed on wafer using an ATN noise measurement system in conjunction with an HP8510 network analyzer, Cascade Microtech AirCoplanar® GSG probes, and a DC probe card with 100pF bypass capacitors located at each of the four power pins. *With the RF pads having been incorporated into the on-chip matching networks, no deembedding is performed on the measured data*—the measurement system is simply calibrated to the probe tips using an impedance standard substrate. The setup from ATN also includes an HP8970B noise figure meter and an automated tuner along with a set of isolators mounted at the probes to provide accurate 50Ω noise figure, minimum noise figure, and Γ_{opt} measurements.

First to come under the microscope is the switched-stage version, for which the noise figure and gain are reported in Figure 5-18. Both stages prove operational;[58] the high performance mode dissipates 8mW when provided a 3V supply, and switching to the low power setting reduces the power by a factor of four. Moving beyond these functionality checks, the first observation from the RF measurements finds the (high performance) noise figure and gain characteristics centered around 5.1GHz—about 12% lower than intended. As the former is tuned by the input match and the latter more predominantly located by the output matching network, these two measures align by virtue of tracking among the spiral inductors.[59] If the models used for representing the inductors have missed, it may be presumed from the data that they have missed in a consistent fashion.

Also falling somewhat short of expectations is the noise figure, coming in at 3.6dB in the high performance state and 10.4dB when the low power stage is selected. The disparity between these numbers and the corresponding simulation values cannot be pinned completely on impedance mismatches however, as the minimum noise figure determined for each mode is also appreciably elevated—by 1.1dB and nearly 4dB, respectively, in the

[58] About this point, the reader may have mixed feelings. This chapter would have been *much* shorter had this not been the case.

[59] The various transistor and MIM capacitances also play a key role in amplifier tuning. In general, however, these components cannot be expected to track across processing variations. Hence, if the input and output impedances are aligned in frequency, the more likely scenario is that the devices on this process run met nominal expectations.

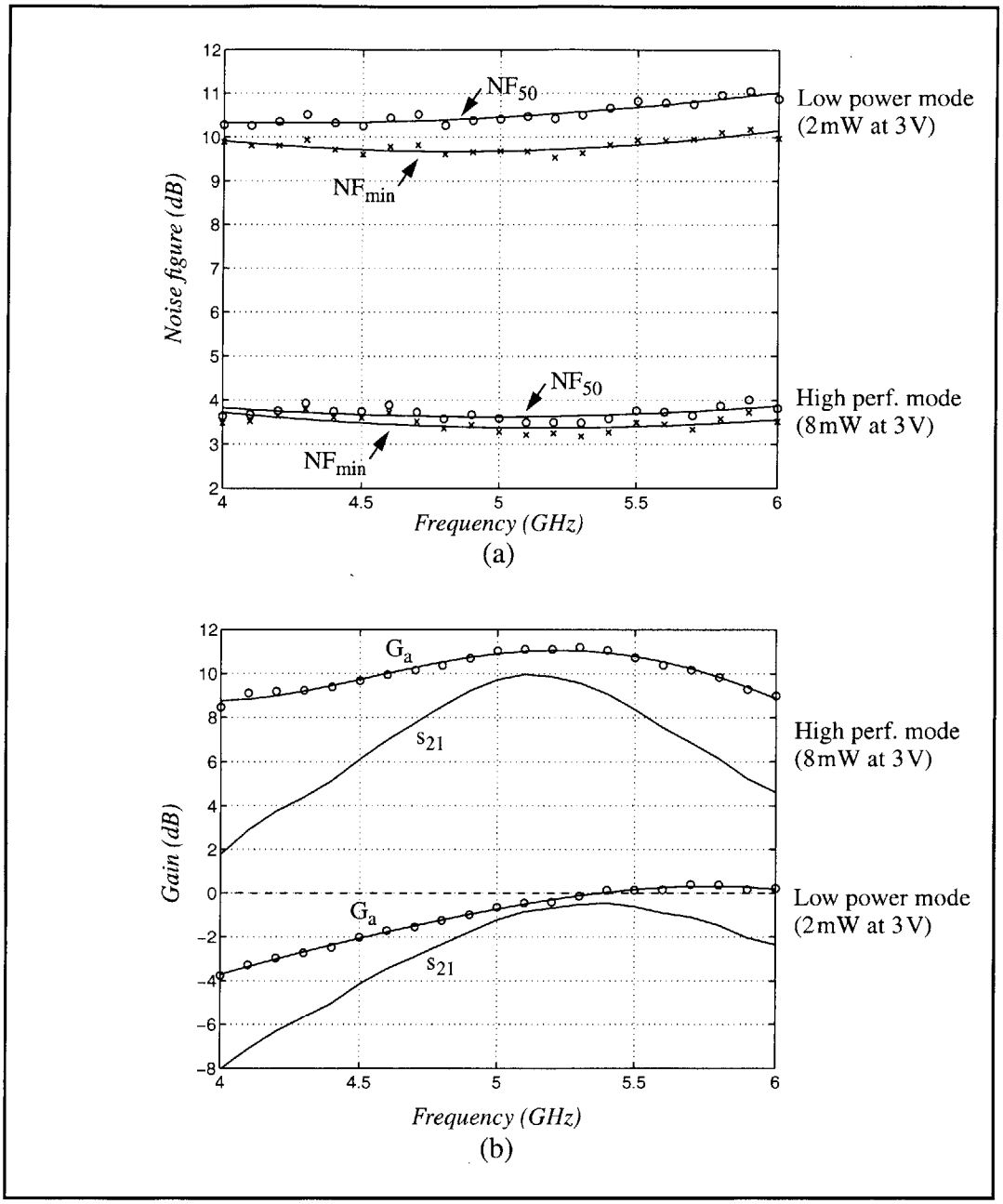

Figure 5-18. Measured noise figure (a) and gain (b) of the 5.8GHz switched-stage bipolar LNA in both the high performance and low power modes. Note the shift in the center frequency of the LNA pass band.

8mW and 2mW settings. In fact, that the 50Ω noise figure (for the high sensitivity stage) comes within 0.2dB of NF_{min} suggests that the optimal noise match is not too far removed from the 50Ω presented by the RF source. Similar trends are observed in the gain characteristics, where the insertion gain peaks at 10dB for the higher gain mode and -0.5dB for the lower. Likewise, the output match is not solely to blame as s_{21} in both modes

approaches to within 1 dB of the available gain measured for the LNA. The G_a numbers here, and as shown throughout this section, have been quantified with the source reflection coefficient set to (the measured) Γ_{opt}.

Further insight into the frequency response of the amplifier can be found in the s-parameter data of Figure 5-19. As seen in part (d) of this figure, the gain characteristics just described mirror the output impedances as reflected in the return loss measurements; the peak in gain is reached at the point where the impedance converges most closely to resonance. The distinction in the depths to which the $|s_{22}|$ curves plunge indicates a dif-

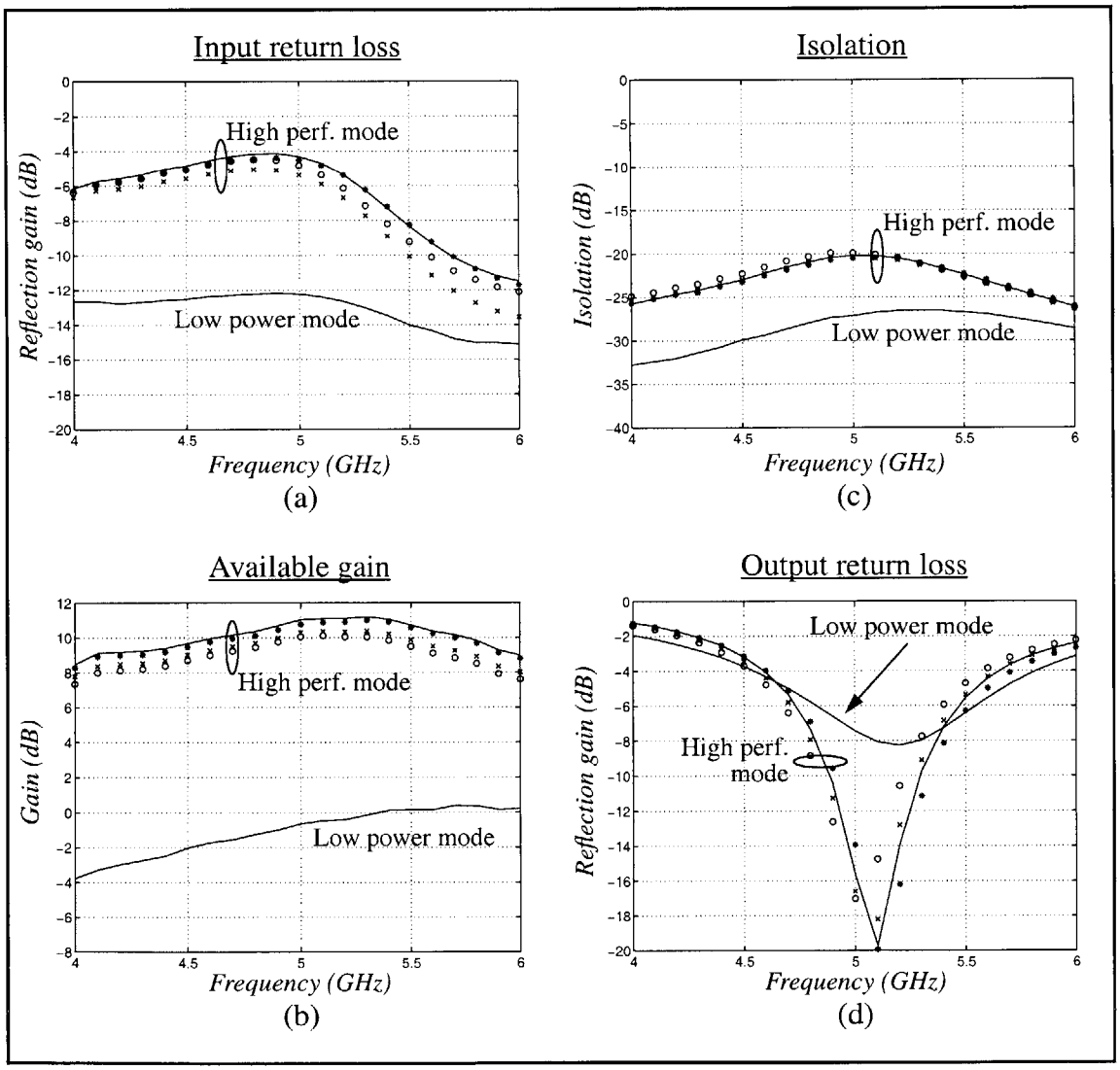

Figure 5-19. Measured s-parameter data for the 5.8 GHz switched-stage LNA; the input return loss (a), gain (b), isolation (c), and output return loss (d) are shown. Solid traces indicate the "default" bias conditions for operation at 3 V. Measurements are also provided in the high performance mode for supply voltages of 2 V ('o'), 2.5 V ('×'), and 4 V ('*').

ference in the effective output resistance between the modes, suggesting that some cancellation of inductor loss is in fact occurring in the high performance state.[60] On the input side, return loss is better than 12dB across the 4-6GHz span when the common-base input stage is active. The more sharply-tuned common-emitter input exhibits its best match near the 6GHz frequency anticipated in the design, although an additional loss mechanism seems to be prevailing for frequencies below 5GHz. That the best noise figure and input match do not arrive at the same frequency had been foreseen during the design and results from correlation between the equivalent input voltage and current noise sources. The remaining s-parameter metric is isolation, and the switched-stage LNA provides more than 20dB in both modes. The proximity of the s_{12} measurements to the simulations—which had only accounted for transistor-related coupling paths—can be taken as a vote of confidence that the amplifier is not being limited by pad-to-pad or inductor-to-inductor mechanisms. Also noteworthy in the isolation behavior is the disparity witnessed between the two modes; as the two stages should be comparable from the transistor perspective, the fact that the reverse gain is significantly smaller in the low power mode lends further credence to the notion of there being more loss in this path.

Finally, measurements are shown for the high performance mode of the LNA operating under a number of different supply voltages. A slight shift in frequency is evident in the output impedance as the change in voltage is impressed across the junctions to either side of the collector region in each cascode transistor. The input return loss moves around by a decibel or two due to the finite supply rejection in bias circuit, but only seems to show improvement relative to the 3V case. Then there is the available gain; the gain that can be squeezed from the LNA increases by about 1dB as the supply is hiked from 2V to 4V, an improvement which can be traced to a deeper depletion of the base-collector capacitance in the cascode devices. Although the 3V trace appears slightly higher than the numbers at 4V, this is not likely a significant juxtaposition: the 3V data presented is an average across several measurements while the voltage dependence of the LNA has been characterized from a single site.

Not too surprisingly, much the same story is told by the CMOS version of the LNA. Similarly depicted in Figure 5-20, the peak amplifier insertion gain and lowest 50Ω noise figure reading are found at 5.4GHz. Afflicted with a higher than anticipated noise figure and a lower than expected gain, the high performance setting for the CMOS LNA

[60] As discussed in Section 5.1.1, the output resistance of the high performance cascode stage is designed to be slightly negative. This helps to improve the gain in the face of high inductor losses.

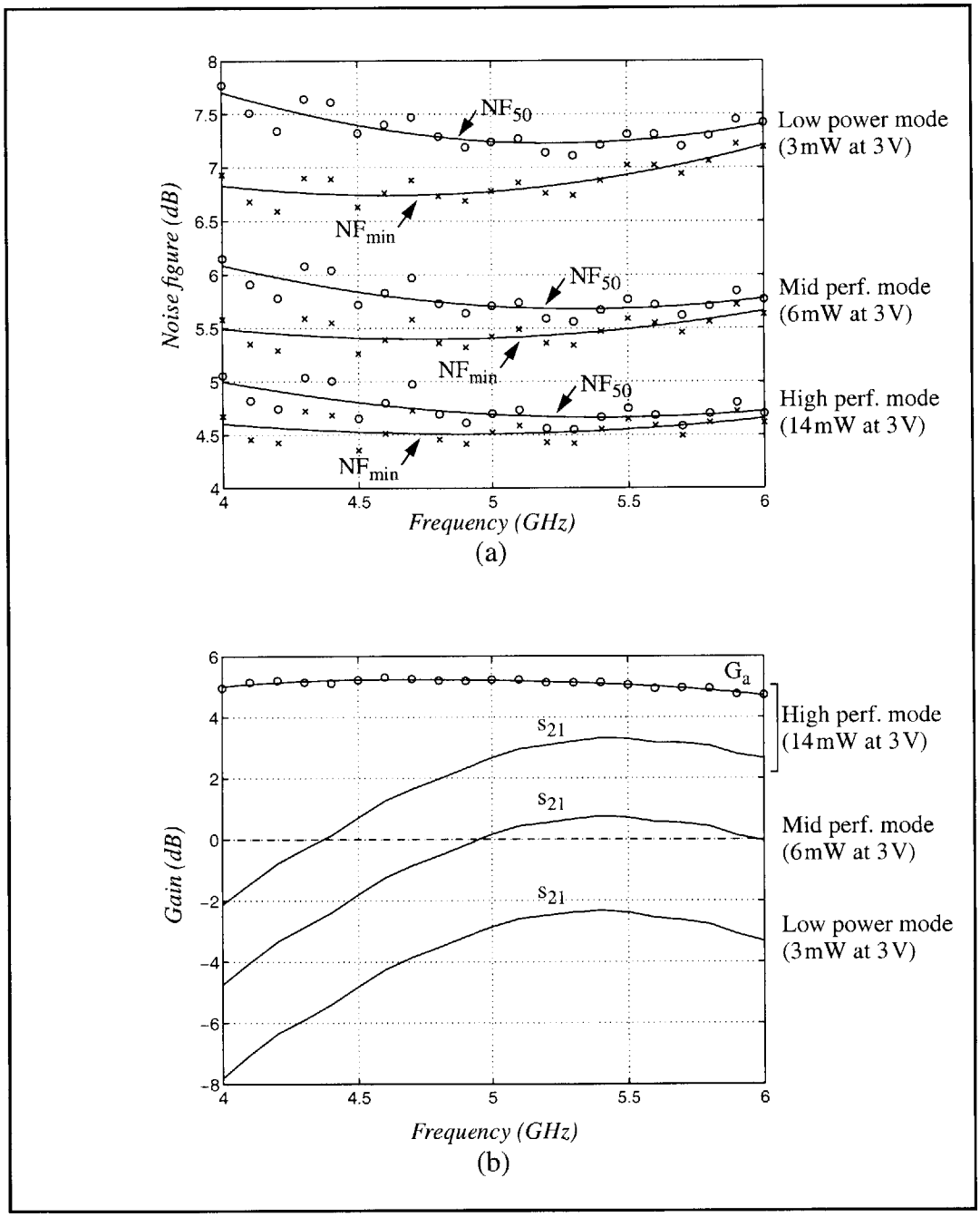

Figure 5-20. Measured noise figure (a) and gain (b) of the 5.8GHz CMOS LNA at three bias current levels. Note the change in scales for the noise figure and gain relative to the switched-stage measurement plots. To improve legibility, G_a has only been shown at the 14mW setting.

manages a 4.7dB NF and an s_{21} of 3.3dB.[61] Not helping the gain situation is that the output impedance is evidently mismatched to a greater degree than manifest by the switched-stage design, a conclusion that follows from having a maximum insertion gain that falls almost 2dB below G_a. The input matching network came closer to the mark, providing a 50Ω noise figure within 0.2dB of NF_{min}. In addition to the 14mW mode, two other bias levels are illustrated. The RF performance of the CMOS amplifier is observed to degrade rather gracefully as its bias current is decreased from 4.67mA to 2mA, at which point sensitivity begins to drop more sharply for further reductions; in light of this behavior, a "mid-performance" label is associated with the 2mA (6mW) condition and the "low power" mode is chosen as having 1mA flowing in the LNA stage. The first factor of two (actually 2.33) cut from the power consumption is obtained at a modest cost (1dB) in noise figure and 2.5dB of gain. The next factor of two (lowering the dissipation to 3mW) leads to a further increase of 1.5dB in the noise figure and another 3dB of gain being lost.

The same three bias levels are represented in the s-parameter measurements shown in Figure 5-21. Remarkable for their austerity, the plots for input return loss, output return loss, and isolation illustrate that the properties of the CMOS LNA—save for gain and noise figure—are virtually unchanged as the bias current is adjusted over its useful range. Consistent with being low Q, the impedance match realized at the input is rather wide if not terribly deep; while the return loss only reaches 10dB at 6GHz, the 3dB bandwidth (assuming a symmetric response) is apparently 2GHz. Also on the low side is the return loss at the output, which never quite exceeds 6dB. A poor match here had previously been surmised in comparing the insertion and available gains of the LNA, and suggests that the output resistance of the CMOS cascode stage is not as high as had been forecast by the BSIM models.

But rather than trying to guess from magnitude plots, a greater level of insight can be gleaned from looking at the complex impedance data when presented on a Smith chart.[62] In this spirit, the input, output, and optimum noise impedances are shown below in Figure 5-22 for the high performance mode of both LNA designs. Adhering to the usual reference directions, s_{11} and s_{22} progress in clockwise directions with increasing frequency; by convention, Γ_{opt} refers to the reflection coefficient that should be presented by the source to minimize the amplifier noise, and thus increasing frequencies carry it

[61] Though designed to provide 5mA, grounding the *bias_adj* pin actually yielded a slightly smaller than nominal 4.67mA bias current in the LNA, resulting in a dissipation of 14mW at 3V.

[62] To borrow a quote which has been attributed to Yogi Berra, "One can see a lot just by looking." Yogi was probably not talking about Smith charts.

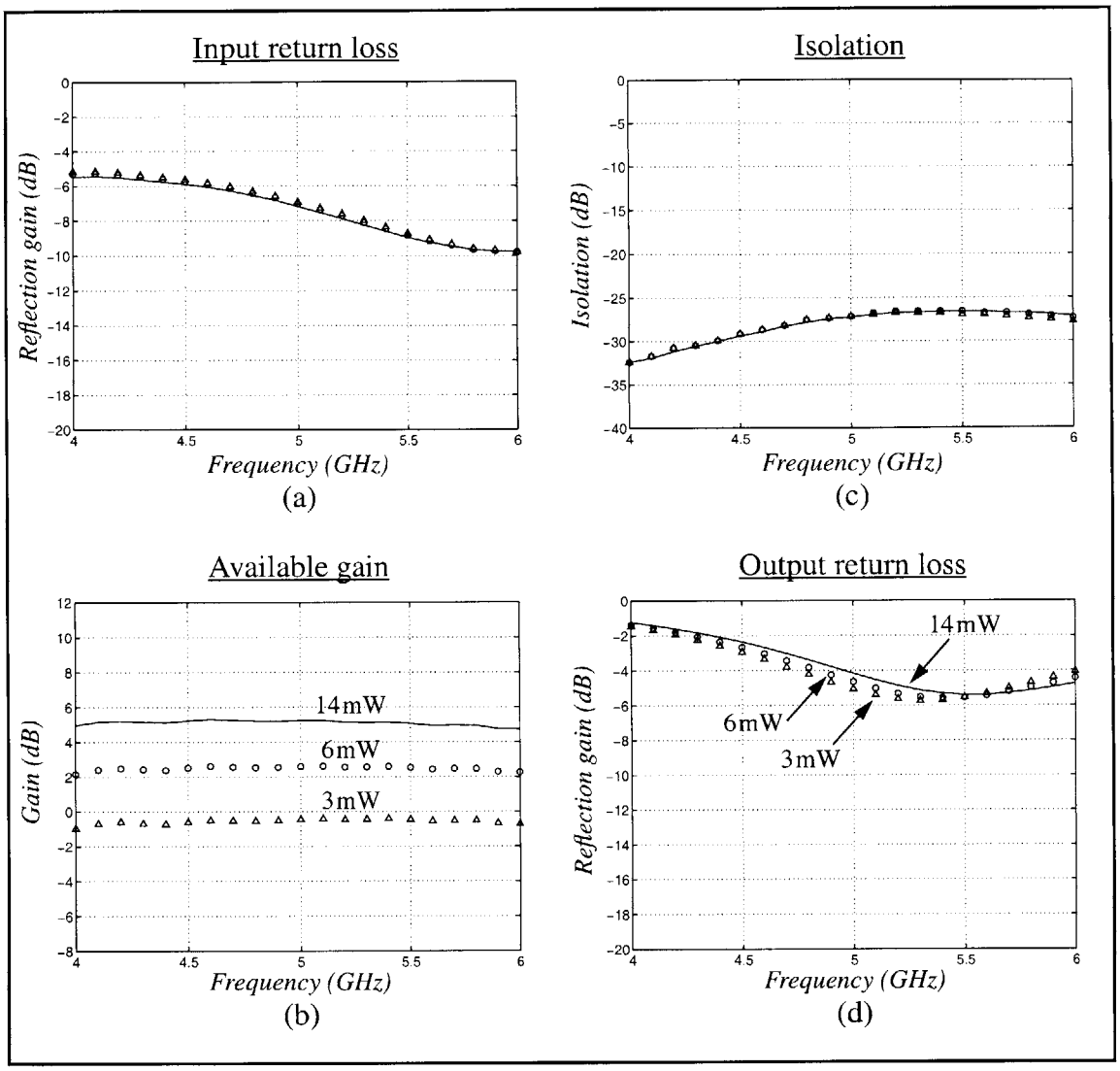

Figure 5-21. Measured s-parameter data for the 5.8GHz CMOS LNA; the input return loss (a), gain (b), isolation (c), and output return loss (d) are shown. Solid traces indicate the "high performance" bias condition (4.67mA) for operation at 3V. Measurements are also provided as the bias current is reduced to 2mA ('o') and 1mA ('▵') with the 3V supply.

counter-clockwise instead. The first observation in the data, pertaining to the output impedances, is that both amplifiers clearly remain in the capacitive half of the chart. For the bipolar version, s_{22} strikes an arc of a fairly large radius—a picture consistent with the high output impedance of the cascode together with the spiral inductor. However, as the frequency increases from 4GHz to 6GHz, the radius decreases steadily as the loss mounts; the impedance never reaches the real axis (i.e. resonance) as more loss is evident in the characteristic than had been predicted. One thought behind the increased loss might be the pads, but measurements of an on-wafer open structure reveal that the RF-shielded pads

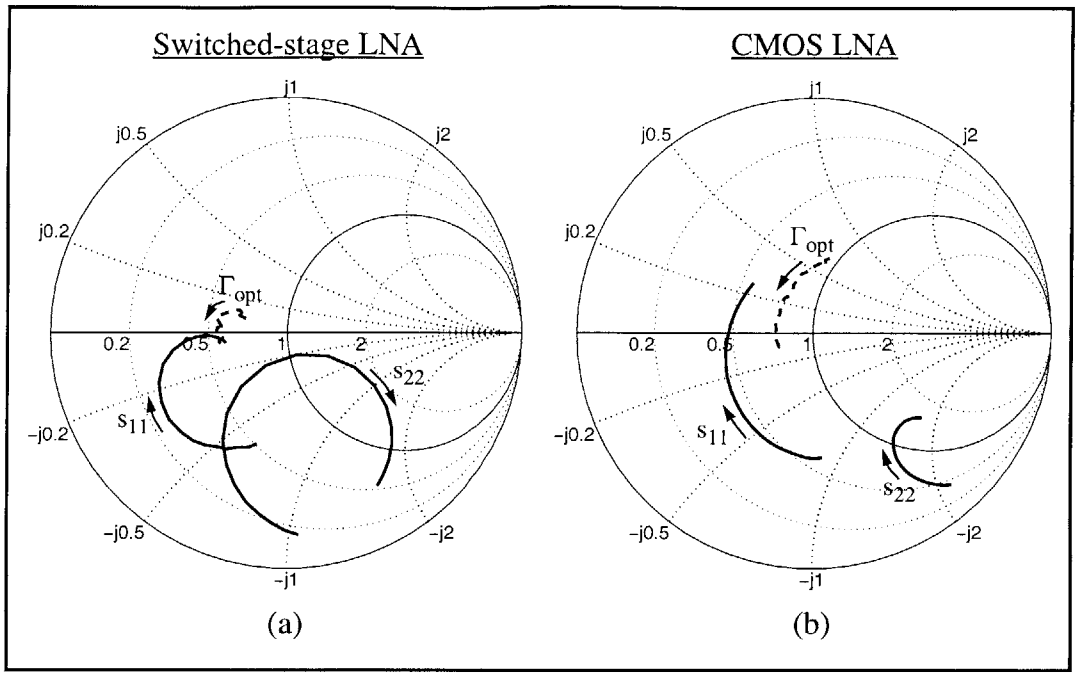

Figure 5-22. Measured input impedance (s_{11}), output impedance (s_{22}) and optimum noise impedance (Γ_{opt}) for the 5.8GHz switched-stage (a) and CMOS (b) LNAs. Arrows indicate the direction of increasing frequency.

consist of only 47fF of capacitance to a ground plane with 25Ω of resistance. Although this is slightly higher on both counts than the model used for the purposes of design, the pad still represents an inconsequential degree of loading.[63] By way of ruling out other possible culprits, wafer kerf data from the process run indicate that the MIM capacitors used in the output matching network are very close to their nominal values, and that the parameters of the NPNs are well within specified limits. Thus, with nowhere else to turn, the finger of blame points squarely at the inductor.

To quantify the differences being seen, a lumped element model for the 2.7nH output matching inductor can be fit to the s_{22} measurements. For the purposes of this parameter extraction, the output impedance of the cascode stage is again represented by the parallel RC network illustrated in Figure 5-2, hence foisting the difference in its entirety onto the inductor. Even with this assumption in hand, there are still more degrees of freedom in the model than can be resolved from the frequency range for which there is available data. This, however, does not prove to be an impediment; irrespective of how the fitting process partitions the loss between the winding and the substrate, the amount of loss being incorporated into the model—and hence the extracted quality factor—is the

[63] The pad capacitance is included output matching network and thus the increase does affect the impedance transformation ratio slightly.

same. Whereas the Q expected of this inductor had been 17, the answer returned by the fit is a considerably lower 5.4 at the same 5.8 GHz.[64] Fixing the series resistance term in the model at the FastHenry-computed value, the capacitive loading that shunts the inductor is determined to be twice that inferred from the scalable modeling approach discussed in Chapter 2.[65] The extra loading also substantially elevates the substrate loss and thus acts to lower the gain as well as the center frequency of the amplifier.

In comparison, the output reflection coefficient of the CMOS LNA—which uses the same inductor—displays a much tighter radius. The high degree of curvature signals the presence of a lower Q matching network, where the lower Q follows from the smaller output impedance exuded by the CMOS cascode stage. In fact, repeating the procedure of fitting the inductor parameters to the s_{22} measurements establishes that the impedance at the drain node of the CMOS amplifier is not even living up to the predictions of the BSIM transistor models provided for this technology. Somewhat artificially forcing the inductor model to accommodate the measured data, a quality factor in the mid-threes is seen rather than the mid-fives as determined from the switched-stage LNA. From this observation it is reasonably safe to assume that the CMOS transistor models have overestimated the output impedance of the cascode, explaining why the return loss and gain characteristics of the CMOS LNA have fallen even farther short of expectations than was seen with the bipolar implementation.[66]

Conversely, the input matching network for both amplifiers consists of a series resonant circuit, for which the trace should appear in the Smith chart along a circle of constant resistance. However, owing to the finite level of isolation between the input and output ports, the impedance at the output can be "felt" at the input.[67] Interaction of this nature is particularly noticeable in the bipolar LNA, where the output network lends a parallel resonance curvature to the input impedance. For the CMOS design, possessing a

[64] This number is further corroborated by the on-wafer characterization of one of the spiral inductors used in the VCO designs presented in the next chapter.

[65] The oxide capacitance in the extracted model is 50% higher than that calculated from the scalable model, a difference that is further magnified by a factor of four reduction in the substrate resistance.

[66] Despite the relatively mild appearance to the decrease in the quality factor assigned to the inductor in this experiment, the drop from 5.4 to 3.4 actually suggests that the output resistance has been substantially overestimated. Even with a comparatively small loss in the inductor—as with the originally expected Q of 17, the transistor stage was expected to play only a small role in the effective output resistance at the drain node. Further increasing the inductor loss (e.g., a Q of 5.4) should have rendered the cascode output resistance inconsequential. As the data illustrate, however, the transistors are clearly having a noteworthy effect.

[67] The reverse case is also true, but the relatively low quality factor of the input matching network does not sway the output impedance as prominently.

lower Q load and an extra 6dB of isolation, the s_{11} curve hews more closely to the classical contour of a series resonator. But regardless of the shape, it is immediately evident that both LNAs have fallen well short of having 50Ω input resistances. There is not much room for confusion about where the responsibility lies for this one, as the real component of the input impedance is the sole province of the degeneration inductances. Both amplifiers present resistances in the 20-25Ω range, meaning that the actual inductive impedance delivered by the inductors L_E and L_S is a factor of two to two and a half less than expected. Judging from the parameters extracted from the s_{22} data, the most likely scenario is that the degeneration inductors are similarly being shunted by more substrate capacitance than had been tendered by the prognostications of the scalable model. In further noting that imaginary components of s_{11} and Γ_{opt} are about right—for both designs, it may even be feasible that the extra capacitance in the spirals is proportionately compensating to keep the series resonant frequency of the input circuit approximately as intended.

Finally, the fact that the dashed lines for Γ_{opt} are both found outside of the 50Ω circle is deserving of a few words. This comes as something of a surprise for the CMOS LNA in particular, with which it had been believed that even a 50Ω $R_{s_{opt}}$ would be out of reach due to a prohibitively high power consumption. Blame cannot rest solely with the inductors however; while increased capacitive loading at the base or gate of the input transistor would act to lower the optimum source resistance, the heightened resistive loss in the spirals would tend to counteract this trend. In the bipolar design, it also is entirely possible that the 15% by which the emitter area of the input transistor had been increased to compensate for exogenous thermal noise sources has overshot the target.[68] For the CMOS implementation, where the power constraint kept the expected $R_{s_{opt}}$ well above 50Ω, the much lower measured value is more likely a sign that the long channel approximation is failing to accurately predict the drain and induced gate noise currents—even in these 0.5μm devices. With the n-channel transistors comprising the LNA having an f_T in the 15GHz range (for the high performance mode), the contribution of *both* current noise sources is magnified as the small-signal current gain in the RF pass band is only three.[69] Unfortunately, the extraction of the excess noise parameters in short channel devices is a very difficult procedure, often leaving guesswork as the only viable route for the purposes of designing low noise stages.

---

[68] The RF input to the switched-stage LNA touches four spiral inductors, each of which contribute thermal noise to the equivalent input noise voltage of the amplifier. In the first iteration of the design (Figure 5-1), the common-emitter device size was set at 6 x 0.5μm x 9.2μm, an area which included an 8% hike to compensate for inductor loss. Once the second (parallel) stage was added, the emitter area increased again, finishing at 6 x 0.5μm x 9.8μm.

However indirectly, the analysis of the impedance measurements has provided some clues as to the whereabouts of the absent gain and the origin of the higher noise in the LNAs. It would appear the that drop in inductor quality factor—from numbers in the high teens to something more along the lines of 5-6—has cost each LNA about 1dB in noise figure and a few decibels of gain. Impedances mismatches at the input have taxed the noise figure another 0.2dB in the high performance modes, and shortfalls of 1-2dB in the insertion gain might be chalked up to imperfect output transformations. But in spite of the issues with the scalable inductor models, inductor Qs that are merely average, and operating the 0.5μm transistors to within a factor of 3 to 4 of their unity current-gain frequencies, the LNAs emerged reasonably well. If nothing else, the designs provide an excellent bipolar versus CMOS comparison in that they operate under the same set of conditions, limitations, and design constraints; the only distinction being that the slightly poorer output match of the CMOS version perhaps cost an additional 1dB of insertion gain. For convenience, a summary of the measured performance is tabulated below.

Table 5.1: Performance Summary for the C-band Monolithic Low Noise Amplifiers

	Bipolar		CMOS		
Power dissipation @ 3V	8mW	2mW	14mW	6mW	3mW
Noise figure	3.6dB	10.4dB	4.7dB	5.7dB	7.2dB
Peak insertion gain	10dB	-0.5dB	3.3dB	0.8dB	-2.3dB
Input return loss	>8dB, 5.5-6GHz*	>12dB, 4-6GHz	>8dB (all modes), 5.2-6GHz*		
Output return loss	>8dB, 4.8-5.3GHz	>6dB, 4.9-5.6GHz	>5dB (all modes), 5.2-5.8GHz		
Isolation	>20dB, 4-6GHz	>27dB, 4-6GHz	>27dB (all modes), 4-6GHz		
Inductor quality factors are in the range of 5 to 6 (at 5.8GHz)					

*Measurement data is only available for frequencies up to 6GHz

[69] This points out a rather important distinction between the sources of noise in CMOS and bipolar devices as the design frequencies begin to push technological limits. Base current shot noise in bipolar transistors does not appreciably increase with frequency, whereas the equivalent induced gate current noise source representation in a MOSFET does increase. Alternatively, the appearance of base shot noise in the output current of a bipolar transistor has a spectral shape that mirrors the current gain—declining for higher frequencies. In a MOSFET, the output spectral density due to induced gate noise is white.

5.4 Lessons in Low Noise: The Sequel

What began as an "exercise" in designing LNAs has turned into a marathon.[70] Device-level technology considerations have been interpreted and applied to two amplifier designs, both of which seek to address some of the more salient features needed for enabling portability through high data rate wireless networks. The focus throughout has been upon comparing CMOS and bipolar technologies for their suitability in RF applications where sensitivity and power considerations push physical device limitations, and to illustrate how intrinsic device measures such as minimum noise figure and maximum available gain translate into "real-world" LNA performance. Although the models of small-signal behavior in field-effect and bipolar transistors are essentially the same, differences in the values and dependencies of the device parameters can lead to significant departures in the limitations that result. While the basic approach to thinking about noise, impedances, and designing low-noise amplifiers remains the same, finding and extracting the optimal performance afforded by each technology is more involved than simply substituting one device for another within a circuit.[71]

When this exercise began, one may have wondered whether the challenges of constructing a usable LNA might wash out the importance of the intrinsic device noise figure. From the examples developed here, this does not seem to be the case. The initial disparity in the minimum achievable noise figure between the 0.5μm CMOS and bipolar technologies was maintained, and possibly even exacerbated, as the design progressed from a bare transistor to a finished amplifier. Given the same load inductor and though consuming nearly twice the power, the noise figure of the CMOS LNA still came in 1.1dB higher, and provided a gain that was 6dB lower, than its bipolar counterpart. Expounding from both simulations and measurements, the gap observed in sensitivity and power appears to hold regardless of whether the available inductors are very good (e.g., a Q of 18 at 5.8GHz) or merely average (e.g., a Q of 6).

Turning the tables for a moment, a CMOS technology may seem to offer a couple of advantages for the LNA in a wideband wireless receiver. First, the gate of a MOSFET provides an open circuit in the DC bias path, easing concerns about the trade-off between noise and linearity in choosing the bias coupling impedance. Another virtue is that the terminal impedances hardly change with bias current, enabling a far greater degree of flexi-

[70] Of course, all of the footnotes have not helped the situation.

[71] The considerations that lead to choices in circuit topology and device operating points may change as technologies evolve. This is as true within a given device style as it is between FET and bipolar processes.

bility in adapting to real-time conditions. The only properties of the CMOS LNA in which any change was elicited were the gain and noise figure, an ideal situation for adaptively trading performance for reduced power consumption within the receiver. These advantages can be replicated, however, where the implementation of a switched current source biasing technique has been shown to furnish similar functionality in a bipolar design while still delivering better RF performance.[72]

Despite all of the attention being garnered by the transistors, the passive devices play an equally prominent role in the capabilities of an LNA. Improving the inductors from the average (the extracted Qs of 5 to 6) to the sublime (the Qs of 17 to 19 expected from the scalable models) projects to lower the noise figure of these designs by approximately 1 dB, a number on the same order as the difference between the CMOS and bipolar versions of the LNA. Perhaps of comparable importance to lessening the loss in the inductors is to facilitate their optimization for each intended application through modeling. Unfortunately, the analytical approach discussed in Chapter 2 has not yielded the level of accuracy required for a first-pass success. Having a large enough set of well-characterized spirals to allow a robust interpolation of inductor parameters based on geometric dependencies may help with this approach. The other option is to use a numerical electromagnetic field solver for each inductive structure being employed.

But even without accurate models, knowledge of the actual loss may have led to rethinking the inclusion of some of the inductors in the designs. In particular, the input matching inductor for the common-base stage (L_{E2}) may have hurt more than it helped. The substrate parasitics of L_{E2} (and also of L_{DC}) provide a path by which signal power can be siphoned away even when the common-base transistor is off, in turn hindering the noise figure and gain of the high performance mode. It is at least mildly feasible that the input return loss of the low power stage could be made to suffice without this matching element. Another interesting alternative would be to explore the performance that could be achieved when a common-emitter input transistor is sized to yield an input resistance of 50Ω—without adding emitter degeneration inductance. Skipping the emitter inductor would forego an optimum noise match, but the resulting higher gain and lower loss might more than make up for the higher $R_{S_{opt}}$.[73]

[72] Nothing is free however; the implementation in the bipolar technology occupied an appreciably larger die area.

[73] While on the topic of pondering alternatives, it may be interesting to reconsider one of the constraints assumed for the switched-stage LNA design. The use of an RF switch was specifically avoided in the signal path due to the loss that would be incurred. Whether the inductors, with Qs of 5, might be exhibiting more loss than an RF switch used to decouple the two stages remains an interesting question.

Several other assumptions made during the design process are also worth revisiting. Among them, it had been believed that the output resistance from the cascode transistor stages would be high enough to be brushed aside by the equivalent parallel resistance afflicted upon the load inductor by loss elements within it. The picture of s_{22} in Figure 5-22, however, tells a different story. For the CMOS LNA, the lower output impedance may have as much to do with drain-substrate parasitics as with channel length modulation, although differences in the output resistance between the stages of the bipolar amplifier were also noted. Invoking some negative resistance to counter inductor loss in the load has evidently proven beneficial. Also deserving of a second look is the guess made during the design that a baseline for the noise manifest by MOSFETs could be established at high bias currents using the BSIM3v3 transistor models, and then adjusted for the effects of induced gate noise using the analytical work proposed by Shaeffer and Lee [76]. From examining the measured data, and after allowing for the higher than expected inductor loss, the best estimate is that the transistor noise figure is somewhere between the predictions offered by the two models. Modeling CMOS devices for RF applications is currently an active area of research [78], and is one that merits continued attention. Finally, the indications that an optimum noise match between a MOSFET and a 50Ω source resistance could not be achieved proved to be inaccurate (even for reasonable power consumption levels). Granted, this is one data point coming from a $0.5\mu m$ technology, and the direction in which optimum source impedances will move with scaling is not clear. The noise power associated with drain and induced gate noise currents does appear to be increasing with smaller feature sizes. However, rising correlation between the two and the higher f_T of scaled devices will both act to reduce the equivalent input noise current. Somewhere in the balance between these trends—along with the degree to which the terminal resistances can scale—an answer might be found.

So, is CMOS a viable alternative for LNAs? Sure. Will it ever "catch up" to bipolar devices for low noise, low power applications? Unless the lithographic dimensions applied to the bipolar processes lag cantankerously, there is no indication that it will. As the transistor unity current gain frequencies become ever higher, terminal parasitics play an increasingly important role. Whereas the gate resistance of field-effect devices tends to suffer with finer geometries, the base resistance in bipolar transistors improves. While the latter might more quickly become limited by contact resistivities, so to will the extrinsic source resistance in FETs [79]. Furthermore, the landscape of noise in CMOS transistors may begin to change as gate lengths reach 100nm. In particular, tunneling through the thin gate oxides will add a shot noise component to the gate current. As tunneling will

increase with V_{GS}, an optimum current density associated with the minimize noise will result, adding one more constraint to the design of CMOS LNAs.[74] Conversely, the absence of new noise sources and excess noise factors has made bipolar transistors seem almost bland in comparison. But when it comes to noise, less is definitely more.

[74] Actually, there is an already an optimum current density that minimizes the noise in FETs, even before considering tunneling current. For a MOSFET (without gate current), the optimum current density is that which maximizes the transistor f_T. When power is constrained, however, a lower noise solution may be reaching by operating a larger device at a sub-optimal bias current.

Chapter 6

An Experiment with Voltage-Controlled Oscillators

Be it for frequency conversion, synthesizers, or modulators, oscillators put the radio frequencies in radios. As the components which generate the carrier frequency needed for efficient transmission of information over a wireless link, oscillators are practically synonymous with RF and microwave. For RF applications, the preferred implementations rely upon resonators—within which the energy storage characteristics determine the frequency of oscillation and lead to lower noise operation than can be achieved with relaxation-based timing circuits. Most incarnations employ an electronically tunable resonant structure so that the oscillator can be made to track an input or reference signal by controlling it through a phase-locked loop. A varactor is typically used as a tunable reactance in the tank, giving rise to the commonly found voltage-controlled oscillator (VCO).

Basic theories regarding the processes that govern behavior in resonant oscillators are well established, and concepts for analyzing the stability of an oscillatory circuit can be traced back to the 1960s in a treatment by Kurokawa [27].[1] Beyond the frequency of oscillation, perhaps the most salient characteristic of an oscillator is its phase noise, which describes the spectrum of the noise that results from perturbations in the phase of the carrier signal. A method of calculating phase noise based on a linear time-variant (LTV) model was presented in Chapter 3 as a framework for thinking about device limitations within the context of oscillators. An experiment in applying this framework and examining the ramifications of the LTV model for technology development is described in this chapter through the design and measurement of a set of four 5.8GHz VCOs. Each of the designs incorporate biasing and buffering circuits on chip, and are constructed to allow an exploration into the extent to which the performance of the oscillator might be dynamically traded for reduced power consumption in a portable, wireless application.

[1] Kurokawa demonstrated that the oscillation amplitude-dependent characteristics of the negative resistance being used to supply energy to the resonant tank need to be matched to the form of the resonance. A more recent review of oscillator theory including Kurokawa's model appeared in [80].

6.1 A Family of Bipolar 5.8GHz VCOs

Whether the medium is discrete or integrated, the Colpitts class of resonant oscillators has seen its share of the spotlight from RF applications. While the topology can take a number of forms, the defining characteristic of a Colpitts oscillator is a capacitive voltage-divider that provides positive feedback around an active gain element. The feedback capacitors can also be employed as the electric energy storage elements in the resonant tank, affording an economy of devices that is difficult to surpass. Fortunately, the economy stops with the device count, as Colpitts circuits have empirically been recognized as offering excellent phase noise performance.[2] Featured in Chapter 3 as an LTV modeling example, the same common-base incarnation of a Colpitts oscillator reappears in Figure 6-1. This configuration possesses a number of attributes that are desirable in the realization of a low power, integrated VCO [81]. Among them, the frequency of oscillation is set by the parallel resonance at the collector node, delivering reliably predictable operation while maximizing the headroom allotted for signal swing across the tank circuit.[3] Another useful benefit is a separation of the bias node from the signal path, eliminating any bias coupling concerns that could otherwise require the addition of ungainly resistors or RF choke inductors.[4] The one possible drawback to this topology might be supply rejection; low frequency ripple in the supply voltage falls directly onto the cathode of the varactor, potentially effecting an undesired modulation of the carrier. However, it need not fall *across* the varactor. By referencing the oscillator tuning voltage (shown in the schematic as an RF ground) to supply, much of the sensitivity in the oscillation frequency to supply variations can be cancelled.

In designing with a Colpitts circuit, the ratio of the capacitors and the total capacitance in the feedback network are degrees of freedom which can be exploited for the purposes of noise reduction. The series combination of C_1 and C_2 conjoins with the inductor and varactor to set the resonant frequency, establishing a constraint on the product of the

---

[2] The reasons behind the prowess of the Colpitts topology were not well understood until linear time-variant (LTV) models were deployed as a means of studying oscillators. As recognized by Hajimiri and Lee [38] and discussed in Chapter 3, the discontinuous conduction of the transistor is timed via the feedback network such that current is only supplied to the resonator during a portion of the period when the oscillator phase is not easily perturbed.

[3] Oscillators can exhibit a variety of unintended behaviors, many of which can be traced to parasitic resonances and feedback paths that compete with the intended mechanism of oscillation. Fairly routine design practices can usually stifle these paths around the common-base topology drawn in Figure 6-1, although the challenge becomes more severe at higher microwave frequencies where the feedback capacitances may be swamped out by transistor parasitics.

[4] Albeit in different ways, the same issues of linearity, noise, and loss discussed for LNAs (Section 5.1.3) can surface in oscillators.

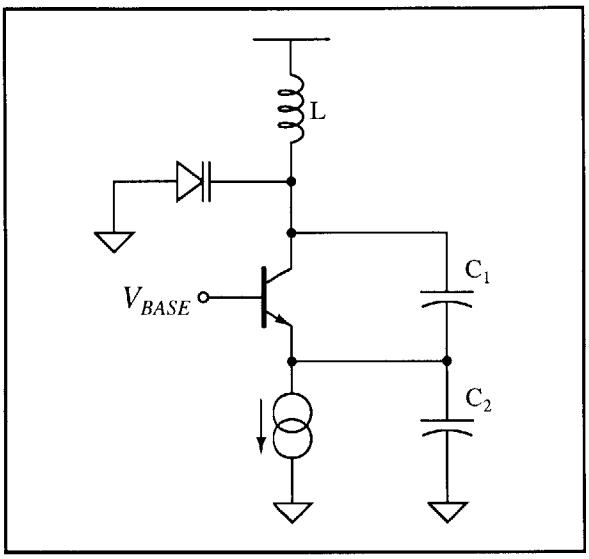

Figure 6-1. A common-base Colpitts-style oscillator circuit.

(total) capacitance and the inductance as $LC = 1/\omega_0^2$. Several approaches have been proposed for partitioning the resonant energy storage into inductive and capacitive components, each involving the maximization of some metric, such as: the characteristic impedance of the tank, the loaded Q of the tank, or the LQ product for the inductor.[5] Pragmatically however, at microwave frequencies, there is precious little room to move with the inductor value, nor much to be won by doing so. For reliable operation, the capacitance of the feedback network should be made large enough relative to the transistor capacitances such that it remains the prevalent signal coupling path in the oscillator—placing an effective upper bound on the inductance of around 1nH at 5.8GHz. Conversely, inductances smaller than a few tenths of a nanohenry are difficult to efficiently realize on chip, leading to the selection of 0.6nH for use in this design.

The other degree of freedom—the ratio of the feedback capacitors—can be used to balance between two competing effects. Applying a large feedback fraction (i.e. $C_1 \gg C_2$) increases the voltage swing at the emitter, thereby narrowing the conduction angle of the transistor and hence the interval over which it contributes shot noise. Unfortunately, as the passive feedback network works in both directions, a higher "turns ratio" in the capacitive

[5] In a simple parallel RLC model for the resonator, the equivalent resistance representing the loss due to a finite Q is determined as $R_{EQ} = QZ_0$; maximizing the characteristic impedance of the tank translates into choosing the largest feasible inductance. Obtaining the highest possible loaded Q for the resonator is just a more involved version of this criterion where the terminal resistances and capacitances of the transistor are also included in the analysis. Maximizing the LQ product results from optimizing across a number of oscillator variables, and was proposed by Ham and Hajimiri [82].

impedance transformer (i.e. a small feedback fraction) is needed to protect the resonator from loading.[6] As a general rule, Colpitts oscillators tend to produce the best results when the voltage division ratio, $n = 1 + C_2/C_1$, is in the range of 3 to 5.[7] The optimum value depends on the quality factor of the resonant tank and a number of other factors, but often lies within a fairly broad minima such that the oscillator phase noise will not change appreciably over this range.

6.1.1 Transistor Considerations in Oscillators

Having gotten the inductor and capacitors out of the way, emphasis can now shift to the active portion of the design. Excepting the current source indicated in Figure 6-1, the common-base Colpitts oscillator relies on but one transistor and thus it had better be chosen carefully. Before the recent embracing of time-variant modeling approaches, efforts in oscillator design rested upon an LTI model which recasts the problem as one of designing a low-noise amplifier with feedback placed around it. In this view, the passive feedback network regeneratively drives the gain stage to maintain an oscillation, doing so through an impedance that is scaled by the transformation ratio of the network. For the common-base implementation being considered here, loss in the resonator results in an equivalent load resistance across the tank of $R_{EQ} = QZ_0$; this resistance is transformed by the capacitive divider, in turn driving the input of the common-base transistor through a resistance R_{EQ}/n^2. From this vantage point, the oscillator gain stage should be constructed as an LNA where the optimum source resistance is designed to be R_{EQ}/n^2 [29]—a problem solved in the preceding chapter for a 50Ω $R_{S_{opt}}$ by sizing an NPN transistor with six 0.5μm x 8.5μm emitters and biasing it at 2.8mA of collector current.[8] Using this device, and assuming an equivalent load resistance of 500Ω for the common-base stage—after a brief consultation with the scalable inductor models suggests a Q of 20 should be feasible, the optimum 50Ω source resistance can be obtained for a turns ratio $n = \sqrt{10}$. Believing however that this ratio may permit loading to excessively mar the resonator Q, a moderately higher divisor is adopted in the design; already having decided upon an inductance of 0.6nH and leaving aside 0.45pF for the varactor, choosing values of 0.75pF and 2pF for C_1 and C_2 (respectively) provides a 5.8GHz center frequency and a turns ratio of $n = 4.3$.[9]

---

[6] Over the interval during which the transistor is conducting, the small-signal emitter resistance ($1/g_m$) of the common-base transistor will also load the resonator.

[7] A number of transistor capacitances are paralleled with C_2 from a small-signal perspective and add to the effective turns ratio. The most notable addition is C_π of the common-base transistor which is a capacitance typically on the order of C_2 for gigahertz frequency designs.

[8] This was the geometry initially calculated, prior to the incorporation of loss in the matching inductances.

As with an LNA, the LTI approach to designing oscillators adjusts the emitter area of the gain transistor while the bias current density is held fixed at that associated with the minimum noise figure (NF_{min}) of the device. In so doing, increasing the transistor size decreases the thermal noise contribution while the shot noise sources become more significant. Minimum noise occurs when these noise mechanisms contribute equally, a condition that is realized when the ratio of the equivalent voltage noise to the equivalent current noise (at the input) equals the source resistance.[10] In the LTI model, satisfying this criterion uniquely determines the optimal transistor size. However, when the same design is subjected to a calculation of phase noise via the LTV model described in Chapter 3, the view that emerges is decidedly different.

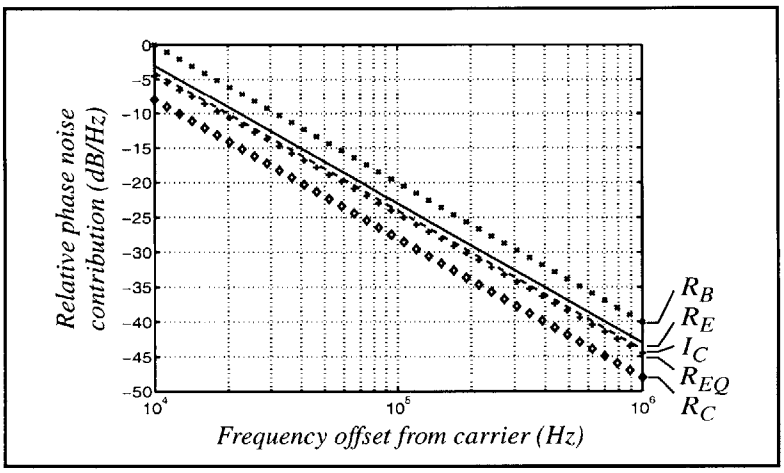

Figure 6-2. Primary contributions to oscillator phase noise from a transistor sized according to an LTI analysis.

Plotted on a relative scale, the five largest components in the f^{-2} region of the phase noise spectrum appear in Figure 6-2. It is evident from this listing that—despite the predictions of the LTI model—thermal noise from the base and emitter resistances remains the largest factor. As discussed in Chapter 3, the reason that the shot noise sources become less heavily weighted is due to the time-variant nature of the circuit. Current only flows in the transistor for the short interval about which the oscillation signal is near its minimum—the very moment that the sensitivity of its phase to the shot noise is at a zero

[9] The turns ratio of 4.3 anticipates an additional 0.5pF of capacitance at the feedback tap from C_π and the current source, and yields a lower than optimal 27Ω for the source resistance. However, judging from the LNA measurements in Section 5.3, this impedance should produce a noise figure within 0.2-0.3dB of the minimum.

[10] In what may be beginning to sound like a broken record, this definition of the optimum source resistance assumes negligible correlation between the equivalent input noise sources.

crossing—and, at all other times, there is no shot noise in the circuit. Conversely, the terminal resistances giving rise to thermal noise are always present. While their effects are also reduced when the transistor is not providing gain, noise from these sources can still couple into the resonant tank through the various capacitances (both explicit and parasitic) in the circuit.

One obvious implication is that the LTI oscillator models do not capture enough information to properly design the transistor geometries. As this example demonstrates, phase noise can be lowered by increasing the device size beyond that suggested by the LTI model. Another misleading conclusion steeped in LTI lore is that the optimum bias current for the oscillator is somehow associated with the transistor NF_{min}. In fact, the situation is quite the opposite: the bias current that minimizes the phase noise of an oscillator actually has very little to do with the transistor, but is rather a function of loss in the resonator and the available headroom for signal swing [82]. Despite all of the apparent complexities within an oscillator, the amplitude of the oscillation remains governed by the most basic of principles—Ohm's law.[11] For a parallel resonator, until limitations from the supply voltage begin to intercede, the oscillation amplitude follows IR_{EQ}, where R_{EQ} is the equivalent load resistance of the resonant tank, and I is the current flowing in the oscillator. In this current-limited regime, the energy stored in the resonator is proportional to I^2 while the shot noise energy in the oscillator increases only by I. For the worst-case scenario, if shot noise is the primary mechanism disturbing the oscillator phase, then the phase noise will improve linearly with the bias current.[12] This dependency prevails until the available voltage headroom prevents the amplitude from realizing further increases with I, at which point the noise performance either flattens out or—if shot noise is the dominant effect—may actually begin to degrade with higher currents. From this argument it is seen that the lowest phase noise is extracted from an oscillator when it is biased at the cusp between the current and voltage-limited modes of operation. Higher bias currents, at best, are just a waste of power. Similarly, increasing the supply voltage without also increasing the current will not materially improve the phase noise. *The key for realizing the best trade-off between power consumption and oscillator performance is to adjust both the bias current and the supply voltage in tandem*, a conclusion which will be directly verified through the measurements presented in Section 6.3.

[11] Implicit in this statement is the assumption that the oscillator is not near the hairy edge of start-up, that the power gain of the transistor is sufficient to cover the losses in the resonator at all of the bias currents being considered. At lower current densities, the oscillator may be operating in a region where the transistors are not switching completely.

[12] This definitely falls under the category of a plausibility argument, as all of the impulse sensitivity functions may also change with the bias current.

The desensitization of the oscillator phase to shot noise also bears consequences for the evolution of device technologies used for wireless applications. As noted during the development of the LTV model (Chapter 3), once the transistor f_T is already high enough relative to the frequency of oscillation, further increases in transistor switching speed will not yield better phase noise performance. As one example stemming from this observation, increasing the base thickness to reduce base resistance at the expense of a longer transit time (and a correspondingly lower f_T) would be a good device trade-off in this case for oscillators. A second possibility for consideration is to reduce the collector doping. This may run counter to the extant dogma of pushing the peak transistor f_T, but will also result in lower base-collector (C_μ) and collector-substrate (C_{jS}) capacitances. The subsequent improvement in phase noise is two-fold. First, in the realm of microwave frequencies, C_μ and C_{jS} form an appreciable fraction of the total capacitance in the resonant tank, lowering the overall Q of the resonance compared to what could be achieved with MIM capacitors. Another subtle—yet important—effect is that C_μ provides an avenue by which thermal noise in the base resistance can couple into the resonator even when the transistor is not conducting.[13] Hence, the second benefit of reducing the base-collector capacitance is in lessening the sensitivity of the oscillator phase to base resistance noise. Both the issue of collector capacitance and the distinction between the LTV and LTI models with regards to transistor sizing are picked up again in the next section, forming the basis of an experiment with three comparative versions of the same VCO circuit.

6.1.2 A Differential Colpitts Oscillator

Whether driving a balanced mixer, a phase splitter, or a frequency divider, having a differential oscillator available generally proves advantageous.[14] A balanced version of the common-base Colpitts circuit can be obtained by coupling together the emitter nodes of two stages as shown in Figure 6-3. The coupling capacitor (C_3) can be considered as the series combination of the grounded feedback capacitors from each single-ended stage, wherein an RF ground is located in the middle.[15] Each of the feedback capacitors (C_1-C_3)

[13] Another path for thermal noise to find its way to the resonator while the transistor is switched off is provided by the base-emitter junction capacitance. In Section 3.3, a non-zero sensitivity to the terminal resistance thermal noise sources was noted to exist throughout the oscillation period.

[14] The differential architecture does cost a factor of two in power consumption, but yields a much smaller phase error over the entire tuning range than could be obtained from a single-ended oscillator and a balun.

[15] An RF ground is only present for the differential mode of oscillation, a point which illustrates the importance of having C_3 be low impedance (relative to transistor and other parasitic capacitances) to prevent a common-mode oscillation from starting. Counting $2C_3$, the 0.75pF base-emitter capacitance in each of Q1-2, plus additional loading by the buffer and the current sources, the effective turns ratio of the capacitive divider works out to... $n = 4.3$.

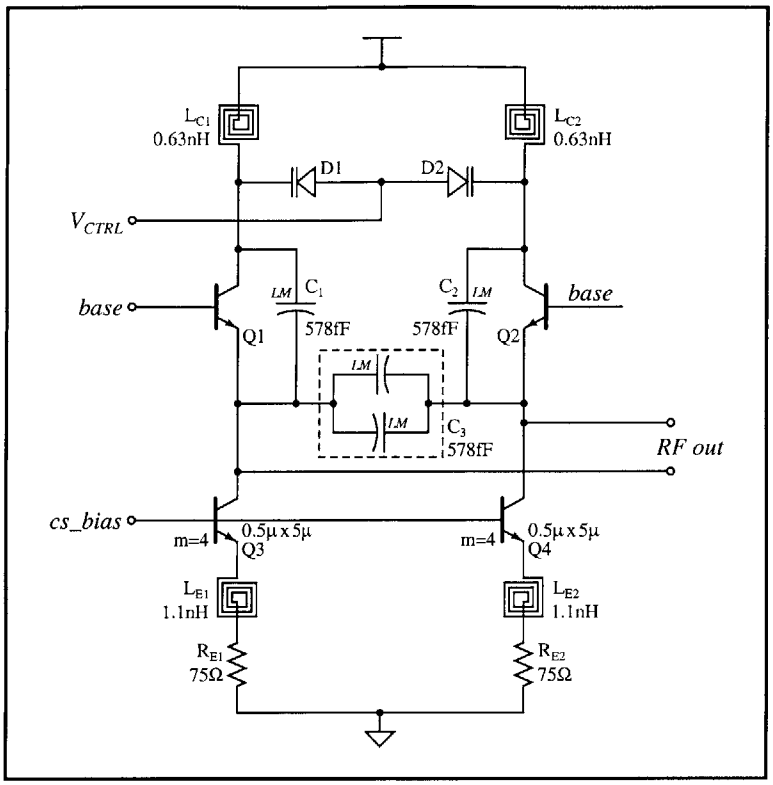

Figure 6-3. The 5.8GHz bipolar oscillator core.

are MIM devices and comprise an integral portion of the resonant tank. The spiral inductors L_{C1}-L_{C2} and a pair of junction varactors complete the resonator, with most of the loss occurring in the inductor, for which a Q of 19 is expected at 5.8GHz.[16] Rather than directly siphoning from the resonator, the output is instead drawn from the capacitive tap in the feedback network. This choice helps to isolate the resonator from any subsequent loading, thereby alleviating the demands on the buffering circuitry which follows the oscillator. Providing the bias current and sitting atop both resistive and inductive forms of degeneration are the current source transistors (Q3-4). By lowering the transconductance and increasing the output impedance, degeneration helps reduce the noise contribution of the current sources and improves the balance of the oscillator in the face of device mismatch.

With a resonant tank Q of almost 18 and a 3V supply, the transition between the current and voltage-limited modes of operation is expected to occur when the bias current in each side of the oscillator core is just over 3mA. An LTV analysis performed on the

[16] Each varactor possesses 450fF of zero-bias junction capacitance, an amount which should yield a tuning range of 600MHz about the 5.8GHz center frequency.

oscillator when biased at this optimum current level indicates that the effects of thermal and shot noise sources can be balanced by sizing the common-base transistors to have six 12.5μm emitters drawn at the minimum width.[17] Compared to the 8.5μm long LTI-inspired device, the 50% increase in emitter area should substantially reduce the base and emitter terminal resistances. The now equal contributions of current and voltage noise render a picture that is very similar to that provided as the completed example of Section 3.4 (plotted in Figure 3-11), except that the higher Q inductors shift each of the traces downward (i.e. in the direction of lower noise) by 10-11dB. From this analysis, the oscillator is expected to deliver a phase noise level of -120dBc/Hz at a 1MHz offset—all for a mere 19mW of power consumption in the differential core.

While this concludes one design of the 5.8GHz oscillator core, several technology angles remain to be explored. First, it has been noted on several occasions that the historical LTI approach and the newer LTV model have yielded differing opinions about how to optimize the transistors. As this appears to be an interesting distinction, a second version of the oscillator is prepared using the LTI-approved geometry, for which simulations suggest that performance should degrade by 1-1.5dB. Another consideration pertains to the base-collector capacitance and whether increasing the collector doping to pave the way for higher peak transistor f_Ts hurts an oscillator more than it helps. Fortuitously, the SiGe graded-base NPN technology being used to implement the designs offers an option to mask one of the self-aligned collector implants, reducing C_μ by 35% at the cost of a factor of 2 in the peak transistor f_T.[18] A third variant of the oscillator core is thus spun for this experiment by masking the extra collector implant in the LTV-sized common-base transistors.

Finally, the theme of being able to dynamically trade away some RF performance for lower power consumption within the oscillator should be examined. As the phase noise is anticipated to vary with supply voltage and bias current, on-chip circuitry that provides independent control of both while establishing the current source reference (*cs_bias*) and the common-base bias voltage (*base*) will prove empirically enlightening. The challenge for the design of the biasing circuitry is to provide an adaptive control and a stable operating point while trying to remain invisible from the standpoint of noise.

[17] The LTV analysis technique was demonstrated for oscillators in Sections 3.3 and 3.4.

[18] Perhaps the more fortuitous aspect is that this has become a fairly standard option in advanced bipolar technologies as a way to offer a higher voltage (breakdown) transistor within the same process. While the peak f_T takes a substantial hit when the collector implant is masked, at the low current densities that should be employed in oscillators, the device with the lower collector doping may actually exhibit a higher f_T. This is not the case with the 0.5μm bipolar technology used in this design, although the reduction in f_T is fairly minimal (about 10%).

6.1.3 Biasing Circuitry for the Bipolar Oscillators

In shaping a topology for the oscillator bias circuit, the primary features to be supported are supply rejection and an ability to adjust the current flowing in the oscillator without necessitating a change in the supply voltage. As a pragmatic matter, furnishing a "default" bias condition in the absence of an exogenous control input is helpful—not only for the purposes of testing, but also in determining any effects real biasing circuitry may have upon the oscillator performance. Noise considerations further imply a limit to how small the devices and currents comprising any bias reference legs can be made, meaning that the number of legs should be kept at a minimum to help reduce power consumption. The oscillator bias network evolving from this compendium of concerns incorporates a current source reference based on a V_{BE}/R relationship and is depicted in Figure 6-4. Diode-tied transistor Q1 and the degeneration resistor R_E provide a 1/4-scale master for the oscillator current sources, where the reference current is determined by:

$$I_{C1} = \frac{V_{BE2} + V_{BE3} - V_{BE1}}{R_{BIAS} + R_E}. \tag{6.1}$$

A fairly large pull-up resistor (R_{REF}) supplies some drive for the V_{BE} stack of Q2-3, and Q4 provides a source for the oscillator reference current. The source transistor Q4 needs to get base drive from somewhere, and a connection back to the collector of Q3 fits the bill

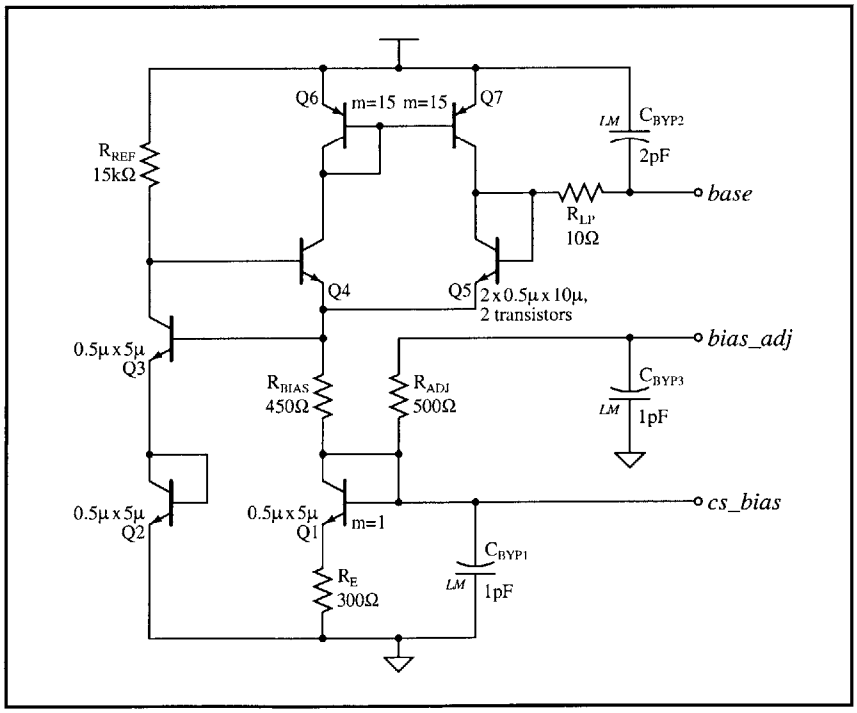

Figure 6-4. Biasing circuitry for the 5.8GHz bipolar oscillators.

while ensuring that the latter device remains in the forward-active region of operation. Adding to the convenience of this connection is that it also establishes a $3V_{BE}$ base bias reference for the oscillator which provides a common-mode output level that will remain constant over supply, process, and temperature.[19] Not yet finished, Q4 takes on a fourth role when transistors Q5-7 are added around it to form a differential pair stage which re-uses the oscillator reference leg as a tail source. By tying together the output and inverting input of the differential pair, a unity-gain buffer is realized to provide the base current drive required for the oscillator core.[20] A suite of bypass capacitors effects some filtering, and an external bias adjustment pin (*bias_adj*) completes the circuit while providing a mechanism for adjusting the current in the oscillator. When the *bias_adj* pin is left floating, the "default" bias condition of $4I_{C1} = 3.1\text{mA}$ prevails in each side of the oscillator core, where the factor of four results from the device multiple used in the current sources.

As a final thought, it could be observed that the base-emitter voltages remain—albeit logarithmically—dependent upon the supply voltage, and thus the degree to which changes in the supply voltage are rejected will be exceedingly finite. Depending on the application, however, independence between the bias current and the supply voltage may not be desired. If a single control were desired to trade performance for power consumption within the oscillator while maintaining the optimum bias current at each supply voltage setting, the appropriate relationship to establish would have the bias current change linearly with the supply voltage with a slope of R_{EQ}^{-1}—one over the equivalent resistance of the resonant tank. But having both degrees of freedom available is better for experimentation, and the bias circuit presented here provides sufficient freedom while getting the oscillator up and running on chip.

6.1.4 An Impedance-Matched Output Buffer

Supplied sufficient current and a resonant tank having a reasonable quality factor, the signal developed in the oscillator core should already possess an amplitude in the vicinity of 1 V. The purpose of buffering the signal is less to increase the swing than it is to

---

[19] As will be seen shortly, a reliable common-mode output voltage is an important feature in enabling the use of a DC-coupled oscillator buffer. The $3V_{BE}$ reference is chosen to provide a sufficient input level for the buffer while maintaining as much supply headroom as possible for signal swing across the resonator.

[20] The base current buffer is one of two feedback loops in the bias circuit for which stability should be verified with a loop gain and phase analysis. While finding the other could be left as a puzzle for the reader, the second is the loop formed by Q3 (a common-emitter stage) and Q4 (an emitter-follower).

provide the power gain required to drive an off-chip 50Ω impedance level while minimally loading the core itself. At the same time, a high degree of isolation is also desirable to avoid having the oscillator frequency "pulled" by loading and interaction with the external environment. As the isolation afforded by an emitter-follower falls perilously low when the frequencies being buffered approach the device f_T, a differential pair topology is adopted for the oscillator output buffer illustrated in Figure 6-5.[21] The high output impedance of the differential pair makes for a poor match to 50Ω (per side), so a tapped-capacitor resonant load is affixed to the differential pair to provide the requisite transformation.[22] Although not shown in the schematic, RF shielded probe pads are used for the outputs and are included in the matching network (effectively in parallel with C_3). To accommodate the contingency that the range of frequencies produced by the oscillator is not as intended, the buffer is designed to deliver a gain within 1dB of its peak value across a 2GHz band centered at 5.8GHz. Having a relaxed bandwidth allows the output inductors to be optimized more for area than quality factor, accomplishing the 2.8nH of inductance in a 4-turn, 150μm on a side, square spiral with a Q expected to be 9.6 in the band of interest. The resonant load is then degenerated by the resistor R_L to further ensure a rea-

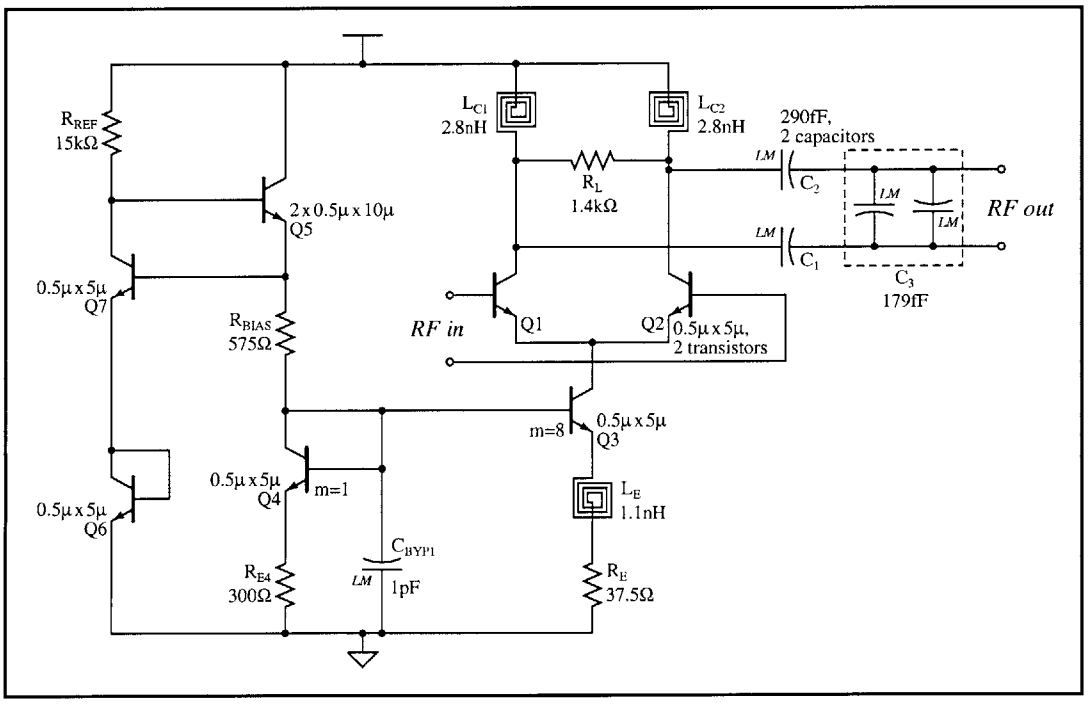

Figure 6-5. Output buffer for the 5.8GHz bipolar oscillator family.

[21] The surprising ineptitude of the emitter-follower configuration for use as a high frequency buffer was described in Section 4.3.

[22] The resonant load for the buffer is a differential form of the output matching network affixed to the LNAs and discussed in Section 5.1.1.

sonably wide response band. Given the impedance afforded by this load, a tail current of 4 mA is chosen for the differential pair so as to provide at least a nominal amount of voltage gain from the input to the collectors.[23] A reference for the tail source is established via a V_{BE}/R loop that mirrors the design used to bias the oscillator core.[24]

While a primary goal for the buffer is to minimize its visibility to the oscillator core, the attainable input impedance becomes limited in that it scales inversely with voltage gain. Designating the load impedance (at resonance) as R_{EQ}, the product of the input impedance and the voltage gain (A_v) is bound by:

$$|A_v| Z_{in} \leq \beta(f) R_{EQ}, \tag{6.2}$$

where $\beta(f)$ denotes the transistor current gain at the frequency of interest.[25] If the impedance at the collector is 300 Ω and a voltage gain of $A_v = 3$ is desired from transistors operating at an f_T/f ratio of 10, then the input impedance can have a magnitude no greater than 1 kΩ—and may be appreciably less considering other device parasitics. Interestingly, the size of the input transistors—and hence the associated junction capacitances—also becomes constrained by gain and the subsequent current density considerations.[26] The collector current must be made large enough to deliver the required transconductance, but then the emitter area needs to be sized to flow this current without causing base pushout in the transistors.[27] The graded-base SiGe NPNs used in this work obtain their peak f_T at a J_C of 1.2 mA/μm^2, and it is important that collector current densities do not extend much beyond this level in the buffer. With this limitation in mind, the input transistors are designed with 0.5 μm by 5 μm emitters to operate at 0.8 mA/μm^2 when the quiescent 2 mA bias flows in each side of the differential pair.[28] Circuit techniques that increase $\beta(f)$—such as the f_T doublers described in Section 4.3.1—could be used to augment the input impedance, but were not seen as being necessary in this design.

[23] After accounting for the impedance transformation, the output buffer does offer a rather substantial power gain. Even with a mismatch at the input, the insertion gain (s_{21}) expected of the buffer in a 50 Ω environment is on par with that of the LNA designs in the preceding chapter.

[24] In fact, the tail source in the buffer could, and probably should, be slaved to the same reference used in the oscillator. However, the ability to independently power down the buffer was deemed to be useful for laboratory purposes.

[25] This relationship ignores base-collector capacitance and the terminal resistances, but can accommodate the small-signal output resistance of the transistor by including it in R_{EQ}.

[26] "Interestingly" is one choice of words. "Painfully" might be another.

[27] Base pushout at high collector current densities is also known as the Kirk effect, and can substantially increase the input capacitance of the transistors so affected.

[28] A square emitter geometry might actually be a better solution for the buffer input transistors as it results in a lower C_μ for any given emitter area. Sadly, however, many existing high performance bipolar technologies only support one emitter width (at least in terms of models).

A determination as to whether the buffer is sufficient, and the biasing sufficiently quiet, can be made with the aid of a circuit simulator. Phase noise calculations using the SpectreRF periodic noise analysis tool have been found in this work to correspond very closely with the final numbers generated by the LTV approach presented in Chapter 3. Assuming a 3 V supply and the associated optimum (core) bias current of 6.25 mA, a simulation is first run for the oscillator core by itself. The spiral inductors are represented by the scalable models—which have suggested Qs of 19 for the 0.63 nH devices—and the frequency control input has been tied to the circuit ground. With no loading imposed and ideal voltage sources supplying the bias nodes *base* and *cs_bias*, the answer provided by SpectreRF is shown with the dashed line in Figure 6-6. At an offset of 1 MHz from the 5.9 GHz carrier, the phase noise spectral density exhibited by the oscillator is calculated to be -120.7 dBc/Hz. Next, the actual biasing circuitry is added and then followed by the buffer, with the full chip simulation (indicated with the solid trace) predicting the noise to be just 0.2 dB higher at the same 1 MHz offset when measured from the output pads. A more noticeable upturn is observed for frequencies within 10 kHz of the carrier, primarily as a result of flicker noise in the output buffer. Unfortunately, quieting this part of the spectrum would likely require (much) larger buffer input transistors, which could in turn compromise the oscillator performance in its own way through increased loading.

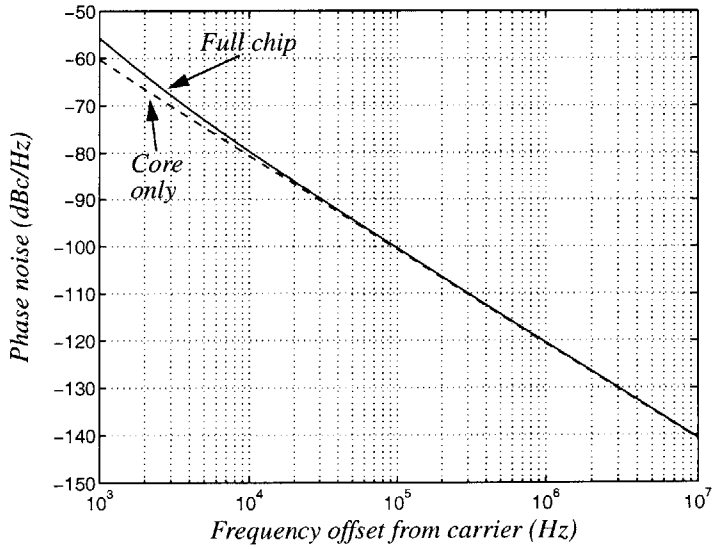

Figure 6-6. Simulated phase noise of the 5.8 GHz bipolar VCO operating at 3 V and with 6.25 mA in the oscillator core.

The simulations shown above were conducted on the oscillator designed around the LTV-sized common-base transistors with 12.5 μm emitter stripes. The other two ver-

sions of the oscillator each use the same buffer and biasing network, and just replace the gain devices in the core. It might be argued that a more fair overall comparison of the high and low f_T variants would involve masking the intrinsic collector implant from all of the transistors being used and not just the two common-base devices. But the purpose of this experiment is more to illustrate some of the distinctions encountered by applying an LTV model to analyze oscillators at the device level, and so a choice was made to minimize the potential for ambiguity by keeping all other variables constant. Granted, the buffer is one example where plenty of current gain and a high peak f_T proves handy. However, for the capacitive transformation ratio and resonator Q in play with this design, loading from the buffer did not seem to appreciably affect the oscillator. A drop in the f_T/f ratio from 10 to 5 is not expected to modify this outcome significantly.[29] The only other usage where the reduced collector doping might have an effect is in the current sources. But employing the more lightly-doped collectors would make for a higher output impedance over all frequencies of interest—a consequence which could only result in improvement. Hence, a design constructed entirely of the lower f_T devices may actually perform better than any of the versions in this comparison.

6.1.5 Layout Design of the Bipolar VCOs

Having settled upon three versions of the oscillator which are very similar, layout efforts can concentrate on a single design. Within the overall dimensions of 1.38mm x 0.97mm, a floorplan emerges through the dictations of accommodating seven spiral inductors as shown in the die photo of Figure 6-7. Each inductor consists of a winding constructed from the top two metal levels and is located over an array of deep trench isolation pockets. Among the seven spirals, the pair on the right belong to the oscillator core wherein the importance of minimizing loss has led to the use of a 25μm trace width. The balanced nature of the oscillator is reflected in its symmetric layout, and ample space is left around the inductors to minimize electromagnetic interaction with adjacent elements. The buffer and output matching network can be found in the lower half of the photo and to the left of the VCO. An impedance match to the (off-chip) load is achieved by the inductors and MIM capacitors in the immediate vicinity of the buffer transistors; the buffer then drives a pair of 50Ω transmission lines which begin as microstrip and then transition to coplanar waveguide before yielding to a set of RF-shielded GSG pads for each of the balanced outputs.[30] This leaves the middle of the die available for residence by the biasing

[29] Additionally, circuit-level design changes can be used to combat the effects of having a lower transistor f_T in the buffer. Even a mild increase in the turns ratio in the capacitive transformer should help, as can f_T doubler topologies for the buffer.

[30] The two ground planes are tied together with underpass straps to prevent the slot line mode.

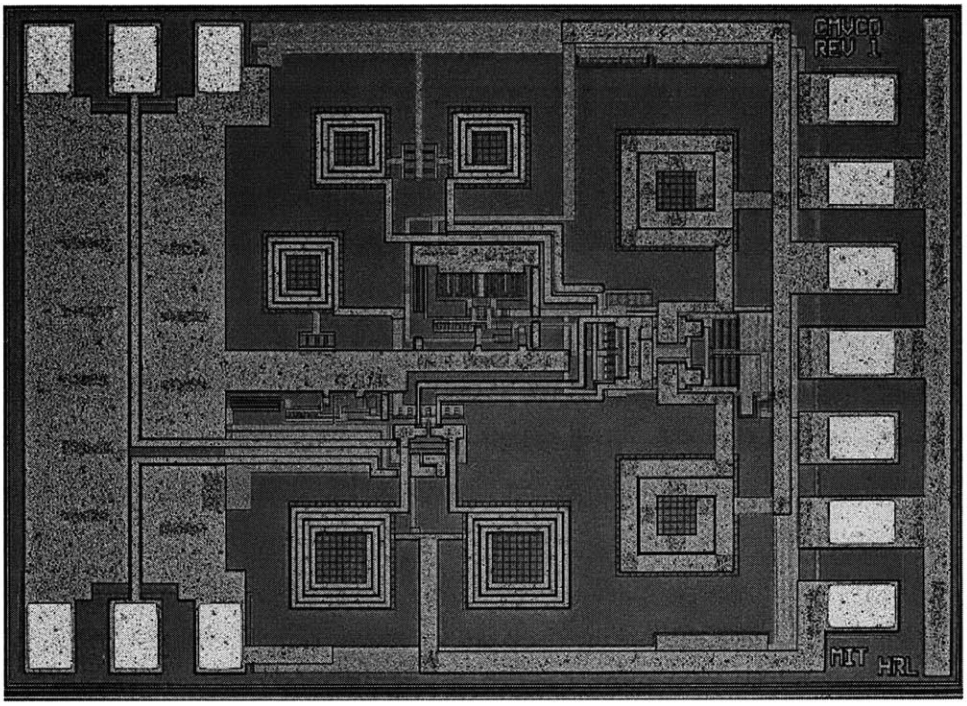

Figure 6-7. Die photo of the 5.8 GHz bipolar VCO IC (CMVCO).

circuitry, around the periphery of which the 1 nH current source degeneration inductors are arranged. To help reduce interconnect loading on the oscillator, the current source transistors are located next to the common-base devices in the core, and then the base and emitter nodes are routed into them. This is opposite of what is encountered in many ICs, but the ability to externally adjust the bias relaxes the degree of matching required between a current source and the reference device to which it is slaved.

On the right and occupying a fairly substantial amount of real estate is a set of pads for a bypassed supply probe having four power pins with three grounds interspersed among them. From the top, the four available pins provide: the *bias_adj* control, the supply voltage for the oscillator and its biasing circuitry, the oscillation frequency control, and a separate supply line for the output buffer. Additional filtering is furnished for each of the four lines with on-chip capacitance. While the oscillator and buffer can be operated from the same supply voltage, they need not be. The second supply pin allows the buffer to be remain a constant as the voltage for the oscillator is adjusted for the sake of phase noise performance.[31] Finally, located above the suite of DC pads, the label CMVCO can be

[31] It is probably worth reiterating at this point: the $3V_{BE}$ base reference established for the oscillator in the biasing circuitry holds the output common mode voltage at a fixed level even as the supply voltage is changed.

seen. Following the convention set forth with the LNAs, this is the C-band Monolithic Voltage-Controlled Oscillator. Suffixes are attached to this name in the two variations on this design to provide a means of visual differentiation. Save for this identifier, the three ICs are not easily distinguishable under a microscope.

This foray into oscillator design began with a Colpitts topology as being readily integrable and amenable to low noise, low voltage operation. A differential form was synthesized, and then biasing and buffering circuits were developed to enhance the functionality without taxing the RF performance. Three versions of the oscillator have been spun, with the nominal design expected to deliver a phase noise spectral density of -121 dBc/Hz at a 1 MHz offset from a 5.8 GHz center frequency. Underscoring the prevailing overemphasis in the RF/microwave field on peak transistor f_T, one of the two oscillator variations should realize a 1-1.5 dB improvement in the phase noise by using devices wherein the attainable f_T is halved due to a lower collector doping concentration. The remaining version strives to illustrate the limitations of assuming time invariance in oscillators. Working with a resonator Q of 18, an LTI model-based approach to setting the device size is anticipated to fall about 1-1.5 dB short of the performance that the nominal design attains merely by using larger transistors. Thus—from the bipolar oscillator design—three data points are established, each the same save for the gain devices in the oscillator core. A fourth reference point, designed entirely in CMOS, is discussed next.

6.2 A CMOS 5.8 GHz VCO Design

Over the course of describing a time-variant model of oscillator behavior and the design of a set of bipolar oscillators, a number of device-level observations have been offered. Of these, one of the more salient findings is that once the transistor f_T is sufficient to support switching at the desired frequency, the dynamics of an oscillator circuit are controlled by the oscillation and not the transistors. As an integral component of the dynamics, the noise currents associated with transistor conduction can be timed by proper circuit design to exist only during intervals in the oscillation period when the phase is not easily disturbed. Together, these findings cast an interesting shadow onto the ever-raging CMOS versus bipolar debate. Clearly, bipolar devices do feature better analog performance in terms of transconductance, output resistance, and high frequency current gain; about this there is little argument.[32] From what has been presented here, however, it may be con-

[32] The comparison of current gain assumes similar technology generations (i.e. a state-of-the-art 0.13 μm CMOS technology is not being flaunted against a 0.5 μm bipolar technology).

cluded that the effect of the transistor on the phase noise performance of an oscillator—given a fixed set of passive elements—essentially reduces to the consideration of terminal parasitics. *To the extent that CMOS devices can be made with comparable terminal resistances and capacitances, the phase noise in the f^{-2} region of the spectrum generated by CMOS and bipolar oscillators should also be comparable.*

Amidst the increased interest in CMOS RF circuits, an oscillator topology that has recently evolved makes use of cross-coupled NMOS and PMOS pairs connected as depicted in Figure 6-8. The cross coupling provides regenerative gain to sustain an oscillation, and having both p-channel and n-channel pairs can balance the pull-up and pull-down drive strengths to lower flicker noise up-conversion [38]. While this topology does not exhibit quite the same level of rejection to the transistor current noise sources that the Colpitts structure achieves, the peak transistor currents are still timed by the extrema in the oscillation signal, and thus the sensitivity of the phase to the drain and gate noise currents will be curtailed.[33] The bias in the oscillator is controlled by tail transistor MP3, and is provided a rather substantial bypass capacitor to filter high frequency noise from the current [84]. But in creating a three device stack, the bigger problem may be to leave sufficient headroom for signal swing across the resonator.

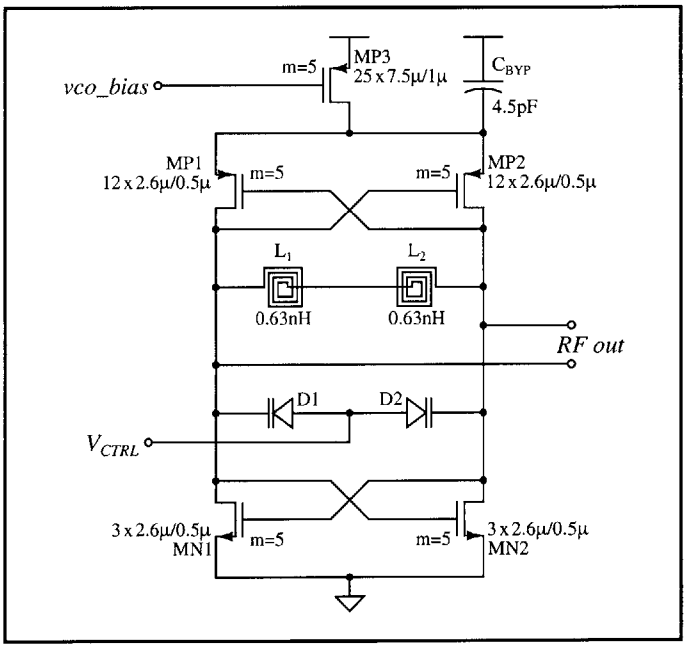

Figure 6-8. The 5.8GHz CMOS oscillator core.

[33] Similar to the case described for bipolar transistors in Section 3.4, the charging currents associated with the gate-drain overlap capacitance and the drain-bulk depletion region need to be discounted from the total drain current for the purposes of determining the time-variant g_{d0}.

Keeping in mind that the limited widths of the transistors in the oscillator core can lead to substantial gate overdrive requirements, it becomes important to operate the tail current source with a minimal voltage drop to increase the range over which the oscillation can swing.[34] One route to realizing a small drain-source potential is to make MP3 very wide so that it can remain subthreshold. Unfortunately, an arbitrarily wide MP3 would have a large transconductance, magnifying the effect of any voltage noise riding on the gate potential (*vco_bias*) as generated by the biasing circuitry. The desired solution then is to use a small tail transistor with a low V_{DS} and a large gate overdrive to support the required bias currents. While optimal for reducing phase noise, these conditions leave the tail device in the triode regime where it ceases to function as a current source; a diode-tied mirror transistor will no longer suffice to accurately control the oscillator current from an established reference. Something more is needed, and will be an important feature in providing the adaptive performance capabilities being sought for wireless networking.

6.2.1 Adaptive Biasing of the CMOS Oscillator

For adaptive biasing to work, the current flowing in the oscillator core must track a reference set by a bias control input. When the tail source is operating in triode, the reference current must flow in a device that is matched to the VCO tail transistor and which has the same gate-source and drain-source potentials. Similar in spirit to a high-swing op-amp design by Gulati and Lee [85], the circuit developed as "something more" to implement this goal is illustrated schematically in Figure 6-9. First, rather than placing any additional capacitive loading directly onto the resonator, a non-oscillating replica of the core is formed by MN1-2 together with MP1-3, wherein the device multiples have been scaled down by a factor of 5 to save power.[35] Replacing the resonant load (that appears in the oscillator) is an RC network; this serves to filter noise while presenting an impedance low enough to keep the loop gain in the replica safely below unity for all frequencies.[36] A reference current leg is then created by duplicating the p-channel half of the replica via MP6-8. The design challenge is to have the drain voltage on the reference transistor (MP8) track that of the tail source (MP3), while still providing the freedom to independently set

[34] Transistor channel widths are limited by the associated oxide and drain-bulk capacitances which become the most significant part of the resonant tank at high RF and microwave frequencies.

[35] More accurately, this should probably be stated in the other direction: the actual oscillator core makes use of five copies of each device in the replica.

[36] Scaling the devices in the replica helps in this regard too; with the bias current and all of the channel widths being smaller by the factor of 5, the transconductance driving the loop is also reduced by the same factor. While a smaller load resistance is better for stability, R_L should be large enough to prevent an offset voltage between the two sides from creating a significant current draw.

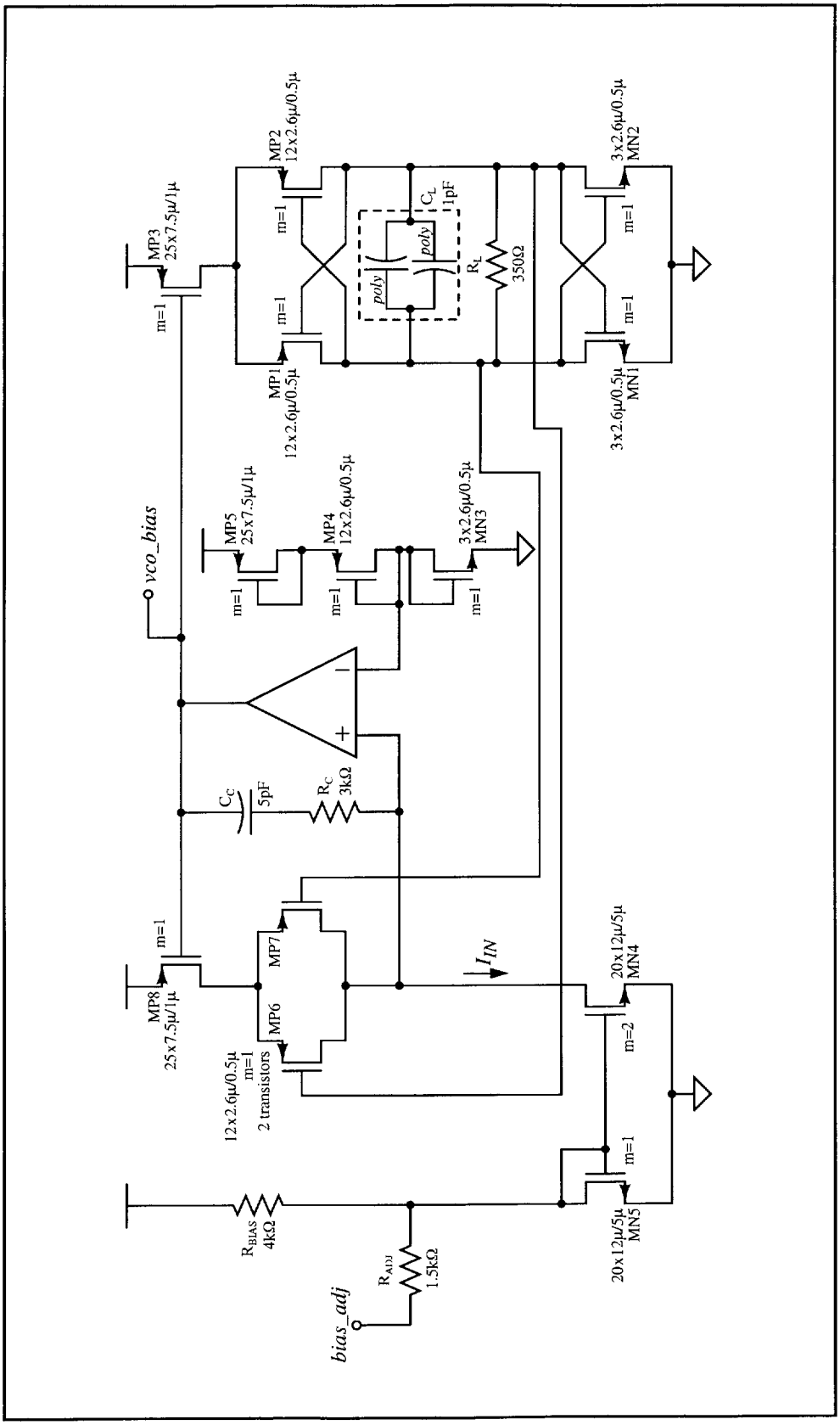

Figure 6-9. Biasing circuitry for the 5.8GHz CMOS oscillator.

the input current (I_{IN}). By bringing the p-channel gate voltages across from the replica, the gate-source potentials of MP6-7 will match those of MP1-2 as long as the MP6-7 pair is maintained in saturation. Under this provision, the drain-source voltages—and hence the drain currents—of the p-channel tail devices (MP8 and MP3) will also match. To keep MP6-7 in the high gain state, the shared drain node is held by an op-amp to a potential set by a diode-tied reference chain that approximates the common-mode output voltage of the oscillator. The op-amp also drives the gate bias on the tail transistors (*vco_bias*) to that required to support the input current (I_{IN}); the same current flows in the DC replica and thus appears in the oscillator core itself (multiplied by a scale factor of m=5). For this application, two resistors and a current mirror (MN4-5) are used to set I_{IN}, although a more elaborate reference could easily be substituted. Resistor R_{BIAS} establishes a "default" bias condition, which may then be adjusted by sourcing additional current through—or pulling current from—the *bias_adj* input.

Operating with a 3V supply and applying a 15:1 range in the input reference current, the simulated behavior is shown in Figure 6-10(a). The oscillator core dons a bias current that matches the input setting to within 4% of the expected ratio of 5—even as the gate voltage on the tail source approaches 0V and the drain-source potential falls to 200mV.[37] While the ratio is exactly 5 for lower currents, the falloff begins at the point where $|V_{GS}|$ exceeds $|V_{DS}|$ by a threshold, marking the onset of triode operation. The finite regulation stems from the limited gain of the p-channel pair being relied upon to reduce the error in approximating the common-mode voltage. The output common-mode level of the CMOS oscillator changes over bias, but the voltage generated by the diode-tied chain to approximate it does not. Although the true DC level of the output is available in the replica, the voltage provided as a reference to the op-amp can not be allowed to change with bias. As the bias resides under control of the op-amp, it would not be possible to arrange for a reference voltage loop and the input current loop to simultaneously have feedback in the negative sense. Hence the need for an open loop voltage reference and its associated error. Nonetheless, a maximum error of 4% in controlling the oscillator bias current is quite reasonable.

Having verified the DC behavior, the next concern is whether the feedback circuit will remain stable over the range of input currents. From the viewpoint of loop transmission, MP8 is a common-source transistor having the MP6-7 pair as a cascode device, pro-

[37] Device mismatch will be another source of deviation from the desired current ratio. However, this deviation tends to be of the "one-for-one" variety. A mismatch of 1% between two devices leads to an error of 1% in the bias current being controlled. For any reasonable set of process tolerances, this should be a smaller error term than the one shown by the simulation.

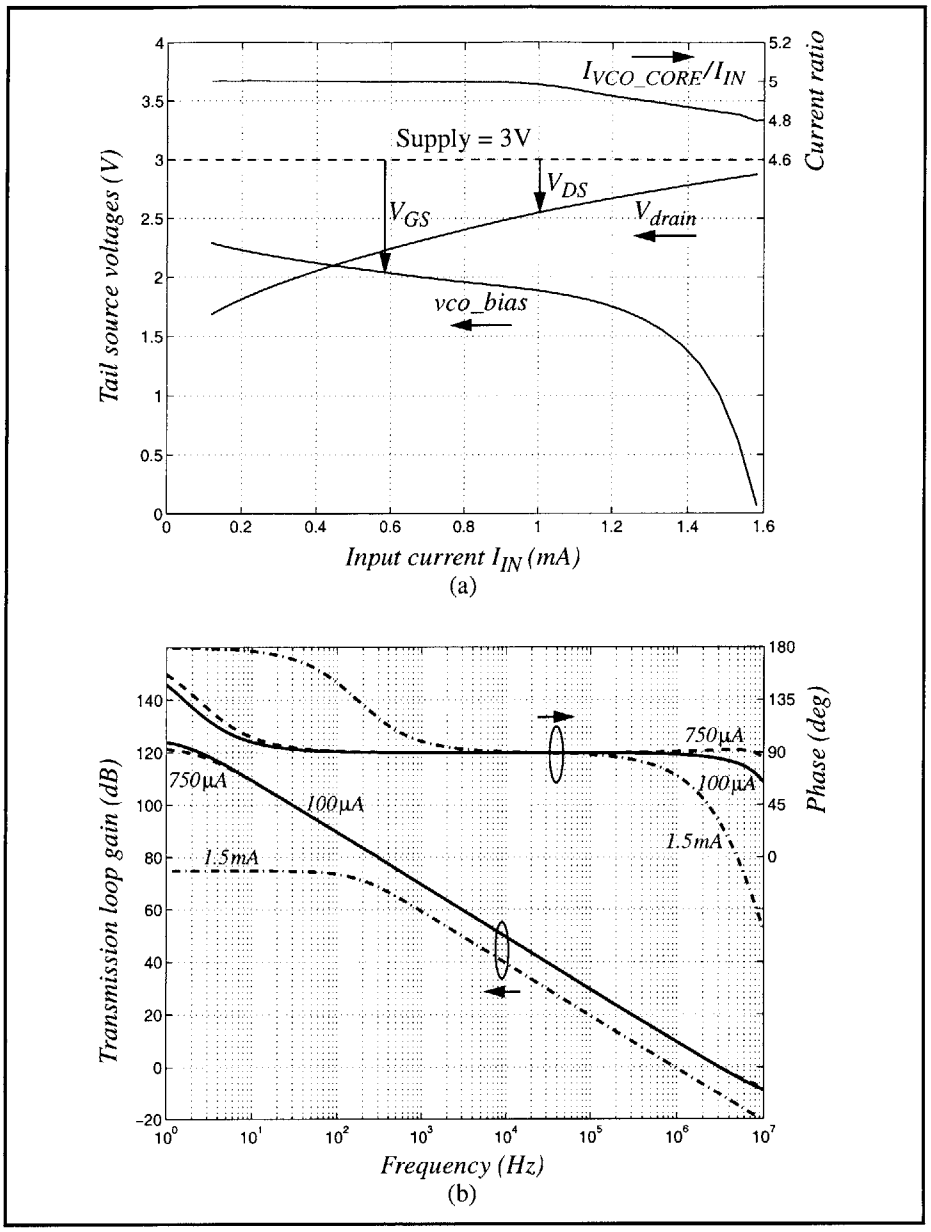

Figure 6-10. Simulation of the DC behavior (a) and loop stability (b) of the CMOS oscillator bias circuit.

viding both an inversion and feedback gain around the op-amp. While the presence of feedback gain eases a number of important design issues,[38] one aspect may be made considerably more difficult: that of loop stability. Compounding this difficulty is that the $g_m r_o$ product in MOSFETs falls with increasing current, varying the loop gain as a function of the input. Mindful of the concern for stability, the op-amp is constructed as a sin-

[38] The equivalent input noise and offset voltage associated with the op-amp are reduced by the feedback gain. Another, perhaps more obvious, benefit is that less gain is required of the op-amp itself.

gle-stage, folded cascode design with high-swing biasing on the cascode devices. The single-stage topology allows the feedback loop to be stabilized by placing a compensation network across the second stage. Via the same pole-splitting mechanism employed in the classical two-stage op-amp, the capacitance C_C sets the dominant pole and the series resistance (R_C) moves the transmission zero to the left half-plane. An additional benefit is that as the loop gain changes with the input current, so too will the dominant pole. Figure 6-10(b) shows the transmission gain and phase for input current levels of 100μA, 750mA, and 1.5mA; with a gain of $(g_m r_o)^3$ on tap for lower currents, a 5pF compensation capacitor is used to place the pole at about 1Hz.[39] Phase margins of greater than 70° are observed over the entire input current range, which should be enough to cover process variations. Similar results are observed for supply voltages from 2.25V to 5V.

Of course, the question that may now arise is whether an oscillation will still be recognizable among the noise emanating from the op-amp, feedback loop, and current mirrors. A number of design features incorporated along the way should help: keeping the tail source transistor as small as possible lowers the conversion of bias noise into oscillator current, the gain of the cascoded common-source input stage reduces the op-amp noise, and bypassing the RF ground node in the oscillator shunts away noise at higher frequencies. But to find a more quantitative answer, SpectreRF is turned loose. A 3V supply is provided for the oscillator, with which the optimum core bias current—using the same

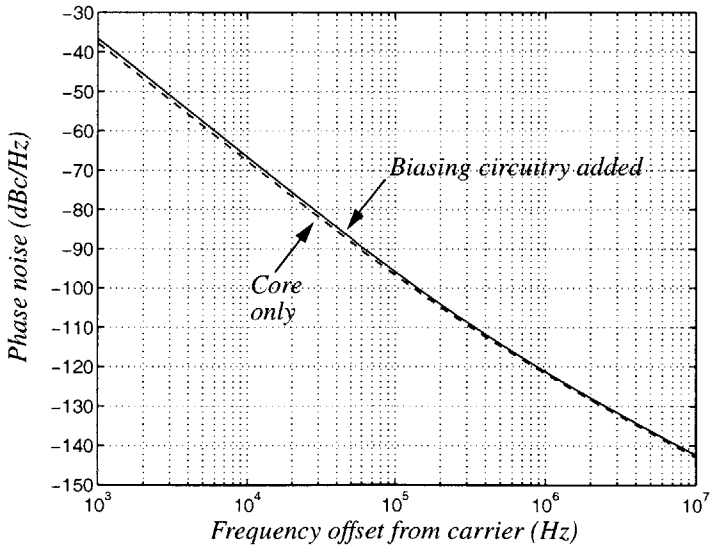

Figure 6-11. Simulated phase noise of the 5.8GHz CMOS VCO operating at 3V and with 5mA in the oscillator core.

[39] As can be noticed in Figure 6-10(b), the loop gain drops considerably when the reference device (MP8) falls out of saturation.

0.63nH inductors as in the bipolar design—is determined to be 5mA.[40] The first simulation is of the oscillator core by itself, fed an ideal voltage source to set the gate bias (*vco_bias*) and having the varactor node tied to ground. Oscillating at 5.9GHz, the CMOS VCO produces the phase noise spectrum plotted in Figure 6-11. The transition between the f^{-3} and f^{-2} characteristic regions occurs at an offset frequency in the 200-300kHz neighborhood—substantially less than the 1MHz flicker noise corner in the n-channel devices, although somewhat higher than the 100kHz exhibited by the p-channel MOSFETs. Beyond the transition frequency, the phase noise within reach of the CMOS oscillator appears to be marginally better than realized by its bipolar brethren. For the core alone, -121.8dBc/Hz is seen at a 1MHz offset along the dashed line. The comparison reverses for frequencies closer to the carrier; at an offset of 100kHz, the phase noise spectral density generated by the CMOS core is -96.7dBc/Hz whereas the bipolar version is 4dB better. Regardless, the initial proposition that the CMOS design should hold its own in the f^{-2} region seems to have merit.

Returning from this comparative aside to the lingering question of bias noise, the results of a second simulation—this time with the entire biasing circuit included—are also shown in Figure 6-11. Although this analysis might not cast the most favorable light, the degradation witnessed in the phase noise is fairly mild; the adaptive biasing circuitry imposes itself to the tune of 0.5-0.6dB in the f^{-2} region, and just slightly more than this at frequencies closer to the carrier. But while contributing its share of noise, the adaptive scheme extends the bias range over which the oscillator can be operated, thereby enabling a higher level of performance to be reached. Extending the performance is the real benefit of adaptive biasing; the only remaining detail is to incorporate an output buffer to protect and share this benefit with the off-chip world.

6.2.2 An Impedance-Matched Output Buffer in CMOS

Following the trail blazed by the design of the bipolar suite of oscillator ICs, a differential pair with a tapped-capacitor resonant load is again called upon as an output buffer. In CMOS, the buffer appears as shown in Figure 6-12, where p-channel devices are used in the signal path for their lower flicker noise.[41] As there are no explicit capacitors in the oscillator from which to tap a signal, the gain and bandwidth provided by the

[40] Even with the tail source operating in triode, the stacked, cross-coupled MOSFET pairs still compress the output swing at a lower bias current than observed in the bipolar design.

[41] The p-channel input is also more amenable to signals at low DC levels. This improves the interface with the oscillator which, due to the p-channel tail source, will provide a common-mode level that is closer to ground than to supply.

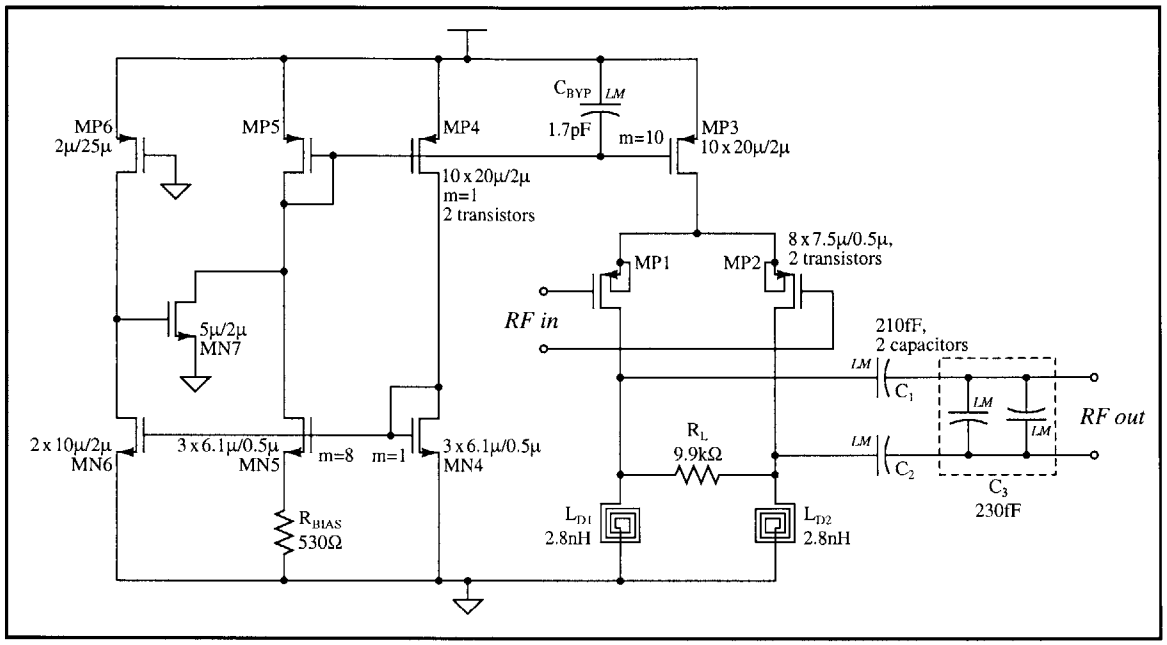

Figure 6-12. Output buffer for the 5.8GHz CMOS oscillator.

buffer will be sacrificed in favor of reducing the loading that it imparts.[42] Small input transistors help in this regard, although the minimum usable device size is determined by current density considerations—just as noted in the bipolar version, albeit for a different reason. Instead of base pushout, the concern with MOSFETs becomes the gate overdrive needed to handle the bias current. A current on the order of milliamps is required to provide a reasonable facsimile of gain, but $V_{GS} - V_t$ for the source-coupled pair must be small enough to yield a viable input common-mode voltage range. As a balance between input impedance and gain, a 4mA tail current is chosen for the buffer: the same bias used in the bipolar design. To help compensate for the lower CMOS transconductance, the load resistance (R_L) is increased substantially—probably to the point where its presence will barely be felt, unless the inductor models prove very far off the mark. Finally, getting the buffer off (the) ground, a reference for the tail current is initiated through a ΔV_{GS} biasing cell, the design of which was covered in Section 5.2.3.

6.2.3 Layout Design of the CMOS VCO

In keeping with the spirit of constructing a consistent set of circuit data points through which to explore device-level technology issues, the layout design of the CMOS

[42] Other possibilities do exist for shielding the resonator from loading. The inductors could be tapped, or a coupled secondary winding could be used. But for this design, the use of a high input impedance buffer should suffice.

oscillator (CMVCO-CM) borrows heavily from bipolar version. The effort in matching is evident in the die photo of Figure 6-13, where the chip dimensions, pad layout, and general floorplan all remain the same. The spiral inductors in the oscillator core and output buffer are also unchanged, as are their locations on the chip. Having no MIM capacitors and smaller varactors in the resonant tank, the layout of the CMOS oscillator is a slightly more compact design than the bipolar core. Also freeing up space is the absence of the current source degeneration inductors, some of which in turn is consumed by the added complexities of the biasing circuit. But serving as yet another illustration of how passives tend to dominate the die area and layout of RFICs, the bias circuit (including the op-amp) fits in a 360μm x 210μm rectangle, entailing not much more area than the two VCO spirals put together.[43]

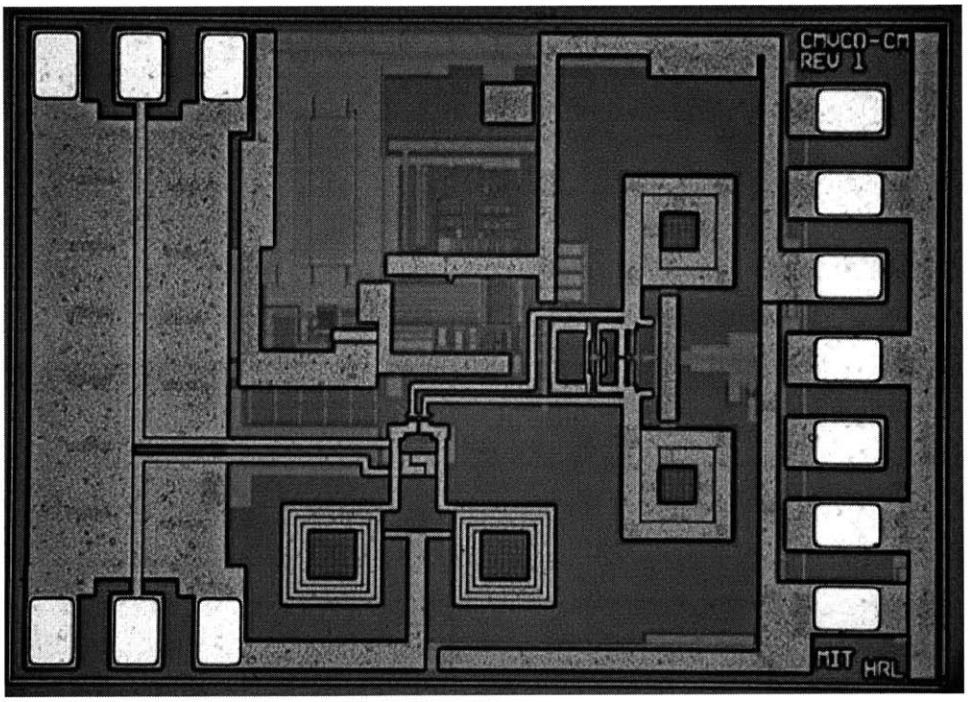

Figure 6-13. Die photo of 5.8GHz CMOS VCO IC (CMVCO-CM).

The design of the CMOS oscillator set out to achieve the same parcel of goals internalized with the bipolar versions: a reasonable tuning range around a 5.8GHz center frequency, the best attainable phase noise within the chosen technology, a mechanism for scaling the performance of the oscillator to meet the real-time demand, and realizing a

[43] The spirals in the VCO are actually 165μm on a side. However, considering the space left around them to provide isolation, the area blocked out by the two inductors is comparable to that occupied by the bias circuit.

fully integrated solution. One of the first issues encountered in the CMOS design was the high flicker noise in lateral field-effect transistors, leading to the choice of an oscillator topology which makes use of the complementary devices to mitigate its upconversion into phase noise. The stacked, cross-coupled pairs were then found to create a performance limitation, motivating the development of a new approach to biasing oscillators. By providing control over the oscillator current using a low V_{DS} tail source, the biasing scheme increases the headroom available for signal swing across the resonator. With this hardware in place, the CMOS design is expected to exhibit a slightly lower phase noise in the f^{-2} region than the bipolar versions achieve using the same inductors—illustrating once again how time-variant oscillator dynamics can reduce the effect of noise sources associated with transistor conduction. In the end, hopefully some lessons can be learned about the role of technology in RF circuits.

The stage has now been set. CMOS versus bipolar. LTV versus LTI. High f_T versus low f_T. Let the games begin.

6.3 Measured Results for the 5.8GHz VCOs

A fleet of four voltage-controlled oscillator ICs has been designed and fabricated in a 0.5μm BiCMOS technology. Along with the same set of passive devices, one variant employs CMOS exclusively and the others are bipolar-only. Each oscillator provides coverage over the 5.8GHz ISM band and supplies a differential output where each side is matched to a 50Ω load impedance through an on-chip buffer. All of the biasing is incorporated on chip, with a bias control input availed as a means of adjusting the performance of the oscillator to dynamically adapt to system requirements for the phase noise and amplitude. Characterized on wafer using Cascade Microtech AirCoplanar® GSG probes and a bypassed DC supply probe card, open-loop oscillator measurements were taken with an HP8563E spectrum analyzer equipped with the optional phase noise "personality" software.[44,45] As spectrum analyzers are not the most precise of instruments, the accuracy of this approach was confirmed by submitting the bipolar oscillators to a phase noise characterization system built by Aeroflex-Comstron. The PN9000B system from Aeroflex-Comstron uses a PLL to lock an internal synthesizer to the oscillator being tested, and then

[44] "Personality" is the description chosen by HP/Agilent; the author disavows any responsibility for the selection of this particular word (although the author may have little room to talk).

[45] A printed circuit board with four low-noise adjustable voltage regulators was used to power the oscillators. No differences were observed in the phase noise measurements between using this supply board and directly powering the oscillators from batteries.

measures the phase noise on the synthesizer while backing out the effects of the loop dynamics. Evidently constructed to test high performance microwave sources, the 10kHz maximum loop bandwidth supported by the PN9000B made acquiring lock very difficult with the fully-integrated CMVCOs.[46] Thus a complete characterization could not be achieved from the closed-loop measurements.

Starting with a 3V supply for the oscillator and buffer, and allowing the on-chip "default" bias conditions to preside (i.e. the *bias_adj* pin was left floating), spectral plots of the outputs produced by the CMVCO bipolar design and the CMOS version are shown in Figure 6-14(a,c). As the output spectrum of each bipolar oscillator essentially looks the same, only the nominal design is illustrated. Notably missing from the plots is any indication of a low frequency spur being modulated onto the carrier—an absence which attests to the stability of the circuits used in biasing the oscillators. The frequency of the carrier can be tuned from 5.4-6.1GHz for a 0-3V range in the control voltage (V_{CTRL}) applied to the bipolar oscillator, and from 5.2-5.9GHz in the CMOS design, where the control voltage is limited to a range of 0-2V. All four oscillator ICs exhibit an output power of -11dBm as a single-ended measurement.[47] The output level remains constant over most of the tuning range, but begins to decline as the varactors approach a forward bias condition. For the bipolar CMVCO IC, the dip amounts to 2.5dB when V_{CTRL} equals the supply voltage. A somewhat more dramatic falloff is noted with the CMOS topology, where the output power has dropped by 2.5dB as the control voltage reaches mid-supply.

Following much the same behavior, the phase noise in the oscillators is invariant with the tuning voltage until the point where the varactors begin to conduct. With the varactors sufficiently reverse-biased, phase noise measurements are taken for the nominal bipolar and CMOS designs—operating at their respective default bias currents—and are included in parts (b) and (d) of Figure 6-14. With the *bias_adj* pin left floating, 6.7mA of supply current is drawn by the bipolar oscillator; of this total, 6mA flows in the core and the remainder in the biasing circuit.[48] Similarly, a current of 5.2mA is established in the

[46] As one example of the sensitivity, the PN9000B could not lock the CMVCO when the frequency tuning voltage was established via a 10kΩ potentiometer connected across a 9V battery. The 10kΩ load discharging the battery led to too much frequency drift for acquisition to be achieved.

[47] A 50Ω termination was applied to the other output in the balanced configuration.

[48] With the limited number of available pins, the current in the oscillator core cannot be directly measured. However, it can be calculated from the total supply current (shared between the core and its biasing circuitry) and the measured current into the *bias_adj* pin—once a few assumptions have been made. Matching among the ratioed transistors (where device multiples are in play) is by far the largest of the assumptions, and it should be a fairly reliable one. For any biasing currents not directly mirrored into the oscillator, nominal values (as determined from simulation) are used in the calculations. These reference currents are comparatively small.

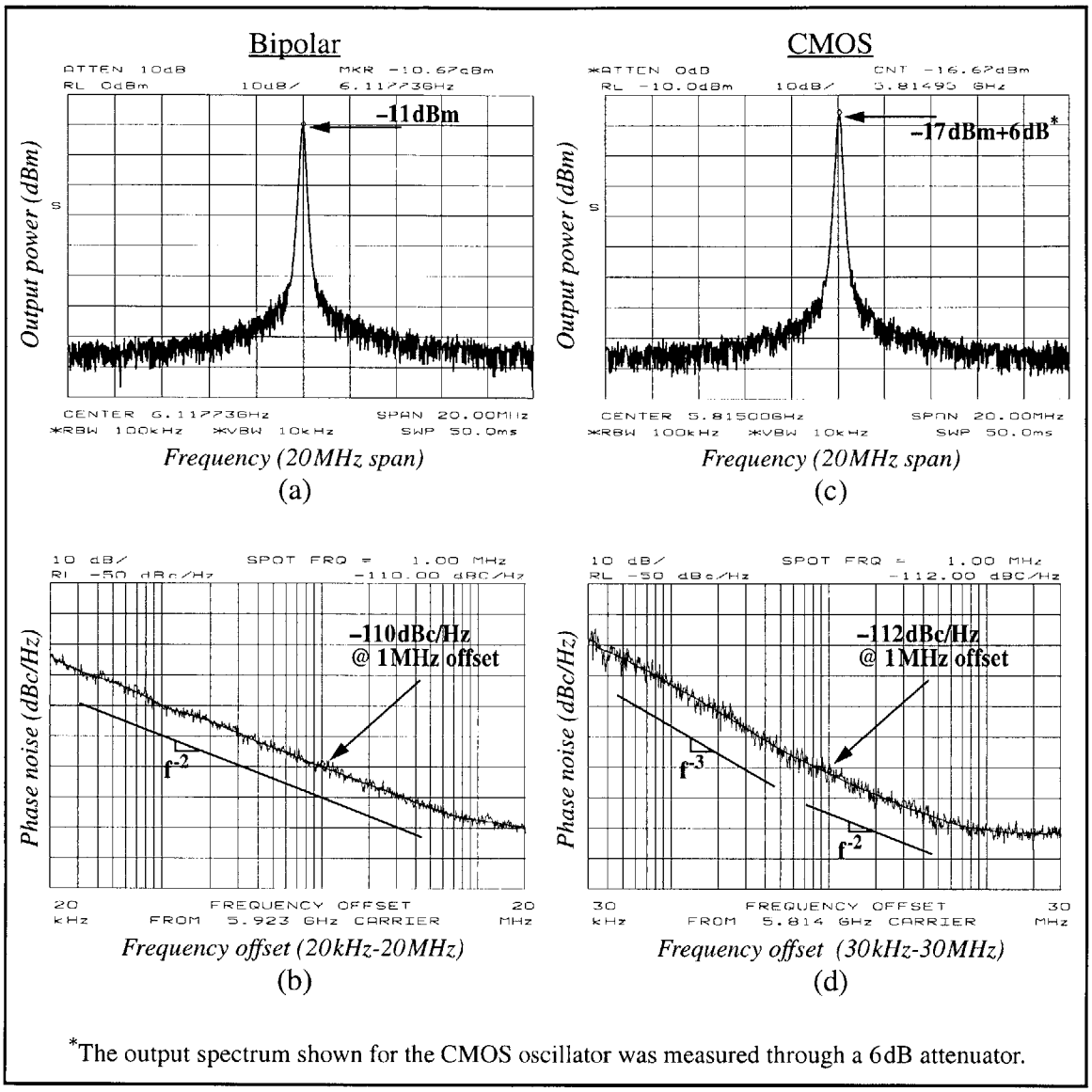

Figure 6-14. Measured spectrum and phase noise plots for the bipolar (a,b) and CMOS (c,d) 5.8GHz VCOs operated at 3V. The bias current in the oscillator core is 6mA for the bipolar VCO and 5.2mA for the CMOS version. Note the change in reference level between the spectrum plots in parts (a) and (c).

oscillator core when the CMOS version is left to its own devices, to which the adaptive biasing circuit adds 3mA of overhead. Given that the default currents are subject to on-chip component tolerances, the bias levels in the oscillator cores are impressively close to the nominally expected 6.25mA and 5mA, respectively. Although the oscillators provide every appearance of operating where intended, the measured phase noise is noticeably higher than had been than anticipated. Interestingly, each design seems to have missed by very nearly the same amount: 10-11dB. This disparity aside, the CMOS oscillator does

eke out better performance in the f^{-2} region of the spectrum, coming in 2dB lower than the bipolar design at a 1MHz offset. However, the f^{-3} characteristic extends farther from the carrier than had been expected with the CMOS oscillator, prevailing for offsets to 500kHz. Meanwhile, the corresponding upturn in the spectra of the bipolar oscillators is nowhere to be found in the open-loop measurements. Consulting the data gathered with the Aeroflex-Comstron system, the corner frequency related to flicker noise in the BJTs is located at 1kHz. Thus, despite the attempts at orchestrating a cancellation of low frequency noise in the CMOS architecture, the bipolar implementations remain substantially quieter near the carrier. Somewhat arbitrarily choosing an offset of 100kHz as another convenient reference point, the CMVCO design exhibits -90dBc/Hz whereas the noise from the CMOS version is 7dB higher.

Before trying to read much further into the measured results, it is important to gain an understanding of the reasons behind the higher than anticipated oscillator phase noise. The first stone has already been cast: by the observation that the CMOS and bipolar oscillators—using different (active) devices and topologies—are all higher than expected by about the same amount. The likely culprit is something in common among the designs, and the spiral inductor used in the resonant tanks is the most prominent carryover. A test structure for this inductor was placed on wafer with the bipolar oscillators, along with a set of calibration pads which allow the inductor to be de-embedded from measured data. While the DC resistance of the winding, at 0.66Ω, seems consistent with a quality factor in the high teens, behavior at RF is more of essence. Characterizing the inductor with s-parameter measurements taken to 6GHz, the loss at 5.8GHz extracts a Q of 6—far below the Q of 19 predicted by the scalable modeling approach proposed in Chapter 2.[49]

As the lumped-element model used to represent the spiral inductors (shown in Section 2.2) possesses more degrees of freedom than can be discerned from the data, a choice was made to fix the substrate resistance (R_{Si}) term to a value interpolated between two of the inductor structures for which circuit models had been provided with the technology. With a value of 497Ω set to each side of the π-model for R_{Si}, an optimizer is then given the freedom to adjust the remaining parameters. A good fit to the data is achieved when the oxide capacitance (C_{ox}) is about as predicted by the scalable models, although the capacitance attributed to the underpass (C_s) is nearly doubled. Compensating in part

[49] The 6GHz restriction was imposed by the available HP8753C network analyzer. Data to higher frequencies would have been useful in uniquely discerning a compact model for the inductor. However, when the quality factor is calculated from the circuit model fit to the data using the definition shown in the Appendix to Chapter 2, the value for Q comes out about the same regardless of how the measured loss and energy storage are partitioned into the elements of the model.

for the increased capacitive energy storage, the series inductance term is reduced by 25%. Most of the additional loss then is forced into the series resistance, which balloons by nearly a factor of 2.5 to match the data at 5.8GHz. A summary of the inductor parameters which compares the model used in the design against that fit to the measurements is provided in Table 6.1.

Table 6.1: Lumped-Element Model Parameter Summary for the Oscillator Inductor

	Scalable model	S-parameter fit
Series inductance (L_s)	0.63nH	0.46nH
Series resistance (R_s)	1.1Ω	2.64Ω
Shunt (underpass) capacitance (C_s)	36.3fF	61.1fF
Oxide capacitance (C_{ox})	42.1fF	44.7fF
Substrate capacitance (C_{Si})	43.4fF	—
Substrate resistance (R_{Si})	3.9kΩ	497Ω
Quality factor at 5.8GHz	19	6

Using the updated circuit model for the resonator inductor, the phase noise calculations are then performed again using the LTV model for the nominal design of the bipolar oscillator (in which the emitter lengths are 12.5μm). The answer from this analysis was provided earlier—as the single-ended example in Chapter 3. As shown in Figure 3-11, the revised expectation for the phase noise is calculated to be -109.6dBc/Hz at a 1MHz offset, very precisely matching the measured result in Figure 6-14(b).[50] The correspondence here between the measurements and the calculations not only lends further credence to the assertion of heavier loss being sustained in the inductors, but also indicates the importance of removing the transistor junction currents in the determination of shot noise. Returning to the argument presented in Section 3.4, appreciable components of the transistor terminal currents were observed to be associated with charging and discharging the junctions—current components which do not generate shot noise. By (erroneously) using the full base and collector terminal currents, the resulting calculation overestimates the actual oscillator phase noise by a rather significant 3dB (as graphically depicted in Figure 3-12).

Although the departure in the loss of the oscillator inductor from that predicted by the model appears responsible for the higher than expected noise, several additional exper-

[50] This may surpass what was previously considered to be the most delayed punchline in modern technical literature (Section 3.4).

iments were conducted to gauge the effects of other elements on the RF performance while the opportunity for exploration was at hand. First to be assessed were the implications of loading and noise imparted upon the oscillator by the on-chip buffer. Using some laser cuts, the buffer was detached on a set of the bipolar oscillator ICs; although cutting the traces attenuated the output by nearly 20dB, a sufficient amount of power made its way to the RF pads to allow a measurement. No change could be observed in the oscillator phase noise compared to when the buffer was attached, demonstrating the efficacy of the chosen topologies. Another set of tests revolved around the question of whether varactor loss was significant enough to be noticed in comparison to the inductors. To reduce the sensitivity of the oscillator to any noise that may exist on the applied tuning voltage, the control signal first passes through a small (10Ω) series resistance followed by a bypass capacitor. Through rework performed by a focused ion beam (FIB) machine, the first modification was to jettison the capacitor in the RC filter, leaving only the 10Ω resistor in series with the varactor. A second alteration deposited a tungsten wire to shunt the resistor, leaving only the filter capacitance. The first of these changes should increase the effective loss in the varactors while the second should reduce the loss. But proving uninteresting, the measured oscillator phase noise was unswayed by either modification, indicting the inductors in limiting the quality factor of the resonant tank. As a third empirical inquiry, the influence of noise in the biasing circuitry is checked by remeasuring the phase noise of the oscillator with the common-base bias voltage (*base*) being supplied by an off-chip voltage source. To provide the needed connection, additional FIB rework has been used to rewire the *base* node in the core to a pad in place of the *bias_adj* control. Yet, once again, there was no appreciable difference to be noted in the phase noise spectrum. While not a complete test, this does provide some assurance that the oscillator performance has not been compromised through the adjustable bias circuits.[51]

Despite the appreciably degraded quality factor, that the same inductor has been used in each design should allow some of the relative comparisons to hold. But the one quality that does change markedly is the equivalent load resistance (R_{EQ}) presented by the resonant tank, in turn affecting the proportionality between the bias current in the oscillator core and the signal swing it develops. This effect becomes evident in the flatter than expected slope exhibited by the data in Figure 6-15, where an inverted measure of phase noise has been plotted against the bias current flowing in the core of the CMVCO bipolar oscillator. The metric being used in this representation is the reciprocal of the phase noise

[51] With all of the effort invested in the experiment, finding that nothing had changed in the least proved mildly bittersweet.

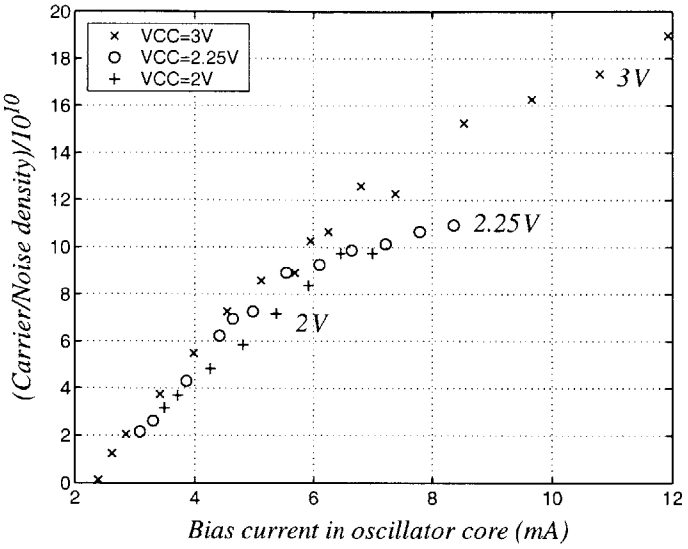

Figure 6-15. CMVCO phase noise behavior as a function of bias current and supply voltage, expressed as ratio of carrier power to noise power in a 1 Hz bandwidth located 1 MHz away from the carrier. Operation at voltages above 3 V did not improve the phase noise due to saturation in the biasing circuit.

spectral density at a 1 MHz offset, yielding the ratio of the signal power to the noise power in a 1 Hz bandwidth—where the bandwidth being measured is located at a frequency displaced from the carrier by 1 MHz. As postulated in Section 6.1.1, when the oscillator is in a regime where its signal swing is limited by the available current, the energy stored in the resonant tank will be proportional to the square of the bias current. But at least some of the noise sources in the oscillator—specifically those related to shot noise—possess an energy that increases linearly with current, creating a plausible scenario in which the carrier to noise ratio (C/N) exhibited by the oscillator follows IR_{EQ}^2 over at least some portion of the operating range. Judging from the measurements taken at supply voltages of 2 V, 2.25 V, and 3 V, this appears to be the prevailing mode for the nominal bipolar oscillator design. When the oscillation is current-limited, increasing the supply alone has little effect on phase noise; however, in creating additional headroom, a higher voltage can pave the way for an increase in the bias current to improve the phase noise. This characteristic begins to be evident in comparing the performance at 2.25 V and 3 V.[52] Unfortunately, the lower R_{EQ} presented by the resonator shifts the onset of voltage-limited operation to cur-

[52] The oscillation amplitude in the bipolar oscillator is limited by the V_{CE} of the common-base transistor, where the collector is at the supply voltage and the emitter rests at a level of $2V_{BE}$ as set by the base bias. Hence the increase in achievable C/N enabled by raising the supply from 2.25 V to 3 V is $(3\,\text{V} - 2V_{BE})/(2.25\,\text{V} - 2V_{BE})$—equaling a factor of 2 for a 0.75 V V_{BE}.

rent levels that are much higher than had been anticipated, leading to a situation where the biasing circuit runs out of base drive well before the optimum phase noise performance of the oscillator is reached.[53] Being thusly inhibited, the bipolar oscillators are unable to demonstrate further reductions in the phase noise with supplies above 3 V. Also being cast into murkiness by the base drive issue is the roll-off indicated in the carrier to noise ratio with higher currents. While headroom limitations could be assumed responsible, base drive starvation in the oscillator may also be playing a role.

Thankfully, there is less ambiguity with the CMOS oscillator. Although saturation in the biasing circuitry again encroaches on the operating range, it happens through a different mechanism and not quite as severely. The maximum bias current occurs when the full supply voltage is applied as the gate-source potential of the tail transistor; as with the bipolar designs, this limit will be encountered in some instances before the voltage-limited regime takes hold of the oscillator, but a greater range in the behavior can be seen. Represented by the more conventional phase noise metric, the spectral density at a 1 MHz offset from the carrier is again chosen as the point of comparison.[54] From this vantage point, the performance of the CMOS design is characterized in Figure 6-16 as the supply voltage is varied from 2.25 V to 5 V and the bias current in the oscillator core is independently adjusted from 2.7 mA to almost 17 mA.

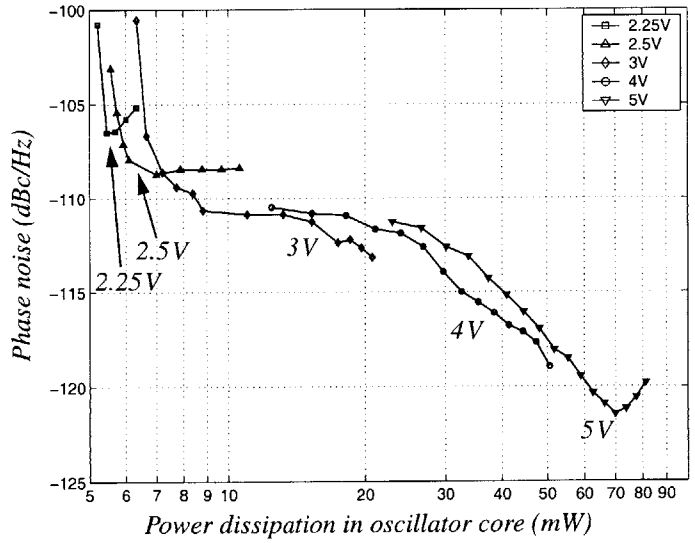

Figure 6-16. CMVCO-CM phase noise behavior as a function of power dissipation.

[53] The limitation flows from the inability of the unity-gain buffer in the bias circuit to supply enough base drive to satisfy the common-base gain transistors when operating at higher collector currents.

[54] Frequencies offset from the carrier by 1 MHz fall safely into the f^{-2} region of the phase noise spectrum, but yet remain well above the noise floor of the spectrum analyzer.

Now readily substantiated with the data provided is the concept of an optimum bias current. On one side of this coin, operating an oscillator in a current-limited region wastes power, as the same phase noise can be obtained with a lower supply voltage. In Figure 6-16, it can be observed that the CMOS oscillator delivers -112dBc/Hz for a 6.5mA bias current whether the supply is 3V, 4V, or 5V. But pushing the current beyond the "knee" into voltage-limited operation actually begins to increase the phase noise, an effect witnessed in the 2.25V and 5V traces. However, the optimal bias current and the realizable phase noise performance both increase with supply voltage, where the adaptive biasing circuitry enables a 15dB range of control in the phase noise to be achieved by adjusting the power consumption in the oscillator—while operating at the optimum current for each supply voltage. Measured at a 1MHz offset from the carrier, the noise improves from -106dBc/Hz for 5.5mW (at 2.25V) to -121dBc/Hz for 70mW (at 5V). A remarkable observation from this data is that the performance versus power consumption trade-off attains a slope that is greater than unity: a 15dB improvement in phase noise comes for an 11dB increase in power. This characteristic seems to suggest that there is not a significant noise penalty associated with the adaptive bias scheme.[55]

Returning to the broader issue of technology, it is interesting to conclude this section by bringing together a direct comparison of the four oscillators. Again using the carrier to noise density metric, the performance of each version is plotted in Figure 6-17 as a function of the bias current in the oscillator core. The first comparison in this experiment is between two sizes of the same bipolar transistor, where the smaller of the two had been suggested as the optimum by a conventional LTI analysis and the larger took into account the time-variant behavior intrinsic to oscillators. Perhaps as the first casualty of the inductor loss being higher than forecast, much of the distinction between the two sizes appears to have been washed out. The measured phase noise performance was nearly the same, with the smaller transistors yielding marginally better numbers. From additional work with SpectreRF, it is believed that the lower Q resonator has broadened the minima in the phase noise as a function of transistor size, and has shifted the optimum device to a size sitting between the two that were selected. Displaying a more noticeable difference, the design which used the NPNs with the more lightly-doped intrinsic collector (and a correspondingly lower peak f_T) outperformed the other two bipolar versions. Also apparent is that the *slope* of the characteristic is steeper with the low f_T devices, indicating that the equivalent load resistance presented by the resonator is slightly higher in this version.

[55] Perhaps a better indication is that the CMOS design, at least in the white noise dominated portion of the phase noise spectrum, outperforms any of the bipolar versions.

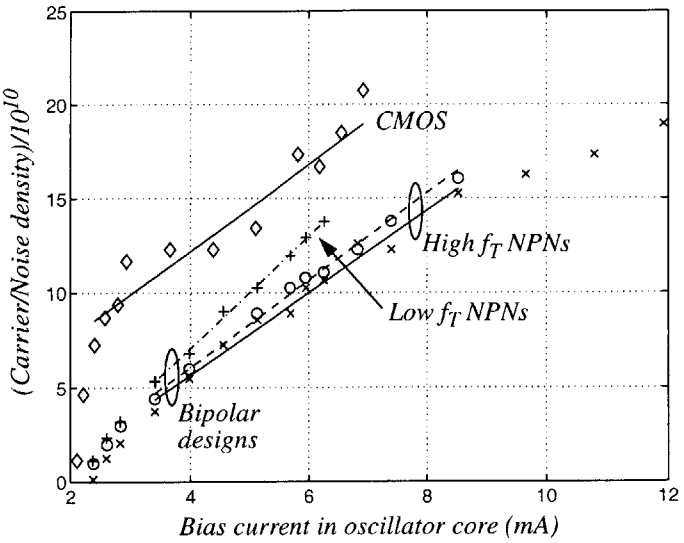

Figure 6-17. Technology comparison of phase noise behavior as a function of bias current in oscillator core, expressed as a ratio of carrier power to noise power in a 1 Hz bandwidth located 1 MHz away from the carrier. A 3 V supply has been used for these measurements.

This supports the notion that the smaller base-collector and collector-substrate capacitances resulting from the lighter doping levels can translate into a palpable improvement in the resonator Q. Finally, given the same set of passive devices and the same 0.5 μm minimum feature size, the best performance was provided by the CMOS oscillator. Despite having the lowest f_T of the bunch, the reduced parasitic resistances and capacitances associated with the device seemed to carry the day—at least for frequencies far enough away from the carrier for flicker noise not to dominate the phase noise spectrum. *It is further interesting to note that the f_T comparison between the three different device types holds not only for the peak value that can be reached, but also for the operating f_T at the (time-averaged) bias current densities reflected in the actual oscillators.* This remains true over the entire range of oscillator core bias currents illustrated by the data in Figure 6-17.

A summary of the findings in the comparison between the bipolar and CMOS 5.8 GHz oscillators is furnished in Table 6.2. Representing the bipolar side are the versions with the higher f_T transistors. The lower f_T NPNs possess a slightly reduced DC current gain, exacerbating the difficulties noted with the limited base drive afforded by the biasing circuitry. Highlights of the oscillators operating with their on-chip established default bias currents are provided for 3 V and 2.25 V supplies. Also noted in the table is the lowest phase noise that could be observed for each oscillator and the bias condition

from which it resulted. But here too, the bipolar versions were hindered by the limited base drive, constraining the bias currents and the phase noise performance well below that associated with the onset of the voltage-limited regime.

Table 6.2: Performance Summary for the C-band Monolithic VCOs

	Bipolar	CMOS
3V operation, "default" bias		
Tuning range	5.4-6.1GHz 0-3V	5.2-5.9GHz 0-2V
Bias current in oscillator core	6.0mA	5.2mA
Biasing circuitry overhead	0.7mA	3.0mA
Phase noise, 100kHz offset	-90dBc/Hz	-83dBc/Hz
Phase noise, 1MHz offset	-110dBc/Hz	-112dBc/Hz
Reduced power: 2.25V, "default" bias		
Bias current in oscillator core	4.6mA	2.8mA**
Phase noise, 1MHz offset	-108dBc/Hz	-105dBc/Hz**
Lowest noise operation (observed)		
Supply voltage	3V*	5V
Bias current in oscillator core	11.9mA*	14mA
Phase noise, 1MHz offset	-113dBc/Hz*	-121dBc/Hz
Output buffer (all modes)		
Bias current @ 3V operation	4.4mA	4.7mA
Q of 6 for oscillator inductors (at 5.8GHz)		

*Performance is limited by saturation in the biasing circuitry.
**The "default" bias in the CMOS oscillator at 2.25V is in the voltage-limited regime.

6.4 Observations on Oscillators

Resting on the foundation of a pair of oscillator topologies known for low noise performance and an ease of integration, a set of four VCOs has been designed by applying insight gleaned from studying a linear time-variant model of oscillator behavior. Circuits were then developed to offer an adjustable bias feature in the oscillators, instilling an

added degree to controllability without degenerating the RF performance. Next, concern for providing isolation and the need to drive a 50Ω load for frequencies near the transistor f_T led to the design of two impedance-matched output buffers based on differential pairs. As with the biasing circuitry, much of the emphasis in the design of the buffers centered around not degrading the performance of the oscillator, wherein a trade-off between input impedance and gain was noted to exist. Favoring the goal of minimizing the loading upon the oscillator cores, the gain provided by the buffers wound up being rather modest.

Using the same suite of passive devices, oscillators were constructed of three transistor types and optimized for the best performance each could attain. Although the resonator Q fell far short of the values predicted during the design, these three versions yielded some interesting comparisons. As an unmistakable exhibition of the role played by transistor terminal parasitics, a ranking of the measured phase noise—in the f^{-2}, white noise dominated, region of the spectrum—went from best to worst as the device f_T went from lowest to highest. This is not to suggest an inverse relationship between switching speed and oscillator noise, but does serve to indicate the relative unimportance of the transistor f_T—once f_T is beyond the oscillation frequency by a factor of 3 to 5. In the search for a relevant figure of merit for transistors, f_{MAX} fares no better as the same ordering (in f_{MAX}) occurs for the devices in this comparison.

In hindsight, the selection of the cross-coupled topology with both n-channel and p-channel pairs might not have been the best choice for the CMOS oscillator. By costing headroom through its stack of three devices, this architecture effectively trades increased noise in the f^{-2} region for lower f^{-3} noise. And unless the match between n-channel and p-channel drive strength is fairly precise, a difficult proposition to ensure over process and temperature, the degree to which the effects caused by low frequency noise are cancelled may not be particularly high. The best reported results have typically observed a reduction of 20-30dB in the upconversion of flicker noise, and it is not likely that any cancellation technique would accomplish much more than this. While significant, for applications where the noise at frequencies farther removed from the carrier is of greater importance, better results probably would have been achieved had only one of the cross-coupled pairs been used.[56]

One other aspect of adaptive performance versus power consumption may be worth considering. The reason that the phase noise improves with higher bias currents (until voltage limitations set in) is that the amplitude of the signal increases faster than the

[56] In addition, the noise in the f^{-3} region may even have been lower were only PMOS devices used in the oscillator core.

noise. A mixer, provided that the oscillator is driving one, may also stand to benefit from the higher signal level. A number of the more salient features of a mixer—including the noise figure, conversion gain, and linearity—generally improve with increased oscillator drive levels. Within the context of a receiver, this may constitute a two for one deal on sensitivity: as more power is spent in the oscillator, the signal to noise ratio supplied by the RF front end will be improved by lowering both reciprocal mixing due to phase noise and the noise figure of the mixer.

Once again, accurate inductor modeling is demonstrated to be central to achieving first-pass success in RFICs. Despite the repeated criticisms leveled at the scalable models, a few words are owed on their behalf: it should be noted that the beleaguered spiral inductor models did not completely fail to find the mark. Although not shown among the measurements in this chapter, the return loss at the output ports of the oscillators reached a very respectable 18 dB, found centered at the targeted 5.8 GHz. As the MIM capacitors for the process run were very near the nominally specified values, the excellent output match indicates that the model for the inductor used in the oscillator buffers was reasonably close. The one difference between this inductor and the others is that the buffer inductor was optimized for area rather than a high quality factor. Thus it appears that the scalable models provide better representations when the desired Qs are more modest.

So, after all of this, what is the best technology for oscillators? When possessing terminal resistances and capacitances on par with bipolar transistors, the CMOS designs have proven themselves to be very competitive in the arena of RF applications where close-in phase noise tends to be less of a concern. Moreover, techniques have been demonstrated to reduce both the effects and the occurrence of flicker noise [38][86]. However, when the device 1/f noise in vertical (silicon) bipolar transistors is three orders of magnitude smaller than that exhibited by MOSFETs, a gap exists which is difficult to close by circuit cancellation techniques. Given this advantage, and when properly scaled in consideration of the lessons presented by this work, a bipolar technology making use of the latest lithography tools will be difficult to beat. The primary challenge will be to avoid getting caught up in an f_T footrace.

Chapter 7

Final Thoughts

Set against the backdrop of high data rate, adaptive wireless networks, a discourse into the concurrent optimization of circuits and technology for RF/microwave ICs has been presented. Approaches to considering the manifestation of noise in linear time-invariant and linear time-variant RF circuits have been described, and were then developed as tools for identifying the physical limitations to circuit performance as constrained by the devices. Design techniques were applied to mitigate these limitations to the extent possible while enabling the concept of adaptive performance, resulting in several new circuits, the illustration of a number of key effects, and the most direct comparison of transistor technologies conducted to date. Completing the story are measured results from the characterization of an experimental set of voltage-controlled oscillators (VCOs) and low-noise amplifiers (LNAs), offering confirmation for many of the observations reached by applying the tools fashioned herein.

This treatise began with a look at the modeling of on-chip spiral inductors. From measurements at the circuit level and of an inductor test cell, it was subsequently determined that an analytical treatment of the substrate (presented in Section 2.2) failed to accurately capture the loss mechanisms inherent to the medium. Instead, interpolating the substrate capacitance and resistance components for spiral inductors within a set of measured data points would be a more effective avenue to realizing a scalable model.[j1] In addition, eddy currents in the substrate (Section 2.3) should be included through a coupled secondary winding, and current crowding in the spiral trace (Section 2.2.1) needs to be incorporated to represent behavior at gigahertz frequencies. Lacking these effects, a scalable model leads an optimization for quality factor astray by overpredicting the reduction in the series resistance and underpredicting the substrate loss associated with using wider metal traces.

[j1] Full electromagnetic simulations tend to be too unwieldy for the optimization of an inductor geometry; however, such simulations could be used to generate the data points for interpolation by the scaling rules discussed in Section 2.2.

One application where inductors prove instrumental is in tuning a narrowband LNA stage for gain and low noise. A number of topologies were explored in Chapter 4 for suitability as RF amplifiers; of those analyzed, the common-emitter configuration was found to work very well when equipped with base and emitter inductors to deliver optimum noise and impedance matches to the RF source. Moreover, it was determined that the noise figure of this stage effectively represents the technological limit to the achievable sensitivity. This limit arrives at an optimum collector current density ($J_{C_{opt}}$); by materializing a balance between shot noise and gain, $J_{C_{opt}}$ represents the current at which the noise of a bipolar transistor is minimized, and it is solely a parameter of the technology.[j2] In this low noise regime, the operation of the device is primarily limited by the junction capacitances. The minimum noise figure, and the power consumption required to bring about this condition, both decrease as the junction capacitance is reduced; the peak transistor f_T is actually of little consequence for an LNA. Given a fixed source resistance, noise figure and power dissipation approximately follow $\sqrt{C'_{jE} + C'_{\mu}}$, where the terms under the radical represent the base-emitter and base-collector junction capacitances per unit emitter area (Section 4.2). Although often the smaller of the two by an appreciable amount, C_{μ} plays a significant role in many other features of an LNA including: stability (Section 4.4), isolation (Section 4.3), gain (Section 4.3.2), and noise figure (Section 4.1). Hence the depletion capacitances at both intrinsic device junctions should be considered important trade-offs at the device design level.

Having located the technology-imposed barriers, the next question to be addressed is the relation between the noise figure of an actual LNA and the minimum noise figure of a transistor. Starting with a 1.5dB NF_{min} in a 5.8GHz design, adding a cascode device resulted in a 0.3dB hit to the noise figure, and incorporating a reduced power mode elevated this figure another 0.25dB (Section 5.1). A base current source bias circuit was then developed to protect linearity while keeping the noise figure degradation due to biasing well under 0.1dB (Section 5.1.3). The remaining notable contribution to the noise figure occurs from loss in the inductors. When the quality factors are very good (17-19), the anticipated noise figure penalty is rather mild (0.2dB). However, with more typical Qs of 5-6, an increase of 1.2dB is attributable to the spiral inductors (Section 5.3). Considering the constant transconductance biasing circuit, useful low power mode, and 50Ω impedances required of a practical design for wireless applications, the LNA in this work demonstrates that a 3.4dB (minimum) noise figure is achievable[j3] in the 5-6GHz range using a

[j2] The optimum current density does increase with the frequency of operation, but is independent of the device size.

[j3] This is assuming an optimum noise match is provided for the LNA.

0.5μm SiGe graded-base bipolar process. This represents a cost of almost 2dB above NF_{min} for an individual transistor in the technology.

An interesting thought to ponder in light of a CMOS implementation is whether the loss and added noise incurred due to the inductors might mask any disparities in NF_{min} at the transistor level. This does not appear to be the case, however, in comparing the bipolar and CMOS LNAs discussed in Chapter 5. Dissipating 14mW from a 3V supply, the CMOS design delivered a minimum noise figure of 4.5dB and an available gain of 5.2dB, falling short on both counts relative to the 3.4dB and 11dB numbers yielded by the bipolar LNA on only 8mW from the same 3V supply.[j4] The 1.1dB disparity is roughly the same as that expected between the intrinsic devices, a correspondence resulting from two differences in the impedance characteristics of the transistors which tend to cancel. The input impedance presented by a MOSFET is relatively insensitive to the bias current, making the adaptive performance concept possible without necessitating a parallel (low power) stage. Unfortunately, power constraints limit the size of the FETs (Section 5.2), thus mandating more inductance to realize an input match; the higher inductor loss associated with the CMOS design offsets the benefit of not having to incur the penalty following from the parallel stage architecture. Another issue with MOSFET impedances pertained to whether an optimum source resistance of 50Ω could be achieved (for reasonable power dissipation levels). Although the $R_{S_{opt}}$ expected from applying the long channel approximation is in the neighborhood of 90Ω, the measured optimum source impedances were found outside the 50Ω circle; the correspondingly low $R_{S_{opt}}$ values suggest that the drain and induced gate noise currents are notably higher than predicted by the long channel model—even in this 0.5μm technology.

While LNAs typify the design of LTI circuits for RF applications, another important class of circuits is exemplified by the periodic LTV behavior of oscillators. A significant distinction between the two classes is that the sensitivity to noise in time-variant circuits is one of the parameters which may vary in time. As a corollary, the noise sources—particularly those associated with transistor conduction—may also exhibit a time-dependent power spectral density. In modeling this behavior, two key extensions to the work of Hajimiri and Lee [38] have been presented. The first extension is in noting that transistors in time-variant systems cannot be represented by equivalent voltage and current noise sources (Section 3.3). Each terminal resistance and conduction noise gener-

[j4] The available gain measurements represent the gain which could be achieved were the source to present a reflection coefficient Γ_{opt}. However, the optimum source impedances determined for the LNAs were not too far removed from 50Ω.

ator must be considered independently; thermal noise in terminal resistances is always present while conduction noise may only exist during a fraction of the period. A second important addition to the LTV model is to recognize that conduction noise is not determined by the total current flowing into the corresponding transistor terminal (Section 3.4). In a bipolar transistor, significant portions of the base and collector terminal currents are associated with charging and discharging the base-collector, base-emitter, and collector-substrate depletion regions—capacitive components which do not generate shot noise. Failure to subtract the displacement currents in the calculation of noise was shown to result in a significant error in a 5.8GHz design, illustrating an effect which will become increasingly profound at higher frequencies.

Resulting from this development is a tool by which the impact of a device technology can be considered for oscillators (and other LTV circuits). Discontinuous conduction in oscillators—when timed appropriately—effectively de-emphasizes the transistor current noise sources. Once the transistor f_T is a factor of 3 to 5 beyond the frequency of oscillation, the time-varying behavior is controlled by the circuit and not the device; beyond this point further increases in f_T have no material effect on oscillator performance. This assertion is demonstrated in a comparison of 5.8GHz oscillators presented in Section 6.3: using the same inductors, a design implemented with 0.5μm CMOS transistors (15GHz peak f_T) produced lower phase noise in the f^{-2} region of the spectrum than did the bipolar incarnations of the oscillator. Among the bipolar versions, the oscillator making use of lower f_T NPNs (peaking at 27GHz) outperformed the same design when NPNs with twice the (peak) f_T were inserted. Of the three, CMOS emerged victorious in the comparison due to having smaller terminal access resistances and parasitic capacitances than the 0.5μm bipolar transistors in this technology. Similarly, lower noise was obtained by the "slower" NPNs in the bipolar oscillator due to a reduction in the base-collector and collector-substrate capacitances—achieved by masking an intrinsic collector implant designed to push out the onset of the Kirk effect (to higher collector current densities).[j5]

Although the RF signal paths have received most of the attention, biasing and output loading can figure prominently in the circuit performance. Noise from the biasing circuitry may easily overwhelm that in the signal path, necessitating careful design and often a non-negligible power consumption (in the bias circuits) to keep the noise low. Several new circuit techniques were developed to mitigate this concern while extending the

[j5] "Slower" is chosen here primarily to reflect the general stigma attached with having a lower peak f_T.

achievable RF performance through adjustability. A switched current source biasing approach was implemented for the bipolar LNA (Section 5.1.3), providing a stable transconductance (over temperature) and a high output impedance to minimize the impact on the noise figure and linearity of the amplifier. Although the topology and the insulating gate made the biasing chore much easier for the CMOS LNA (Section 5.2.3), a temperature-stabilized transconductance was also realized within it by incorporating a ΔV_{GS} bias reference.[j6] The circuit used for biasing the bipolar oscillators was designed to reuse reference legs to reduce power consumption while providing low noise (Section 6.1.3). Finally, a feedback biasing scheme was developed for the CMOS VCO to allow control of the current in the oscillator core through a tail source device which operates in triode to maximize the voltage headroom afforded to the oscillation (Section 6.2.1). Each of these circuits provides a "default" bias condition established by an on-chip resistor, but can be adjusted via an external input to support the adaptive operation which will become a cornerstone of future, bandwidth on demand, wireless networks.

Through the designs presented in this work, an approach has been illustrated for thinking about the ramifications of a technology upon the LTI and LTV components of an RF system. From these examples in a 0.5μm BiCMOS process, a number of generalities have been extended to guide the development of semiconductor technologies for wireless communications. One of the most poignant findings is that peak transistor f_T is overemphasized for RF circuits. Simply put, a higher device f_T does not automatically improve everything; in fact, depending on which trade-offs are invoked, increasing the f_T may actually translate to lower performance. Junction capacitances play a far more important role in low noise, low power applications than does base transit time (τ_F), bringing into question the trend toward thinner base devices. Accompanying the thinner base layers has been an increase in the doping concentrations applied to the (intrinsic) base and collector regions, leading to higher junction capacitances and hindering the RF performance being sought. The real improvements being observed from advanced silicon-based bipolar technologies can be traced to the continued shrinking in the lithographic dimensions, effectively overcoming the penalties associated with recent directions in the stack design for bipolar transistors. Another choice which should be questioned is the graded germanium profile incorporated into most modern SiGe bipolar technologies. A graded profile is adopted to reduce τ_F by creating a drift field for minority carriers in the base. But for use in radio circuits, a box profile can provide advantages by creating a valence band disconti-

[j6] Both LNAs will exhibit some residual temperature dependence as the on-chip resistors used to establish the current possess a non-zero T_C. However, the first-order effects in the transconductance have been compensated as illustrated through simulation in Figure 5-11.

nuity at the base-emitter junction. With an energy barrier in place to block the injection of holes from the base, the emitter doping concentration can be lowered to reduce C_{jE}.

For RF applications, CMOS has demonstrated performance that is competitive with implementations in mature bipolar processes—even winning along some oscillator metrics. However, while the noise figure and f_T in MOSFETs have seen marked improvements with scaling, the question going forward will be the extent to which this trend will continue in the face of high field mobility degradation and increases in the drain and induced gate noise currents which seem to be outpacing the gains in transconductance. Furthermore, the higher flicker noise in CMOS appears to remain a significant issue, as low frequency noise can be addressed only to a degree through circuit cancellation techniques at RF—and not likely by enough to close the gap with bipolar devices. But perhaps the better question is to ask which device technology has more room for improvement. Keeping in mind the lessons advanced through this work, as CMOS and bipolar processes are aggressively scaled, the performance of a bipolar technology will prove difficult to beat in the RF arena.

Bibliography

Wireless Systems

[1] Federal Communications Commission News Report No. DC 96-36, *Commission Proposes to Make Spectrum Available for Use by New Unlicensed Equipment; NII/ SUPERNet Devices Would Facilitate Interconnection to NII*, April 25, 1996.

Inductor Modeling

[2] T. H. Lee, *A Fully Integrated Inductorless FM Receiver*, Sc.D. Thesis, Massachusetts Institute of Technology, December 1989.

[3] K. W. Kobayashi and A. K. Oki, "A Novel Heterojunction Bipolar Transistor VCO Using an Active Tunable Inductance," *IEEE Microwave and Guided Wave Letters*, July 1994, pp. 235-237.

[4] S. Hara and T. Tokumitsu, "Monolithic Microwave Active Inductors and Their Applications," *IEEE International Sympoisum on Circuits and Systems*, June 1991, pp. 1857-1860.

[5] P. R. Nuytkens, et al., "A Monolithic Tunable High-Q InAlAs/n^+–InP–HFET Resonator," *Indium Phosphide and Related Materials Conference*, April 1993, pp. 389-392.

[6] C. H. Leong and G. W. Roberts, "A Sixth-Order UHF Bandpass Filter using Silicon Bipolar Active Inductors," *International Symposium on Circuits and Systems*, June 1998, pp. 127-130.

[7] D. A. Hitko, *A Low Power, Low Noise, 1.8GHz Voltage-Controlled Oscillator*, S.M. Thesis, Massachusetts Institute of Technology, February 1997, chapter 3.

[8] M. Steyaert and J. Craninckx, "1.1GHz Oscillator Using Bondwire Inductance," *IEE Electronics Letters*, February 3, 1994, pp. 244-245.

[9] T. I. Ahrens and T. H. Lee, "A 1.4-GHz 3-mW CMOS LC Low Phase Noise VCO Using Tapped Bond Wire Inductances," *International Symposium on Low Power Electronics and Design*, August 1998, pp. 16-19.

[10] P. J. Sullivan, et al., "An Integrated CMOS Distributed Amplifier Utilizing Packaging Inductance," *IEEE Transactions on Microwave Theory and Techniques*, October 1997, pp. 1969-1976.

[11] D. C. Laney, et al., "Lateral Microwave Transformers and Inductors Implemented in a Si/SiGe HBT Process," *IEEE MTT-S International Microwave Symposium*, June 1999, pp. 855-858.

[12] J. R. Long and M. A. Copeland, "The Modeling, Characterization, and Design of Monolithic Inductors for Silicon RF IC's," *IEEE Journal of Solid-State Circuits*, March 1997, pp. 357-369.
[13] F. W. Grover, *Inductance Calculations: Working Formulas and Tables*, D. Van Nostrand Company, New York, 1946.
[14] H. Hasegawa, M. Furukawa, and H. Yanai, "Properties of Microstriplines on Si-SiO_2 System," *IEEE Transactions on Microwave Theory and Techniques*, November 1971, pp. 869-881.
[15] E. Pettenpaul, et al., "CAD Models of Lumped Elements on GaAs up to 18GHz," *IEEE Transactions on Microwave Theory and Techniques*, February 1988, pp. 294-304.
[16] C. P. Yue and S. S. Wong, "Physical Modeling of Spiral Inductors on Silicon," *IEEE Transactions on Electron Devices*, March 2000, pp.560-568.
[17] A. M. Niknejad and R. G. Meyer, "Analysis, Design, and Optimization of Spiral Inductors and Transformers for Si RF IC's," *IEEE Journal of Solid-State Circuits*, October 1998, pp. 1470-1481.
[18] H. M. Greenhouse, "Design of Planar Rectangular Microelectronic Inductors," *IEEE Transactions on Parts, Hybrids, and Packages*, June 1974, pp. 101-109.
[19] S. S. Mohan, et al., "Simple Accurate Expressions for Planar Spiral Inductances," *IEEE Journal of Solid-State Circuits*, October 1999, pp. 1419-1424.
[20] W. B. Kuhn and N. M. Ibrahim, "Approximate Analytical Model of Current Crowding Effects in Multi-Turn Spiral Inductors," *IEEE Radio Frequency Integrated Circuits Symposium*, June 2000, pp. 271-274.
[21] C. P. Yue and S. S. Wong, "On-chip Spiral Inductors with Patterned Ground Shields for Si-based RF IC's," *IEEE Journal of Solid-State Circuits*, May 1998, pp. 743-752.
[22] H. G. Booker, *Energy in Electromagnetism*, Peter Peregrinus Ltd., New York, 1982.
[23] C. A. Desoer and E. S. Kuh, *Basic Circuit Theory*, McGraw-Hill, Inc., New York, 1969; refer to the applications of Tellegen's Theorem in Chapter 9, Section 5.4.
[24] K. O, "Estimation Methods for Quality Factors of Inductors Fabricated in Silicon Integrated Circuit Process Technologies," *IEEE Journal of Solid-State Circuits*, August 1998, pp. 1249-1252.
[25] MATLAB, Mathworks, Inc., *http://www.mathworks.com*, Natick, MA.
[26] FastHenry, Research Laboratory of Electronics, Massachusetts Institute of Technology, *ftp://rle-vlsi.mit.edu/pub/fasthenry*, Cambridge, MA.

Oscillator and Mixer Modeling

[27] K. Kurokawa, "Some Basic Characteristics of Broadband Negative Resistance Oscillator Circuits," *Bell System Technical Journal*, July-August 1969, pp. 1937-1955.
[28] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proceedings of the IEEE*, February 1966, pp. 329-330.
[29] H. K. Lindenmeier, "Noise Matching Techniques in Transistor Oscillators," *IEEE Circuits and Systems Symposium* , 1980, pp. 1052-1055.

[30] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, New York, 1998, chapter 12.
[31] J. Groszkowski, "The Interdependence of Frequency Variation and Harmonic Content, and the Problem of Constant-Frequency Oscillators," *Proceedings of the IRE*, vol. 21, no. 7, 1934, pp. 958-981.
[32] Private communication with Dan McMahill.
[33] G. R. Olbrich, et al., "Calculation of HEMT Oscillator Phase Noise Using Large Signal Analysis in Time Domain," *IEEE MTT-S International Microwave Symposium*, June 1994, pp. 965-968.
[34] A. Hajimiri and T. H. Lee, *The Design of Low Noise Oscillators*, Kluwer Academic Publishers, Norwell, MA, 1999, Chapter 4.
[35] C. D. Hull and R. G. Meyer, "A Systematic Approach to the Analysis of Noise in Mixers," *IEEE Transactions on Circuits and Systems I*, December 1993, pp. 909-919.
[36] F. X. Kärtner, "Analysis of White and $f^{-\alpha}$ Noise in Oscillators," *International Journal of Circuit Theory and Applications*, September-October 1990, pp. 485-519.
[37] A. Hajimiri and T. H. Lee, *The Design of Low Noise Oscillators*, Kluwer Academic Publishers, Norwell, MA, 1999, Appendix B.
[38] A. Hajimiri and T. H. Lee, "A General Theory of Phase Noise in Electrical Oscillators," *IEEE Journal of Solid-State Circuits*, February 1998, pp. 179-194.
[39] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase Noise in Oscillators: A Unifying Theory and Numerical Methods for Characterization," *IEEE Transactions on Circuits and Systems I*, May 2000, pp. 655-674.
[40] J. Rael and A. Abidi, "Physical Processes of Phase Noise in Differential LC Oscillators," *IEEE Custom Integrated Circuits Conference*, May 2000, pp. 569-572.
[41] Q. Huang, "Phase Noise to Carrier Ratio in LC Oscillators," *IEEE Transactions on Circuits and Systems I*, July 2000, pp. 965-980.
[42] J. Philips and K. Kundert, "Noise in Mixers, Oscillators, Samplers, and Logic: An Introduction to Cyclostationary Noise," *IEEE Custom Integrated Circuits Conference*, May 2000, pp. 431-438.
[43] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York, 1998, chapter 7.
[44] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, New York, 1998, chapter 4.

Low-Noise Amplifiers

[45] E. Taniguchi, et al., "Dual Bias Feed SiGe HBT Low Noise Linear Amplifier," *IEEE International Microwave Symposium (MTT-S)*, 2001, pp. 285-288.
[46] D. Yongwang and R. Harjani, "A +18 dBm IIP3 LNA in 0.35μm CMOS," *IEEE International Solid-State Circuits Conference*, 2001, pp. 162-163.
[47] G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, John Wiley & Sons, New York, 1990, chapter 4.
[48] A. van der Ziel, Noise: *Sources, Characterization, and Measurement*, Prentice Hall, Englewood Cliffs, NJ, 1970, chapter 7.

[49] S. P. Voinigescu, et al., "A Scalable High-Frequency Noise Model for Bipolar Transistors with Application to Optimal Transistor Sizing for Low-Noise Amplifier Design," *IEEE Journal of Solid-State Circuits*, September 1997, pp. 1430-1439.

[50] M. Schröter and H.-M. Rein, "Transit Time of High-Speed Bipolar Transistors in Dependence on Operating Point, Technological Parameters, and Temperature," *IEEE Bipolar Circuits and Technology Meeting*, 1989, pp. 250-253.

[51] P. R. Gray, et al., *Analysis and Design of Analog Integrated Circuits*, 4th edition, John Wiley & Sons, New York, 2001, chapter 11.

[52] E. G. Nielsen, "Behavior of Noise Figure in Junction Transistors," *Proceedings of the IRE*, July 1957, pp. 957-963.

[53] D. K. Lynn, C. S. Meyer, and D. J. Hamilton, editors, *Analysis and Design of Integrated Circuits*, McGraw-Hill, New York, 1967, chapter 12.

[54] Carl Battjes, Tektronix Inc., Beaverton, OR, *Monolithic Wideband Amplifier*, U.S. Patent 4,236,119, November 25, 1980.

[55] D. L. Feucht, *Handbook of Analog Circuit Design*, Academic Press, San Diego, CA, 1990, chapter 10.

[56] C. T. Armijo and R. G. Meyer, "A New Wide-Band Darlington Amplifier," *IEEE Journal of Solid-State Circuits*, August 1989, pp. 1105-1109.

[57] S. Krishnan, et al., "Broadband Lumped HBT Amplifiers," *Electronics Letters*, March 2000, pp. 466-467.

[58] J. S. Yuan, *SiGe, GaAs, and InP Heterojunction Bipolar Transistors*, Wiley & Sons, New York, 1999, chapter 2.

[59] A. Schmidt and S. Catala, "A Universal Dual Band LNA Implementation in SiGe Technology for Wireless Applications," *IEEE Journal of Solid-State Circuits*, July 2001, pp. 1127-1131.

[60] R. S. Carson, *High-Frequency Amplifiers*, Second Edition, Wiley & Sons, New York, 1982, chapter 7.

[61] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, Second Edition, McGraw-Hill, Boston, MA, 1999, chapter 8.

[62] B. Sklar, *Digital Communications*, Second Edition, Prentice Hall, Upper Saddle River, NJ, 2001, chapter 4.

[63] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, New York, 1998, chapter 4.

[64] J. Colvin, S. Bhatia, and K. O, "A Bond-Pad Structure for Reducing Effects of Substrate Resistance on LNA Performance in a Silicon Bipolar Technology," *Bipolar Circuits and Technology Meeting*, 1998, pp. 109-112.

[65] C. Biber, et al., "Technology Independent Degradation of Minimum Noise Figure Due to Pad Parasitics," *IEEE Microwave Theory and Techniques Symposium*, 1998, pp. 145-148.

[66] R. Fried, Y. Blecher, and S. Friedman, "CMOS ESD Protection Structures—Characteristics and Performance Comparison," *Physical and Failure Analysis of Integrated Circuits Symposium*, 1995, pp. 106-109.

[67] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York, 1998, chapter 5.

[68] B. Razavi, R.-H. Yan, and K. F. Lee, "Impact of Distributed Gate Resistance on the Performance of MOS Devices," *IEEE Transactions on Circuits and Systems I*, November 1994, pp. 750-754.

[69] A. van der Ziel, "Thermal Noise in Field-Effect Transistors, *Proceedings of the IRE*, August 1962, pp. 1808-1812.

[70] A. van der Ziel, "Gate Noise in Field Effect Transistors at Moderately High Frequencies," *Proceedings of the IEEE*, March 1963, pp. 461-467.

[71] D. K. Shaeffer and T. H. Lee, "A 1.5-V, 1.5-GHz CMOS Low Noise Amplifier," *IEEE Journal of Solid-State Circuits*, May 1997, pp. 745-759.

[72] N. Sano, et al., "Physical Mechanism of Current Fluctutation under Ultra-Small Device Structures," *IEEE International Workshop on Computational Electronics*, 1998, pp. 112-115.

[73] Y. C. Yeo, et al., "Direct Tunneling Gate Leakage Current in Transistors with Ultrathin Silicon Nitride Gate Dielectric," *IEEE Electron Device Letters*, November 2000, pp. 540-542.

[74] T. C. Sepke, *Investigation of Noise Sources in Scaled CMOS Field-Effect Transistors*, S.M. Thesis, Massachusetts Institute of Technology, to be published in June 2002.

[75] A. van der Ziel, Noise: *Sources, Characterization, and Measurement*, Prentice Hall, Englewood Cliffs, NJ, 1970, chapter 6.

[76] D. K. Shaeffer and T. H. Lee, *Low-Power CMOS Radio Receivers*, Kluwer Academic Publishers, Norwell, MA, 1999, chapter 4.

[77] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, New York, 1998, chapter 11.

[78] C. Enz, "An MOS Transistor Model for RF IC Design Valid in All Regions of Operation," *IEEE Transactions on Microwave Theory and Techniques*, January 2002, pp. 342-359.

[79] S.-D. Kim, C.-M. Park, and J. C. S. Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime," *IEEE Transactions on Electron Devices*, March 2002, pp. 457-472.

Oscillator Design

[80] D. A. Hitko, *A Low Power, Low Noise, 1.8GHz Voltage-Controlled Oscillator*, S.M. Thesis, Massachusetts Institute of Technology, February 1997, chapter 2.

[81] D. A. Hitko, *A Low Power, Low Noise, 1.8GHz Voltage-Controlled Oscillator*, S.M. Thesis, Massachusetts Institute of Technology, February 1997, chapter 4.

[82] D. Ham and A. Hajimiri, "Concepts and Methods in Optimization of Integrated LC VCOs," *IEEE Journal of Solid-State Circuits*, June 2001, pp. 896-909.

[83] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, New York, 1998, chapter 17.

[84] E. Hegazi, et al., "A Filtering Technique to Lower Oscillator Phase Noise," *IEEE International Solid-State Circuits Conference*, 2001, pp. 364-365.

[85] K. Gulati and H-S. Lee, "A High-Swing CMOS Telescopic Operational Amplifier," *IEEE Journal of Solid-State Circuits*, December 1998, pp. 2010-2019.

[86] E. A. M. Klumperink, et al., "Reducing MOSFET 1/f Noise and Power Consumption by Switched Biasing," *IEEE Journal of Solid-State Circuits*, July 2000, pp. 994-1001.