UNIVERSITY OF CALIFORNIA SAN DIEGO

Electronic Circuits for Communications and Sensing Applications

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Electronic Circuits and Systems)

by

Cooper Sinclair Levy

Committee in charge:

Professor Patrick P. Mercier, Chair
Professor James F. Buckwalter, Co-Chair
Professor Peter M. Asbeck
Professor Shadi A. Dayeh
Professor Roland R. Lee

2019

The dissertation of Cooper Sinclair Levy is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California San Diego

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

This dissertation covers two somewhat related topics, and a third very disparate topic. I chose to include all three topics in the dissertation because a significant amount of my graduate studies was spent on each. The dramatic differences would make an introductory chapter quite disjointed, so I have not included one. In lieu of an introduction, I hope that all of the chapters are somewhat self-contained and give enough references to help the reader fill in the gaps as needed. I have also included this preface to give specific callouts to material I have found particularly useful over the years.

For the work on distributed amplifiers, the primary reference I have used over the years is the wonderful *Microwave Engineering* by David Pozar. I haven't spent much time on distributed amplifiers over the last few years, but I can still almost drop the book on a table and have it open to the microwave filters chapter. The distributed amplifier papers by Amin Arbabian were also useful references over the years.

The work on Doherty power amplifiers occupied a significant amount of my efforts during graduate school, but there are a few standout works that have provided insight over the years. In particular *RF Power Amplifiers for Wireless Communications* by Steve Cripps gives an extremely insightful coverage of power amplifiers. It is accessible to someone just staring out, and yet is still a valuable reference for experts in the field. Cripps' other book *Advanced Techniques in RF Power Amplifier Design* is also very useful, in particular for someone interested in Doherty, outphasing, predistortion, and other techniques that may become relevant to 5G power amplifiers. If you can get you hands on a copy of Peter Asbeck's lecture notes for the power amplifier course he has taught over the years at UCSD, they are wealth of knowledge. Every time I open them up, I find something I had forgot about or an interesting detail I had always missed before. Both of these references give a clearer explanation of power amplifiers that anything else I have come across. The material on envelope tracking in these references would also be relevant for the distributed amplifier work of the first chapter. I also recommend perusing the original paper by Doherty. It's not the most accessible to the modern designer given that it was written with tube amplifiers in mind, but it is

the only reference I am aware of that gives an explanation similar to my own understanding of the Doherty power amplifier. In particular, it shows the various versions of the Doherty that are covered in this dissertation.

For the work on atomic magnetometry, I was much less familiar with the background physics of this topic than for the circuits of the previous works. For someone with a strong background in quantum mechanics, the material will likely not be too difficult. For those like myself, the dissertation by Scott Seltzer was an immensely useful resource. Scott's explanations along with his clear derivations and great figures make his dissertation a great place to start for someone with no background in atomic magnetometry. Once you have some background in the subject, the book *Optical Magnetometry* by Budker and Kimball is a useful jumping-off point for a wide range of magnetometry topics.

ACKNOWLEDGEMENTS

I would first like thank the numerous advisers that have supported me throughout my time at UCSD. During the first few years of my PhD, I was fortunate to have the opportunity to work with Jim Buckwalter. His laid-back approach to mentorship allowed me to explore many ideas that would probably have never come up in a more structured environment. While most of the material we discussed in our weekly brainstorming sessions never directly went anywhere, those sessions spawned ideas that led to almost three PhDs worth of work, with ideas that still have not been explored. While Jim has not received as much credit as I believe he should for his insightful direction, I will always be grateful for his guidance.

For the last few years, I have been fortunate enough to work with Patrick Mercier on a second ”chapter” of my PhD. Our work together started out as a seemingly straightforward problem to address, but ran into a number of obstacles, both technical and logistical. Throughout, Patrick has been a calm and understanding adviser. I have appreciated his patience and perseverance.

While never officially my adviser, Peter Asbeck has been a indispensable mentor throughout my time at UCSD. I have had the opportunity to work closely with him and his students on a number or projects. Professor Asbeck is always willing to discuss just about any topic, be it circuits, device physics, or personal. His kindness and enthusiasm have made him a wonderful role model.

For the better part of my time at UCSD I have had a variety of collaborations with Shadi Dayeh. While some have been more productive than others, I have enjoyed getting to expand my knowledge of device physics while also working to show applications of devices. The opportunity to work with Shadi and many of his students over the years has been a lot of fun.

In the middle of my PhD, I was a teaching assistant for four quarters, one with Professor Bang-Sup Song, and three with Professor Ian Galton. Before meeting Professor Song I thought I understood analog circuits, but he taught me I didn’t. I imagine he would say, most of what I know about analog circuits is still probably wrong, but maybe less wrong than what others know. I will always be grateful for his insight. Professor Galton also taught me a great deal about analog circuits, but from a completely different viewpoint. I appreciate the rigor he brings to circuits. I have also

appreciated his honest and useful advice over the years.

I would also like to thank Professor Roland Lee for agreeing to serve on my defense committee. While we were unfortunately never able to collaborate on MCG measurements with the magnetometer work, I am appreciative of his time and feedback.

Having taken seven years to complete my PhD, I have made many friends at UCSD.

In the High-Speed Integrated Circuits group, I was very fortunate for the mentorship and/or friendship of Jun Li, Tissana Kijsanayotin, Po-Yi Wu, Arpit Gupta, Wei Wang, Kristian Madsen, Cheng-Kai Luo, Mohammad Mehrjoo, and Najme Ebrahimi. In the last few years, as the group dwindled in size, I was extremely grateful for the friendship of Bagher Rabet, Po-Han Wang, and Kelvin Fang. What little bit of sanity I had left was in part due to our conversations, ranging from deeply technical and squabbles over useless minutiae, to lab gossip and what little bit of pop culture and current events I follow.

While I spent less time with the Energy-Efficient Microsystems group, my adopted group, I am grateful for the friendship and welcoming of all the group members, Hui Wang, Julian Warchall, John Louie, Somayeh Imani, Sally Safwat-Amin, Jiwoong Park, Bao Lam, Xiaoyang Wang, and Jason Huang. In particular I have been thankful for the friendship of Dhongue Lee, one of the few people I have known my entire time at UCSD. We have both had plenty of ups and downs on the journey, but have finally made it to the end.

There have been a number of other students from the circuits and devices community at UCSD that I have been fortunate enough to work with and call friends. In no particular order, Chung-Lun Hsu, Haowei Jiang, Li Gao, Ben Dhaljelm, Narek Rostomyan, Varish Diddi, Woojin Choi, Hasan Al-Rubaye, Samet Zihir, Hsin-Chang Lin, Chih-Hsiang Ko, Yu-Hsin Liu, and Kerim Kibaroglu.

I have also had a number of friends outside of circuits in the ECE community, Paolo Gabriel, Peter Weigel, Hannah Grant, Nikki Hsueh, Michael Ostertag, and Andrew Saad. I have most certainly omitted many people, and I apologize if your name isn't here, but know this does not mean I have not appreciated your camaraderie, just that there are so many people I am thankful to have

known at UCSD that it's hard to not miss someone when making a list from memory.

My roommates during both graduate school and undergrad have been a constant source of encouragement and collaboration. Alex Sun is by far the most relaxed person I know. While our technical background is what brought us together, I'm thankful for the balance Alex brought to my life during grad school. Voravit Vorapipat has been a great friend and collaborator. He gets where I come from on a technical level. It's still hard for me to believe that one fateful night of discussion with Vit in the kitchen led to at least five papers and defined a good amount of grad school for both of us. Jefy Jayamon has also provided a great source of entertainment and curiosity to my life. I've been grateful for our scattered yet insightful technical discussions, often going late into the night. To my undergrad roommates Joey Goodkight and Joey Hsiao, it has been great being able to commiserate over the ups and downs of both grad school and life.

Finally, the support of my family has meant everything to me. In many ways, I think they were more anxious to see that I finish than I was. My sisters Meghan, Brooke, and Ashley have provided a familiar comfort of childhood and escape from lab, Meghan and Ashley during the time that we overlapped living in San Diego, and Brook during visits. Most of all, I am thankful for the continual support of my parents Steve and Sheryl. Most of my hobbies, cooking, quilting, crocheting, woodworking, and home repairs to prevent the landlord showing up, to name a few, are due to their creative, eclectic, hard-working personalities. My dad Steve has supported me in so many ways. His ability to understand where my mind is most of the time, and willingness to listen to likely multiple thousands of hours of circuits rambling, and follow most of it, is one of the main reasons I made it through grad school. One of the other reasons I've made it is the completely opposite personality of my mom Sheryl. Knowing that there's no need to connect with her on a technical level has often been a source of relief. I'm not sure what I would have done without her constant encouragement and love.

Chapter 1, in full, is a reprint of the material as it appears in IEEE Journal of Solid-State Circuits, 2016, Fang, Kelvin; Levy, Cooper; Buckwalter, James. The dissertation author was one of the primary authors of this work.

Chapter 2, in part, is a reprint of the material as it appears in the Proceedings of the Compound Semiconductor Integrated Circuit Symposium, 2016, Levy, Cooper; Vorapipat, Voravit; Buckwalter, James. The expanded presentation in this dissertation may appear as another publication in the future with the additional contributor Mercier, Patrick. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, has been accepted for publication, as it will appear in IEEE Transactions on Instrumentation and Measurement, Levy, Cooper; Kornack, Thomas; Mercier, Patrick. The dissertation author was the primary investigator and author of this material.

VITA

2010	Bachelor of Sciences, University of California Berkeley
2010	Bachelor of Arts, University of California Berkeley
2013	Master of Sciences, University of California San Diego
2019	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

C. S. Levy, T. W. Kornack, P. P. Mercier, "Bell-Bloom Magnetometer Linearization by Intensity Modulation Cancellation," Transactions on Instrumentation and Measurement. (Accepted)

C. Hill, C. S. Levy, H. AlShammary, A. Hamza and J. F. Buckwalter, "RF Watt-Level Low-Insertion-Loss High-Bandwidth SOI CMOS Switches," IEEE Transactions on Microwave Theory and Techniques, vol. 66, no. 12, pp. 5724-5736, Dec. 2018.

C. Hill, C. S. Levy, H. AlShammary, A. Hamza, J. F. Buckwalter, "A 30-dBm, 300-MHz 45-nm SOI CMOS Power Modulator for Spread-Spectrum Signal Processing at the Antenna," IMS 2018, Philadelphia, PA.

S. Vishniakou, R. Chen, Y. G. Ro, C. J. Brennan, C. Levy, E. T. Yu, S. A. Dayeh, "Improved Performance of Zinc Oxide Thin Film Transistor Pressure Sensors and a Demonstration of a Commercial Chip Compatibility with the New Force Sensing Technology," Adv. Mater. Technol. 2018, 1700279.

V. Vorapipat, C. S. Levy, P. M. Asbeck, "A Class-G Voltage-Mode Doherty Power Amplifier," IEEE Journal of Solid-State Circuits, vol. 52, no. 12, pp.3348-3360, Dec. 2017.

V. Vorapipat, C. Levy, P. Asbeck, "2.8 A Class-G Voltage-Mode Doherty Power Amplifier," ISSCC 2017.

V. Vorapipat, C. S. Levy, P. M. Asbeck, "Voltage Mode Doherty Power Amplifier," IEEE Journal of Solid-State Circuits, vol. 52, no. 5, pp.1295-1304, May 2017.

C. S. Levy, V. Vorapipat, J. F. Buckwalter, "A 14-GHz, 22-dBm Series Doherty Power Amplifier in 45-nm CMOS SOI," 2016 CSICS, Oct 2016.

K. Fang*, C. S. Levy* and J. F. Buckwalter, "Supply-Scaling for Efficiency Enhancement in Distributed Power Amplifiers," IEEE Journal of Solid-State Circuits, vol. 51, no. 9, pp. 1994-2005, Sept. 2016. (*Authors contributed equally to this work)

V. Vorapipat, C. Levy and P. Asbeck, "A Wideband Voltage Mode Doherty Power Amplifier," 2016 RFIC, San Francisco, CA, 2016, pp. 266-269. (Best student paper award)

K. Fang, C. Levy and J. F. Buckwalter, “A 105-GHz, supply-scaled distributed amplifier in 90-nm SiGe BiCMOS,” BCTM, 2015 IEEE, Boston, MA, 2015, pp. 182-185.

C. S. Levy, J. F. Buckwalter, “A 6-41 GHZ Distributed Amplifier with Supply Scaling for Efficiency Enhancement,” Compound Semiconductor Integrated Circuit Symposium (CSICS), 2015 IEEE, 11-14 Oct. 2015.

C. S. Levy, P. M. Asbeck, J. F. Buckwalter, “A CMOS SOI Stacked Shunt Switch with Sub-500ps Time Constant and 19-Vpp Breakdown,” Compound Semiconductor Integrated Circuit Symposium (CSICS), 2013 IEEE, 13-16 Oct. 2013.

C. Levy, P. Asbeck, J. Buckwalter, “A Stacked-FET Harmonic-Reject Mixer in SOI CMOS,” Power Amplifier Symposium, 2013 IEEE, 9-10 Sept. 2013.

E. Y. Chin, C. S. Levy, A. R. Neureuther, “Variability Aware Timing Models at the Standard Cell Level,” SPIE Design for Manufacturability through Design-Process Integration, Proc. SPIE, Vol. 7641, 76410H, 2010.

ABSTRACT OF THE DISSERTATION

Electronic Circuits for Communications and Sensing Applications

by

Cooper Sinclair Levy

Doctor of Philosophy in Electrical Engineering (Electronic Circuits and Systems)

University of California San Diego, 2019

Professor Patrick P. Mercier, Chair
Professor James F. Buckwalter, Co-Chair

This dissertation addresses two distinct topics, namely circuits for radio-frequency and millimeter-wave transmitters with emphasis on power amplifiers, and control circuits and system design for linearizing atomic magnetometers.

Power amplification for wireless transmitters, despite receiving myriad attention over the last few decades, is still one of the main bottlenecks in terms of complete transmitter integration and reducing system power dissipation. First, a distributed amplifier architecture aiming to improve peak efficiency by voltage supply scaling will be presented. By using multiple supplies, wasted headroom is eliminated in the early stages of the distributed amplifier where the output voltage swing is relatively low. Second, a class of Doherty power amplifiers that was rediscovered by the

author by reverse engineering the canonical Doherty power amplifier, and a modern implementation, will be presented. The implementation stacks the voltage swings of the main and peaking amplifiers of the Doherty power amplifier, allowing increased output power in scaled CMOS without concern of breakdown. Finally, atomic magnetometers have shown promise as replacements in many applications where SQUIDs are currently used, with the benefits of no supercooling required, and the ability to operate in Earth's geomagnetic field. At the same time, operation outside of a magnetically-shielded environment has numerous side-effects. The last section will present a technique for linearizing the Bell-Bloom atomic magnetometer, improving its performance in interference-rich environments. The technique notes that the detected output signal of the magnetometer contains not only information about the spin precession of the optically pumped atoms, but also a large component due to the pumping laser modulation. By subtracting this known pumping modulation signal from the detected output, the system linearity can be significantly improved.

Chapter 1

Supply-Scaling for Efficiency Enhancement in Distributed Power Amplifiers

1.1 Introduction

Efficient utilization of millimeter-wave (mm-wave) bands will emphasize amplification across several frequency bands in a single amplifier. Ultra-wideband amplifiers have applications in high-speed data links, broadband transceivers, high-frequency instrumentation, and high-resolution imaging. For instance, fine spatial resolution in imaging systems requires narrow pulses to produce a wide range of frequency content [1]. Amplification of a 10 ps Gaussian pulse for sub-millimeter accuracy requires bandwidth on the order of 100 GHz. Conventional tuned amplifiers have difficulties satisfying such large bandwidth requirements due to their inherent gain-bandwidth tradeoff. On the other hand, distributed amplifiers (DAs) provide an effective solution with their large bandwidth, and low gain variation and sensitivity to mismatch [2].

While silicon technology scaling has improved transistor cutoff frequency f_T to the hundreds of GHz, transistor scaling does not tend to improve the efficiency or the output power of distributed amplifiers. If silicon-based processes can supplant III-V technologies in mm-wave systems, higher reliability and yield of silicon must be leveraged against the lower intrinsic gain and breakdown

of CMOS/BiCMOS processes. A number of DAs with bandwidths in excess of 80 GHz have been demonstrated in silicon [3–8]. However, conventional DAs suffer from poor power efficiency, making these designs unattractive for broadband power amplification. To address the efficiency issues, previous attempts at DA scaling have been realized by impedance tapering of the loaded collector-line and scaling of the gain stage device sizes [9, 10]. Unfortunately, this incurs greater resistive line losses and high-frequency reflections due to impedance mismatch, degrading the gain as well as limiting the number of stages that can be implemented. Therefore, the design of distributed amplifiers for high output power and efficiency over wide bandwidth remains an open challenge.

This paper presents a supply-scaled distributed amplifier that offers improved collector efficiency (CE) and power-added efficiency (PAE). The analysis investigates load modulation at each stage within the distributed amplifier. An analysis of supply scaling indicates how this technique performs load pulling analogous to impedance tapering but does not incur the same passive losses or frequency dependency. By feeding separate dc supply voltages through high-pass constant-k filter sections, improved power efficiency is achieved while maintaining a constant 50-Ω line impedance within the amplifier bandwidth. An 8-stage supply-scaled DA is demonstrated in a 90-nm SiGe BiCMOS process. This work expands upon the design presented in [11] to detail the analysis of interstage load modulation due to traveling waves and the design methodology of the band-pass DA. New measurements of the DA with uniform biasing are also presented to verify the supply-scaling theory. Section II presents an overview of the limitations of conventional DA designs and tapered lines. Section III introduces the concept of supply-scaling and discusses its advantages over impedance tapering techniques. The design and analysis of a band-pass DA to enable independent supply biasing is presented in Section IV. Measurements of the fabricated DA and comparison with previous works are included in Section V.

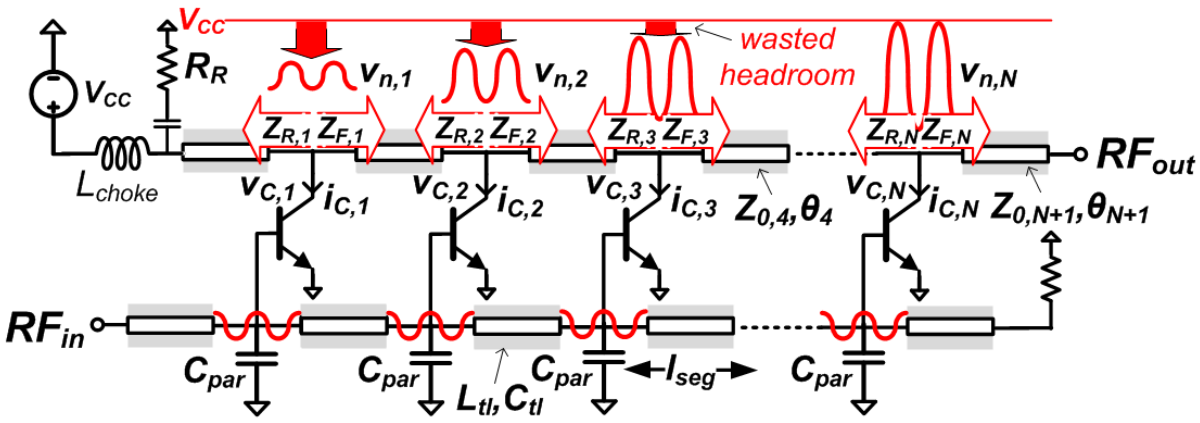


Figure 1.1: Schematic of a conventional low-pass DA.

1.2 Distributed Amplifier Efficiency Limitations

Fig. 1.1 shows the schematic of a conventional uniform DA. Distributed amplifiers constructively add the output current from each gain stage in the collector transmission line as the RF input signal travels along the base line. Neglecting losses, the DA exhibits gain that linearly increases with the number of stages while maintaining bandwidth in contrast to cascading amplifier stages.

In a DA, transistor parasitic capacitances are absorbed into the input and output lines to create lumped-element T-section constant-k filters. The cascade of T-sections forms an artificial transmission line whose cutoff frequency determines the bandwidth of the DA [12]. For transmission line segments of length $l_{seg}/2$, with inductance-per-length L_{tl} and capacitance-per-length C_{tl}, to each side of the gain stage, loaded by parasitic capacitance C_{par}, the T-section characteristic impedance $Z_{0,l}$ and low-pass cutoff frequency $f_{c,l}$ are given by

$$Z_{0,l} = \sqrt{\frac{L_{tl} \times l_{seg}}{C_{tl} \times l_{seg} + C_{par}}} \tag{1.1}$$

$$f_{c,l} = \frac{1}{\pi\sqrt{(L_{tl} \times l_{seg})\,(C_{tl} \times l_{seg} + C_{par})}}. \tag{1.2}$$

While DAs achieve large gain-bandwidth (GBW) product, conventional topologies exhibit poor power efficiency due to a number of factors. Since the collector of each transistor sees the

same impedance in both directions, half of the collector current from each stage travels towards the reverse termination (these reverse currents generally do not cancel, and the power is lost). Secondly, the wideband nature of the amplifier prohibits harmonic tuning of transistor outputs, preventing waveform engineering for higher-efficiency classes of amplifier operation. Finally, the voltage swing at each stage is not uniform, with later stages having a larger swing due to voltage summing along the collector transmission line. Since the dc collector bias is shared amongst all stages, this results in a large amount of wasted headroom. The inefficiency is evident from Fig. 1.1, where the impedance seen at the collector of the n^{th} stage in an N-stage DA with amplitude of v_n and current swing of $i_{C,n}$ is

$$Z_{C,n}(\omega) = \frac{v_n(\omega)}{i_{C,n}\prod_{m=1}^{n} e^{-j\theta_{B,m}(\omega)}} = \frac{v_{F,n}(\omega) + v_{R,n}(\omega) + v_{C,n}(\omega)}{i_{C,n}\prod_{m=1}^{n} e^{-j\theta_{B,m}(\omega)}}, \tag{1.3}$$

where

$$v_{F,n}(\omega) = \sum_{k=1}^{n-1}\left[v_{C,k}(\omega)\sqrt{\frac{Z_{F,n}(\omega)}{Z_{F,k}(\omega)}}\prod_{m=k+1}^{n} e^{-j\theta_{C,m}(\omega)}\right] \tag{1.4a}$$

$$v_{R,n}(\omega) = \sum_{k=n+1}^{N}\left[v_{C,k}(\omega)\sqrt{\frac{Z_{R,n}(\omega)}{Z_{R,k}(\omega)}}\prod_{m=n}^{k} e^{-j\theta_{C,m}(\omega)}\right] \tag{1.4b}$$

$$v_{C,n}(\omega) = i_{C,n}\left(Z_{F,n}(\omega) \parallel Z_{R,n}(\omega)\right)\prod_{m=1}^{n} e^{-j\theta_{B,m}(\omega)}\,. \tag{1.4c}$$

The forward traveling voltage wave $v_{F,n}$ is due to the current of the preceding stages (zero for the first stage), while $v_{R,n}$ is the reverse wave from subsequent stages (zero for the last stage), and $v_{C,n}$ is the voltage induced by the transistor. The latter is determined from the small-signal impedance seen by collector n looking toward the reverse termination (R_R),

$$Z_{R,n}(\omega) = Z_{0,n}\frac{Z_{R,n-1}(\omega) + jZ_{0,n}\tan\theta_{C,n}(\omega)}{Z_{0,n} + jZ_{R,n-1}(\omega)\tan\theta_{C,n}(\omega)}\,, \tag{1.5}$$

and toward the load (R_L),

$$Z_{F,n}(\omega) = Z_{0,n+1}\frac{Z_{F,n+1}(\omega) + jZ_{0,n+1}\tan\theta_{C,n+1}(\omega)}{Z_{0,n+1} + jZ_{F,n+1}(\omega)\tan\theta_{C,n+1}(\omega)}, \tag{1.6}$$

where $Z_{0,n}$, θ_n describe the characteristic impedance and electrical length of the transmission line section to the left of stage n, and $Z_{R,0}$, $Z_{F,N+1}$ equal R_R and R_L, respectively. In the case of a single stage amplifier, the impedance seen at the collector is not affected by traveling waves and $Z_{C,n} = v_{C,n}/i_{C,n}$, which would be recognized as the impedance to optimally match the transistor for the transfer of power into a load. For the general case of an N-stage DA, however, this is not true.

1.2.1 Uniform Distributed Amplifier

For a uniform DA, $Z_{F,n} = Z_{R,n} = Z_0$ and $\theta_{B,n} = \theta_{C,n} = \theta$. Therefore, the impedance seen at the collector simplifies to

$$Z_{C,n}(\omega) = \frac{Z_0}{2}\left(\sum_{k=1}^{n-1}\frac{i_{C,k}}{i_{C,n}} + 1 + \sum_{k=n+1}^{N}\frac{i_{C,k}}{i_{C,n}}e^{-2j(k-n)\theta(\omega)}\right). \tag{1.7}$$

When each device contributes the same current, *i.e.* $i_{C,k} = i_C$, this impedance is further simplified to

$$Z_{C,n}(\omega) = \frac{Z_0}{2}\left(n + e^{-j(N+1-n)\theta(\omega)}\frac{\sin(N-n)\theta(\omega)}{\sin\theta(\omega)}\right). \tag{1.8}$$

In this case, the impedance seen by the collector has a linearly increasing real component, as well as a complex component that leads to frequency dependent amplitude and phase variation. Fig. 1.2 shows the variation in $Z_{C,n}$ with respect to electrical length θ, and Fig. 1.3 shows collector voltages v_n. It can be seen that all but the final gain stage in a uniform DA have output voltages and impedances that change periodically with frequency. The preceding theory has been corrobrated by electro-optic measurements on signal propagation internal to distributed amplifiers [13]. More generally, when each stage of the DA is not uniform, the n^{th} transistor not only sees frequency dependent load modulation due to $v_{F,n}$ and $v_{R,n}$ from (1.3), but also frequency-varying $Z_{F,n}$ and $Z_{R,n}$,

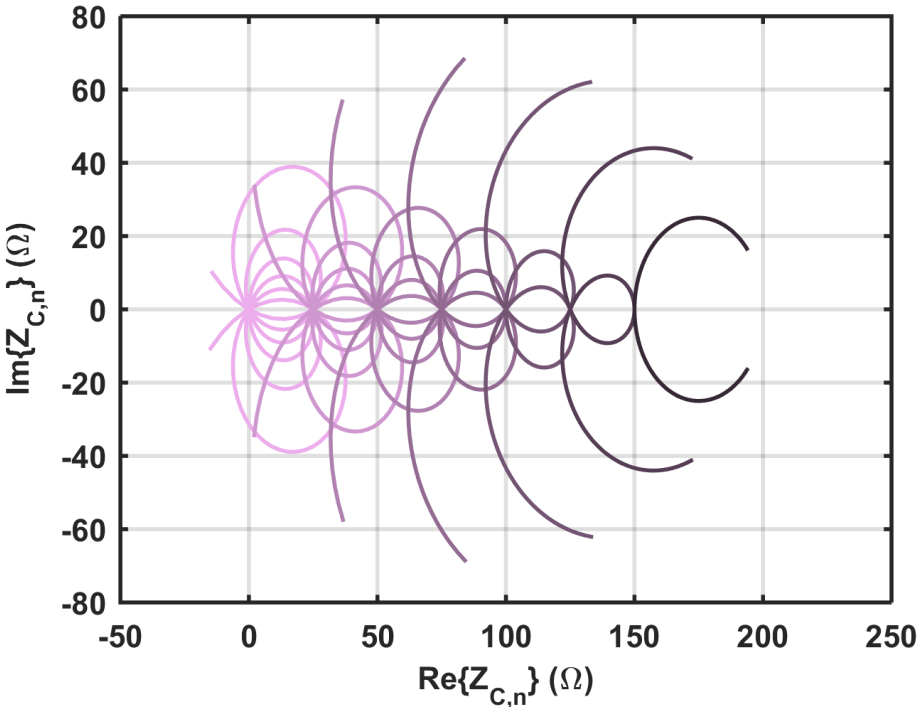


Figure 1.2: Frequency variation in real and imaginary collector impedances in a uniform DA for $\theta = 0.36 - 2.79$.

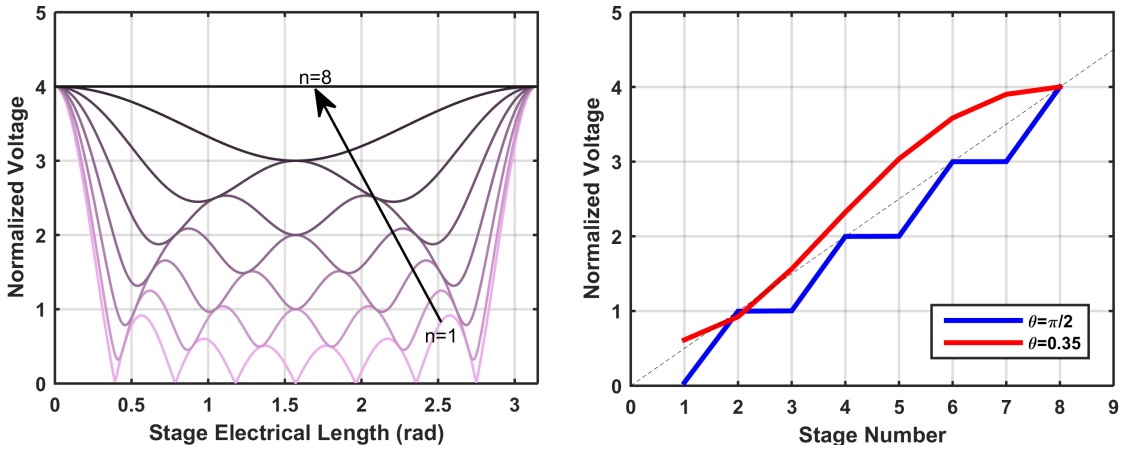


Figure 1.3: Successive collector voltage magnitudes versus transmission line electrical length in a uniform DA.

which impact $v_{C,n}$ as described in (1.4c). The ability to control the impedance at each collector forms the basis of loadline modulation.

1.2.2 Optimal Loadline Matching

In a conventional uniform DA, the collector of each transistor is fixed to $V_{C,n} = V_{CC}$ and the loadline impedance is

$$Z_{OPT,n} = \frac{V_{CC} - V_K}{I_C} , \tag{1.9}$$

where V_K is the knee voltage of the technology and I_C is the dc bias current (constant across stages). For a uniform DA, this loadline impedance is always larger than or equal to the optimum impedance seen in (1.3), since the reverse traveling voltages do not always add constructively. For class-A operation, the maximum amplitudes for the voltage and current are $v_n = V_{CC} - V_K$ and $i_{C,n} = I_C$. We desire to set the collector impedance according to the loadline impedance:

$$Z_{OPT,n} = \frac{max(v_n)}{max(i_{C,n})} = Z_{C,n} . \tag{1.10}$$

However, it is obvious that the voltage swing at each transistor is different even as the current through each stage is fixed. This leads to a non-optimal loadline matching for uniform DAs.

One solution to the DA efficiency problem is impedance tapering along the output transmission line as proposed by [2]. Using this approach, collector line impedances and lengths are set such that $v_{R,n}$ (or $i_{R,n}$)$= 0$ and all the generated power travels to the RF output, circumventing the need for a reverse termination. The tapered line load pulls each transistor n to see $Z_{opt,n} = (V_{CC} - V_K)/I_{C,n}$ with a constant voltage swing. A number of impedance-tapered DAs have been demonstrated in silicon and III-V processes with efficiency gains [14, 15]. However, due to the frequency dependency of the load pulling mechanism, these designs can only achieve narrowband efficiency enhancement through careful optimization of unequal-length sections. Additionally, the high load impedances at early stages require narrow-width transmission lines that are lossy and difficult to synthesize, even with III-V back-end-of-line (BEOL) processes [9].

Recently, efforts have been made in [10] to design an enhanced-efficiency DA with over 100 GHz bandwidth, utilizing simultaneous scaling of device size with output line impedance. The low degree of tapering, however, dictates the need for an explicit reverse termination ($Z_{R,1}(\omega) \neq \infty$ as in ideal impedance tapering) and sacrifices the perfect cancellation of reflected waves. The resulting mismatches along the output line, combined with losses from the high-impedance early stages, limit the output power and overall PAE, especially at higher frequencies. It is evident that the inability to synthesize a large range of transmission line impedances with low loss is a major detriment to attempts at efficiency improvement using these techniques.

1.3 Efficiency Enhancement through Supply Scaling

To avoid tapered transmission lines, we propose a supply-scaling technique for enhancing DA efficiency while maintaining a constant 50-Ω characteristic impedance along the synthesized collector line. From Fig. 1.3, it can be seen that the voltage at successive collectors increases along the output line monotonically, and more accurately, the average voltage increases linearly inside the amplifier pass band. This feature contrasts with a stage-scaled DA (shown in Fig. 1.4 with transmission line scaling of 1.1 and current scaling of 0.96), which exhibits larger variation in voltage and impedance with respect to frequency.

By independently setting the dc collector voltages $V_{C,n}$ to match the maximum v_n in each section of a standard DA, we eliminate the wasted headroom present at each but the last stage. This approach effectively moves the loadline of each transistor to an optimal point for dc power consumption without requiring any change in the passive component parameters from stage to stage:

$$Z_{OPT,n} = \frac{V_{C,n} - V_K}{I_{C,n}} = \frac{v_{F,n} + v_{R,n}}{i_{C,n}} + \frac{Z_0}{2} \,. \tag{1.11}$$

While supply-scaling performs load modulation analogous to a tapered line (Fig. 1.5), it offers a number of advantages for wideband operation. Not only are high-impedance transmission lines avoided, but the sensitivity of efficiency-enhanced operation to frequency and process variation is

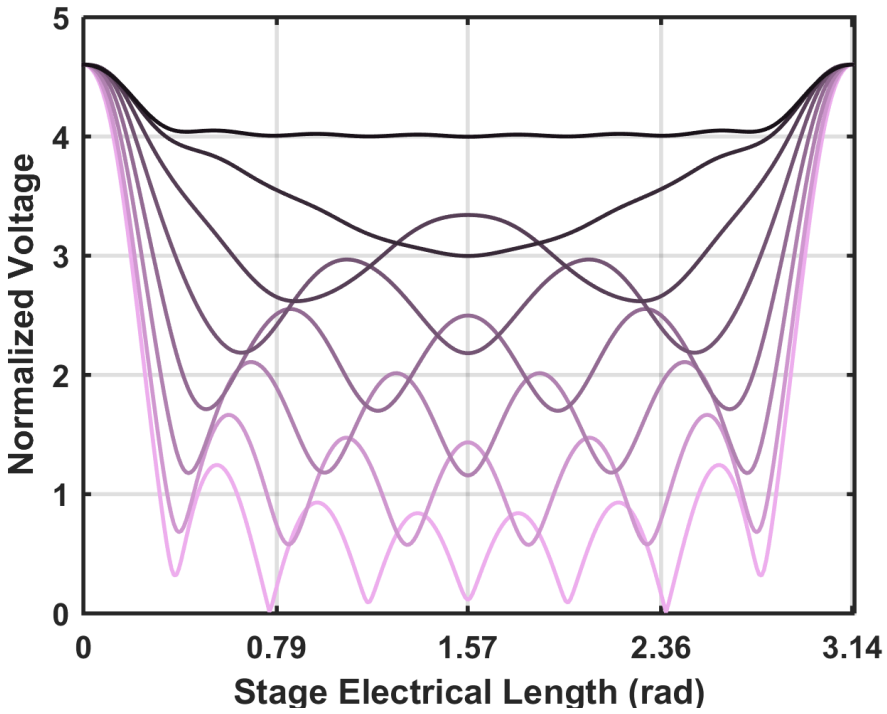


Figure 1.4: Simulated collector voltage magnitudes of a stage-scaled DA with 1.1 impedance scaling and 0.96 current scaling.

lower compared to that of impedance tapering as well. Looking at the N^{th} collector in Fig. 1.3, the peak output power for an ideal lossless non-tapered DA operating under class-A bias is constant across all frequencies and given by

$$P_{out} = \frac{N i_{C,n}}{2\sqrt{2}} \times \frac{v_{C,N}}{\sqrt{2}} = \frac{1}{8} N^2 I_C^2 Z_0 \,. \tag{1.12}$$

The dc power consumed per stage is

$$P_{DC,n} = V_{C,n} I_C \,. \tag{1.13}$$

In a uniform DA (i.e. $V_{C,n} = V_{C,N}$), the theoretical collector efficiency is therefore $CE = P_{out}/NP_{DC,n} =$ 25%, half that of a conventional class-A amplifier.

On the other hand, if the supply voltages are scaled such that $V_{C,n} = n\left(V_{C,N} - V_K\right)/N$, the

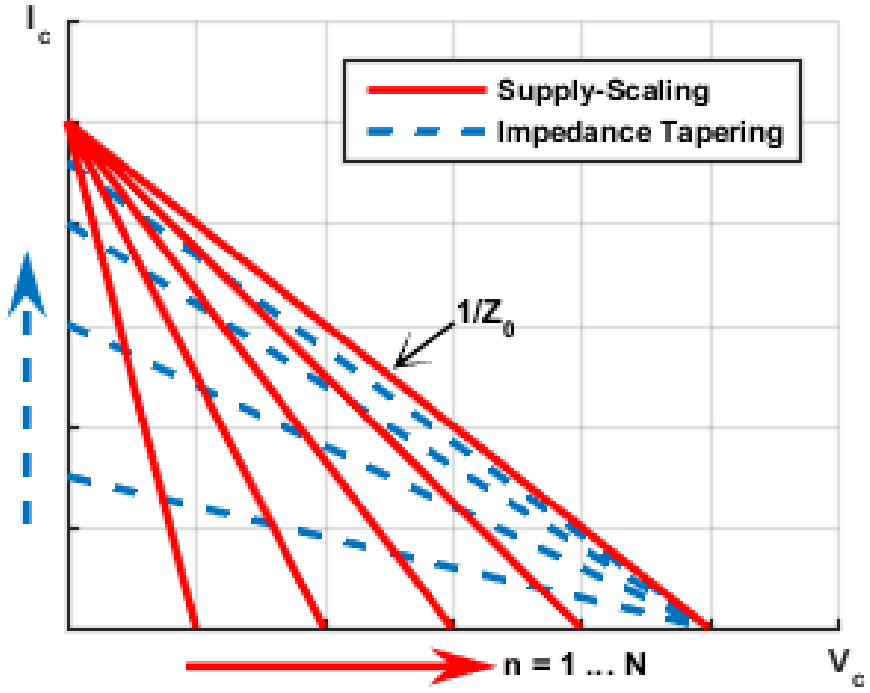


Figure 1.5: Loadlines of subsequent stages in a DA under supply-scaling and impedance tapering schemes. Whereas successive stages maintain constant voltage swing under impedance tapering, a supply-scaled architecture matches the optimum loadlines through independent voltage biasing.

maximum voltage swing at each stage within the pass band, the collector efficiency becomes

$$CE = \frac{\frac{1}{4}N^2 I_C V_{C,1}}{I_C V_{C,1} \sum_{n=1}^{N} n} = \frac{N}{2(N+1)} . \tag{1.14}$$

As shown in Fig. 1.6, the theoretical efficiency of a supply-scaled DA approaches 50% as N becomes large. In reality, a number of factors prevent maximum efficiency operation, including collector line losses and nonzero knee voltage. Providing individual dc supplies to each gain stage may prove to be impractical for real systems as well.

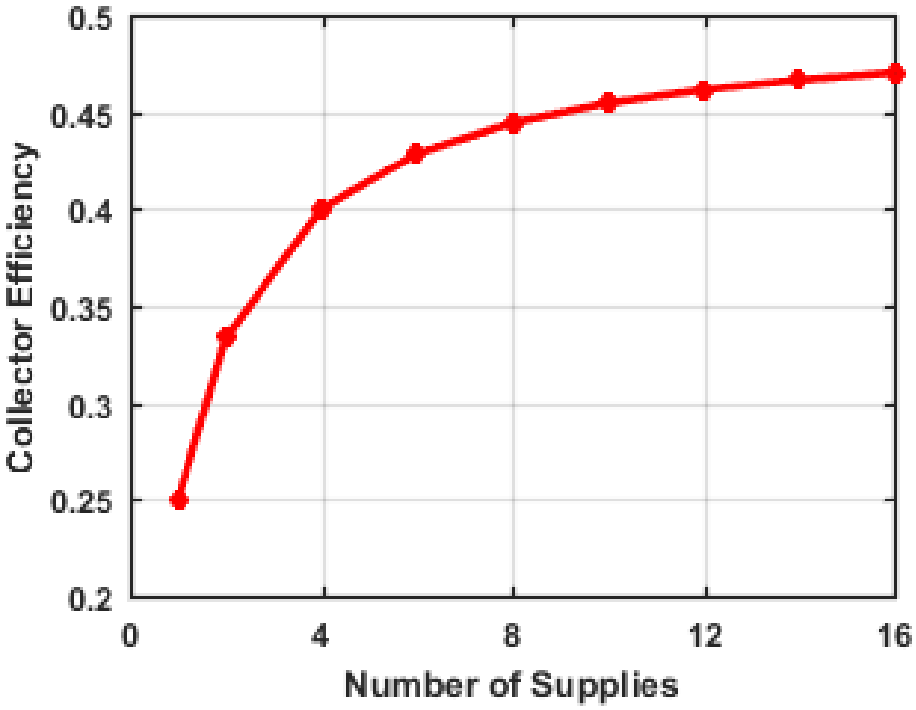


Figure 1.6: Peak collector efficiency of an ideal supply-scaled DA as a function of number of supplies.

1.4 Design Methodology of Band-pass DA

Conventional DA designs feature base and collector transmission lines with low-pass characteristics and a shared dc bias across all gain stages. To avoid I^2R loss, efficient DAs must supply the collector bias through an off-chip bias tee or choke, whose low-frequency cutoff prevents

true dc performance, rather than through the reverse termination. In some applications, such as odd-derivative Gaussian pulse generation and wideband RF communications, it is not necessary to provide amplification down to dc. Thus, the bias voltage levels can be isolated between DA sections. To realize independent biasing of the supply voltages along the DA and eliminate the need for a bulky bias-tee, a band-pass topology is chosen, which introduces dc-blocking capacitors and dc-feed inductors in between transmission line segments as parts of a high-pass T-section filter.

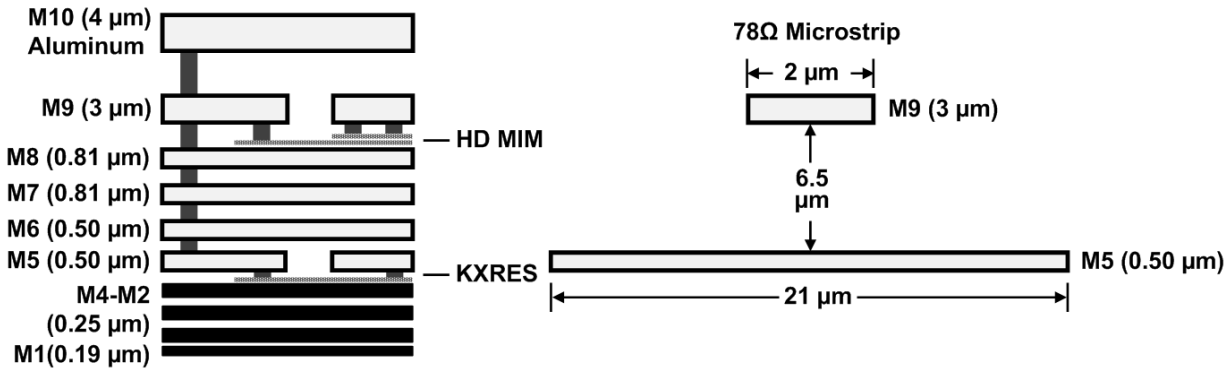


Figure 1.7: BEOL stackup and cross-section of microstrip T-line in 90-nm SiGe BiCMOS process.

1.4.1 Passive Element Design

To achieve $Z_{0,l} = 50\ \Omega$ in (1.1), the transmission line $Z_{0,T} = \sqrt{L_{tl}/C_{tl}}$ must be greater than 50 Ω since the device parasitic capacitance lowers the final characteristic impedance. In addition, losses in the transmission line, expressed per-length as α_{tl} in the propagation constant $\gamma_{tl} = \alpha_{tl} + \beta_{tl}$, limit the marginal gain of each additional stage [16]. We seek to minimize the total attenuation factor $\alpha_{tl} l_{seg}$ while maximizing Z_0 to allow for the largest parasitic capacitance loading, and thus, gain per stage. Fig. 1.7 shows the BEOL stackup for this process. Since the dielectric stack height is sufficiently large, a microstrip line is used as the transmission line element to ease access to the device. Optimizing the shunt capacitance loading budget with respect to line resistance results in a 2 μm-wide line on M9 layer, with $Z_{0,T}$ of 78.6 Ω and less than 0.7 dB/mm loss up to 110 GHz. Keeping $Z_{0,l} = 50\ \Omega$ and setting our target bandwidth $f_{c,l} = 110$ GHz, the total series inductance and shunt capacitance per stage are 145 pH and 58 fF, respectively. For comparison with Fig. 1.3,

this results in a transmission line θ of $\pi/2$ at 55 GHz.

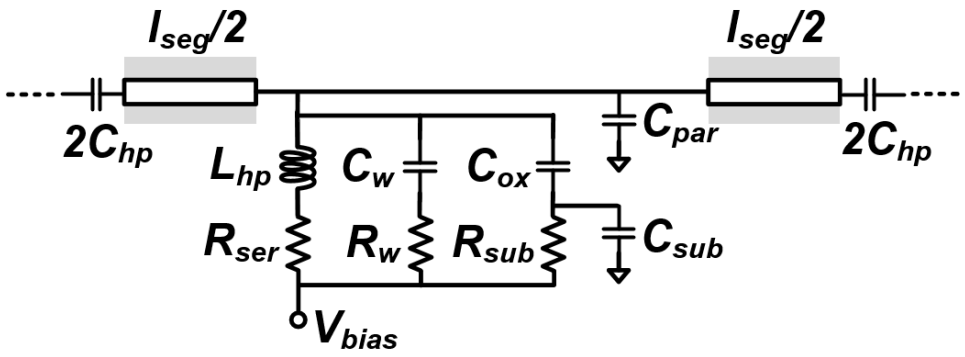


Figure 1.8: Synthesized band-pass transmission line including inductor parasitics.

The band-pass DA also includes a high-pass constant-k section to decouple the dc level of each stage. Fig. 1.8 shows the embedded high-pass T-section within the standard low-pass filter. For shunt inductance L_{hp} and series capacitors $2C_{hp}$, the characteristic impedance and low-frequency cutoff are

$$Z_{0,h} = \sqrt{\frac{L_{hp}}{C_{hp}}} \tag{1.15}$$

$$f_{c,h} = \frac{1}{4\pi\sqrt{L_{hp} \times C_{hp}}}. \tag{1.16}$$

Matching $Z_{0,h} = Z_{0,l} = 50\ \Omega$ and choosing $f_{c,h} = 8$ GHz to cover X-band frequencies, (1.15) and (1.16) give $L_{hp} = 500$ pH and $C_{hp} = 200$ fF.

As the additional components of the high-pass T-section contain parasitic elements, the effect on DA performance should be considered. The shunt inductors contribute not only inductance, but also shunt capacitance and conductance due to winding and substrate leakage. Shunt capacitance sets a self-resonant frequency for the inductor in many applications, but in the artificial transmission line of a DA, it can be included as part of the low-pass T-section. Thus, whereas the parasitic capacitance degrades the high-frequency performance in a purely high-pass DA [17], it is absorbed here with C_{par} from the transistor. While this means the capacitance will not modify the response of the high-pass constant-k sections, it does lead to a reduction in the allowable device size, which may result in a reduction in either f_T or P_{1dB} and P_{sat}, depending on how the device biasing is optimized.

As such, it is imperative to make the inductor footprint as small as possible to occupy less of the shunt capacitance budget. Generally, reducing the line width and turn diameter to decrease the capacitance to ground results in higher series resistive losses through the inductor. However, due to the inductor's already high-impedance nature at RF frequencies, this parasitic resistance is not critical in comparison to capacitive effects. Assuming a fixed resistance R_S over frequency, and an inductor component quality factor greater than 10 in the band of interest,

$$G_{Shunt} = \frac{1}{R_{Shunt}} = \frac{1}{R_S\left[1+\left(\frac{\omega L_S}{R_S}\right)^2\right]} = \frac{R_S}{R_S^2+(\omega L_S)^2} \approx \frac{R_S}{(\omega L_S)^2}. \tag{1.17}$$

From (1.17), it is clear that with increasing frequency, the impact of the inductor losses becomes minimal, justifying the use of a low-Q, compact, high turn count, square spiral. Electromagnetic simulations [18] of the inductor show the conductance is less than 2 mS above 20 GHz, significantly less than that of the transistor, indicating negligible impact on DA performance. Fitting the simulated behavior to the lumped-element model shown in Fig. 1.8, it is found that R_{ser} dominates the resistive performance, and thus, the capacitive elements are absorbed into the constant-k section with minimal effect. The spiral inductor achieves peak Q of 16 after absorption of 10 fF total shunt capacitance into the artificial transmission line (Fig. 1.9).

In a similar manner as the shunt inductor, the shunt capacitance to ground for the series blocking capacitors of the high-pass T-section can be included in the low-pass constant-k line. High density metal-insulator-metal (MIM) capacitors in the SiGe technology used exhibit shunt parasitic capacitance of only 2.5 fF, and hence do not significantly impact the device size. The series inductance of the capacitor can be absorbed into the inductance of the low-pass constant-k section, and the resistive losses are directly included in the series loss of the artificial transmission lines. Care must be taken in design of the connections to the series capacitors to minimize this resistance. Capacitors from adjacent high-pass sections can be combined to reduce area and loss.

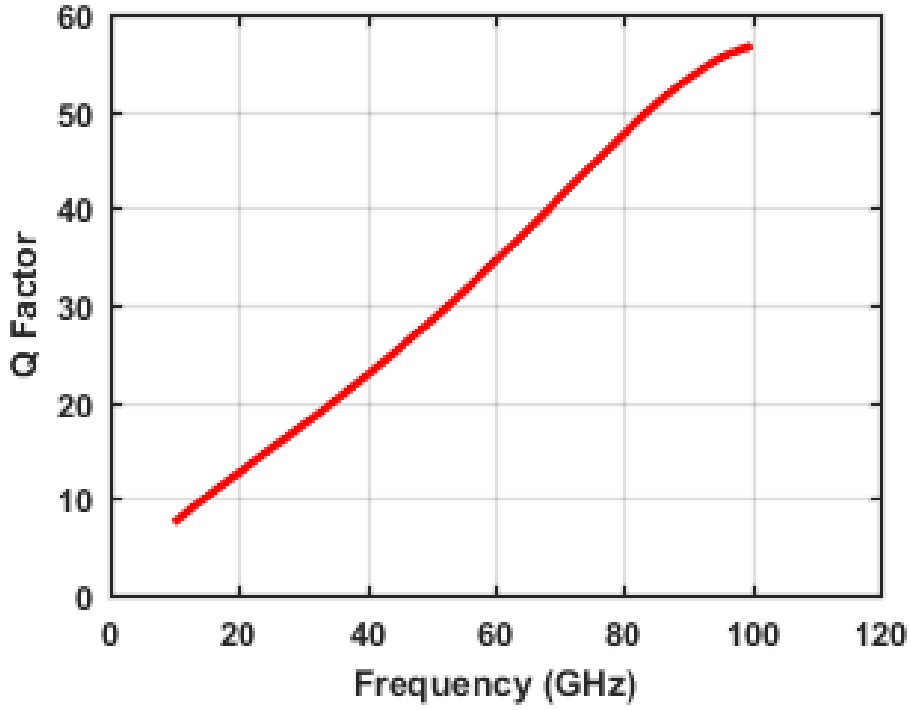


Figure 1.9: Simulated shunt inductor Q factor after absorption of capacitive elements into constant-k section.

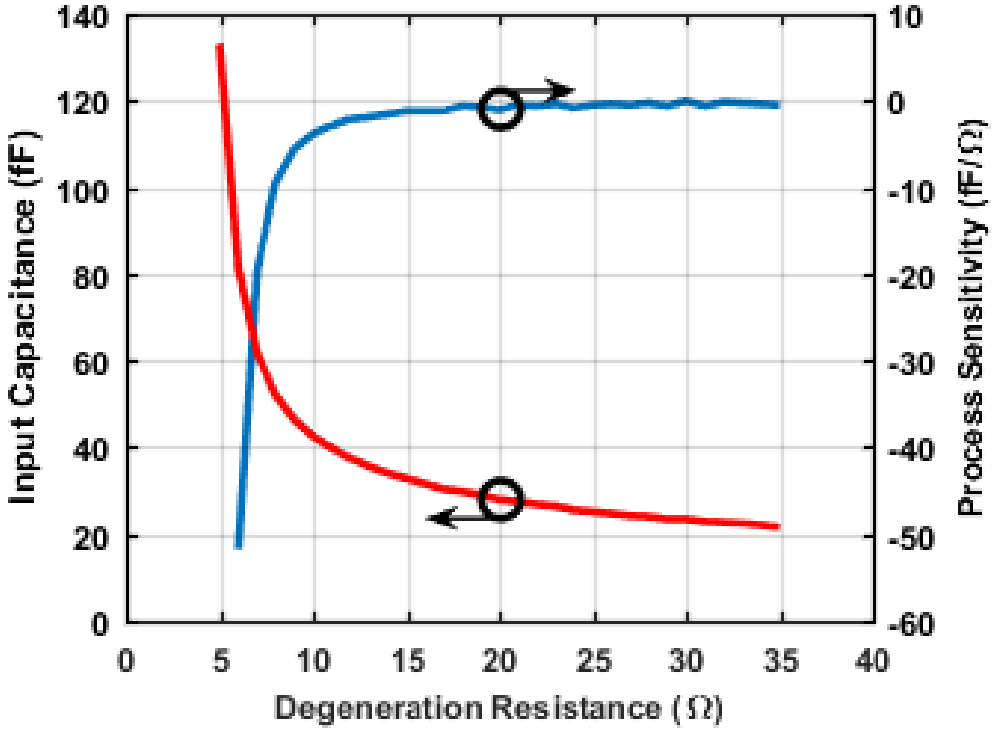


Figure 1.10: Input capacitance and sensitivity to degeneration resistance R_E of a 6μm/0.09μm HBT.

1.4.2 Active Element Design

In the 90-nm SiGe BiCMOS process, HBTs have simulated peak transit frequency f_t upwards of 300 GHz [19, 20]. However, base resistance in bipolar devices creates shunt losses in the synthesized input line, degrading the gain-bandwidth (GBW) product of the DA. Furthermore, as detailed in [10], the input conductance and capacitance of a common-emitter amplifier increase with frequency, incurring extra loss and impedance mismatch. To counteract these effects, resistive emitter degeneration flattens the input characteristic for ultra-wideband operation. For base-emitter capacitance C_{be}, base resistance r_b, transconductance g_m, and degeneration resistance R_E, the capacitance and conductance looking into the base are given by [10]:

$$C_{in}^{'} = \frac{C^{'}}{1+\omega^2 C^{'2}\left(r_b+R_E\right)^2} \tag{1.18}$$

$$G_{in}^{'} = \frac{\omega^2 C^{'}\left(r_b+R_E\right)}{1+\omega^2 C^{'2}\left(r_b+R_E\right)^2}\,, \tag{1.19}$$

where $C^{'} = C_{be}/\left(1+g_m R_E\right)$. As shown in Fig. 1.10, increasing emitter resistance R_E has diminishing returns on the decrease of input capacitance - this in turn reduces the transmission line loading sensitivity to process variation. On the other hand, while larger R_E linearly reduces the effective G_m of the transistor, the smaller input capacitance and conductance allows for the use of a larger device and lowers the shunt loss in the T-section. With these considerations, a 20 Ω resistor is chosen to maintain high overall gain for the target operation bandwidth. The bandwidth gained comes at the expense of a slight degradation in efficiency, as there is power lost across R_E. As shown in Fig. 1.11, we trade only 1% in PAE to improve the bandwidth by more than 80 GHz.

The final DA gain stage employs an HBT cascode to mitigate the Miller effect, increase the input and output impedances of the stage, and improve the isolation between base and collector lines. High-performance HBTs in this process achieve maximum f_t at 2 mA/μm current density. We bias the common-emitter transistor at 1.8 mA/μm to avoid the f_t-rolloff associated with Kirk effect at high current swing. Because of the difference in capacitance seen at the base and collector, emitter lengths

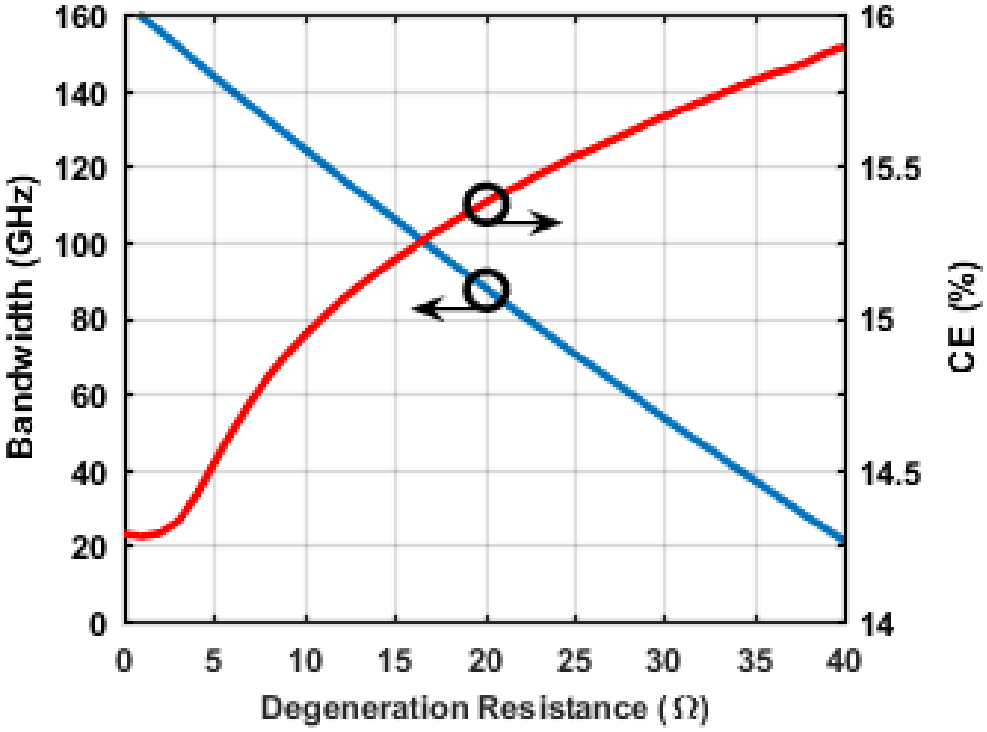


Figure 1.11: Bandwidth-efficiency tradeoff of the DA versus gain stage degeneration resistance R_E.

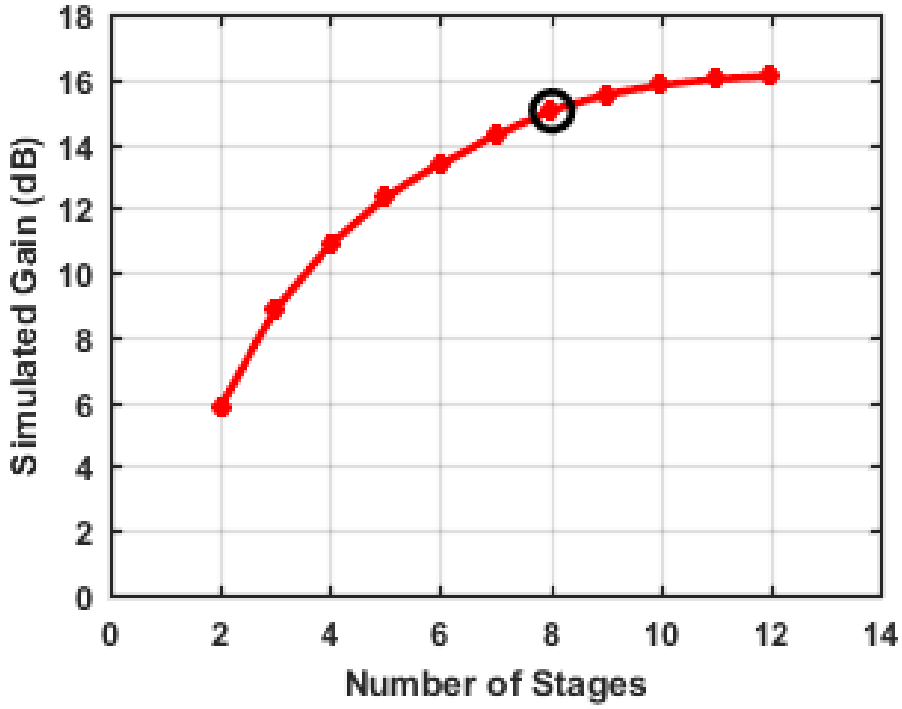


Figure 1.12: Simulated gain as a function of number of band-pass DA stages.

of 12 μm and 6 μm are chosen for the common-base and common-emitter devices, respectively, to satisfy $Z_{0,l} = 50\ \Omega$ for both input and output lines. Additionally, a degeneration capacitor of 40 fF is included in parallel with R_E to introduce a high-frequency zero at $\omega_z = 1/(R_E C_E)$ for gain peaking near the low-pass cutoff. To ensure good decoupling of the dc bias network, the base of the cascode device is biased at 2.6V through a combination of MOS and MIM RC low-pass filters. The MIM capacitor is placed as closely to the device as possible to limit the parasitic inductance of the bias path, preventing high-frequency instability.

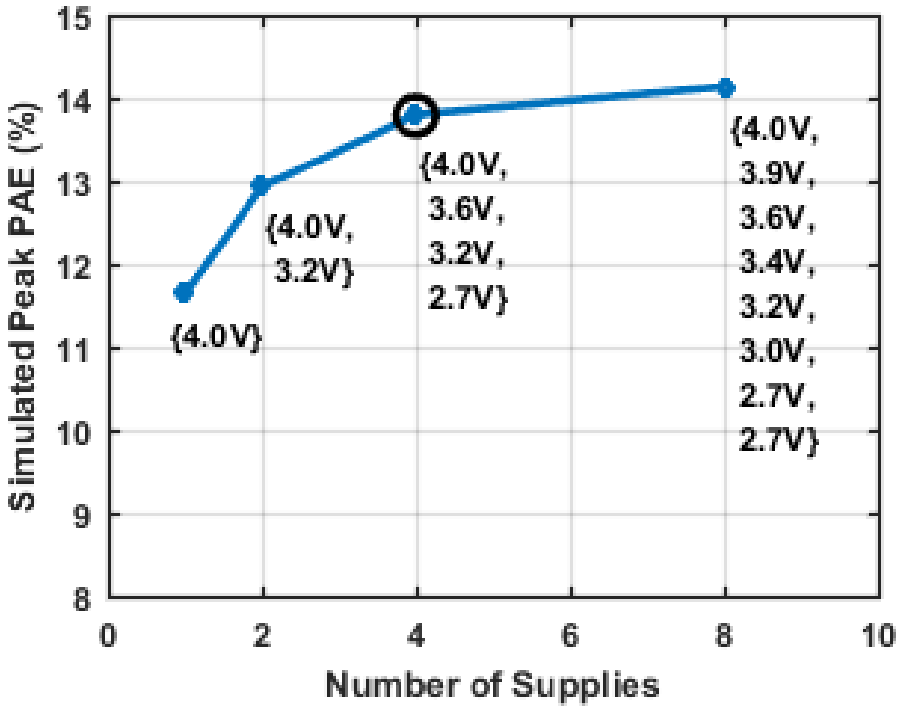


Figure 1.13: Simulated optimal PAE versus number of independent scaled supplies for an 8-stage DA.

Though an ideal DA exhibits unbounded GBW product as more stages are added, attenuation of the signals along the base and collector transmission lines limits the achievable gain in practicality. Fig. 1.12 shows the simulated gain versus number of stages. An 8-stage DA is found to offer a good GBW-to-chip area ratio. A sweep of the number and values of collector dc voltages in Fig. 1.13 then reveals the optimal supply-scaling for peak PAE. Since the improvement in efficiency between four and eight independent supplies is small, we opt to tie the supplies of every two stages

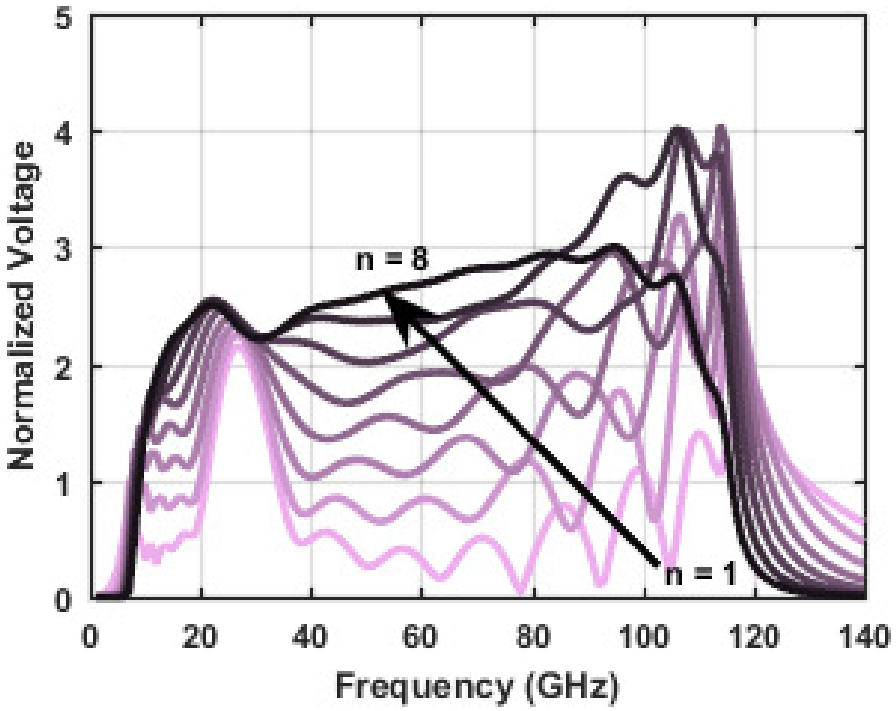


Figure 1.14: Simulated collector voltage magnitudes in the 8-stage supply-scaled DA.

together as a tradeoff between efficiency and chip complexity. This results in a simulated 20% improvement in PAE over a uniform DA with negligible impact on gain. Though the peak CE for an ideal supply-scaled DA with four independent supplies is 40% from (1.14), we are only able to scale from 2.7 to 4.0 V, giving a peak CE of 29.6% - transistor knee voltage and emitter resistance and transmission line losses further reduce the efficiency. Fig. 1.14 shows the simulated collector voltage magnitude for each successive gain stage. Compared to the voltage distribution found for an ideal low-pass DA in Fig. 1.3, the band-pass DA sees a shift in the zero electrical length frequency due to the high-pass T-section. Additionally, the collector voltages exhibit non-idealities as the DA approaches the cutoff frequency of the low-pass artificial transmission line. However, one may observe the expected monotonically increasing voltage distribution in the mid-band.

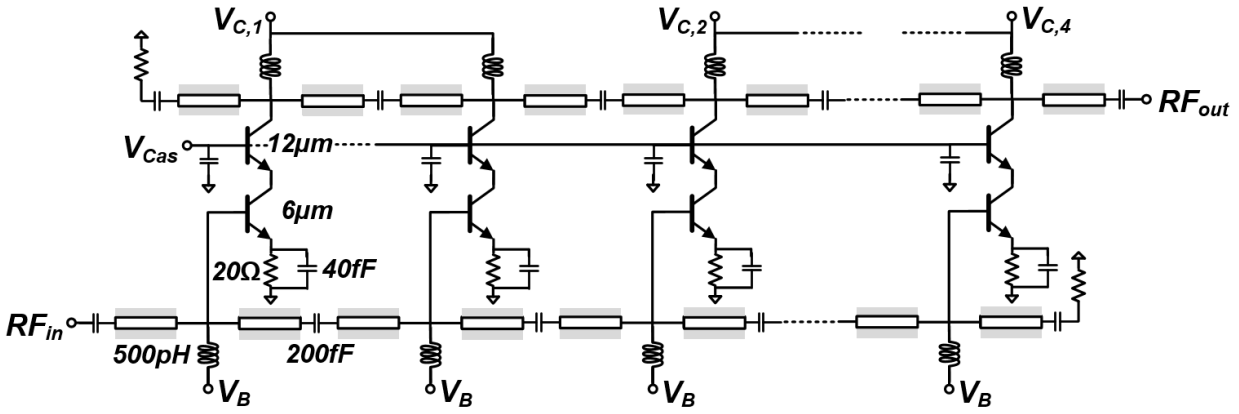


Figure 1.15: Schematic of the fabricated 8-stage supply-scaled DA.

1.5 Measurement Results

The schematic of the fabricated 8-stage supply-scaled DA is shown in Fig. 1.15, and a chip microphotograph is shown in Fig. 1.16. The amplifier occupies an area of 2.65 mm $\times$ 0.57 mm, including pads. Measurement of the DA is performed with on-wafer probing, and no de-embedding of pad parasitics is done. Forward and reverse terminations are provided on-chip, and the high-pass T-section filters allow for supplying of dc biases without the need for bias-tees. Four supply voltages of 2.7, 3.2, 3.6, and 4.0 V draw 21.6 mA of nominal bias current each, resulting in total dc power consumption of 297 mW for small input signal. Fig. 1.17 shows the measured and simulated S-parameters and stability factor μ of the DA. The amplifier achieves a peak small-signal gain of 12.0 dB with a 3 dB pass-band bandwidth from 14-105 GHz (91 GHz), corresponding to a GBW product of 362 GHz. Measured gain is 2.5 dB lower than simulated, which is consistent with slow HBT process corners on this fabrication run. Additionally, the S-parameters show a pronounced ripple in gain and impedance match around 30 GHz. This degradation is mainly caused by imperfect modeling of the dc supply distribution network, whose extra inductance becomes manifest in the high-pass T-section filters near the low-frequency cutoff. Except for this ripple, the input and output rejection ratios S11 and S22 are better than 9 dB from 10-90 GHz. The reverse isolation is less than -24 dB, and the amplifier is unconditionally stable across the entire measured frequency range.

Large-signal measurements are performed across the entire operating frequency band.

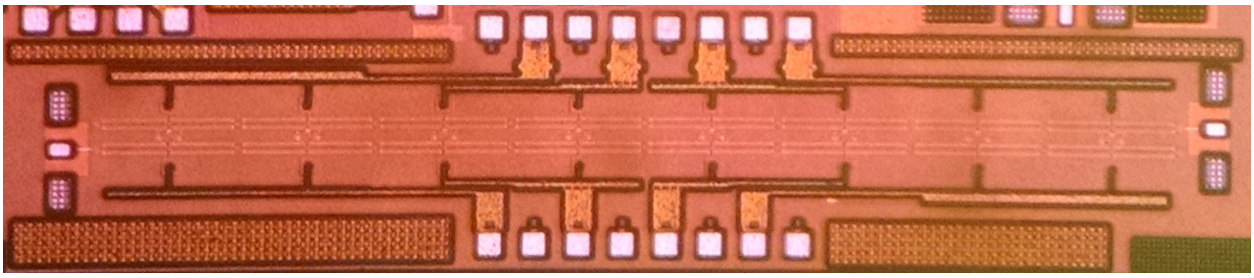


Figure 1.16: Chip microphotograph of the supply-scaled DA.

Measured and simulated power, gain, and efficiency at the midband frequency of 50 GHz are shown in Fig. 1.18. The DA has a 1 dB gain-compressed output power P_{-1dB} of 14.9 dBm and saturated output power P_{sat} of 17.0 dBm. Peak CE and PAE are 15.1% and 12.6%, respectively. The output power characteristics match simulation results reasonably well, though maximum PAE suffers by 1.2% due to the lower gain. Fig. 1.19 shows measured and simulated P_{sat}, P_{-1dB}, and peak CE and PAE at various frequencies across the bandwidth. The 3 dB power bandwidth is greater than 70 GHz (15-87 GHz), and PAE is better than 8.5% up to 80 GHz, except for the aforementioned gain degradation at 30 GHz. As a control experiment, the DA was also measured with all supply voltages set equal to the final stage value of 4 V (Fig. 1.20). The uniform DA exhibited 17% reduced CE and PAE and consumed the same current compared to supply-scaled operation, while only a 0.2 dB increase in gain and saturated output power on average was observed.

Supply-scaling comes at the disadvantage of requiring multiple supplies. Nonethelesss, given approximately 95% efficiency of modern switched dc-dc converters, the supply-scaled DA efficiency would be reduced by 5.3% due to converter inefficiency. Even accounting for the converter, the supply-scaled DA maintains an overall 11.1% efficiency advantage over the uniform DA. In addition, the introduction of switched-mode power converters could suggest the possibility to explore dynamic supply modulation for envelope-tracking techniques under back-off conditions. This verifies the supply-scaling theory and design as an effective method of efficiency enhancement without DA performance degradation.

Table 1.1 summarizes the performance of similar published wideband DAs. The supply-scaled DA achieves the largest reported GBW product of any single-stage silicon-based distributed

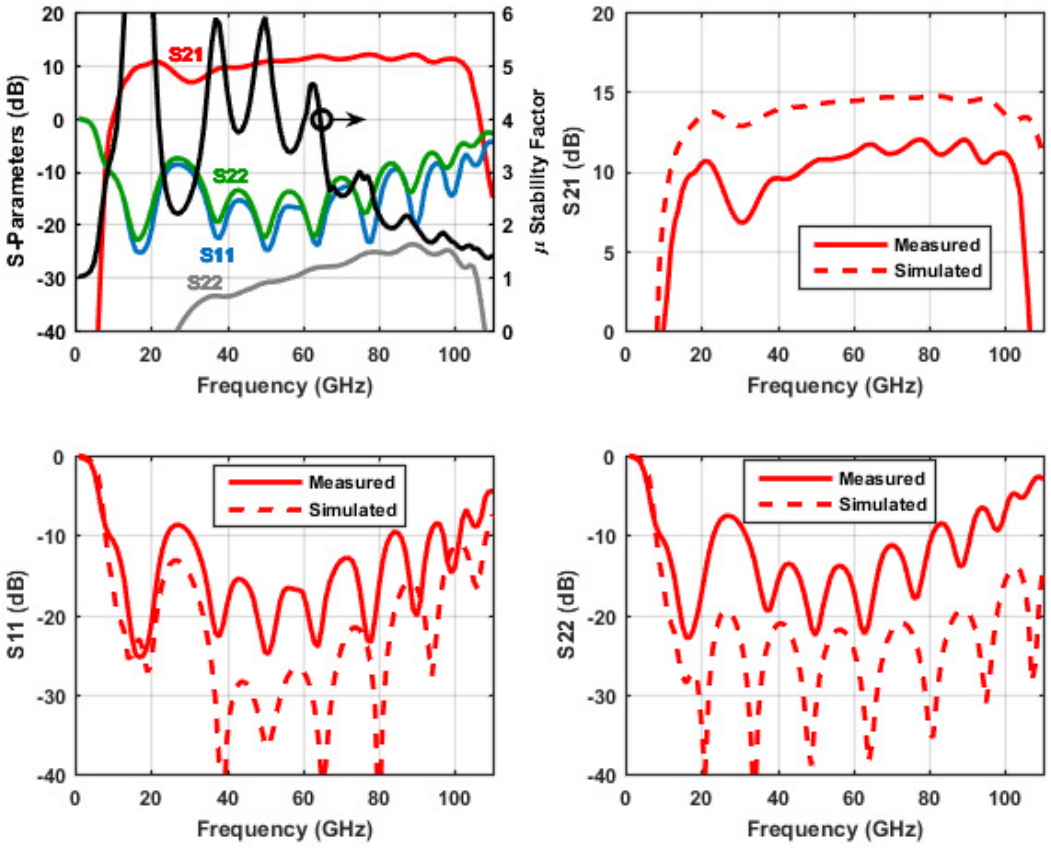


Figure 1.17: Measured and simulated S-parameters and μ stability factor of the supply-scaled DA.

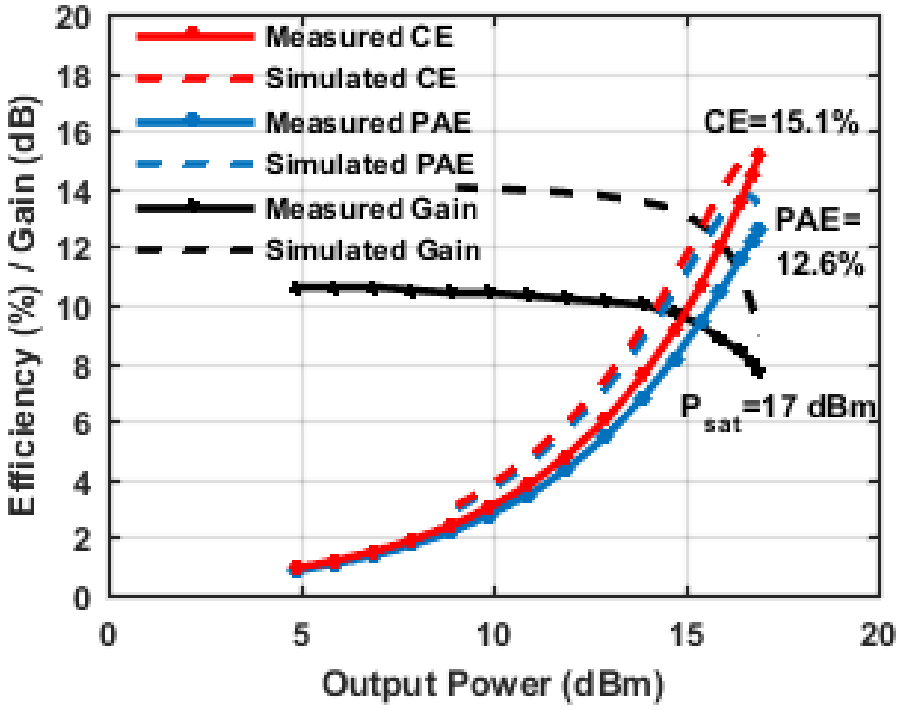


Figure 1.18: Measured and simulated power gain, collector efficiency, and power-added efficiency at 50 GHz.

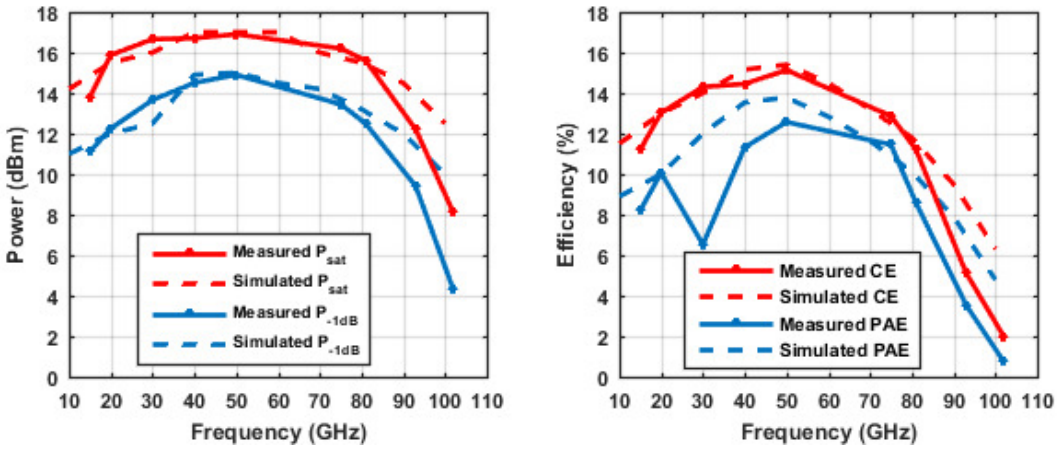


Figure 1.19: Measured and simulated output power, 1 dB gain-compressed power, and peak CE and PAE over the operating bandwidth.

Table 1.1: Comparison To Other Published DAs

Reference	Gain (dB)	BW (GHz)	GBW (GHz)	Peak P_{-1dB} (dBm)	PAE at P_{-1dB} (%)	Area (mm^2)	Technology
[3]	15	80	450	N/A	N/A	0.31	40-nm digital CMOS
[4]	8.5	135	360	10	7.9	0.36	55-nm SiGe BiCMOS
[5]	9	92	259	N/A	N/A	0.45	45-nm CMOS SOI
[6]	11	85 (5-90)	320	12	6.8	1.28	0.12-μm CMOS SOI
[7]	7.4	80	190	N/A	N/A	0.72	90-nm CMOS
[8]	10	170	537	7.5	4.5	0.38	0.13-μm SiGe BiCMOS
[10]	10	110	348	16.7*	11.5	2.18	0.13-μm SiGe BiCMOS
[21]	4	87 (4-91)	138	9	4.4	0.80	0.12-μm CMOS SOI
[22]	14	74	370	3.2	2.4	1.72	90-nm CMOS
[23]	20	39	394	6.5	1.8	2.24	0.18-μm CMOS
[24]	7	43	105	N/A	N/A	1.80	0.13-μm CMOS SOI
This Work	12	91 (14-105)	362	14.9	9.7	1.51	90-nm SiGe BiCMOS

*Pseudo-differential operation

Table 1.2: Comparison To Other Published Wideband mm-Wave PAs

Reference	Gain (dB)	BW (GHz)	Peak P_{-1dB} (dBm)	Peak P_{sat} (dBm)	Peak PAE (%)	P_{sat} BW (GHz)	Technology
[10]	10	110	16.7*	17.5*	13.2	77	0.13-μm SiGe BiCMOS
[25]	9	N/A	N/A	21	3.5	24 (73-97)	0.13-μm SiGe BiCMOS
[26]	15	56	14.4	17.2	9.2	15 (77-92)	0.13-μm SiGe BiCMOS
[27]	17	15	14.5	17.5	12.8	23 (65-88)	0.12-μm SiGe BiCMOS
[28]	12	27	12.5	14.8	8.7	25 (80-105)	65-nm CMOS
[29]	18	33	12.0	14.0	4.5	33 (77-110)	65-nm CMOS
This Work	12	91	14.9	17.0	12.6	70 (15-85)	90-nm SiGe BiCMOS

*Pseudo-differential operation

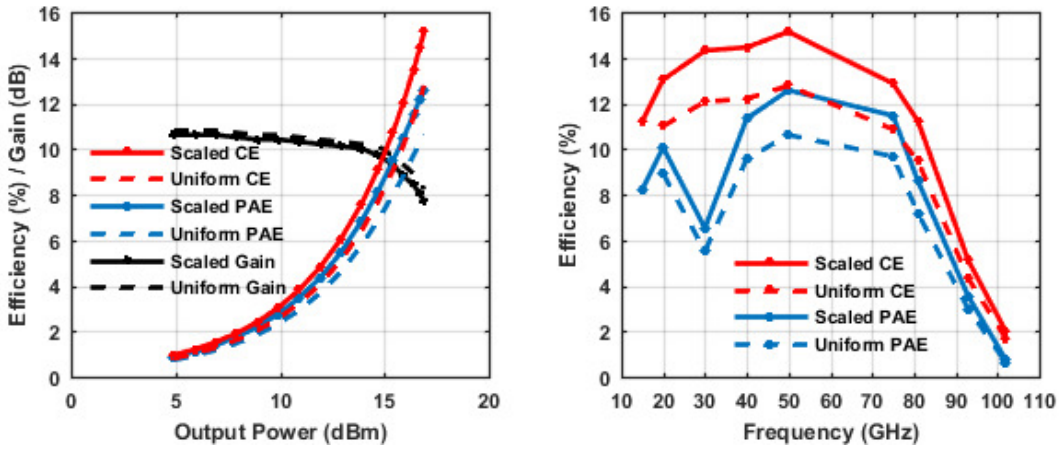


Figure 1.20: Measured power gain, CE, and PAE at 50 GHz, and peak CE and PAE across the bandwidth, of the DA with uniform supply biasing.

power amplifier ($>$ 10 dBm) to the author's knowledge. Among amplifiers shown, this work also has the largest single-ended output power, with comparable 3 dB power bandwidth. In particular, the supply-scaled DA exhibits nearly 3 dB greater output power and 3% higher efficiency above 70 GHz than the stage-scaled SiGe BiCMOS DA in [10]. While this work implemented a single-ended amplifier as opposed to [10], supply-scaling could also be applied to a differential DA to double the output power. A differential design might also be utilized to simplify the biasing circuits. However, differential DAs in the millimeter-wave bands require an output balun with a bandwidth that matches the amplifier bandwidth, and this poses a significant challenge. Compared to silicon-based W-band tuned PAs in Table 1.2, the DA achieves much greater bandwidth while maintaining similar peak power, efficiency, and gain.

1.6 Conclusion

This paper has introduced supply-scaling as a technique to achieve wideband power efficiency enhancement in DAs, and its advantages over impedance tapering techniques are discussed from the point of view of interstage load modulation. Design methodology of a band-pass DA to enable the technique is presented, with focus on the effects of high-pass constant-k filter element parasitics. To verify the theory of efficiency enhancement, an 8-stage supply-scaled DA is demon-

strated in a 90-nm SiGe BiCMOS process with bandwidth greater than 90 GHz. Peak saturated output power of 17 dBm is measured with a relatively high PAE of 12.6% using four supply voltages from 2.7 V to 4.0 V. Compared to previously reported mm-wave silicon-based DAs, the presented amplifier demonstrates superior power and efficiency performance above 70 GHz.

Acknowledgment

The authors would like to acknowledge Integrand Software for the use of EMX, the Trusted Foundry Access Program for chip fabrication, and the National Science Foundation for support through a CAREER research award and Graduate Fellowship. Additionally, the support of the Office of Naval Research is appreciated. They also thank Bagher Rabet for valuable discussions.

This chapter, in full, is a reprint of the material as it appears in IEEE Journal of Solid-State Circuits, 2016, Fang, Kelvin; Levy, Cooper; Buckwalter, James. The dissertation author was one of the primary authors of this work.

Bibliography

[1] P. Indirayanti, W. Volkaerts, P. Reynaert, and W. Dehaene, "Picosecond pulse generation with nonlinear transmission lines in 90-nm CMOS for mm-wave imaging applications," in *19th IEEE Int. Conf. on Electronics, Circuits and Systems*, pp. 885–888, Dec 2012.

[2] E. L. Ginzton, W. R. Hewlett, J. H. Jasberg, and J. D. Noe, "Distributed amplification," *Proc. IRE*, vol. 36, pp. 956–969, Aug 1948.

[3] P.-H. Chen, K.-S. Yeh, J.-C. Kao, and H. Wang, "A high performance DC-80 GHz distributed amplifier in 40-nm CMOS digital process," in *IEEE MTT-S Int. Microwave Symp.*, pp. 1–3, Jun 2014.

[4] J. Hoffman, S. P. Voinigescu, P. Chevalier, A. Cathelin, and P. Schvan, "A low noise, DC-135GHz MOS-HBT distributed amplifier for receiver applications," in *IEEE Comp. Semicond. Integr. Circuit Symp.*, pp. 1–4, Oct 2015.

[5] J. Kim and J. F. Buckwalter, "A 92 GHz bandwidth distributed amplifier in a 45 nm SOI CMOS technology," *IEEE Microwave Wireless Compon. Lett.*, vol. 21, pp. 329–331, May 2011.

[6] J. Kim, J.-O. Plouchart, N. Zamdmer, R. Groves, M. Sherony, Y. Tan, M. Talbi, J. Safran, and L. Wagner, "A 12dBm 320GHz GBW distributed amplifier in a 0.12um SOI CMOS," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, vol. 1, pp. 478–540, Feb 2004.

[7] R.-C. Liu, T.-P. Wang, L.-H. Lu, H. Wang, S.-H. Wang, and C.-P. Chao, "An 80GHz travelling-wave amplifier in a 90nm CMOS technology," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, vol. 1, pp. 154–590, Feb 2005.

[8] P. V. Testa, G. Belfiore, D. Fritsche, C. Carta, and F. Ellinger, "170 GHz SiGe-BiCMOS loss-compensated distributed amplifier," *IEEE J. Solid-State Circuits*, vol. 50, pp. 2228–2238, Jul 2015.

[9] C. Campbell, T. Lee, V. Williams, M.-Y. Kao, H.-Q. Tserng, and P. Saunier, "A wideband power amplifier MMIC utilizing GaN on SiC HEMT technology," *IEEE J. Solid-State Circuits*, vol. 44, pp. 2640–2647, Oct 2009.

[10] J. Chen and A. M. Niknejad, "Design and analysis of a stage-scaled distributed power amplifier," *IEEE Trans. Microwave Theory Tech.*, vol. 59, pp. 1274–1283, Apr 2011.

[11] K. Fang, C. Levy, and J. F. Buckwalter, "A 105-GHz, supply-scaled distributed amplifier in 90-nm SiGe BiCMOS," in *IEEE Bipolar/BiCMOS Circuits and Tech. Meeting*, pp. 182–185, Oct 2015.

[12] T. T. Wong, *Fundamentals of Distributed Amplification*. Boston, MA: Artech House, 1993.

[13] M. Rodwell, M. Riaziat, B. Auld, and D. Bloom, "Internal microwave propagation and distortion characteristics of traveling-wave amplifiers studied by electrooptic sampling," *IEEE Transactions on Microwave Theory and Techniques*, vol. 34, pp. 1356–1362, dec 1986.

[14] J. Roderick and H. Hashemi, "A 0.13μm CMOS power amplifier with ultra-wide instantaneous bandwidth for imaging applications," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, pp. 374–375, 375a, Feb 2009.

[15] B. Sewiolo, G. Fischer, and R. Weigel, "A 12-GHz high-efficiency tapered traveling-wave power amplifier with novel power matched cascode gain cells using SiGe HBT transistors," *IEEE Trans. Microwave Theory Tech.*, vol. 57, pp. 2329–2336, Aug 2009.

[16] K. B. Niclas, R. R. Pereira, and A. P. Chang, "On power distribution in additive amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 38, pp. 1692–1700, Nov 1990.

[17] T. D. Gathman and J. F. Buckwalter, "A Ka-band high-pass distributed amplifier in 120nm SiGe BiCMOS," in *IEEE Int. Microwave Symp. Dig.*, pp. 952–955, May 2010.

[18] "Emx," 2003-2016.

[19] J. J. Pekarik, J. Adkisson, P. Gray, Q. Liu, R. Camillo-Castillo, M. Khater, V. Jain, B. Zetterlund, A. DiVergilio, X. Tian, A. Vallett, J. Ellis-Monaghan, B. J. Gross, P. Cheng, V. Kaushal, Z. He, J. Lukaitis, K. Newton, M. Kerbaugh, N. Cahoon, L. Vera, Y. Zhao, J. R. Long, A. Valdes-Garcia, S. Reynolds, W. Lee, B. Sadhu, and D. Harame, "A 90nm SiGe BiCMOS technology for mm-wave and high-performance analog applications," in *IEEE Bipolar/BiCMOS Circuits and Tech. Meeting*, pp. 92–92, Oct 2014.

[20] H.-C. Lin and G. M. Rebeiz, "A 110-134-GHz SiGe amplifier with peak output power of 100-120 mW," *IEEE Trans. Microwave Theory Tech.*, vol. 62, pp. 2990–3000, Oct 2014.

[21] J.-O. Plouchart, J. Kim, N. Zamdmer, L.-H. Lu, M. Sherony, Y. Tan, R. A. Groves, R. Trzcinski, M. Talbi, A. Ray, and L. F. Wagner, "A 4-91-GHz traveling-wave amplifier in a standard 0.12-μm SOI CMOS microprocessor technology," *IEEE J. Solid-State Circuits*, vol. 39, pp. 1455–1461, Sep 2004.

[22] A. Arbabian and A. M. Niknejad, "Design of a CMOS tapered cascaded multistage distributed amplifier," *IEEE Trans. Microwave Theory Tech.*, vol. 57, pp. 938–947, Mar 2009.

[23] J.-C. Chien and L.-H. Lu, "40-Gb/s high-gain distributed amplifiers with cascaded gain stages in 0.18-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, pp. 2715–2725, Dec 2007.

[24] C. Pavageau, M. Si Moussa, J.-P. Raskin, D. Vanhoenaker-Janvier, N. Fel, J. Russat, L. Picheta, and F. Danneville, "A 7-dB 43-GHz CMOS distributed amplifier on high-resistivity soi substrates," *IEEE Trans. Microwave Theory Tech.*, vol. 56, pp. 587–598, Mar 2008.

[25] E. Afshari, H. Bhat, X. Li, and A. Hajimiri, "Electrical funnel: A broadband signal combining method," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, pp. 751–760, Feb 2006.

[26] A. Arbabian, B. Afshar, J.-C. Chien, S. Kang, S. Callender, E. Adabi, S. D. Toso, R. Pilard, D. Gloria, and A. Niknejad, "A 90GHz-carrier 30GHz-bandwidth hybrid switching transmitter with integrated antenna," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, pp. 420–421, Feb 2010.

[27] A. Komijani and A. Hajimiri, "A wideband 77-GHz, 17.5-dBm fully integrated power amplifier in silicon," *IEEE J. Solid-State Circuits*, vol. 41, pp. 1749–1756, Aug 2006.

[28] K.-J. Tsai, J.-L. Kuo, and H. Wang, "A W-band power amplifier in 65-nm CMOS with 27GHz bandwidth and 14.8dBm saturated output power," in *IEEE Radio Freq. Integr. Circuits Symp.*, pp. 69–72, Jun 2012.

[29] K.-L. Wu, K.-T. Lai, R. Hu, C. F. Jou, D.-C. Niu, and Y.-S. Shiao, "77110 GHz 65-nm CMOS power amplifier design," *IEEE Trans. Terahertz Sci. Technol.*, vol. 4, pp. 391–399, May 2014.

Chapter 2

Stacked-Series Doherty Amplifier for mm-Wave Power Amplification

2.1 Introduction

While the Doherty power amplifier (DPA) has been in existence since 1936 [1], and saw almost immediate use for AM broadcast [2], it has continued to see widespread use and innovation to present day [3–5]. In the last few years there has been growing interest in mm-wave Doherty power amplifiers, particularly as mm-wave communications with consumer devices in cellular networks becomes inevitable. Many previous uses for mm-wave radios have utilized signals with low peak-to-average power ratio (PAPR), limiting the benefit that can be provided by a DPA, especially given the increased complexity. However, upcoming communications standards will use multi-carrier waveforms with PAPR of at least 6-8dB [6], necessitating power amplifiers with high efficiency not only at peak power, but also at significant power backoff (PBO), near the signal average power.

While there have been multiple notable efforts to implement mm-wave DPAs in compound semiconductor technologies [7, 8], many recent efforts have been directed at CMOS implementations [9–13]. CMOS PAs have struggled to gain acceptance for applications other than WiFi below 6 GHz,

but the lower required output power and complete integration offered by implementation in CMOS will likely make CMOS PAs a viable competitor for mm-wave phased arrays. At the same time, many of the issues with CMOS below 6 GHz persist at higher frequencies. In particular passive losses, high power handling, and tolerance to load impedance variation are of critical concern. For lower frequencies, the justification for needing high efficiency in CMOS PAs has been user equipment (UE) battery life, as CMOS has not been utilized for basestation (BTS) applications. However in mm-wave phased arrays, the individual radios in both UE and BTS equipment are quite similar, with the primary difference being that basestations will have larger arrays. For basestations, heating from DC power dissipation is of more concern than power consumption itself. Many early mm-wave handset demonstrations have included fans or large copper heatsinks [14]. Improving PA efficiency not only reduces the amount of DC power consumed to produce a given PA output power, but it also reduces the amount of DC power dissipated as heat, helping to address both UE and BTS concerns. As drafts of standards and FCC regulations have evolved during the development of mm-wave consumer communications, there has also been a trend towards higher average output power in more recent products, meaning the PA deisgn which was originally thought to be important, but not vital to a successful mm-wave phased-array product, is again, becoming a key show-stopper.

DPAs in CMOS at mm-wave frequencies must address all aforementioned concerns of CMOS PAs, but also have additional issues. The success of DPAs in RF basestations has been largely enabled by the use of digital predistortion (DPD) to correct for the poor linearity common to most DPAs near the point where the peaking amplifier turns on (to be discussed in more detail in the next section). Unfortunately DPD beomces increasingly difficult at the higher bandwidths of emerging mm-wave communications standards, and even more challenging given that each individual element of a phased array would likely require its own predistortion. In summary, a successful mm-wave Doherty power amplifier should have reasonably high average efficiency, robustness to load impedance variation, sufficiently high output power, and inherent linearity.

This paper expands upon the conference paper previously presented at the Compound Semiconductor Integrated Circuit Symposium (CSICS) [15]. This paper includes a detailed look at

the canonical DPA in contrast to the series DPA to better understand Doherty's original appraoch to designing the DPA. This gives insight into how one might design PAs with the same efficiency characteristic as the DPA, but using different architectures. Our approach to implementing a DPA at mm-wave will be shown, with additional explanation of the operation of the stacked-seried DPA. Additionally, measurement results showing the PA with modulation will be presented. This modification of an intermediate architecture in Doherty's original derivation of the canonical DPA architecture connects the amplifiers and load in a series configuration to address some of the common concerns of CMOS PAs. In section II, the canonical Doherty will be reviewed and the differences of a series Doherty will be highlighted. Section III will describe the stacked series architecture used in this work, and section IV will give details of the implementation, followed by measurement results in section V.

2.2 Comparison of the Canonical and Series Doherty Architectures

While the underlying method for efficiency enhancement is the same for both the canonical and series DPAs, there are important differences between the architectures which make the series Doherty an attractive architecture for implementation in CMOS. In this section the key principles of Doherty operation for comparing the canonical and series architectures will be reviewed. The series architecture will then be introduced, and it's operation explained. Finally, the advantages and drawbacks of implementing each architecture, particularly in CMOS, will be discussed.

While most explanations of the DPA focus on the change in impedance seen by the main amplifier as output power increases, this is more a consequence of a practical implementation of the DPA rather than the original goal Doherty aimed to achieve. Doherty noted that the path to high efficiency in a typical linear PA is to maximize the voltage swing at the plate of the tube (drain for a MOSFET, collector for a BJT), regardless of power level [1]. This means to achieve high efficiency at lower power, a higher load resistance is required. But with an already maximized drain

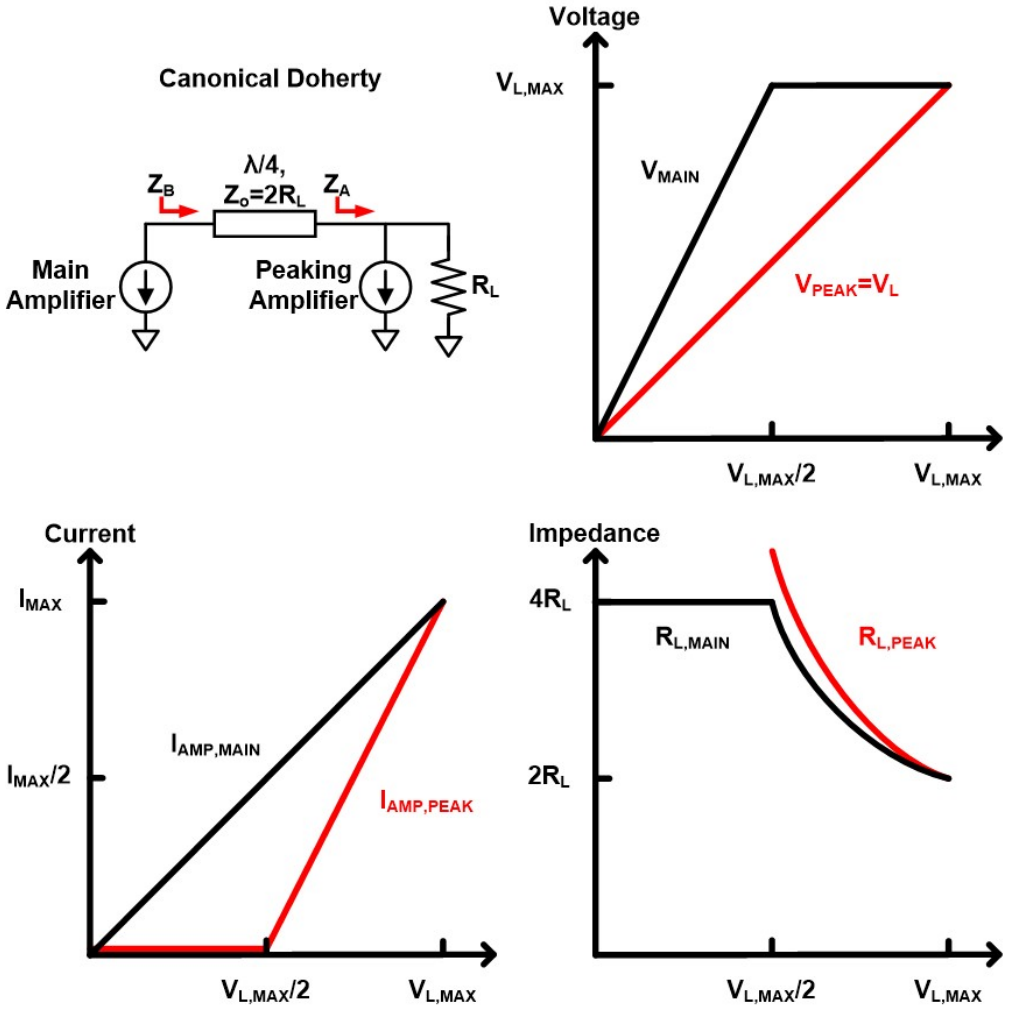


Figure 2.1: The canonical Doherty PA with voltage, current, and load resistances for the constituent amplifiers versus the load voltage swing.

voltage swing, the power cannot increase further with the same load resistance. Doherty's solution to this problem was to decouple the drain voltage of the main power amplifier from the load voltage. In the canonical DPA shown in Fig. 2.1, this is accomplished by placing a $\lambda/4$ transmission line (hereafter referred to as an impedance inverter for its ability to mirror impedances across the Smith chart) between the main amplifier and load resistance. In the canonical Doherty, the characteristic impedance of the impedance inverter is twice that of the load impedance, meaning that when only the main amplifier is active it sees an effective load impedance of $4R_L$, leading to peak voltage at lower power. Once the main amplifier has reached peak voltage swing, the peaking amplifier begins to provide current with amplitude as shown in Fig. 2.1. The peaking amplifier current is in phase with the load current contributed by the main amplifier, leading to a larger voltage swing at the load, while the main amplifier current does not change. This higher voltage at the load with the same current provided by the main amplifier looks like a higher effective impedance Z_A in Fig. 2.1, and after the impedance inverter, a lower impedance Z_B seen by the main amplifier. This lower impedance reduces the voltage swing at the drain of the main amplifier, meaning to maintain maximum peak voltage and high efficiency, the current of the main amplifier can be increased. At peak power, both the main and peaking amplifiers are saturated and provide equal power. In this case each amplifier sees an impedance equal to twice the load impedance ($Z_A = Z_B = 2R_L$). As mentioned previously, the key point noted by most is that the impedance at the drain of the main amplifier decreases, but the more fundamental point is that the voltage and current at the drain of the main amplifier can be controlled independently. If they are controlled such that the drain voltage is maintained near saturation independent of drain current, the amplifier can maintain high efficiency over a wide range of output power.

The concept of independent voltage and current was conveyed by Doherty in Figure 2 as the motivation for his approach to designing the canonical Doherty [1]. In his derivation of the canonical DPA, an intermediate transmission line transform is used which gives Figure 3 of the same paper (Fig. 2.2). This intermediate architecture is what has widely been termed the series DPA. The operation of this structure in terms of voltage and current is more straightforward as shown in

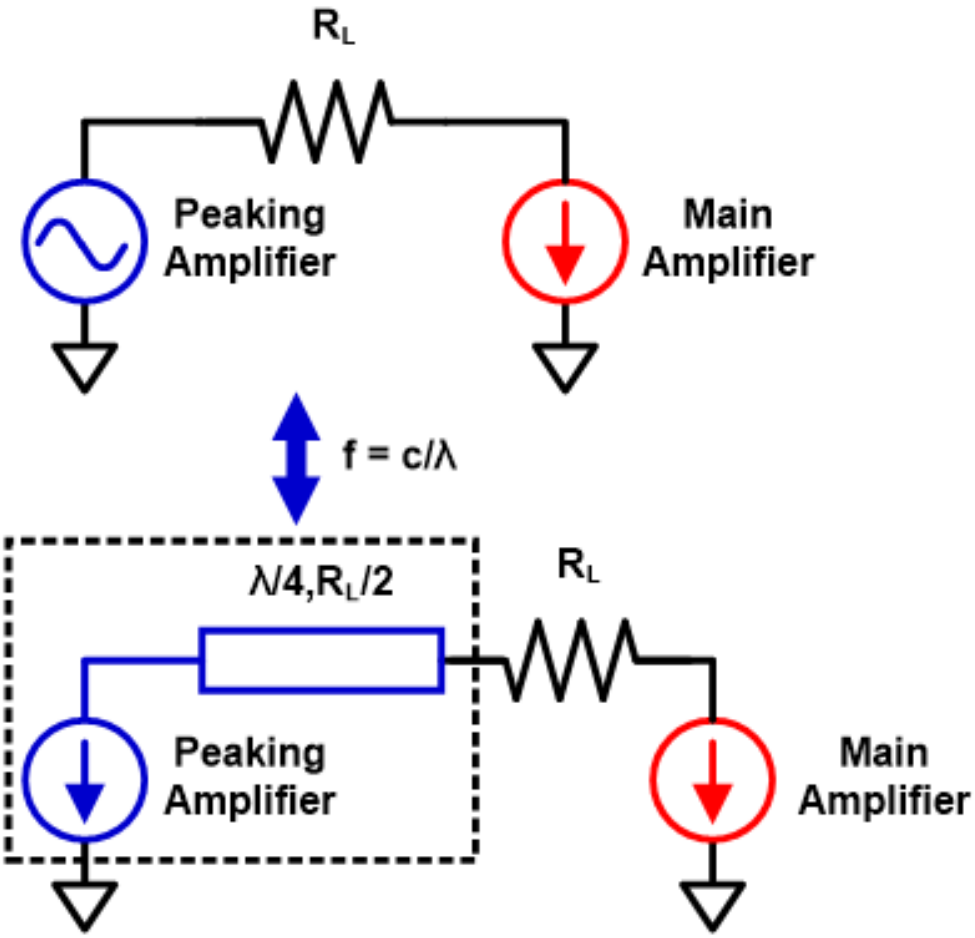


Figure 2.2: The current-voltage DPA from Figure 2 of Doherty's original paper which is then transformed into the series DPA and shown in Figure 3.

Fig. 2.3. At low power, the high impedance of the drain of the peaking amplifier seen through the impedance inverter (for series DPA the characteristic impedance on the impedance inverter is half of the load impedance) presents a virtual ground at the load allowing the main amplifier to directly drive the load. Once the main amplifier voltage swing is maximized, the peaking amplifier begins to provide current. Even if there is voltage swing at the drain of the peaking amplifier, ideally, the impedance seen looking into the drain is quite high if the peaking amplifier is maintained in a linear mode of operation. This means the current provided by the peaking amplifier is transformed to a voltage, with the corresponding low impedance of a voltage source for Z_A. By setting the voltage on one side of the load, and the current on the other, the voltage of the main amplifier can be made independent of the main amplifier current as desired.

The important differences between the canonical and series DPAs can now be examined. Unlike the canonical DPA, the main amplifier of the series DPA drives R_L instead of $4R_L$ at low power. At high power, the main and peaking amplifier each drive $R_L/2$, again, one fourth the impedance of the canonical Doherty. This would indicate that for a fixed load resistance, the series Doherty should be capable of providing four times the output power for the same voltage swing across the MOSFETs. This can also be seen by the voltage across the load being the series combination of the voltages of the two devices (doubled voltage gives four times the power). In CMOS power amplifiers, the output power that can be achieved is typically limited by the MOSFET breakdown voltage. Reduced breakdown voltage in scaled CMOS is leading to reduced output power in CMOS power amplifiers. One way to overcome this is a matching network to reduce the impedance seen at the drain of the device, but this comes with increased losses due to the higher transformation quality factor with limited component quality factor in CMOS. Alternatively, a series power combining structure can be used to sum the voltages of multiple low voltage, high current devices. The series DPA accomplishes series power combining in combination with the efficiency enhancement at backoff provided by the DPA architecture.

Another approach to addressing the voltage-handling issue of the canonical DPA would be to implement the main and peaking amplifier each as a stacked amplifier [9]. While this appears

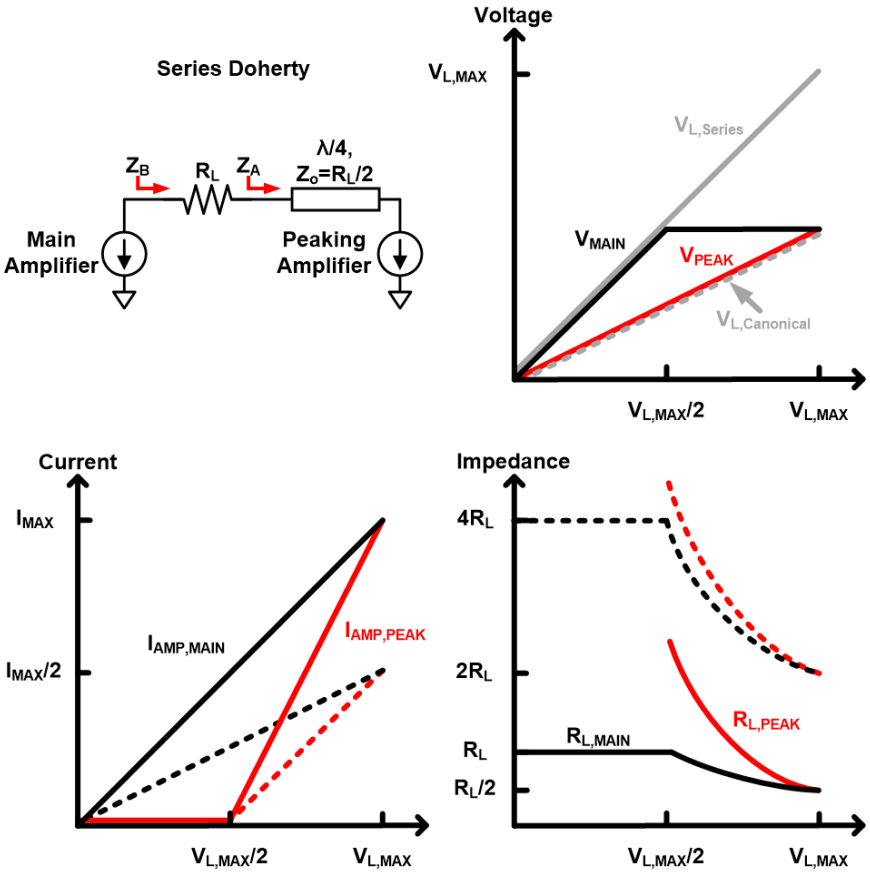


Figure 2.3: The series Doherty PA with voltage, current, and load resistances for the constituent amplifiers versus the load voltage swing. When different from the series DPA, canonical DPA curves are shown with dashed lines for comparison.

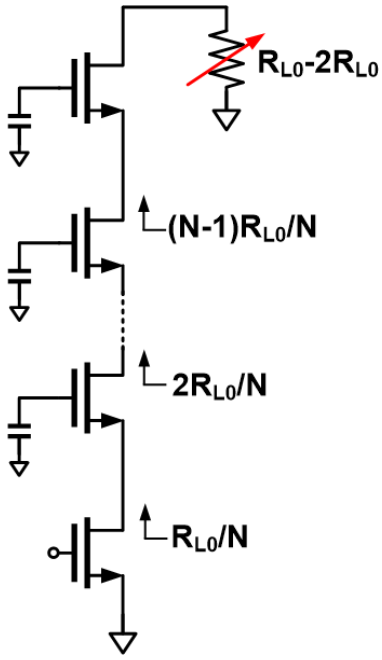


Figure 2.4: Stacked amplifier impedances with load modulation. This model assumes C_{gd} of the MOSFETs is negligible. While this assumption is conservative, [9] does indicate that there is some increase in voltage swing on devices higher in the stack with higher load impedance.

to be a viable approach, theoretically the suitability of a stacked amplifier for power amplifiers that experience load modulation such as the DPA is questionable. A stacked amplifier is typically designed such that the impedance seen looking into the source of each stacked device is

$$Z_{s,n} = \frac{n-1}{N} R_L, \; 2 \leq n \leq N \tag{2.1}$$

where n is the device number with $n = 1$ being the common source device at the bottom of the stack, and N the number of devices in the stack (Fig. 2.4). This allows the voltage swing seen at the top drain to be equally divided across all of the devices in the stack. If the C_{gd}, and C_{ds} of the MOSFETs are assumed to be negligible, the impedance looking into the source of each device is independent of the actual load impedance. Typically the stacked amplifier is designed to provide equal voltage division at peak power, the most critical condition in terms of device degradation from hot carrier injection (HCI), meaning the load resistance used when sizing capacitors for a stack in a canonical DPA would be $2R_L$. If we then consider the case of 6 dB backoff, where the load

impedance seen by the main amplifier is $4R_L$, the votlage swing at the top drain is the same, but the impedances at each source in the stack are still related to the $2R_L$. This means

$$\begin{aligned} V_{ds,n} &= I_{ds}Z_{s,n+1} - \sum_{k=1}^{n-1} V_{ds,k}, 1 \leq n \leq N-1 \\ V_{ds,N} &= I_{ds}R_{L,mod} - \sum_{k=1}^{n-1} V_{ds,k}. \end{aligned} \tag{2.2}$$

It can be seen that $V_{ds,n}$ is constant for $1 \leq n \leq N-1$, but if $R_{L,mod}$ is increased from $R_{L,P_{max}}$, the resulting $V_{ds,N}$ is

$$\begin{aligned} V_{ds,N} &= I_{ds}R_{L,mod} - \frac{N-1}{N} I_{ds}R_{L,min} \\ &= V_{L,max}(\frac{1}{2} + \frac{1}{2N}). \end{aligned} \tag{2.3}$$

For a cascode or two-stack amplifier, (2.3) suggests that there is a 50% increase in voltage dropped across the common gate device at the higher load impdeance, but for a four-stack, attempting to deliver the same voltage and power as two two-stacks with series power combining, the voltage dropped across the top device of the stack is increased by 150% at the higher load impedance, significantly impacting device reliability. As indicated by the simulated waveforms in [9], the C_{gd}, and C_{ds} of the devices in the stack can help to distribute the voltage equally across the devices of the stack. But this does come at the cost of reduced impedance looking into the drain of the device, limiting the efficacy of the Doherty principle, as it is based on seeing a high impedance looking into the drain of each amplifier.

The series DPA also has a concerning issue, namely the load is driven from both terminals, but in a non-balanced fashion. If the series DPA were to directly drive a dipole antenna, the non-balanced driving would lead to asymmetric illumination of the antenna, altering the input resistance seen by the DPA as well as the radiation pattern. A balun can be used to convert the dual-port output to a single-ended output but introduces additional undesirable loss in the output network.

In summary, the series DPA operates on the same principle of the canonical DPA, with the

Figure 2.5: Transformation of the series DPA into a single-ended series DPA with the main and peaking amplifiers stacked.

peaking amplifier providing a means to control the drain voltage of the main amplifier independently of the main amplifier current. One of the reasons Doherty avoided use of the series DPA is that it suffer from a dual-port output, and practical implementation would require a balun in the output netowrk of the DPA [16]. At the same time, the series voltage combining of the series DPA architecture provides tremendous benefit in the pursuit of high output power from a CMOS PA without significantly sacrificing device reliability. Because of the reliability and power benefits, this paper addresses the issue of asymmetric load drive by a rearrangement of the elements in the series DPA, as will be presented in the next section.

2.3 Stacked Series Doherty PA

As was mentioned, the series DPA suffers from its two-port load and non-balanced driving of the load. In this section it will be shown that the elements of the series DPA can be rearranged, transforming the series DPA to have a single-ended output while maintaining the benefits of the series combining of the main and peaking amplifiers. The rearranged circuit has a floating main amplifier leading to somewhat different waveforms than those of a typical PA. Additionally, special attention is required in designing the input and output networks.

Conceptually, the conversion of the series DPA to a single-ended output is a simple application of Kichhoff's Current Law. It only requires that the position of the load and main amplifier are switched. As the two elements are in series, they both carry the same current. By switching their positions, both their currents and hence voltages are unchanged. Since the same voltage and

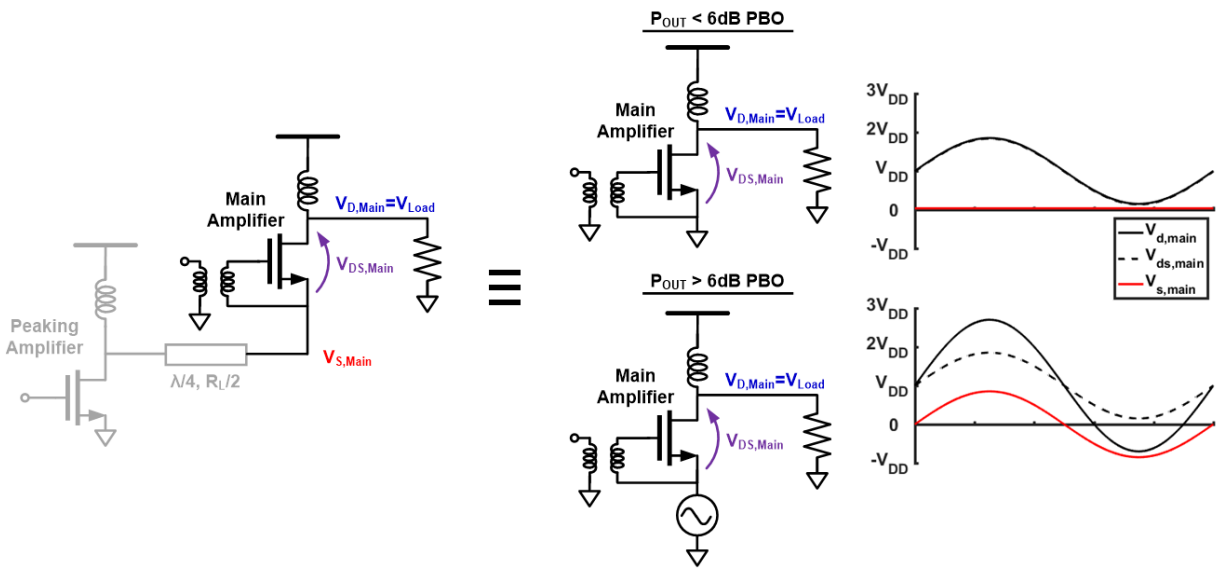


Figure 2.6: Series DPA with equivalent schematics of the low- and high-power modes of operation. Along with voltage waveforms showing the main amplifier remains safe and in saturated operation.

current appear at the transmission line, the peaking amplifier cannot differentiate between the two arrangements and it's operation is unaffected. Given that all element voltages and currents are unchanged, theoretically the circuit operation is unchanged. But from the standpoint of practical implementation, the two arrangements are quite different. In particular, the main amplifier in the modified arrangement floats, and the peaking amplifier drives the source of the main amplifier through the impedance inverter. As will be explained, this means the amplifier can be viewed as stacking the voltages of the two amplifiers, hence the terming stacked-series DPA.

The ac operation of the stacked-series DPA can be understood by considering the operation before and after the main amplifier reaches its peak voltage swing. For low power, before the main amplifier reaches peak voltage swing, the peaking amplifier remains off, with a high output impedance. The high impedance is transformed to a low impedance, close to 0 Ω by the impedance transformer, and the source of the main amplifier sees an ac short as shown in Fig. 2.6. This allows the main amplifier to operate like a standard linear-mode PA driving the load resistance.

Once the voltage swing of the main amplifier approaches saturation, the peaking amplifier begins to provide current with a phase of $+90^{\circ}$ relative to the main amplifier. Since the main and

peaking amplifier currents are correlated, to understand the impedance and voltage seen at the source of the main amplifier, the two amplifiers must be considered together. Assuming the main amplifier is an ideal current source, the resulting waveform at the source is

$$v_{s,main} = i_{d,peak} Z_0. \tag{2.4}$$

This voltage is independent of the current of the main amplifier, meaning we can model the peaking amplifier and impedance inverter as a voltage at the source of the peaking amplifier as shown in Fig. 2.6. This voltage swing at the source of the main amplifier combines the voltage contribution of the two amplifiers is a configuration similar to a stacked power amplifier, hence why this topology is referred to as a stacked-series DPA. For valid operation of the main amplifier, $V_{ds,main} < 2V_{DD}$, so if $v_{d,main} = i_{d,main} R_L$, then $v_{s,main} = max(0, v_{d,main} - 2V_{DD})$ and $i_{d,peak} = max(0, \frac{i_{d,main} R_L - 2V_{DD}}{Z_0})$ to maintain the main amplifier operating as a linear transconductor. In the present design, the standard DPA backoff, with peaking at 6 dB PBO, is targeted. Therefore, at peak power the main amplifier current and power should double from their values at 6 dB backoff. If $v_{ds,main}$ is maintained constant, the effective impedance seen by the main amplifier must drop by a factor of two. The peaking amplifier should also provide the same power as the main amplifier at peak power. If both amplifiers are operated from the same supply voltage, this means they both should provide the same current and see the same effective load impedance. Since $v_{s,main} = v_{ds,main}$ at peak power, $v_{ds,peak} = v_{s,main}$, meaning the impedance inverter does not perform impedance inversion at peak power, or $Z_0 = R_L/2$ as originally stated. By similar analysis to that used for the voltage at the source of the main amplifier, the voltage at the drain of the peaking amplifier is only dependent on the main amplifier current. As long as $i_{d,main} Z_0 \leq 2(V_D D - V_{knee})$, the peaking amplifier will remain in saturation.

While the description of the stacked-series DPA to this point has covered the RF operation of the output network of the amplifier, it has neglected the practical issues of DC biasing as well as the need for a transformer coupled input to the main amplifier. These two points are related, as they are both consequences of the choice to have a floating main amplifier rather than a balun in the output network.

The balun is needed on the input of the main amplifier to maintain the desired V_{gs} independent of the output power of the amplifier. At low power, prior to the peaking amplifier turn-on, the source voltage of the main amplifier is held close to constant, with only small variation due to resistive losses to ground. In this scenario, the input balun simply transformer-couples a single-ended input to the gate of the main amplifier. However, when the peaking amplifier begins to generate voltage swing at the source, if there were no input balun, the input to the main amplifier would only set V_g, not V_s. To properly set the current in the main amplifier, we must control V_{gs}. If instead the input of the main amplifier is transformer coupled, the intended input appears across the balun's secondary terminals, adding the desired V_{gs} in series with the V_s provided by the peaking amplifier.

The reason for accepting the complexity of a floating main amplifier is the improved efficiency from removing the loss of an output balun. If we consider the drain efficiency (DE) of the floating amplifier case as a baseline, the DE is

$$DE_{float} = \frac{P_{out,float}}{P_{DC}}, \tag{2.5}$$

while with an output balun there is additional loss factor

$$DE_{balun} = \frac{P_{out,float} IL}{P_{DC}}. \tag{2.6}$$

With a floating main amplifier, a balun is stll required, but on the input of the main amplifier. If we consider the PAEs of the two options, the input balun will have an impact due to the reduced gain of the amplifier

$$PAE_{float} = \frac{P_{out,float} - P_{in}}{P_{DC}} = \frac{P_{out,float} - P_{out,float}\frac{IL}{G}}{P_{DC}}, \tag{2.7}$$

$$PAE_{balun} = \frac{P_{out,float} IL - P_{out,float}\frac{IL}{G}}{P_{DC}}. \tag{2.8}$$

The additional loss term in (2.8) significantly impacts the efficiency as shown in Fig. 2.7.

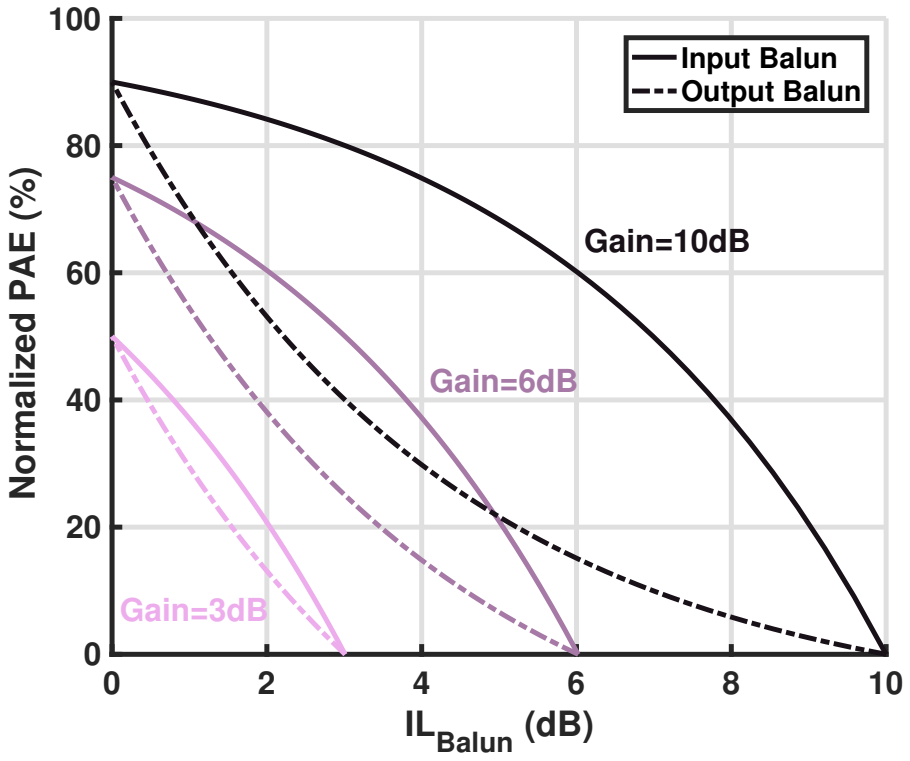


Figure 2.7: PAE as a function of balun loss, comparing the cases with balun on the input and output.

For example, with a gain of 9 dB before balun loss, if an output balun is used and introduces 0.5 dB insertion loss, relatively low, the same PAE could be achieved with a floating amplifier and input balun with 2.7 dB loss. This would allow for a more compact lossy balun, and/or higher PAE, if the floating main amplifier architecture is used.

Design of the input balun itself is straightforward. The gate capacitance of the main amplifier is resonated with the magnetizing inductance of the balun, and a series capacitor is added on the input to resonate the leakage inductance. The one challenge is the DC biasing of the gate of the main amplifier. Since the input balun is tied between the gate and source of the main amplifier common source device, a dc blocking capacitor must be added in series with the secondary winding of the balun. Ideally this capacitor should be made as large as is practical, with area, shunt parasitic capacitance to the substrate, quality factor, and self-resonance frequency limiting the size.

The RF operation of the stacked series DPA to this point has assumed that the source

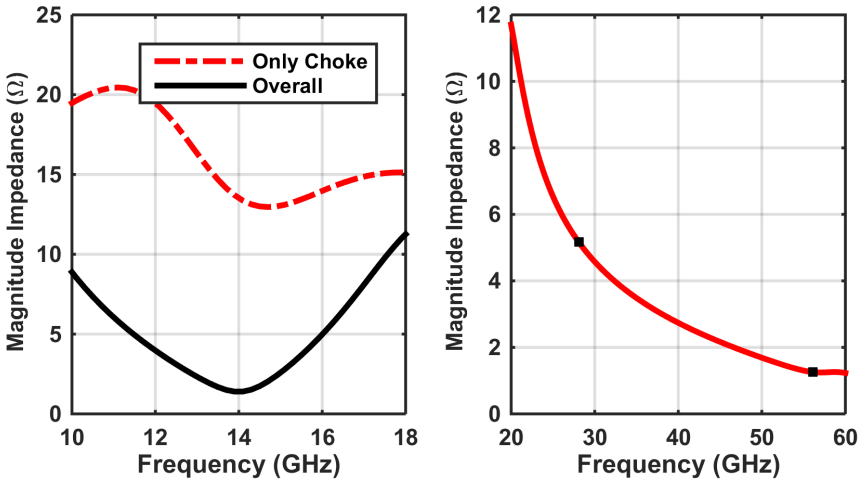


Figure 2.8

current of the main amplifier is sunk by the virtual ground provided by the impedance inverter. Unfortunately the dc current of the linear main amplifier cannot use the impedance inverter as a path to ground. This is because a dc block is inserted at the drain of the peaking amplifier to allow the main and peaking amplifiers to be properly biased and both operate from the same single supply voltage. For this reason an RF choke from the source of the main amplifier to ground is introduced. The choke is resonated with shunt parasitic capacitance at the source of the main amplifier to present high impedance at the RF carrier frequency (Fig. 2.8). Likely the largest concern with the stacked series DPA architecture is the increased possibility for instability introduced by the source choke. While the impedance inverter maintains a low resistance at the RF center frequency, off resonance the impedance seen at the source has a considerable reactive part - capacitive below resonance and inductive above. In combination with the input balun, if the quality factor of these networks is too high, instability can occur, as the main amplifier resembles a Colpitts or Hartley oscillator. While low quality factor typically implies loss and reduced efficiency, in this design, introducing loss with the main amplifier input match and balun can ensure stability without a significant impact on efficiency.

Finally, the implementation in this paper was at 14 GHz. As a quarter wave impedance inverter implemented with a transmission line is impractical on chip at this frequency, a C-L-C

Π-network is used to implement the impedance inverter. A future implementation could use a high-pass L-C-L Π-network which would simplify the matching network and further reduce component count, as the shunt inductors can be implemented as part of the peaking amplifier drain and main amplifier source chokes, and the series capacitor already exists for DC blocking between the main and peaking amplifiers. Stability would again need to be carefully considered, as the high-pass L-C-L Π-network has a somewhat different impedance over frequency.

2.4 Dual- and Single-Input DPAs

Traditionally, the DPA is implemented as a single-input structure, where a power splitter, quarter-wave delay line on the main amplifier path (for series DPA), and appropriate main and peaking amplifier transconductance versus output power for efficiency peaking at backoff are implemented as part of the design. To allow for flexibility in configuration and testing of this first prototype of the SSDPA structure, the inputs to the main and peaking amplifiers were separated. While the two inputs of this prototype require separate input signals for the main and peaking amplifiers, this provides additional freedom in that the input amplitudes and phases can be controlled independently, with previous demonstrations of a linear dual-input DPA at VHF [17], and digital DPAs at low GHz frequencies [18, 19]. Dual vector modulators for use with dual-input outphasing [20] and Doherty PAs [10, 21] have been demonstrated, supporting the dual-input DPA as a viable approach even at higher microwave and mm-wave frequencies. At the same time, one of the goals of this work is to show that given appropriate splitter and drivers, this prototype could be made into a single-input DPA. As a compromise between completely independent vector inputs and a single input (with approximately 90° phase shift at the main amplifier input), the input amplitudes are allowed to be somewhat independent, while their relative phase is held constant regardless of input amplitude. This is a reasonable tradeoff since achieving different absolute gain and gain profile across power is feasible with properly designed predrivers.

To determine the relative input amplitudes, the desired output currents are a good starting

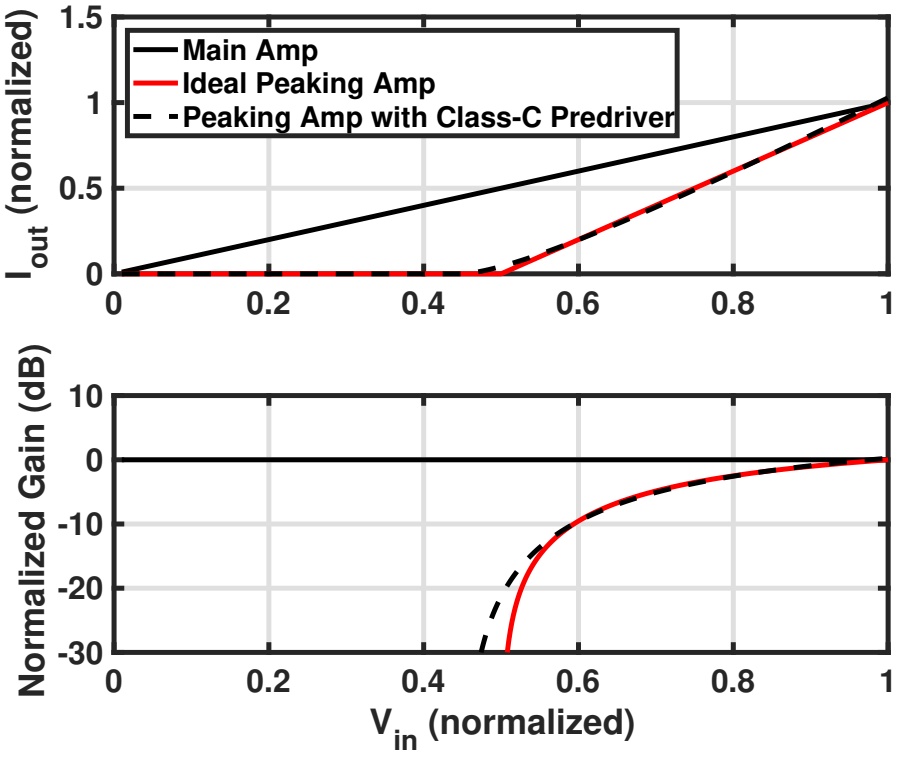


Figure 2.9: Comparison on the desired output current and gain of the peaking amplifier with those provided by an ideal class-C predriver. The main amplifier desired characteristics are shown for reference.

place. If we attempt to replicate the traditional DPA profile, and if the devices of both the main and peaking amplifiers are assumed to be reasonably linear transconductors, both of the same size, then the input voltages should approximately reflect the desired output currents. To achieve this, the main amplifier input will vary linearly with output current, and the peaking amplifier input will be zero at low power levels and then increases linearly with higher slope than the main amplifier above some threshold, as shown in Fig. 2.3, replacing I_{out} with $g_m V_{in}$. With this configuration, and assuming independent predrivers for the main and peaking amplifiers, the required gain of the two predrivers, assuming an identical input driving both, can be determined (Fig. 2.9). As seen, the main amplifier predriver gain is that of a standard class-A/AB amplifier, while the peaking amplifier predriver has gain similar to a class-C amplifier.

Classically, a similar gain profile can be implemented by a class-C bias and doubling the

size both on the peaking amplifier device [22]. This has a few undesirable consequences, the most significant being the reduced gain and gain variation of the final stage due to the class-C bias. By moving the required gain variation introduced by the class-C amplifier to the predriver, the peaking amplifier can still have delayed turn-on with a class-B/AB bias. Pushing the class-C bias to the predriver and getting more gain from the output stage of the peaking amplifier path can also lead to improved PAE. This also allows the peaking amplifier device to be the same size as the main amplifier, reducing the capacitive loading at the drain node and increasing the output impedance of the peaking amplifier, making it look more like the desired ideal current source. Another important problem is the very low bias required on the input of the class-C peaking amplifier to maintain the conduction angle near zero until approximately 6 dB PBO, challenging with the large input seen by the PA only 6 dB from peak power. If the class-C bias is moved to the predriver, the swing at the input of the predriver will be significantly smaller, making the required class-C gate bias voltage at the predriver input more practical.

2.5 Measurement Setup and Results

The stacked-series DPA was implemented in a 45 nm SOI CMOS process (Fig. 2.10). The schematic of the implemented PA is shown in Fig. 2.11. PA characteristics were measured including continuous wave (CW) and modulated power, efficiency, and linearity. The dual-input PA requires coordination between the two inputs. In particular, phase coherence and relative amplitude control must be maintained for proper operation.

The relative phase between the two inputs is set applying large, near P_{sat}, inputs to both PAs. The relative phase is swept and CW output power versus phase data is collected. The relative phase is held constant at the value that gives the highest output power. The exact value of the relative phase is not meaningful, as the cable lengths are not perfectly matched. It should also be noted that the efficiency and output power is not terribly sensitive to the exact relative phase. Offsets of $\pm 5^{\circ}$ can still provide good output power and efficiency.

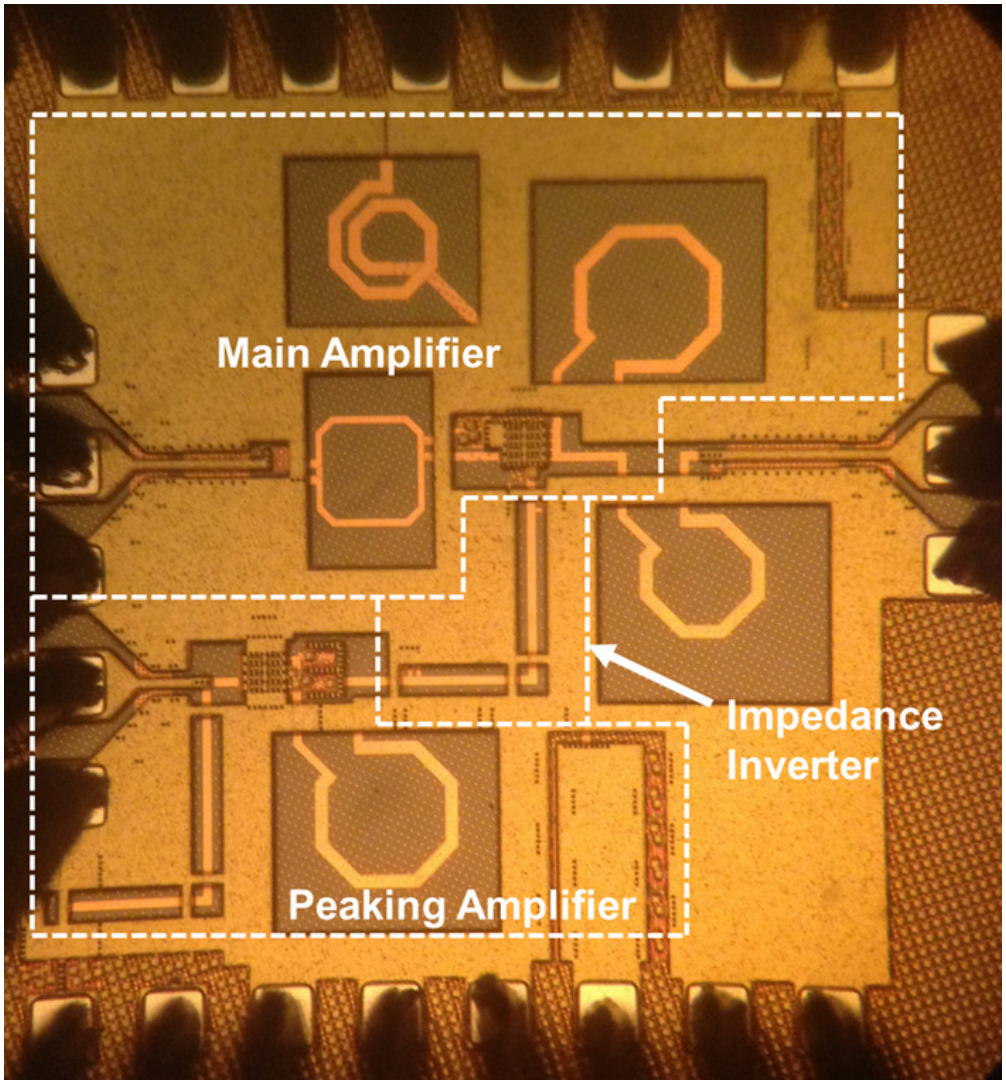
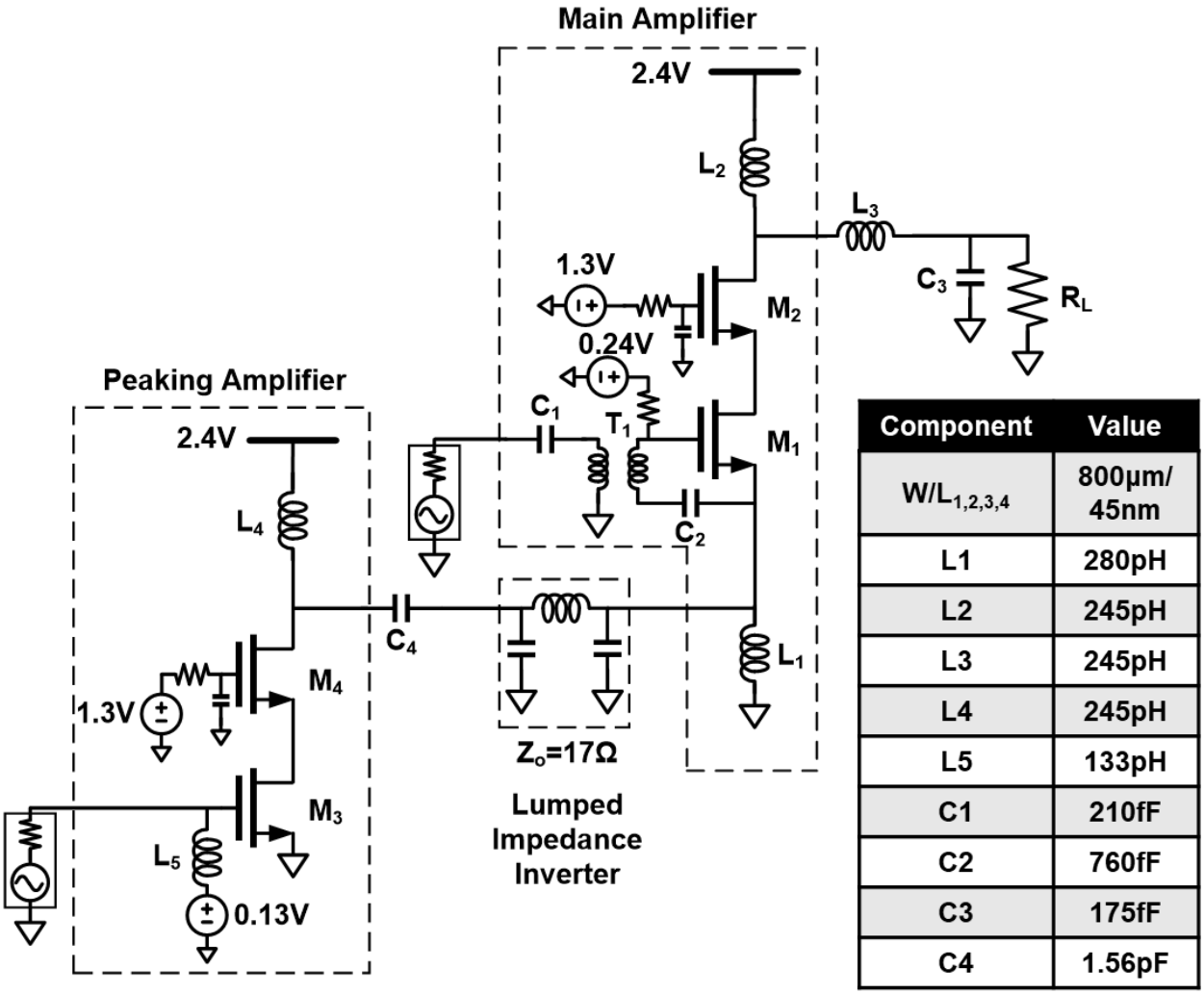


Figure 2.10: Photomicrograph of the SSDPA prototype, fabricated in a 45 nm SOI CMOS process.



Component	Value
$W/L_{1,2,3,4}$	800µm/ 45nm
L1	280pH
L2	245pH
L3	245pH
L4	245pH
L5	133pH
C1	210fF
C2	760fF
C3	175fF
C4	1.56pF

Figure 2.11: Schematic of the implemented SSDPA.

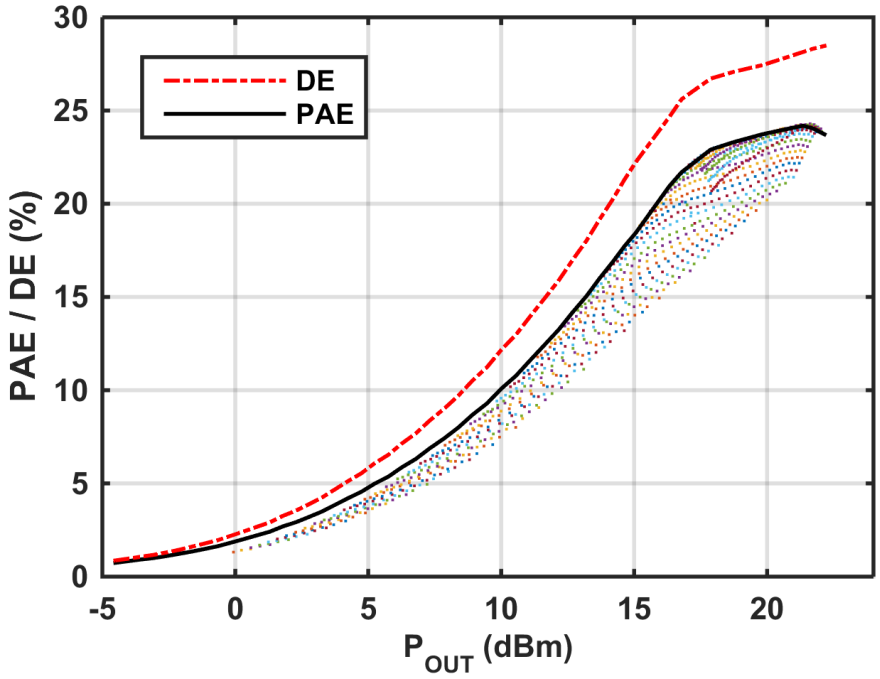


Figure 2.12: PAE and DE for swept combinations of main and peaking amplifier input amplitude. The solid black and red curves represent PAE and DE for input voltages on the ideal DPA input loci, and give roughly the best efficiency of any input, validating Doherty operation.

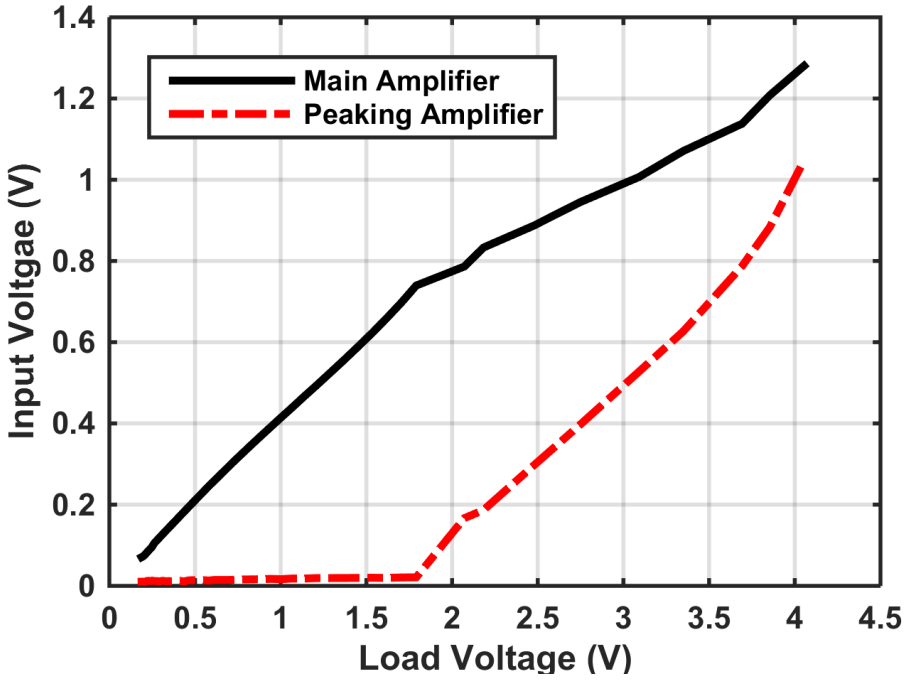


Figure 2.13: Input amplitudes of the main and peaking amplifiers from the swept dataset of fig. 2.12 most representative of the desired ideal DPA inputs. These inputs give the efficiency versus output power characteristic shown by the black and red lines in fig. 2.12.

To determine the maximum efficiency achievable with the amplifier, CW efficiency versus output power data is collected with independently swept amplitudes for both inputs (Fig. 2.12). As can be seen, there is a locus of input power combinations that provides significant efficiency enhancement at PBO. Among these combinations, input amplitudes representative of the inputs described in the previous paragraph are chosen, as shown in Fig. 2.13. With these chosen inputs, the solid black line in Fig. 2.12 is achieved, showing that the devices are nearly linear transconductors, and inputs representative of the ideal DPA main and peaking output currents are desirable. With these inputs, a peak DE and PAE of 28% and 24%, and saturated output power of 22 dBm is achieved. At 6 dB PBO, the DE and PAE are maintained at 23% and 20% respectively, showing significant improvement over class-B PBO. It is also of interest to understand the gain profile of the amplifier. To do this, the input power is assumed to be the sum of the two input powers, $P_{in,main} + P_{in,peak}$. Using the inputs corresponding to the high-efficiency, ideal input loci, the gain is calculated to be as shown in Fig. 2.14. There is less than 1.25 dB gain ripple up to the saturated output power, showing that the main amplifier is not compressed significantly to achieve good efficiency at PBO.

Wideband modulated data was also collected with the stacked-series DPA. For modulation, the relative phase between the two inputs is still held constant and the amplitudes are set as described for the CW high-efficiency locus. Two channels of a Keysight 8195A arbitrary waveform generator are used to directly generate the modulated data at 14 GHz. While the input amplitude control could be considered a form of digital predistortion, it is memoryless and extremely simple to implement, with one possible implementation, requiring only an n-bit multiplexer, shown in Fig. 2.15. Assuming the main amplifier amplitude is available as a digital word, the MSB determines whether or not a signal is sent to the peaking amplifier. For large enough inputs, the actual word sent to the peaking amplifier is simply a bit-shifted version of the main amplifier input. Note, while the peaking amplifier does have one less bit of resolution due to the LSB always being zero, this is not of concern, as the main amplifier is responsible for setting the current, and hence output power, and small variations in the output of the peaking amplifier alone have negligible impact on the DPA

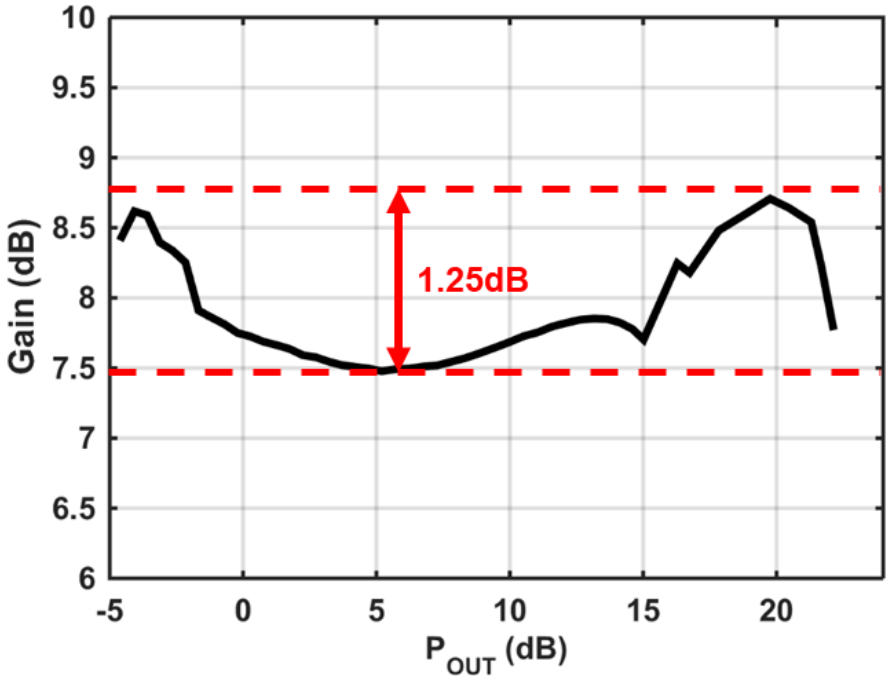


Figure 2.14: Measured gain of the SSDPA with the inputs shown in fig. 2.13.

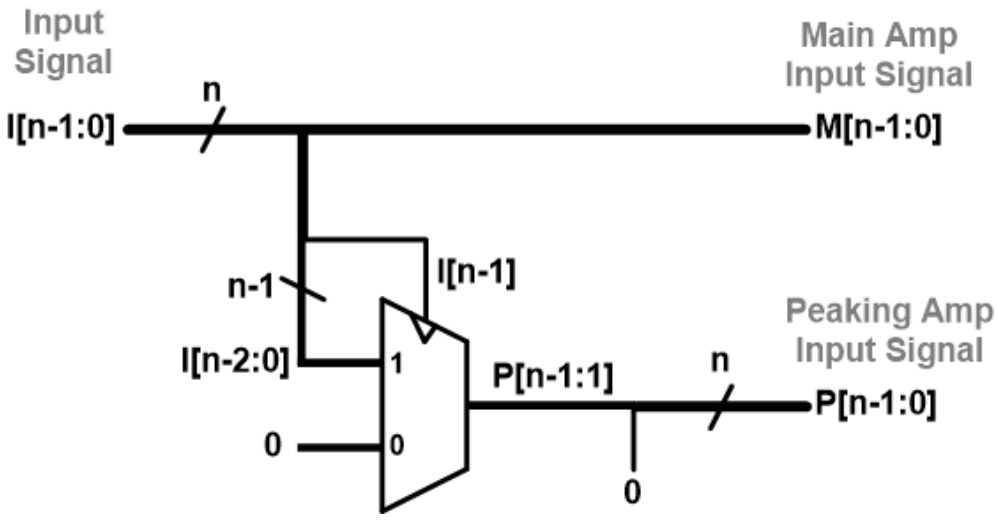


Figure 2.15: Logic to implement the quasi-DPD required to generate the peaking amplifier amplitude input signal for a dual-input DPA.

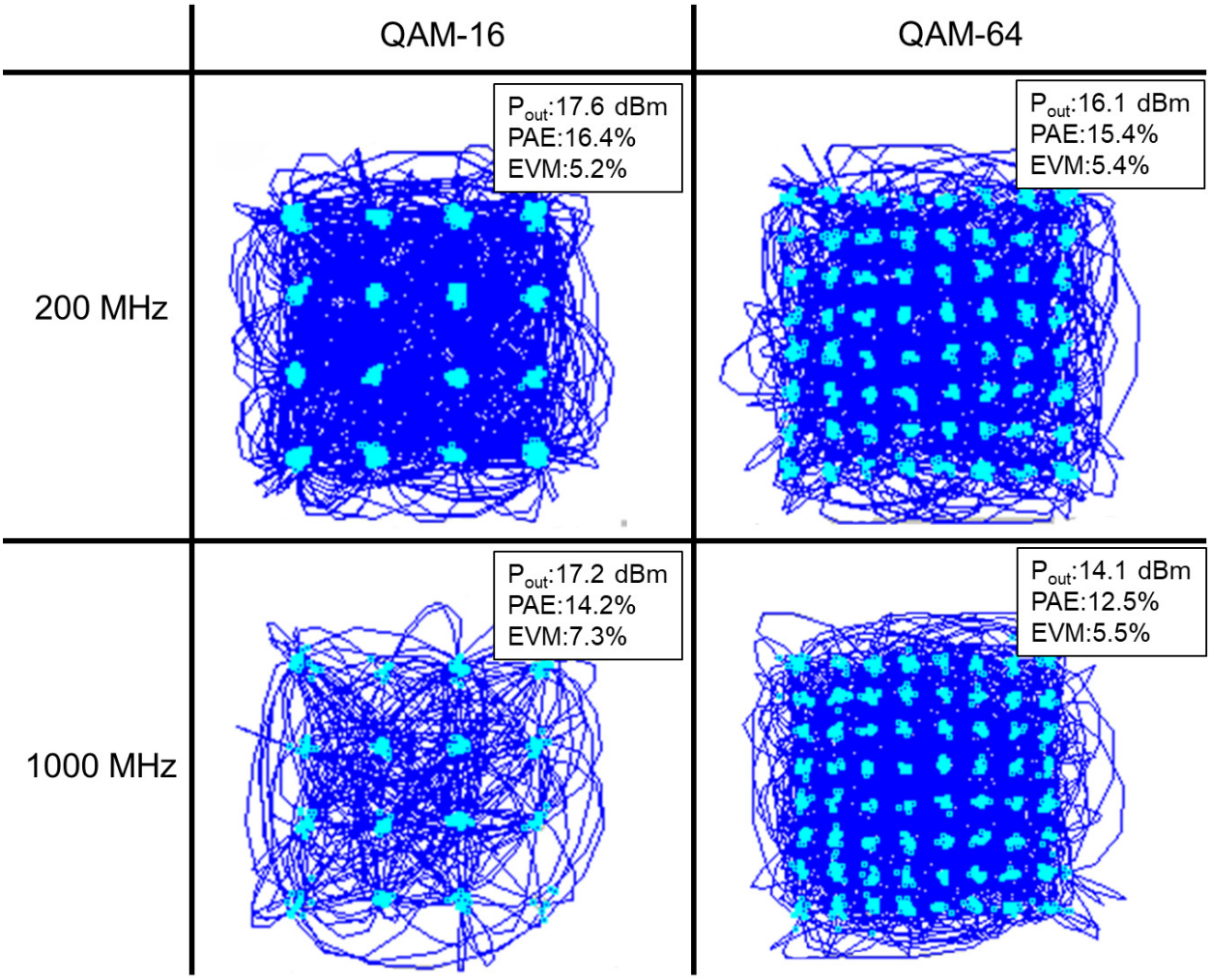


Figure 2.16: QAM-16/-64 constellations at 200/1000 MHz bandwidths.

output.

A variety of single-carrier QAM-16 and QAM-64 measurements are collected at 200 MHz and 1 GHz bandwidths (Fig. 2.16). In QAM-16/-64 measurments, EVM is maintained below -22 dB / -25 dB (8% / 5.6%) respectively. For the narrower bandwidth 200 MHz signals, output power and PAE of 17.6/16.1 dBm and 16.4/15.4% are achieved for QAM-16/-64 respectively. At the wider 1 GHz bandwidth, output power and PAE of 17.2/14.1 dBm and 14.2/12.5% are achieved for QAM-16/-64 respectively, with the sizeable reduction in efficiency and output power for QAM-64 likely due to the memory from the amplifier bandwidth. While the peak power and efficiency in this work is somewhat lower than more recent results [9], the high backoff efficiency and linearity still

give competitive modulated efficiency and output power results compared to other silicon-based DPAs at this frequency.

2.6 Conclusion

This paper has presented a method for implementing a series Doherty power amplifier as a single-ended PA. The series architecture acts as a voltage combiner making it attractive for implementation in CMOS where low breakdown voltages often limit the achievable output power of PAs. It is also particularly amenable to implementation at mm-wave, where passives can be integrated in the semiconductor process back-end-of-line with a reasonable footprint, important for implementing the input transformer and main amplifier source RF choke. Assuming there will be minimal drop in achievable output power when scaling to either 28 GHz or 39 GHz [12], the bands that have been chosen for 5G cellular applications in the time since this work was originally designed, the output power and achievable efficiency presented make a compelling argument for use of the SSDPA at mm-wave frequencies.

Acknowledgment

The authors would like to thank GlobalFoundries for fabrication, and Integrand Software for the use of EMX. They would also like to thank B. Rabet, N. Rostomyan, and P. Asbeck for useful discussion and suggestions with the modulated data measurements.

This chapter, in part, is a reprint of the material as it appears in the Proceedings of the Compound Semiconductor Integrated Circuit Symposium, 2016, Levy, Cooper; Vorapipat, Voravit; Buckwalter, James. The expanded presentation in this dissertation may appear as another publication in the future with the additional contributor Mercier, Patrick. The dissertation author was the primary investigator and author of this material.

Bibliography

[1] W. H. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proceedings of the Institute of Radio Engineers*, vol. 24, pp. 1163–1182, Sept. 1936.

[2] W. H. Doherty and O. W. Towner, "A 50-Kilowatt Broadcast Station Utilizing the Doherty Amplifier and Designed for Expansion to 500 Kilowatts," *Proceedings of the IRE*, vol. 27, pp. 531–534, Sept. 1939.

[3] J. Staudinger, "Doherty history, principles and key developments," in *2016 IEEE MTT-S International Microwave Symposium (IMS)*, pp. 1–4, May 2016.

[4] R. Pengelly, C. Fager, and M. Ozen, "Doherty's Legacy: A History of the Doherty Power Amplifier from 1936 to the Present Day," *IEEE Microwave Magazine*, vol. 17, pp. 41–58, Feb. 2016.

[5] A. Grebennikov and S. Bulja, "High-Efficiency Doherty Power Amplifiers: Historical Aspect and Modern Trends," *Proceedings of the IEEE*, vol. 100, pp. 3190–3219, Dec. 2012.

[6] "Specification # 38.211."

[7] J. Curtis, A. Pham, M. Chirala, F. Aryanfar, and Z. Pi, "A Ka-Band doherty power amplifier with 25.1 dBm output power, 38% peak PAE and 27% back-off PAE," in *2013 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 349–352, June 2013.

[8] D. P. Nguyen, B. L. Pham, and A. Pham, "A compact 29% PAE at 6 dB power back-off E-mode GaAs pHEMT MMIC Doherty power amplifier at Ka-band," in *2017 IEEE MTT-S International Microwave Symposium (IMS)*, pp. 1683–1686, June 2017.

[9] N. Rostomyan, J. A. Jayamon, and P. M. Asbeck, "15 GHz Doherty Power Amplifier With RF Predistortion Linearizer in CMOS SOI," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, pp. 1339–1348, Mar. 2018.

[10] K. Greene, A. Sarkar, and B. Floyd, "A 60-GHz Dual-Vector Doherty Beamformer," *IEEE Journal of Solid-State Circuits*, vol. 52, pp. 1373–1387, May 2017.

[11] S. Hu, F. Wang, and H. Wang, "2.1 A 28ghz/37ghz/39ghz multiband linear Doherty power amplifier for 5g massive MIMO applications," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 32–33, Feb. 2017.

[12] N. Rostomyan, M. zen, and P. Asbeck, "28 GHz Doherty Power Amplifier in CMOS SOI With 28% Back-Off PAE," *IEEE Microwave and Wireless Components Letters*, vol. 28, pp. 446–448, May 2018.

[13] A. Agah, H. Dabag, B. Hanafi, P. M. Asbeck, J. F. Buckwalter, and L. E. Larson, "Active Millimeter-Wave Phase-Shift Doherty Power Amplifier in 45-nm SOI CMOS," *IEEE Journal of Solid-State Circuits*, vol. 48, pp. 2338–2350, Oct. 2013.

[14] J. Horwitz, "MediaTeks prototype 5g phone has a fan, but final design supposedly wont," Sept. 2018.

[15] C. S. Levy, V. Vorapipat, and J. F. Buckwalter, "A 14-GHz, 22-dBm Series Doherty Power Amplifier in 45-nm CMOS SOI," in *2016 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, pp. 1–4, Oct. 2016.

[16] Y. Cho, K. Moon, B. Park, J. Kim, and B. Kim, "Voltage-Combined CMOS Doherty Power Amplifier Based on Transformer," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, pp. 3612–3622, Nov. 2016.

[17] R. Beltran and F. H. Raab, "VHF Doherty amplifier with GaN FETs and independent drive-signal control," in *2011 IEEE MTT-S International Microwave Symposium*, pp. 1–4, June 2011.

[18] V. Vorapipat, C. Levy, and P. Asbeck, "A wideband voltage mode Doherty power amplifier," in *2016 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 266–269, May 2016.

[19] S. Hu, S. Kousai, J. S. Park, O. L. Chlieh, and H. Wang, "A +27.3dbm transformer-based digital Doherty polar power amplifier fully integrated in bulk CMOS," in *2014 IEEE Radio Frequency Integrated Circuits Symposium*, pp. 235–238, June 2014.

[20] M. S. Mehrjoo, S. Zihir, G. M. Rebeiz, and J. F. Buckwalter, "A 1.1-Gbit/s 10-GHz Outphasing Modulator With 23-dBm Output Power and 60-dB Dynamic Range in 45-nm CMOS SOI," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, pp. 2289–2300, July 2015.

[21] K. Greene and B. Floyd, "Dual-vector phase rotator for Doherty beamformers," in *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 331–334, May 2015.

[22] S. Cripps, *RF Power Amplifiers for Wireless Communications, Second Edition*. RF Power Amplifiers for Wireless Communications, Second Edition, 2006.

Chapter 3

Bell-Bloom Magnetometer Linearization by Intensity Modulation Cancellation

3.1 Introduction

Atomic magnetometers have found use in a variety of magnetometry applications [1–10]. In particular, atomic magnetometers prove useful for making high-sensitivity measurements in non-zero ambient fields, such as the geomagnetic field of Earth where superconducting quantum interference devices (SQUIDs) will not operate. In everyday life, many settings such as urban environments and office spaces experience significant environmental noise from AC electrical power distribution, as well as low frequency mechanical movements, and other possible sources that are not easily identified as shown in Fig. 3.1, [1, 11]. Oftentimes in such scenarios, two magnetometers can be deployed in a gradiometer configuration to make background noise a common-mode signal, thereby facilitating measurement of the local spatial variation in the magnetic field. If the environmental noise is too large, the linearity of an atomic magnetometer can limit the usefulness of even a gradiometer.

If scalar measurements suffice, M_x and M_z atomic magnetometers [12, 13], which measure magnetic fields by sensing atomic spin procession of an optically-probed, radio frequency (RF)

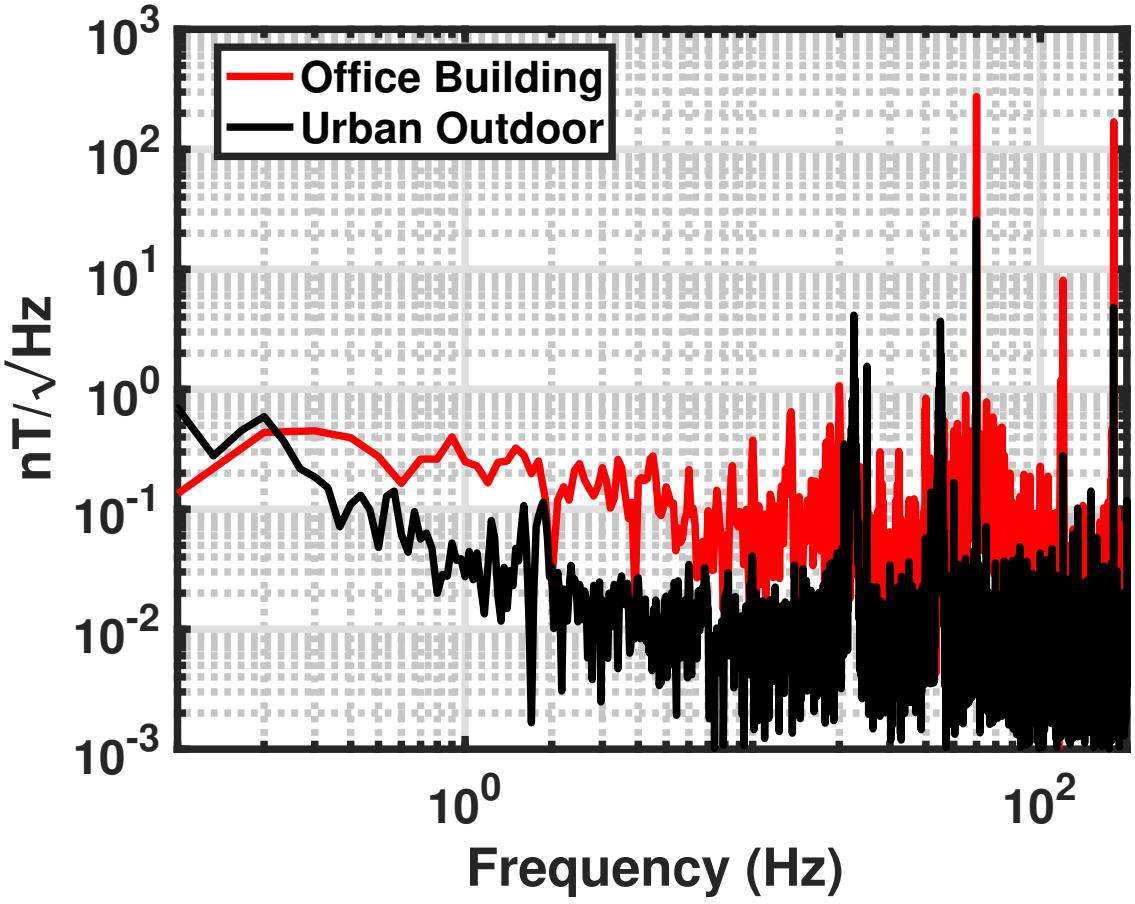


Figure 3.1: Spectra of the magnetic field in typical office and outdoor urban environments. AC power distribution and harmonics are clearly visible, as well as other unattributed signals and noise.

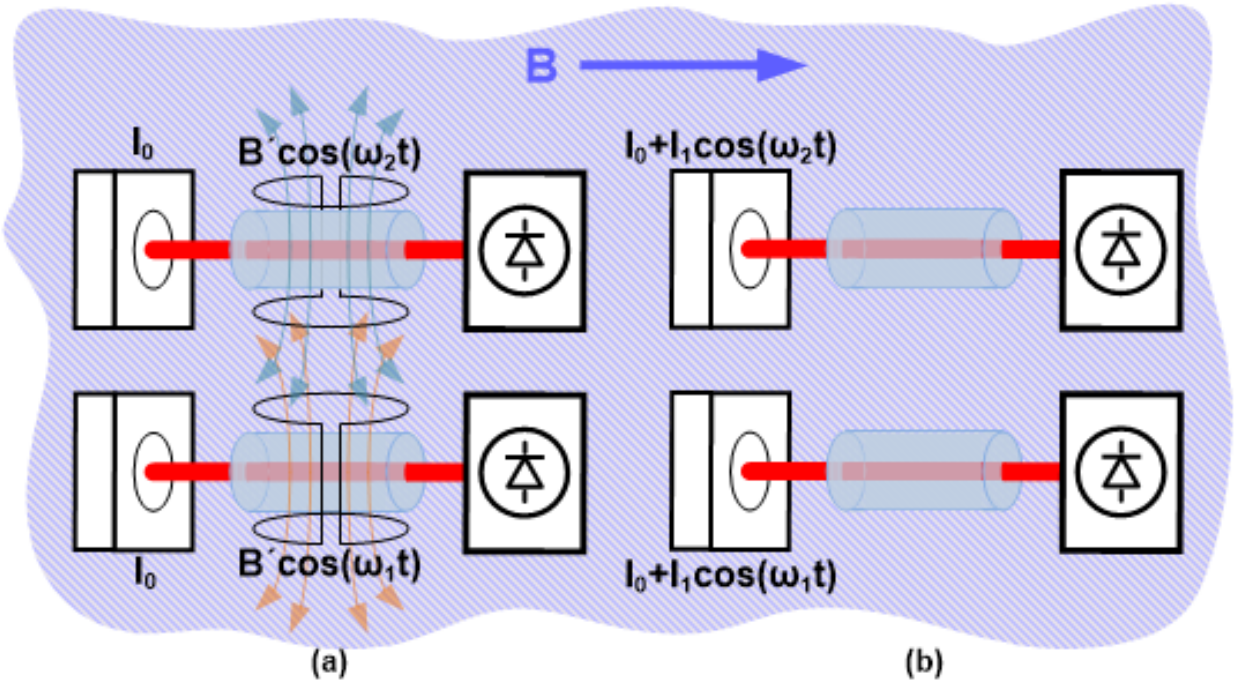


Figure 3.2: Comparison between the Mx and Bell-Bloom modes of operation. As seem in (a) the coils that modulate the field applied to the Mx magnetometer generate fields that can interfere with adjacent magnetometer. With Bell-Bloom operation, shown in (b), since the modulation is applied optically, there are no locally generated RF magnetic fields, eliminating crosstalk.

modulated vapor cell, are straightforward, and offer excellent sensitivity. While both M_x and M_z magnetometers can be made to work in gradiometers, coupling of the fields generated by the RF coils of each magnetometer, as depicted in Fig. 3.2(a), can lead to sensitivity and linearity degradation for adjacent magnetometers. Alternatively, the Bell-Bloom magnetometer [14], depicted in Fig. 3.2(b), does not suffer from this interference, as the modulation is applied to the pumping light source rather than an RF field, and is thus more appropriate for gradiometer applications. The Bell-Bloom architecture is also more robust in terms of orientation, having only a dead axis, compared to a dead plane and dead axis in M_x. However, the Bell-Bloom architecture does have its own issues, such as reduced sensitivity and linearity due to the modulated pumping rate and laser power. If these issues are addressed, there is a compelling case for the Bell-Bloom magnetometer.

For applications outside of the laboratory environment, reasonably low power consumption, compact size, and complete integration of laser, cell, photodetector, and control electronics are

also important. The trend towards microfabricated atomic magnetometers has led to complete physics packages (vapor cell, lasers, photodiodes, heaters, and other miscellaneous components) with volumes well under one cubic centimeter. In these small form factors, simplifying the physics package aides in fabrication and robustness. The simplicity of the single pump/probe Bell-Bloom physics package, with only one laser and one photodiode, is ideal for a microfabricated atomic magnetometer. While M_x and M_z magnetometers have similar requirements to Bell-Bloom in terms of control, they require an additional RF coil, which is not always trivial to add given the myriad of other signals and biases already being routed in the small form factor.

Another important consideration outside of the laboratory environment is variations in temperature and average magnetic field intensity. Laboratory characterization can often be performed in an open-loop configuration [14–17], but for robustness to environmental variation, pragmatic magnetometers must be operated in a closed-loop fashion, such as that shown in Fig. 3.3, representative of a typical Bell-Bloom magnetometer operated in a closed-loop configuration. Here, the variation in quadrature amplitude or phase seen at the output of the demodulator is used as an error signal to control the driving frequency, similar to an FM demodulator.

This paper will focus on the reduced linearity of the Bell-Bloom magnetometer from modulation of the laser power. All atomic magnetometers suffer from poor linearity when large fields (variation in Larmor frequency comparable to the magnetic linewidth) are applied. Modulating the laser power, particularly in a single pump/probe beam Bell-Bloom magnetometer, degrades the linearity compared to an M_x magnetometer. To combat this, a technique is proposed to cancel the laser power modulation. Section II will first examine the response of the M_x magnetometer because of its more-ideal response and the simpler equations describing its operation, and then the analysis will be extended to the Bell-Bloom magnetometer, highlighting the issues it brings. In Section III, the proposed cancellation technique and the theoretical improvement it affords the Bell-Bloom magnetometer will be discussed. Section IV will explain the implementation of a developed prototype, followed by measurement results in Section V.

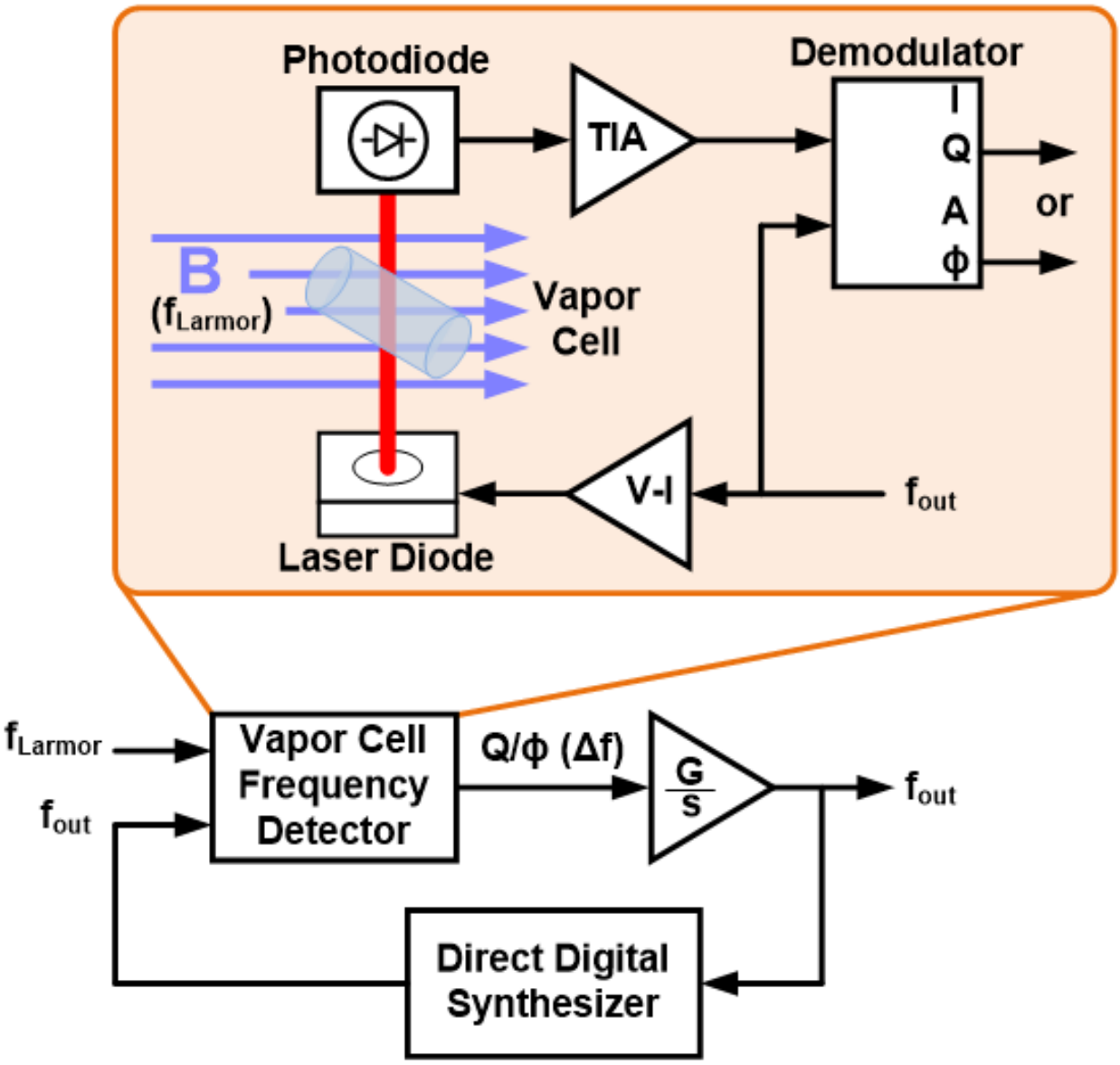


Figure 3.3: Bell-Bloom magnetometer operated in closed-loop. Here the phase of the output signal s_{out} is multiplied by a constant, integrated, and the resulting signal controls the input frequency f_{in}, similar to an FM demodulator.

3.2 M_x and Bell-Bloom Magnetometer Operation

3.2.1 M_x Mode

A common mode of operation for scalar atomic magnetometers is the so-called M_x mode. In this mode, a field is applied to an RF coil wound perpendicular to the pump/probe beam of the cell, while the laser beam is held at constant intensity and wavelength. By modulating the applied field at the Larmor frequency of the alkali metal vapor in the cell, the spin of the atoms in the cell can be coherently driven. The resulting photodetector signal intensity is:

$$I_{pd,Mx} = I_0 e^{-nL(\frac{1}{2}-\frac{1}{2}S_x)\sigma_0}, \tag{3.1}$$

where I_0 is the laser intensity, n is the gas density in the cell, L is the cell path length, S_x is the spin polarization, and σ_0 is the alkali atom absorption cross section.

Following the derivation of the spin polarization from the Bloch equation as performed in [18], the spin polarization with a sinusoidal field $B'\cos\omega t$ generated by the RF coil is:

$$S_x = \frac{1}{2} S_0 \gamma B' \frac{T\cos\omega t + (\omega-\omega_L)T^2\sin\omega t}{1+[(\gamma B'/2)^2+(\omega-\omega_L)^2]T^2}, \tag{3.2}$$

with equilibrium spin polarization in the absence of magnetic field S_0, gyromagnetic ratio γ for the alkali atoms, T the spin coherence lifetime, limited by spin exchange, and ω_L the Larmor frequency of the alkali atoms. As can be seen from (3.2), the spin has in-phase and quadrature components which vary with the difference in frequency between the Larmor frequency and that of the field generated by the RF coil. If the attenuation due to absorption is relatively small, the photodetector signal can be linearized:

$$I_{pd,Mx} \approx I_0(1 - nL(\tfrac{1}{2} - \tfrac{1}{2}S_x)\sigma_0). \tag{3.3}$$

The spin gives rise to a frequency-dependent amplitude and phase of the photodetector intensity (Fig. 3.4) with

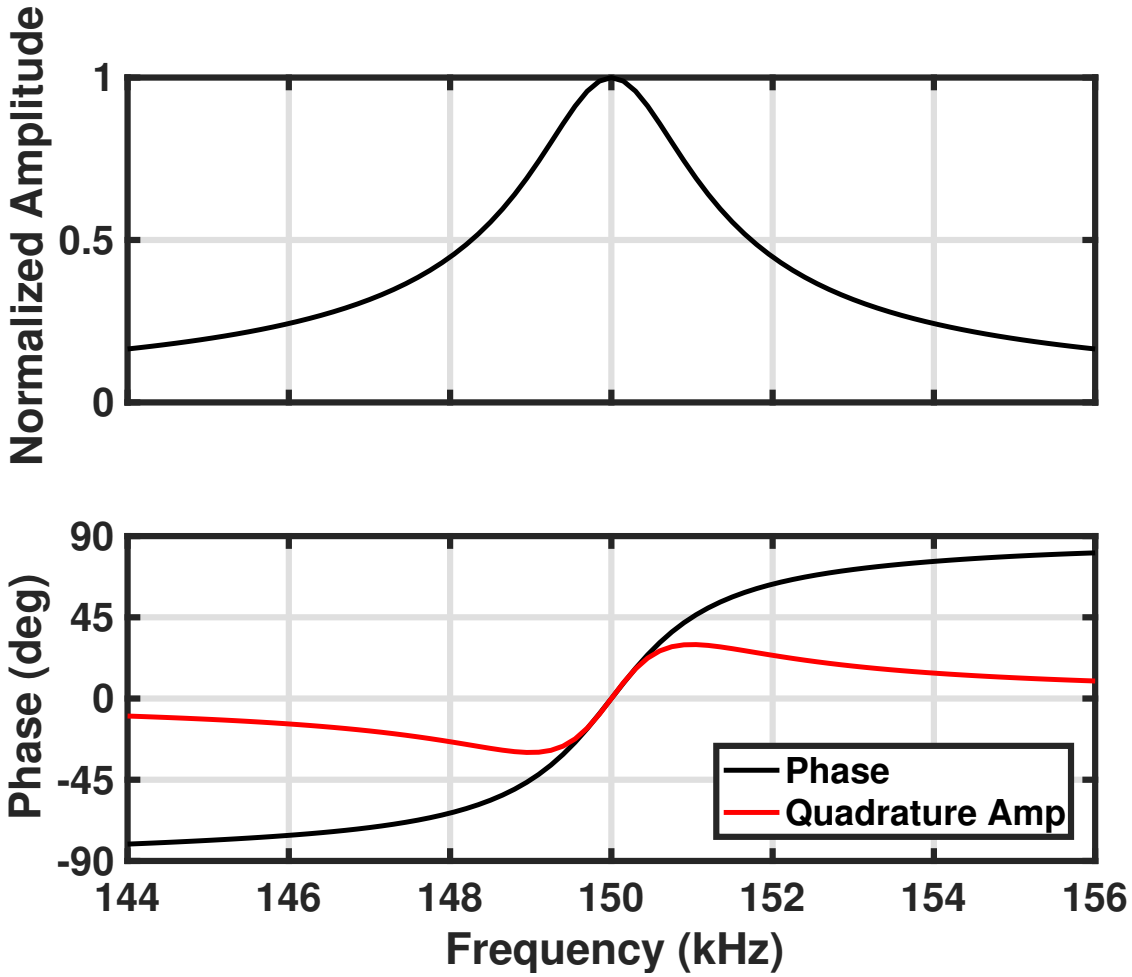


Figure 3.4: M_x magnetometer normalized amplitude, and phase and quadrature amplitude response. The quadrature amplitude response is scaled to match the slope of the phase response at the center of the dispersion.

$$\angle I_{pd,Mx} \approx \arctan\left(\left(\omega - \omega_L\right) T\right). \tag{3.4}$$

This allows the Larmor frequency to be determined by tracking the zero crossing in the dispersive phase response. Another common way to track this zero crossing is using the quadrature ($\sin \omega t$) signal amplitude, which is also dispersive and has the same zero crossing [1, 15, 17]. Using the quadrature amplitude requires only a single demodulator compared to the two demodulators and a coordinate rotation digital computer (CORDIC) required to detect phase. However, if the response of the phase and quadrature amplitude signals for the M_x magnetometer (Fig. 3.4) are compared, it becomes clear that the phase signal has wider bandwidth and a more linear response over frequency. It has also previously been noted [19] that from a control perspective, the phase signal leads to a more robust control system with the same sensitivity as the quadrature amplitude.

The higher amplitude of the phase signal at large frequency offsets translates to faster slewing and a wider range of frequency over which lock can be maintained if there is a sudden change in the magnetic field. Maximizing the loop gain, and bandwidth, while maintaining stable operation is the most straightforward method to improving the linearity of the magnetometer, but at some point the nonlinear response of the magnetometer's dispersive characteristic will limit the system linearity. When the dispersive characteristic impacts performance, the higher linearity of the phase response as compared to the quadrature amplitude leads to reduced harmonic distortion. This makes the magnetometer more robust and sensitive in the presence of large-amplitude time-varying fields such as 50 and 60 Hz power distribution.

3.2.2 Bell-Bloom Mode

The cell can also be operated as a Bell-Bloom magnetometer, where the pumping laser is modulated at a frequency near the Larmor frequency, and the corresponding amplitude and phase variation at the photodetector is measured. Separating the pump and probe beams can avoid some of the complications of a single pump/probe beam, but the need for two lasers makes this approach

less amenable to integration in a simple, small, low power form-factor.

In compact magnetometers with a cubic-millimeter-volume cell, such as the one used in this paper, a vertical cavity surface emitting laser (VCSEL) is used for the pump/probe beam [16]. Unfortunately, modulating the VCSEL leads to a change in the number of photons emitted and their wavelength. This means there is not only a time varying pumping rate, but also a time-varying absorption cross section. The response of the Bell-Bloom magnetometer can be described by:

$$I_{pd,BB} = I_i(t)e^{-nL(\frac{1}{2}-\frac{1}{2}S_x(\lambda(t),t))\sigma_0(\lambda(t))}, \tag{3.5}$$

where the laser output power and wavelength as a function of time are given by:

$$I_i(t) = I_0 + I_1\cos\omega t, \tag{3.6}$$

$$\lambda(t) = \lambda_0 + \lambda_1\cos\omega t. \tag{3.7}$$

The observed spin and absorption cross section are given by:

$$S_x \approx \frac{1}{4}RS_0\left[\frac{(\Delta\omega\cos\omega t + (\omega-\omega_L)\sin\omega t}{(\Delta\omega)^2+(\omega-\omega_L)^2}\right] \tag{3.8}$$

$$\sigma_0(\lambda(t)) = cr_e f_{\mathrm{D1}}[\mathcal{P}(D1a)\mathcal{L}_{D1a}(\lambda(t)) + \mathcal{P}(D1b)\mathcal{L}_{D1b}(\lambda(t))] \tag{3.9}$$

$$\mathcal{L}_{D1a}(\lambda) = \frac{\Gamma/2}{(\frac{c}{\lambda}+\frac{c}{\lambda_{D1a}})^2+(\frac{\Gamma}{2})^2}, \tag{3.10}$$

with R the pumping rate if the atoms are continuously pumped (no wavelength modulation), $\Delta\omega$ the magnetic linewidth, c the speed of light, r_e the classical electron radius, f_{D1} the fraction of the classical cross-section associated with the D1 resonance ($^2S_{\frac{1}{2}}$ to $^2P_{\frac{1}{2}}$ transition), $\mathcal{P}(D1a)$ and $\mathcal{P}(D1b)$ the relative strengths of the D1a and D1b hyperfine transitions, and λ_{D1a} and λ_{D1b} the wavelengths of photons with energy of the two hyperfine transitions.

Simulations of the magnetometer presented in this paper are based on (3.5)-(3.10), but to understand the approximate effect of each term on the observed response of the photodetector, it

is useful to linearize parts of the system and separate the terms. Assuming the magnitude of the argument of the exponential is much smaller than 1, a linear approximation can be made for the attenuation due to the near-resonant absorption of the alkali vapor:

$$I_{pd,BB}(t) \approx [I_0 + I_1 \cos \omega t][1 - nL(\tfrac{1}{2} - \tfrac{1}{2}S_x)\sigma_0(\lambda)] \tag{3.11}$$

$$I_{pd,BB}(t) \approx I_0 + I_1 \cos \omega t + (I_0 + I_1 \cos \omega t)(A + B\cos \omega t + C \sin \omega t + D \cos 2\omega t + E \sin 2\omega t) \tag{3.12}$$

$$\Rightarrow I_{pd,BB,fund}(t) = (I_0 B + I_1(1 + A + \tfrac{1}{2}D)) \cos \omega t + (I_0 C + \tfrac{1}{2} I_1 E) \sin \omega t. \tag{3.13}$$

Expressions for the coefficients A, B, C, D, and E are not given, as the inability to write (3.10) as a closed-form Fourier series in ω makes the expansions quite complicated and does not give much intuition. The constant and second harmonic terms of (3.12) will be filtered out by the demodulator at frequency ω, thereby resulting in (3.13). Then the phase of (3.13) is:

$$\angle I_{pd,BB,fund}(t) \approx \arctan\left(\frac{I_0 C + \frac{1}{2} I_1 E}{I_1 + I_0 B + I_1(A + \frac{1}{2}D)}\right), \tag{3.14}$$

with A, B, C, D, and E ω-dependent values that are small compared to 1. Hence, the laser intensity modulation, which can be seen as the leading I_1 term in the denominator of (3.14), overshadows the components of the phase response contributed by the spin projection and absorption cross section.

Figure 3.5 shows the corresponding amplitude and phase response of the Bell-Bloom magnetometer with only a limited change in phase near resonance. The effect can also be observed in the time domain as shown in Fig. 3.6. The interaction of the laser with the alkali atoms only produces a small change in the waveform observed at the photodetector. Just as in (3.14), the signal at the photodetector is instead dominated by the intensity modulation which gives no information about the spin. The large intensity modulation reduces the phase response versus frequency of the Bell-Bloom magnetometer to have roughly the same dispersion characteristic as that of the quadrature signal. Unless the intensity modulation is mitigated the single pump/probe Bell-Bloom

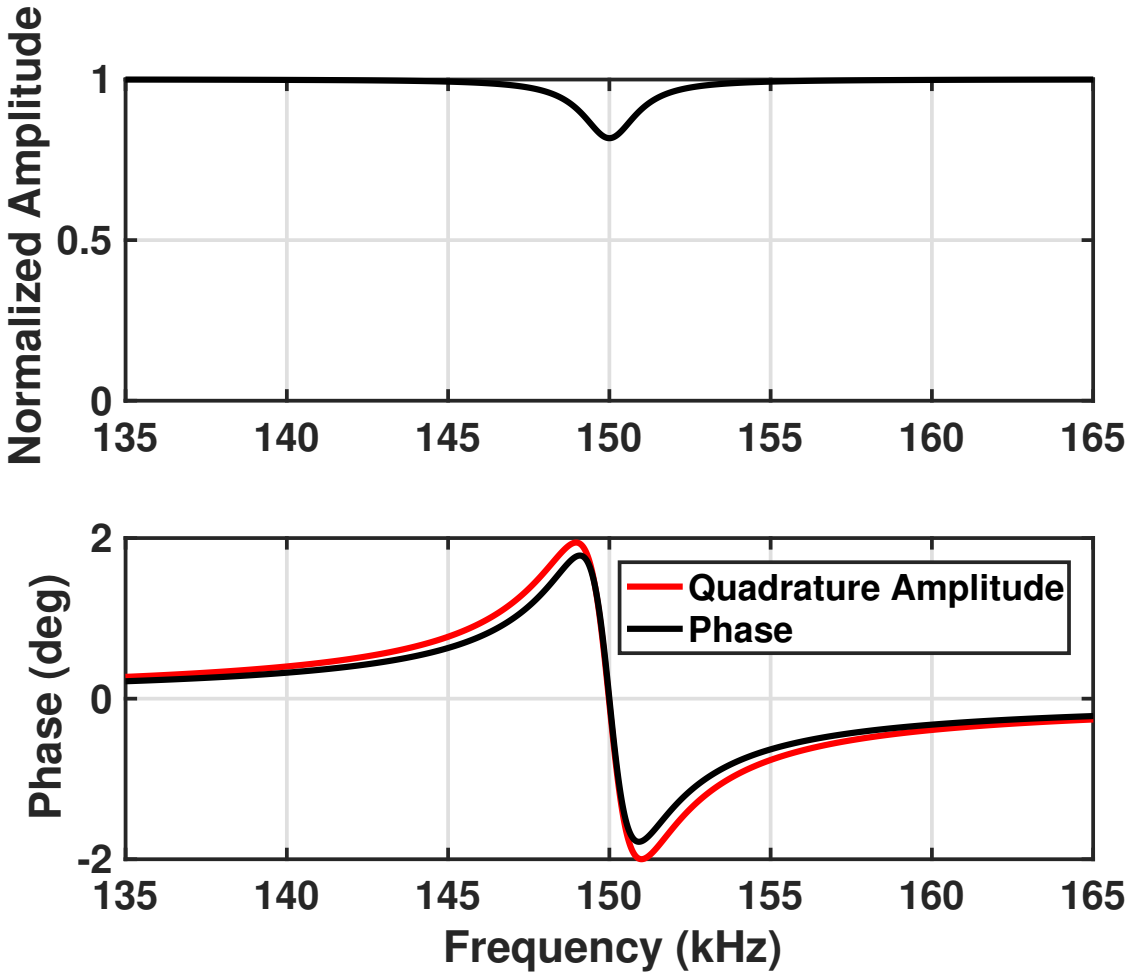


Figure 3.5: Bell-Bloom magnetometer normalized amplitude, and phase and quadrature amplitude response. The quadrature amplitude response is scaled to match the slope of the phase response at the center of the dispersion.

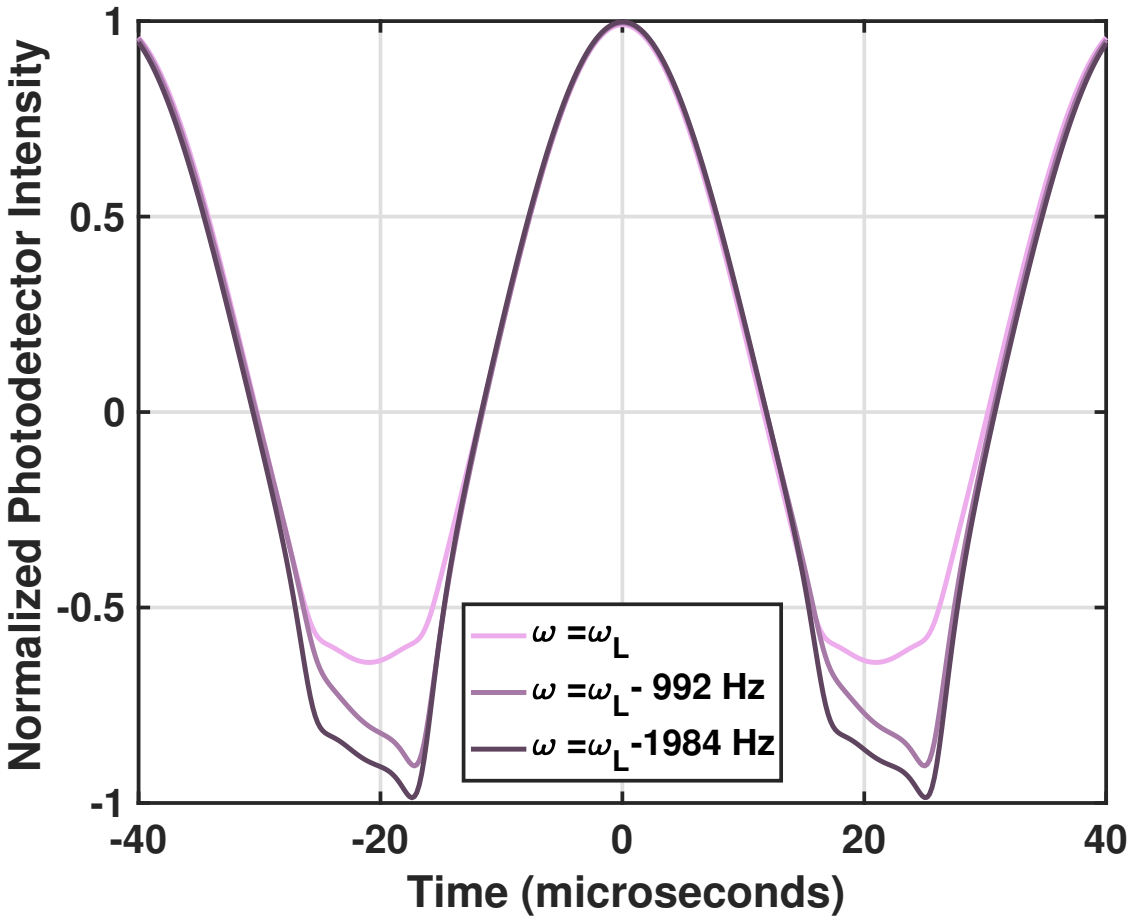


Figure 3.6: Simulated time domain waveforms from Bell-Bloom magnetometer. The waveforms show laser modulation at the Larmor frequency and increasing offsets from the Larmor frequency. At all frequencies, the interaction with the alkali atoms only contributes a small amount to the overall signal at the photodiode which is dominated by the laser intensity modulation.

magnetometer linearity and slew rate are roughly that of the quadrature amplitude of the M_x magnetometer, the worst case performance.

It should be mentioned that the intensity modulation can be avoided if the pump and probe beams are separated, giving the improved performance of the single pump/probe M_x magnetometer. For a microfabricated magnetometer this is impractical as it requires the use of two VCSELs, two photodiodes, and two laser heaters. Additionally, another loop is required to control the heater for the second VCSEL, and due to the small volume the two heaters will likely interact leading to possible stability concerns in the temperature control loops. These extra complications suggest that the single pump/probe Bell-Bloom is much more practical to implement than the separate pump/probe version and support an investigation into improving the single pump/probe response.

The reason for this dwarfing of the atomic response is that the Bell-Bloom magnetometer only pumps the atomic absorption for a fraction of the Larmor period, while the M_x magnetometer pumps at all times. Two unfortunate consequences of this are reduced sensitivity and reduced linearity. The sensitivity reduction is essentially due to the fact that the system is not always pumped, leading to a smaller signal. As shown in [15], a carefully chosen pumping signal can mitigate the sensitivity reduction.

The reduced linearity can be observed by comparing the phase response of the two magnetometers. Assuming a constant gain term can be placed in either system, the phase response of the Bell-Bloom magnetometer can be scaled to achieve similar $\frac{d\phi}{df}$ at the center of the resonance and then the derivatives of the resulting phase curves can be compared across frequency (Fig. 3.7). Since the phase output describes the error between the Larmor frequncy and modulation frequency of the pump beam, the nonlinearity of the phase response is equivalent to that generated by the frequency detector in a traditional FLL. The increased amplitude of the higher-order derivatives in the Bell-Bloom magnetometer leads to inferior linearity. While the sensitivity degradation can be partially addressed by choosing specific driving waveforms, fundamentally, it cannot be equivalent to that of the M_x case since the atoms can not be continuously pumped. However, it would seem that if the large signal describing the variation in pumping wavelength and/or power could be removed,

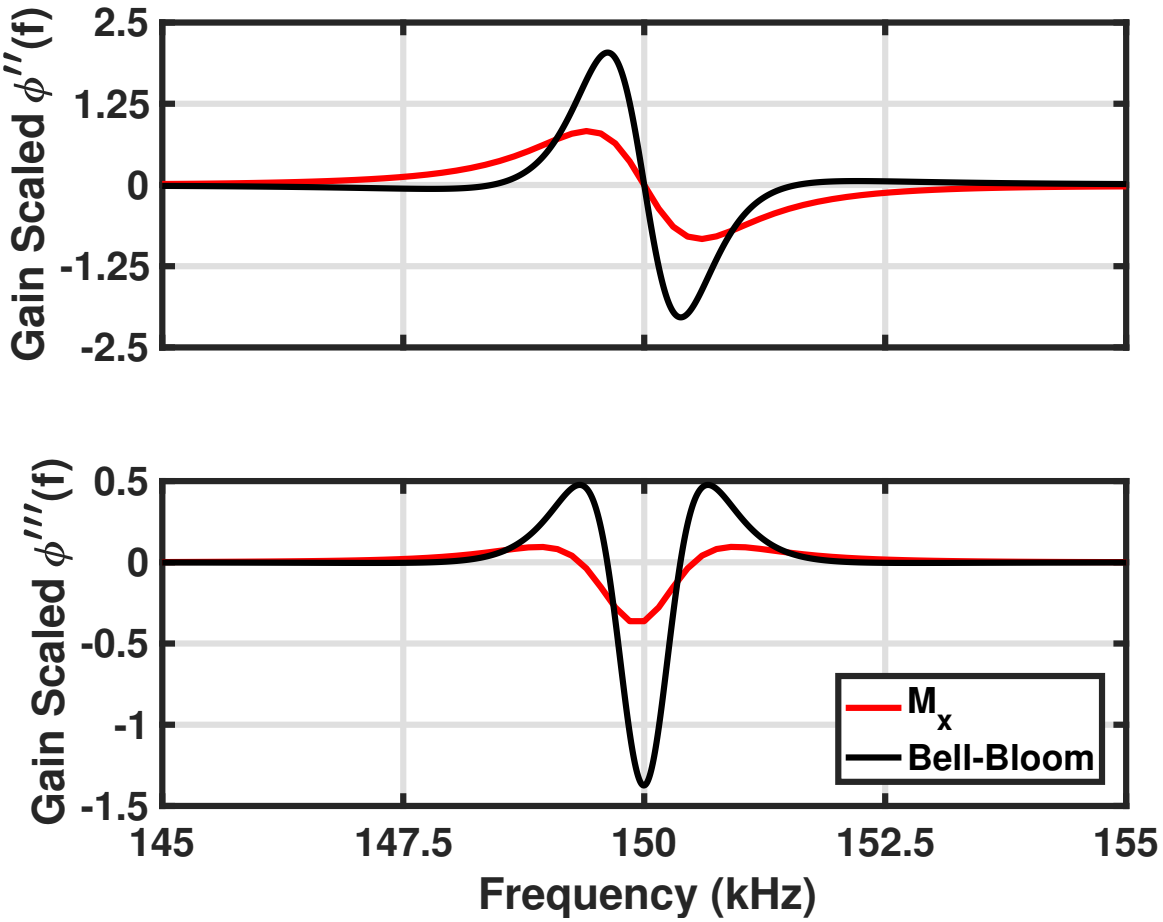


Figure 3.7: Comparison between the second and third derivatives of the normalized phase response of M_x and Bell-Bloom magnetometers. The larger derivatives for Bell-Bloom lead to more nonlinearity.

the original phase response of the M_x magnetometer could be recovered. Consequently, the better linearity of the M_x magnetometer could be obtained with the simpler Bell-Bloom magnetometer, motivating the cancellation technique introduced in this work.

3.3 Cancellation for Bell-Bloom Magnetometers

As was shown in the previous section, the laser intensity modulation dominates the response of a single pump/probe Bell-Bloom magnetometer, and in particular, the intensity modulation dominates the phase response. From the approximate response of (3.13), the power modulation, I_1, is seen to be roughly additive to the total response. Since the input frequency is set by the system, for example via the direct digital synthesizer in Fig. 3.3, and since the amplitude of the laser intensity modulation as observed at the photodiode can be measured, it is possible to subtract a sinusoid with the same frequency and phase, and with amplitude, I_2, comparable to I_1. The system response then becomes:

$$\begin{aligned} I_{out}(t) =& I_{pd,BB}(t) - I_2\cos\omega t \\ \approx& [I_0 + I_1\cos\omega t][1 - nL(\tfrac{1}{2} - \tfrac{1}{2}S_x)\sigma_0(\lambda)] - I_2\cos\omega t. \end{aligned} \tag{3.15}$$

Simplifying as in (3.13), and looking at the frequency-domain representation gives

$$I_{out,fund}(\omega) = (I_0 B(\omega) + I_1(1 + A(\omega) + \tfrac{1}{2}D(\omega)) - I_2) + j(I_0 C(\omega) + \tfrac{1}{2}I_1 E(\omega)) \tag{3.16}$$

where as mentioned in the previous section, A, B, C, D, and E are functions of ω. A, B, and D can also have a constant term, and can be expressed as $A(\omega) = a_0 + a(\omega)$ for example, with the constant terms giving no information about the response to magnetic field, and the frequency-dependent terms tending to zero for frequencies far from the Larmor frequency. By the odd symmetry of the coefficients of the sine term in (3.8), C, and E do not have a constant term. The phase response is

then approximately:

$$\angle I_{out} \approx \arctan\left(\frac{I_0 C(\omega) + \frac{1}{2} I_1 E(\omega)}{I_1(1 + a_0 + \frac{1}{2} d_0) + I_0 b_0 - I_2 + I_0 b(\omega) + I_1(a(\omega) + \frac{1}{2} d(\omega))}\right), \tag{3.17}$$

which becomes comparable to the response of the M_x magnetometer for I_2 close in value to $I_1(1 + a_0 + \frac{1}{2} d_0) + I_0 b_0$.

For the approximate solution of (3.16) and (3.17), the carrier-to-modulation-amplitude ratio (CMAR) is defined as

$$CMAR = \frac{I_2}{I_1(1 + a_0 + \frac{1}{2} d_0) + I_0 b_0}. \tag{3.18}$$

In practice, the denominator of (3.18) is not known exactly ahead of time, and is determined experimentally by measuring or simulating the fundamental component of the photodiode response when the magnetometer is driven at a frequency sufficiently far from the Larmor frequency, such as near the left or right edges of Fig. 3.5, where the response due to the precession of the atoms is negligible. For the Bell-Bloom magnetometer without cancellation, the CMAR is 0, and perfect cancellation would have a CMAR of 1.

Ideally the intensity modulation would be perfectly canceled ($CMAR = 1$); however, this is not practical, as any noise or other drift and variation over time would lead to a change in the sign of the denominator of (3.17) for frequencies away from the center of the dispersion, and hence a 180° change in the phase, putting the loop in positive feedback. The phase and amplitude response for a variety of values of CMAR are shown in Fig. 3.8. To make cancellation effective, the CMAR needs to be in the range 1-1.05. (Fig. 3.9).

In this work, I_2 has always been chosen such that CMAR ≥ 1. It is possible to choose CMAR $< \min(\mathrm{Re}\{I_{out,fund}(\omega)\})$, or $\min(\mathrm{Re}\{I_{out,fund}(\omega)\}) <$ CMAR < 1, but these choices introduce other complications. If CMAR $< \min(\mathrm{Re}\{I_{out,fund}(\omega)\})$, depicted by CMAR=0.8 in Fig. 3.10, the signal amplitude approaches zero at the Larmor frequency, and there is an abrupt change in phase from $+90°$ to $-90°$, actually reducing the linear range of the magnetometer. Alternatively, $\min(Re\{I_{out,fund}(\omega)\}) <$ CMAR < 1 can be chosen, represented by CMAR=0.98 in Fig. 3.10.

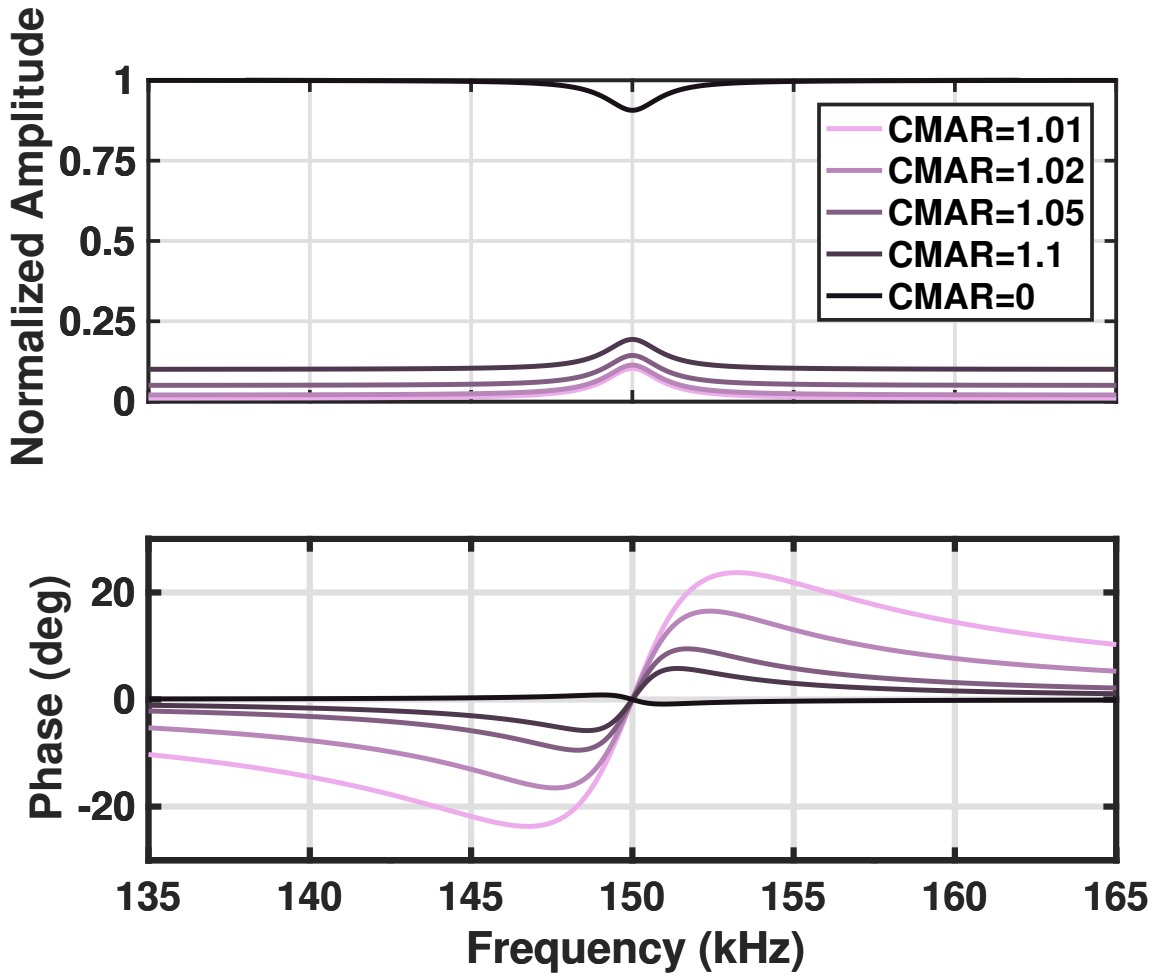


Figure 3.8: Amplitude and phase response showing increasing phase as CMAR approaches 1.

While this gives comparable linearity to CMAR=1.02 near the Larmor frequency, away from the resonance, there are regions where the frequency detector gain is actually higher than at the center of the resonance. Without additional logic to account for this and perform limiting or additional digital linearization, the increased gain will lead to instability with large input signals. The issues associated with these two alternatives justify the choice of CMAR ≥ 1.

The benefits of cancellation can be seen in the width and shape of the dispersion response. Looking at the relative distance between the maximum and minimum of the dispersion curve, we see that width increases as the CMAR approaches 1, as desired. This allows the system to track the sensed magnetic field, even with large sudden jumps. As we move beyond the peaks of the dispersion, the output phase, and hence gain, drops leading to undesirable slower settling until the

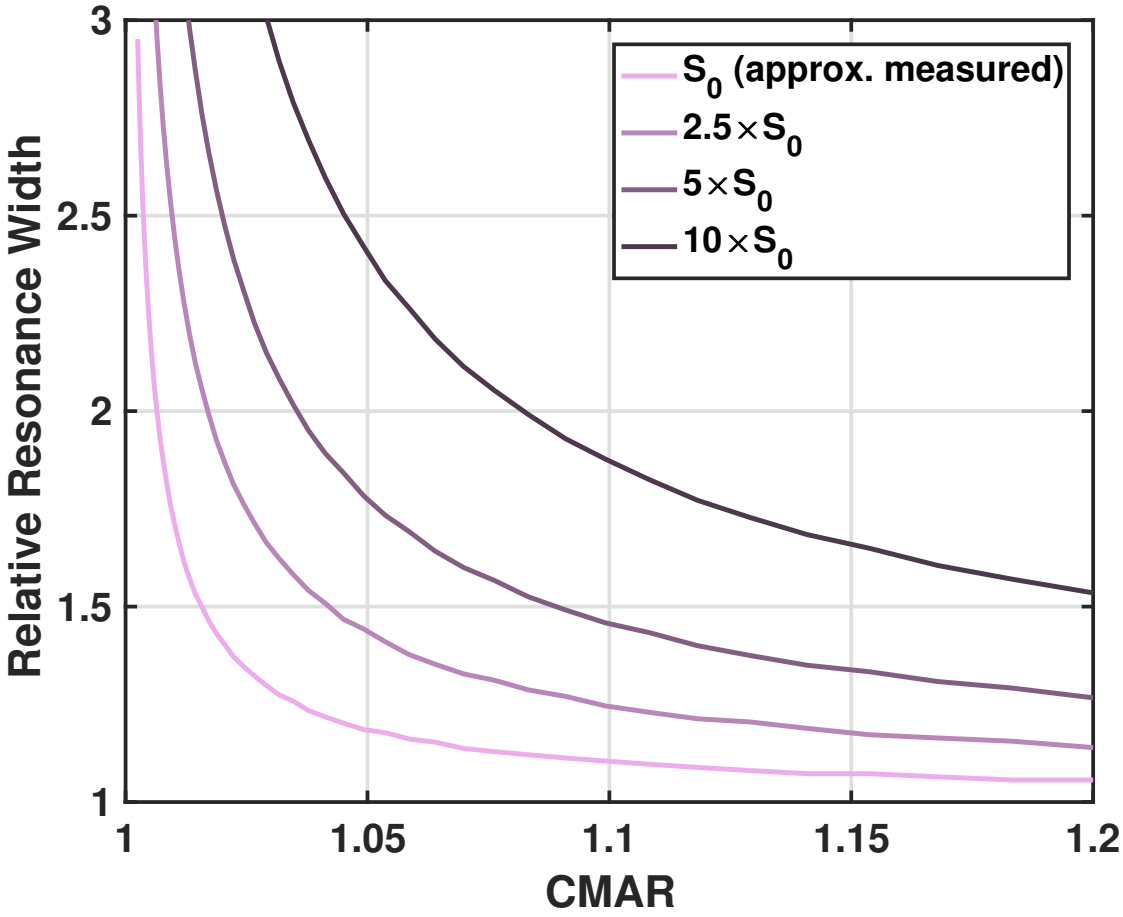


Figure 3.9: Resonance width as a function of CMAR, relative to width with no cancellation. Darker curves represent increasing peak spin polarization and absorption with the lightest curve representative of the vapor cell used in the experiments of this paper.

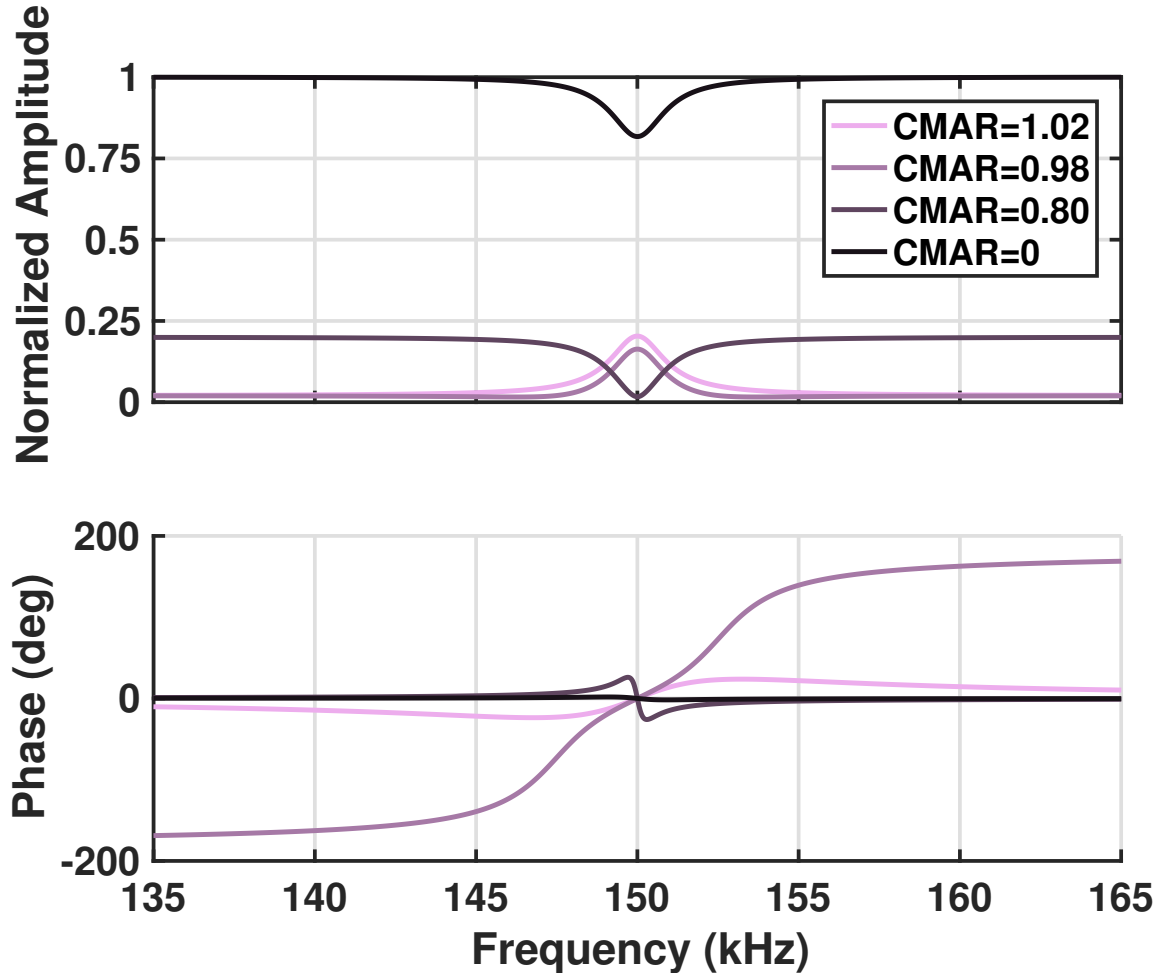


Figure 3.10: Amplitude and phase response showing the need for CMAR ≥ 1. With CMAR $< \min(\mathrm{Re}\{I_{out,fund}(\omega)\})$, the linearity is actually worse than with no cancellation, and with $\min(\mathrm{Re}\{I_{out,fund}(\omega)\}) < \mathrm{CMAR} < 1$ the increased gain off resonance can lead to instability.

driving frequency approaches the new Larmor frequency. Figure 3.9 shows the distance between the minima and maxima of the dispersion curves as the cancellation amplitude is varied relative to the intensity modulation amplitude. Each curve represent a different amount of spin polarization which is affected mainly by the pumping rate for a fixed vapor cell. For the cell and laser used in the experiments described in this paper, the pumping rate is somewhat limited. Based on matching the measured absorption minimum from measurement to the simulation, the lightest curve is most representative of the used cell. With this, theoretically a factor of 1.5-2 improvement in the width of the resonance is expected for practical cancellation values in the CMAR=1-1.05 range. This also shows that improving the pumping rate makes cancellation more effective, or equivalently, allows for less accurate cancellation.

Also of concern is the linearity of the phase response. As is well known in PLLs and FLLs, one of the limiting factors in spur suppression is the linearity of the phase or frequency detector respectively. If these blocks introduce nonlinearity, it can be suppressed by at most the loop gain of the system. If the magnetometer is operated in a large time-varying field, such as 60 Hz background in a home or office environment, the corresponding large variation in field will cause the system to drift from the zero crossing towards the peaks of the dispersion due to finite bandwidth of the system. The nonlinearity of the phase response will lead to harmonic distortion, limiting the dynamic range of the magnetometer. As shown in Fig. 3.11, as the cancellation approaches the ideal value, the magnitude of the second and third derivatives of the phase response versus frequency drop, leading to a reduction in harmonic distortion.

It is worth noting that additional linearity improvement could be achieved by determining a quasi-inverse function for the phase response. This work does not explore this quasi-inverse approach, as the implementation is quite different, requiring significant adaptive filtering to track the time-varying absorption, linewidth, and pumping rate, all dependent on temperature, and weakly dependent on the center frequency. The reason the correction function is only a quasi-inverse, is that at large offsets from the Larmor frequency, the signal amplitude goes to zero. This makes determining the offset difficult since noise in the system dominates the signal. For noise-limited

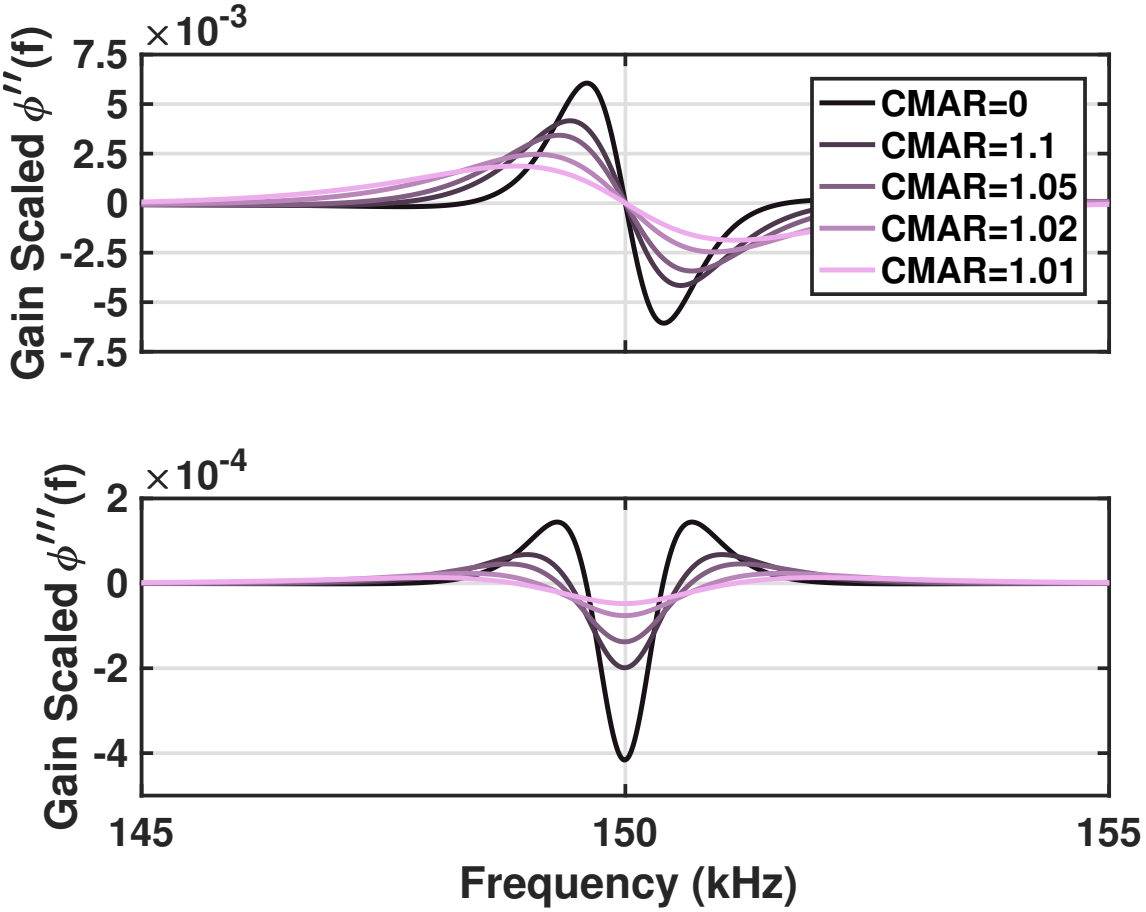


Figure 3.11: Second and third derivatives of phase response with different cancellation amplitudes. The phase responses were scaled linearly to give equal gain at the center of the dispersion before taking the derivatives, giving equal system bandwidth for all cases.

frequencies, all methods of cancellation, no cancellation, and M_x suffer from similar limitations.

3.4 System Implementation

The implementation of cancellation for the Bell-Bloom magnetometer is quite straightforward in theory. Starting with the system implementation shown in Fig. 3.3, a variable gain path is added in parallel with the magnetometer, and summed with the output of the magnetometer, as shown in Fig. 3.12.

In practice, there are a number of nonidealities introduced by the magnetometer and elec-

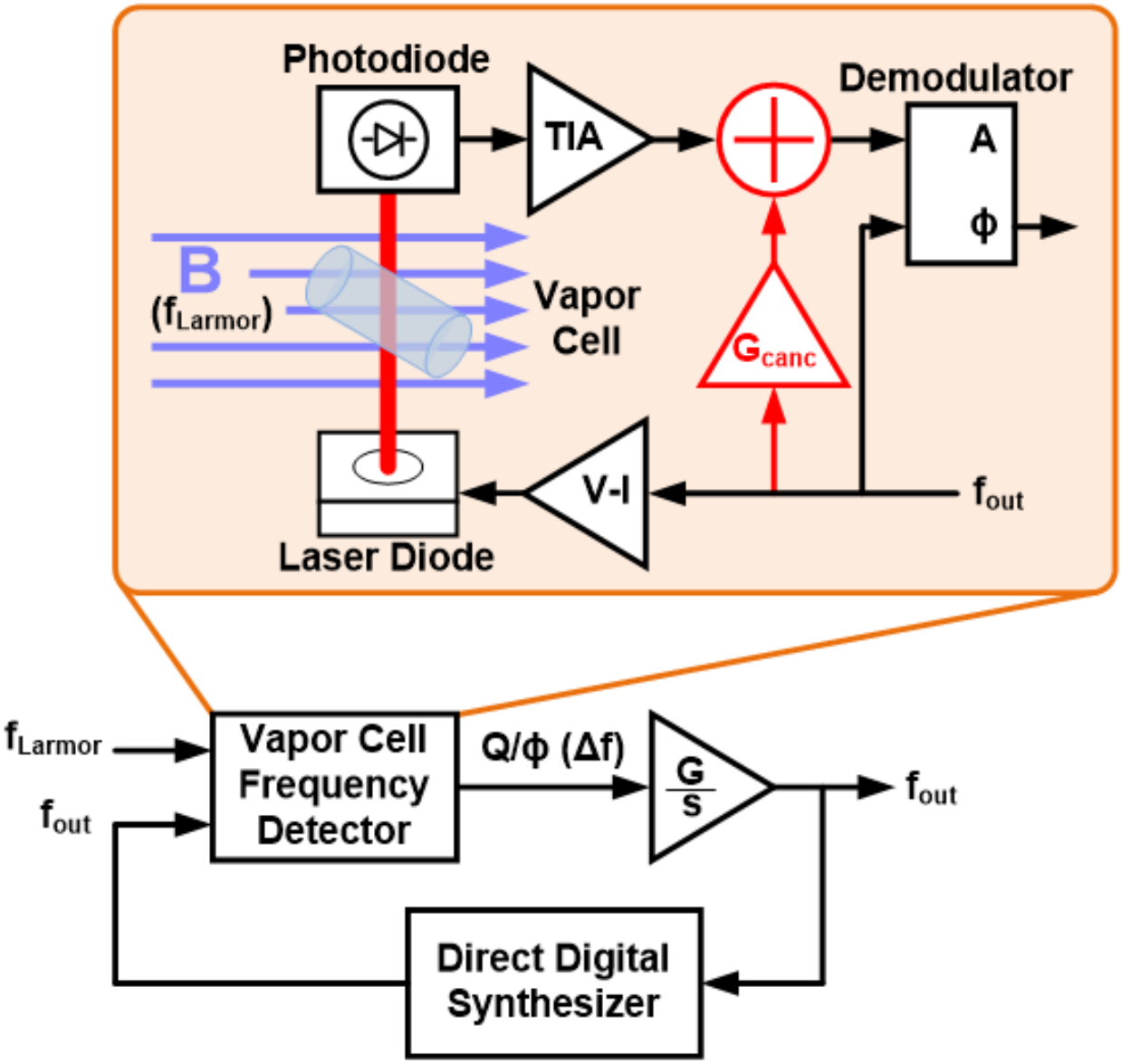


Figure 3.12: Implementation of cancellation in a Bell-Bloom magnetometer. The added blocks to implement cancellation are shown in red.

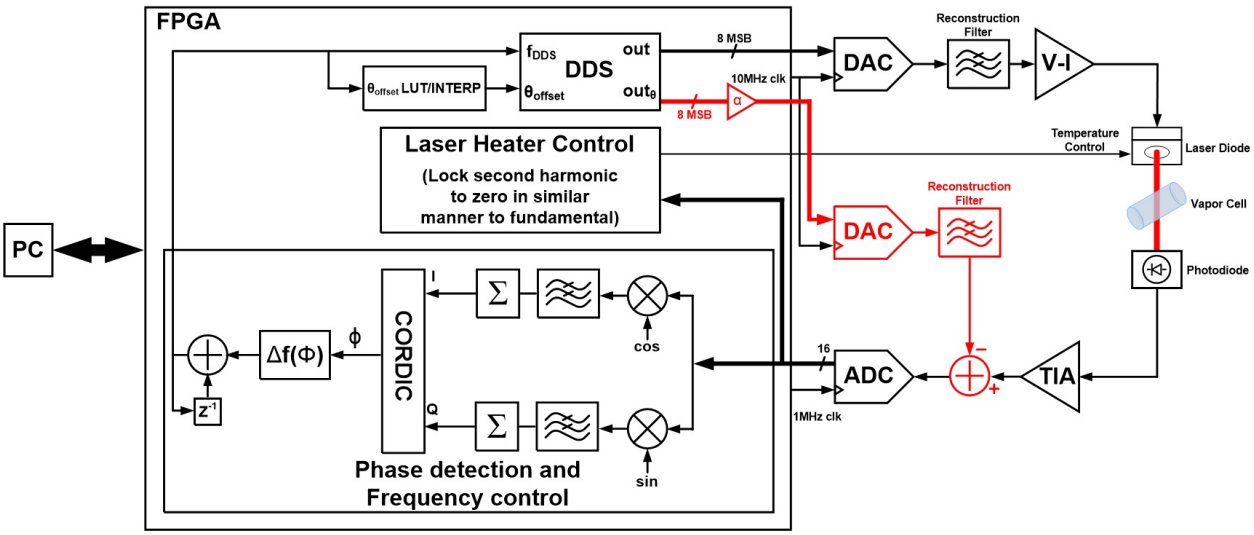


Figure 3.13: Detailed block diagram of the closed-loop Bell-Bloom magnetometer implementation in this work.

tronic control loop which must be corrected to implement the Bell-Bloom magnetometer, even without cancellation. A more detailed diagram of the implementation used in this work is shown in Fig. 3.13. As operation is desired over the entire range of Earth's magnetic field ($\approx 25\ \mu\text{T} - 65\ \mu\text{T}$), and the system contains delays and nonlinear phase and amplitude response over frequency, these issues must be corrected to achieve reasonable performance.

Prior to data collection, an open-loop foreground calibration is performed. The calibration is performed in two steps. First, the cancellation path shown in red is disabled, and the input frequency driving the VCSEL is swept from $70 - 230$ kHz to cover the expected range of magnetic fields. If not using cancellation, the calibration is finished. If cancellation is desired, the magnetometer path is then disabled, and the frequency driving the cancellation path is swept over the same frequency range. If cancellation is not being used, the phase data from the first sweep is first processed to locate the absorption maxima from alkali atom interaction, and the sweep data for a suitable range around this frequency is removed. The remaining data is fit to a polynomial, and a look-up-table (LUT) with linear interpolation is populated. This LUT is used to appropriately phase shift the reference sine and cosine signals in the demodulator of the digital lock-in amplifier implemented on the FPGA. If cancellation will be used in the system, the amplitudes and phases of the magnetometer

and cancellation paths are compared. The cancellation path amplitude is adjusted to be slightly larger than the in-phase component of the magnetometer path. The cancellation path phase is also adjusted to match the delay of the magnetometer path by adding a phase delay equivalent to the difference in the magnetometer and cancellation paths calibration sweep phases.

Additionally, the VCSEL wavelength is quite sensitive to temperature variation with a temperature coefficient of 0.06 $nm/^{\circ}$C. With this variation, a 1 $^{\circ}$C change in temperature leads to a change in wavelength comparable to that from the Bell-Bloom VCSEL current modulation and significantly alters the absorption of photons by the alkali atoms. This means the VCSEL temperature must be controlled to well within 0.1 $^{\circ}$C to avoid large changes in the signal amplitude over time. This can be accomplished by tracking the amplitude of the second harmonic of the magnetometer waveform. The asymmetry introduced by this second harmonic content (Fig. 3.6) roughly describes the duty-cycle of the pumping in the Bell-Bloom magnetometer, which we would like to remain constant over time. Since cancellation only introduces a tone at the fundamental, the second harmonic amplitude is unaffected by cancellation, so no change is required in the laser temperature control loop when cancellation is applied. The heating applied to the laser is varied with a PI controller with bandwidth on the order of 0.1 Hz, sufficient for the experiments in this work.

It is also worth noting that while cancellation was done in the analog domain in this work, it could easily be implemented in the digital domain. In the digital case, only the single lookup table for correcting the phase nonlinearity of the magnetometer path is required. This also removes the need for a second DAC, filter, and the adder, which can all be implemented with very little power in the digital domain. A digital implementation also lends itself to increased reconfigurability such as would be used in an attempt to linearize beyond the simple phase cancellation approach. The analog implementation was used in this work to allow for variable gain in the adder. This gives the option of increasing the signal amplitude after cancellation to use more of the full-scale input-range of the ADC.

The digital components of the system (lock-in, filtering, direct digital synthesizers (DDS))

are implemented on FPGA to allow for ease of reconfiguration in the prototype. The system is controlled through a user interface on a computer connected to the FPGA, and data is streamed back to the computer for analysis. The control loop for operation when the system is locked is fully contained on the FPGA and could be synthesized for implementation on an ASIC in the future.

The FPGA is placed on a printed circuit board (PCB) along with all mixed-signal blocks and power management to run the system from a battery. The physics package (vapor cell, VCSEL, photodiode, and heaters) are packaged together and connected by a short length of flexible PCB to an analog frontend (AFE) board. The AFE contains the voltage-to-current converter, transimpedance amplifier, and summer. Buffering is added between the AFE and mixed-signal boards, allowing for them to be separated by a few feet to minimize electromagnetic coupling from the mixed-signal board to the physics package.

3.5 Measurement Results

In this work, a Cesium microfabricated atomic magnetometer with cell volume 2 mm^2 is used. For controlled characterization purposes the system is measured with the physics package placed inside a 4-layer magnetic shield. Coils inside the shield allow for generation of static and modulated fields. The AFE is placed just outside the shield. The dispersion width, sensitivity, and linearity of the Bell-Bloom magnetometer are characterized both with and without cancellation to compare performance.

The measured dispersive phase response of the system is compared with a variety of cancellation amplitudes, shown in Fig. 3.14. As expected, as the cancellation accuracy improves, the peak-to-peak width of the dispersion increases. By measuring the variation in absorption without cancellation (This is the relative magnitude of the dip as found in measurement; the simulated dip is shown in Fig. 3.8), the peak spin polarization used in the theoretical model can be modified to match the cell used in this setup. With this fit, there is reasonable agreement between the width of the measured dispersion curves and the model as shown in Fig. 3.15. In the actual system, to maintain

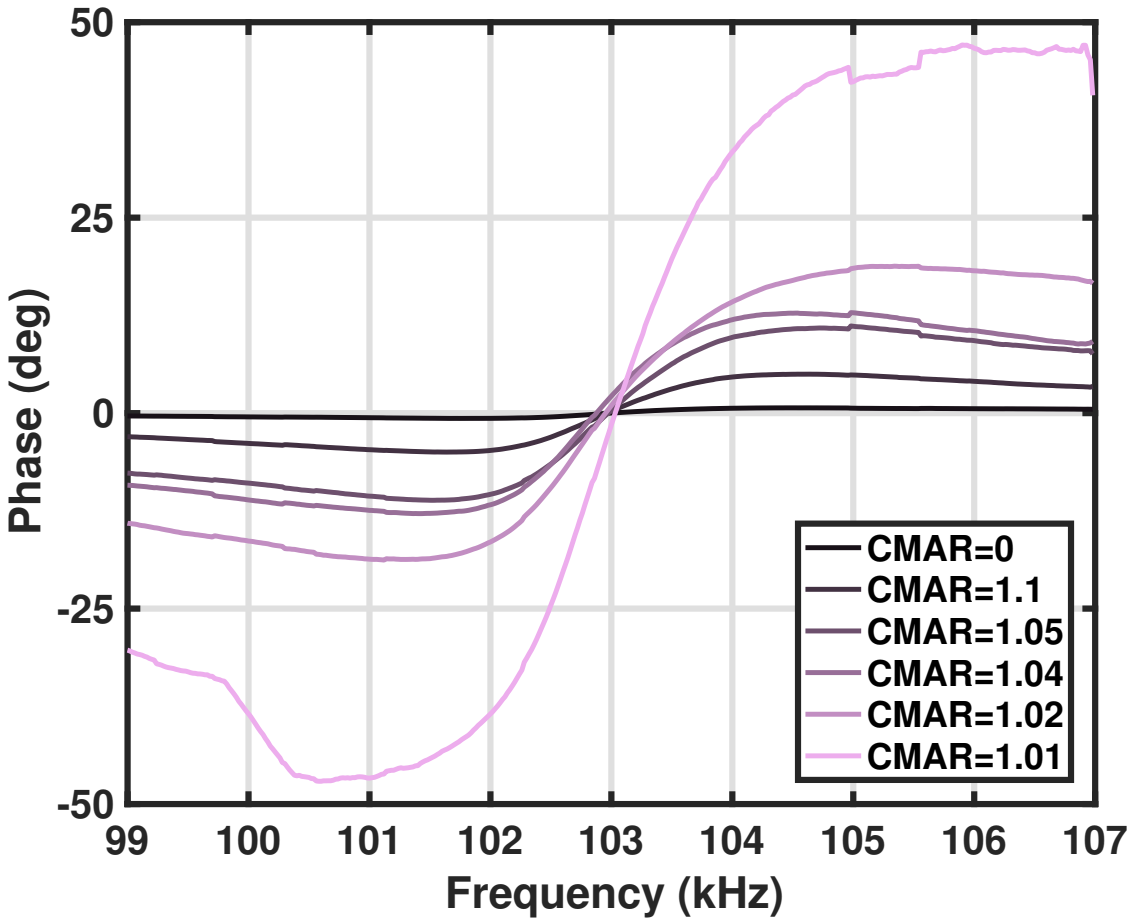


Figure 3.14: Comparison of the dispersion curves generated from frequency sweeps. The curve with the least variation is with no cancellation (CMAR=0), and the curve with most variation has CMAR=1.01.

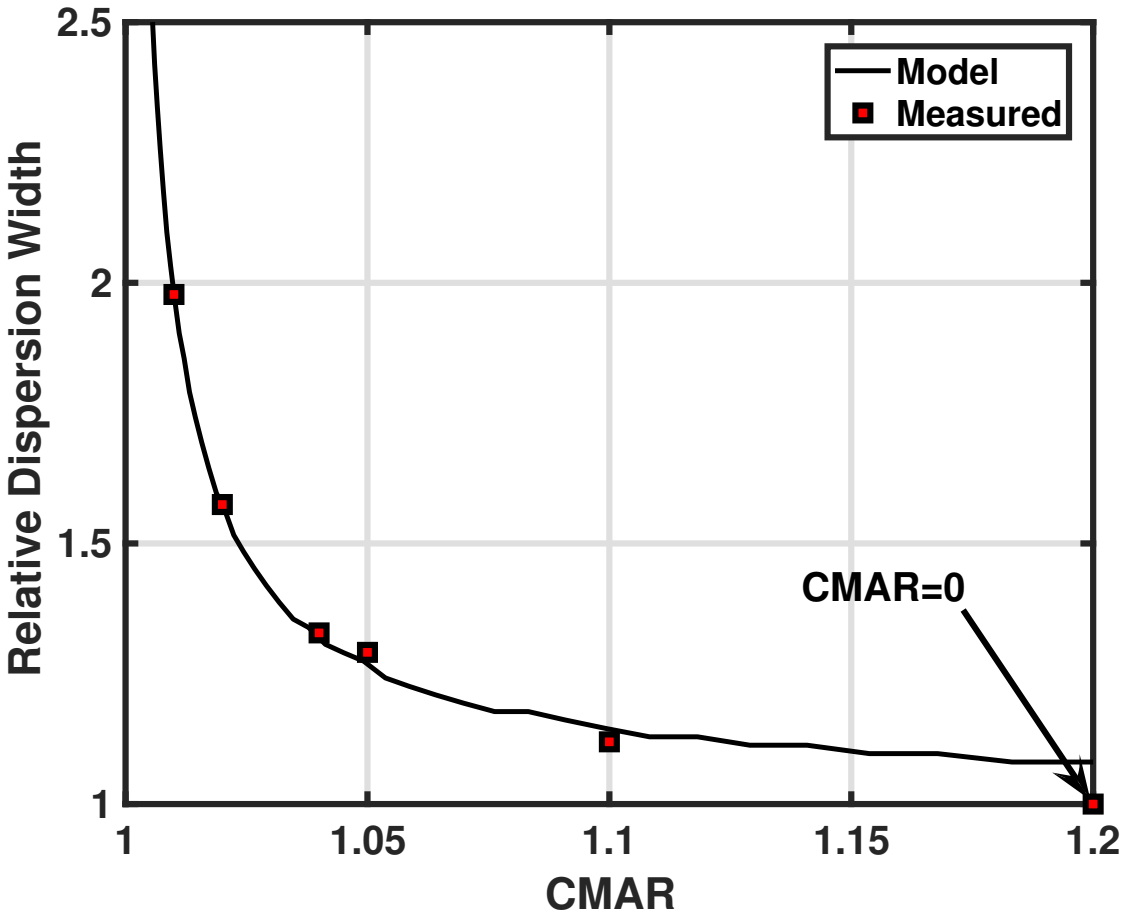


Figure 3.15: Dispersion peak-to-peak width for different cancellation amplitudes. When the maximum spin-polarization in the model is matched to the measured absorption from a single sweep with no cancellation, there is good agreement between the measured data and model.

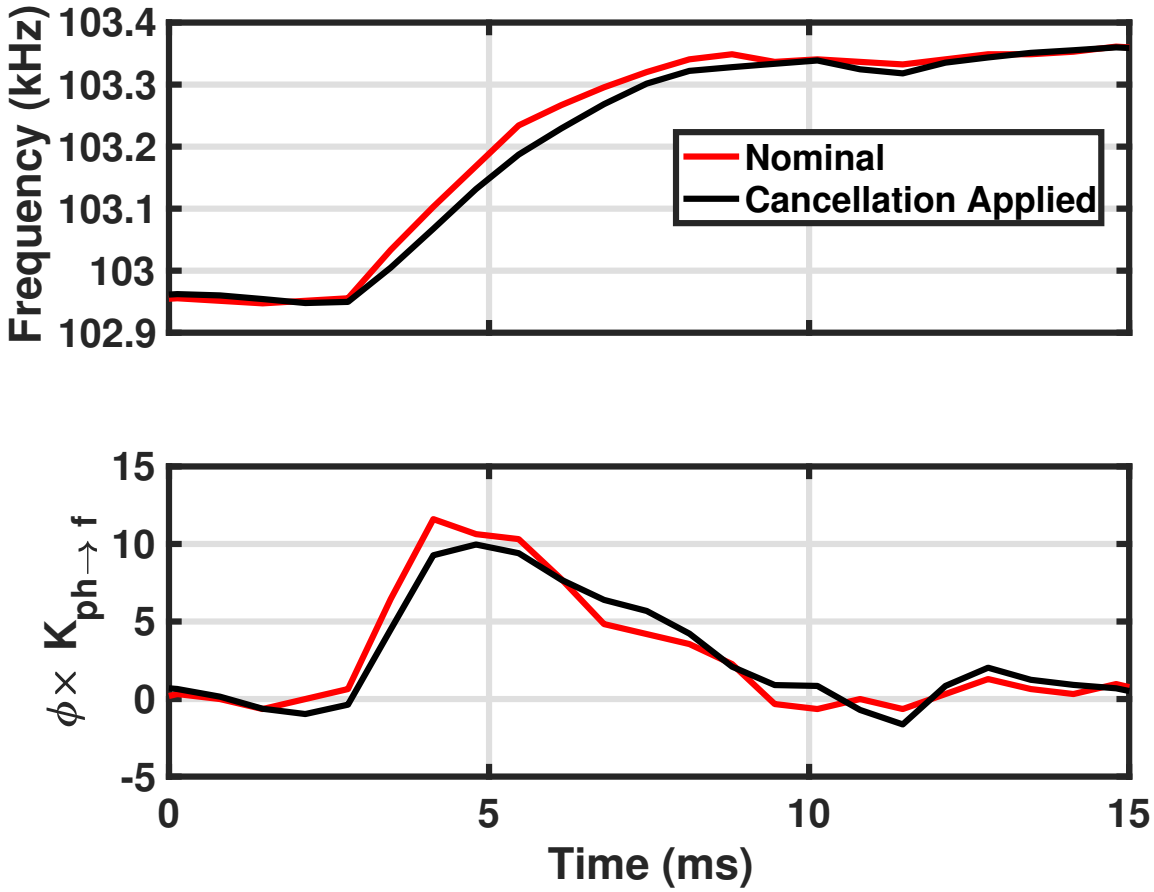


Figure 3.16: Magnetometer response to 100 nT step change in the magnetic field. The top plot shows similar settling with and without cancellation. The bottom plot show that since the magnetometer is still operating in the linear regime, the loop gain of both loops is roughly equal .

a constant bandwidth independent of cancellation amplitude, the loop gain is maintained constant by digitally adjusting the gain of the $\Delta f(\phi)$ block shown in Fig. 3.13. As cancellation amplitude is increased, the phase response of the magnetometer increases and hence $\Delta f(\phi)$ is reduced for constant loop gain. This leads to similar stability as well as bandwidth for the loop, independent of the cancellation signal amplitude.

One characteristic often of interest is the response time of the magnetometer with a large step change in the magnetic field. For steps that stay within the range where the phase detector response is relatively linear, the magnetometer is expected to have exponential settling. For steps that exceed this linear range, the response will show slewing. For large steps, the increased phase

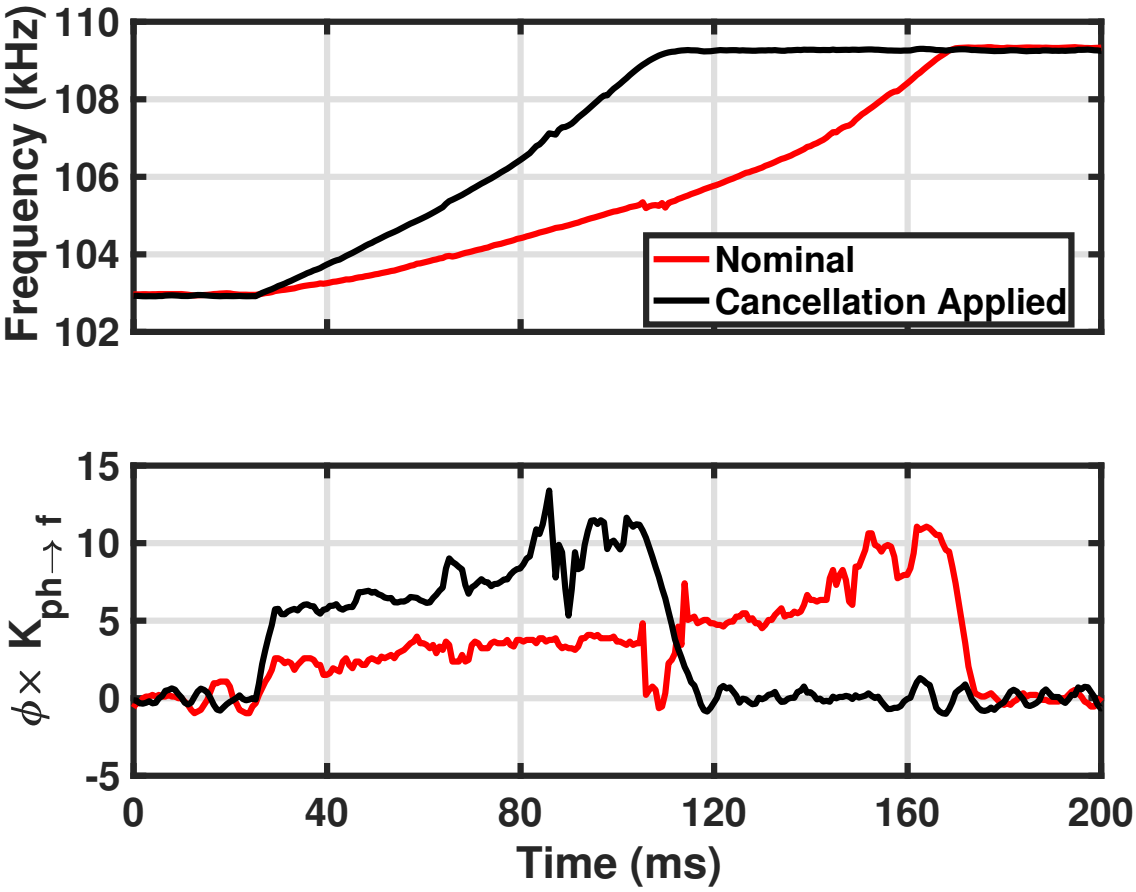


Figure 3.17: Magnetometer response to a 1.8 μT step change in the magnetic field. The top plot shows that with cancellation the magnetometer slews much faster. The bottom plot show that the faster slewing can be attributed to the larger phase signal at large offsets from the center of the dispersion curve when cancellation is applied.

output of the magnetometer when cancellation is applied should lead to a faster response. Figure 3.16 demonstrates that for a small step of 100 nT, the system settles at a similar rate regardless of cancellation. This is expected since the small-signal bandwidth of the magnetometer is constant independent of cancellation. On the other hand, when a 1.8 μT step in the magnetic field is applied, the magnetometer with cancellation does indeed slew faster as shown in Fig. 3.17. For step changes in the magnetic field in excess of roughly 10 μT, even with cancellation, the phase detector response becomes minimal. In this extreme case, a loss-of-lock detection that sweeps the input frequency to find the new field value will likely respond faster than waiting for the loop to re-achieve lock on its own.

The sensitivity of the magnetometer is measured both in the open-loop mode, with the slope of the dispersive phase response relating the observed phase variation to frequency deviation, as well as in a closed-loop mode where the center of the dispersion is tracked as in FM demodulation with a FLL. The magnetometer achieves a measured sensitivity of approximately 50 $\mathrm{pT}/\sqrt{\mathrm{Hz}}$ over a 1-100 Hz frequency range. This is a limitation of the specific physics package used in these experiments, and similar packages that are not available for these measurements have shown sensitivities better than 5 $\mathrm{pT}/\sqrt{\mathrm{Hz}}$. At the same time, the sensitivity does not significantly impact the linearity, which is the emphasis of this study. Where needed, averaging is used to improve the noise-limited sensitivity allowing for observation of distortion.

The linearity of the system is measured by generating a 10 Hz magnetic field with increasing amplitude using the coils in the magnetic shield (in addition to an arbitrarily chosen static field of approximately 30 μT), and observing the distortion in the measured output. To characterize the distortion, the total harmonic distortion (THD) is measured and multiplied with the signal amplitude to give a measure of the system sensitivity. As these large tones are typically an undesired environmental noise, the concern is not so much how the distortion affects the environmental noise itself, but how it degrades the system sensitivity. With the significant averaging used in these measurements, the distortion observed when the environmental noise signal strength is small is limited by differential nonlinearity (DNL) in the LUT. For small to medium amplitude 10 Hz tones,

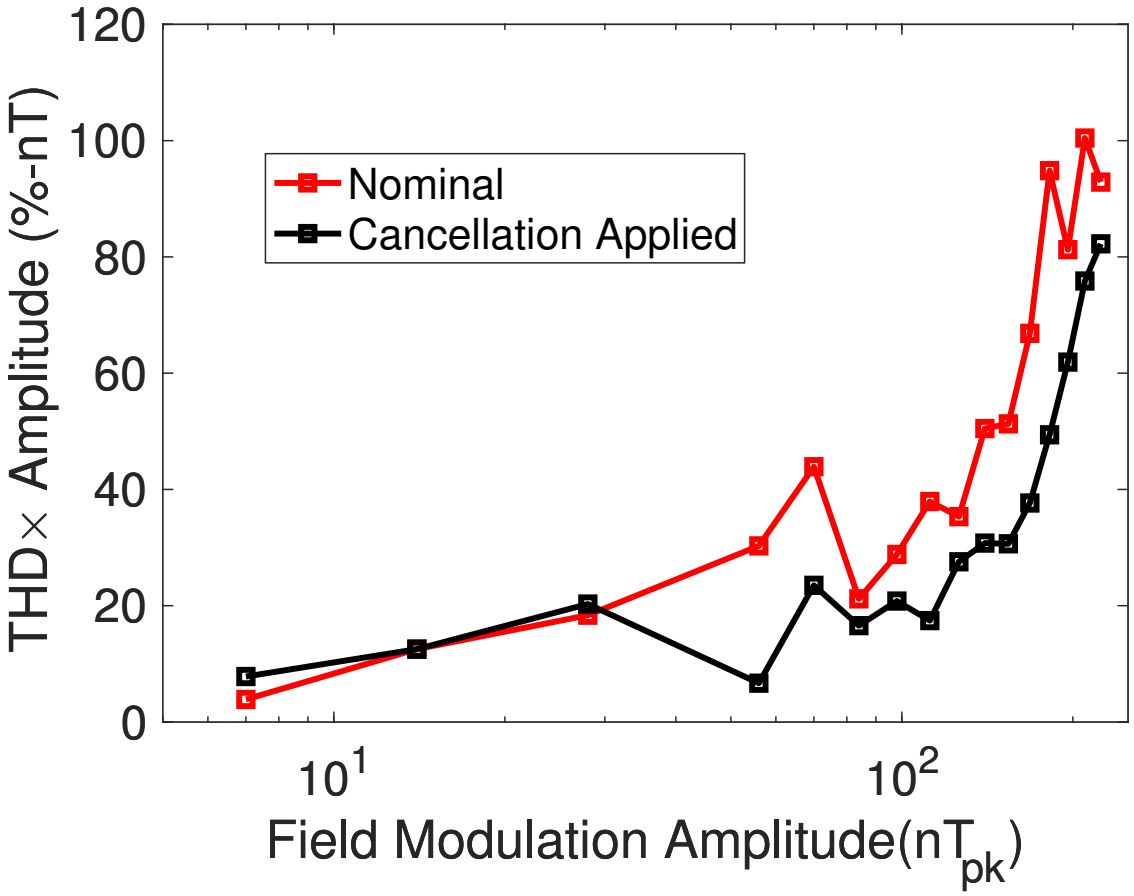


Figure 3.18: Measured sensitivity for a variety of environmental noise tone amplitudes. Cancellation does provide a reduced level of distortion compared to the nominal mode of operation.

the distortion does not vary much with signal amplitude and sets a minimum for the linearity-limited sensitivity.

Figure 3.18 shows $THD \times A_{pk}$ where A_{pk} is the amplitude of the 10 Hz tone. We see that as the environmental noise tone increases, the distortion does indeed increase. But as expected, cancellation does provide an improvement in distortion-limited sensitivity compared to the nominal mode of operation. This means that with cancellation, a larger environmental interference tone can be tolerated without degrading the system system sensitivity, improving the performance of the Bell-Bloom magnetometer.

3.6 Conclusion

This paper has presented a technique for enhancing the linearity of a Bell-Bloom atomic magnetometer. By analyzing the mathematical model of the magnetometer, it is demonstrated that use of the phase response instead of the quadrature amplitude theoretically gives enhanced linearity. Additionally, by electronically canceling the large signal at the photodetector due to light that is not absorbed by the magnetometer, the slew rate and linearity of the magnetometer can be further improved. With large applied fields, measurements of a cesium magnetometer demonstrate the slew-rate and linearity enhancement of the presented cancellation technique. The cancellation approach described in this paper should prove a useful technique for atomic magnetometers in magnetic-interference-rich environments. In particular the technique could assist in sensitive biomagnetic measurements of humans in non-magnetically-shielded environments.

Acknowledgment

The authors would like to thank the Fetzer Franklin Fund for support, and D. Newby for assistance.

This chapter, in full, has been accepted for publication, as it will appear in IEEE Transactions on Instrumentation and Measurement, Levy, Cooper; Kornack, Thomas; Mercier, Patrick. The

dissertation author was the primary investigator and author of this material.

Bibliography

[1] G. Bison, R. Wynands, and A. Weis, "A laser-pumped magnetometer for the mapping of human cardiomagnetic fields," *Applied Physics B: Lasers and Optics*, vol. 76, pp. 325–328, Mar 2003.

[2] K. Kim, W.-K. Lee, I.-S. Kim, and H. S. Moon, "Atomic vector gradiometer system using cesium vapor cells for magnetocardiography: Perspective on practical application," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, pp. 458–462, Apr 2007.

[3] S. Xu, M. H. Donaldson, A. Pines, S. M. Rochester, D. Budker, and V. V. Yashchuk, "Application of atomic magnetometry in magnetic particle detection," *Applied Physics Letters*, vol. 89, p. 224105, Nov 2006.

[4] M. P. Ledbetter, I. M. Savukov, D. Budker, V. Shah, S. Knappe, J. Kitching, D. J. Michalak, S. Xu, and A. Pines, "Zero-field remote detection of NMR with a microfabricated atomic magnetometer," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 2286–2290, Feb 2008.

[5] H. Xia, A. B.-A. Baranga, D. Hoffman, and M. V. Romalis, "Magnetoencephalography with an atomic magnetometer," *Applied Physics Letters*, vol. 89, p. 211104, Nov 2006.

[6] V. Mathé, F. Lévêque, P.-E. Mathé, C. Chevallier, and Y. Pons, "Soil anomaly mapping using a caesium magnetometer: Limits in the low magnetic amplitude case," *Journal of Applied Geophysics*, vol. 58, pp. 202–217, Mar 2006.

[7] J. Belfi, G. Bevilacqua, V. Biancalana, S. Cartaleva, Y. Dancheva, and L. Moi, "Cesium coherent population trapping magnetometer for cardiosignal detection in an unshielded environment," *Journal of the Optical Society of America B*, vol. 24, p. 2357, Aug 2007.

[8] M. N. Nabighian, V. J. S. Grauch, R. O. Hansen, T. R. LaFehr, Y. Li, J. W. Peirce, J. D. Phillips, and M. E. Ruder, "The historical development of the magnetic method in exploration," *Geophysics*, vol. 70, pp. 33ND–61ND, Nov 2005.

[9] H. Korth, K. Strohbehn, F. Tejada, A. G. Andreou, J. Kitching, S. Knappe, S. J. Lehtonen, S. M. London, and M. Kafel, "Miniature atomic scalar magnetometer for space based on the rubidium isotope 87 Rb," *Journal of Geophysical Research: Space Physics*, vol. 121, pp. 7870–7880, Aug 2016.

[10] A. David, M. Cole, T. Horsley, N. Linford, P. Linford, and L. Martin, "A rival to Stonehenge? Geophysical survey at Stanton Drew, England," *Antiquity*, vol. 78, pp. 341–358, Jun 2004.

[11] D. Cohen, "Large-volume conventional magnetic shields," *Revue de Physique Appliquée*, vol. 5, no. 1, pp. 53–58, 1970.

[12] A. L. Bloom, "Principles of operation of the rubidium vapor magnetometer," *Applied Optics*, vol. 1, p. 61, Jan 1962.

[13] D. Budker and D. F. J. Kimball, eds., *Optical Magnetometry*. Cambridge University Press, 2013.

[14] W. E. Bell and A. L. Bloom, "Optically driven spin precession," *Physical Review Letters*, vol. 6, pp. 280–281, Mar 1961.

[15] R. Jimenez-Martinez, W. Griffith, Y.-J. Wang, S. Knappe, J. Kitching, K. Smith, and M. Prouty, "Sensitivity comparison of mx and frequency-modulated bell–bloom cs magnetometers in a microfabricated cell," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, pp. 372–378, Feb 2010.

[16] P. D. D. Schwindt, S. Knappe, V. Shah, L. Hollberg, J. Kitching, L.-A. Liew, and J. Moreland, "Chip-scale atomic magnetometer," *Applied Physics Letters*, vol. 85, pp. 6409–6411, Dec 2004.

[17] P. D. D. Schwindt, B. Lindseth, S. Knappe, V. Shah, J. Kitching, and L.-A. Liew, "Chip-scale atomic magnetometer with improved sensitivity by use of the Mx technique," *Applied Physics Letters*, vol. 90, p. 081102, Feb 2007.

[18] S. J. Seltzer, *Developments in alkali-metal atomic magnetometry*. PhD thesis, Princeton University, 2008.

[19] G. Bison, R. Wynands, and A. Weis, "Optimization and performance of an optical cardiomagnetometer," *Journal of the Optical Society of America B*, vol. 22, pp. 77–87, Jan 2005.