

UCLA

UCLA Electronic Theses and Dissertations

Title

Envelope Tracking Supply Modulator with Trellis Search-Based Switching and 160 MHz Capability

Permalink

<https://escholarship.org/uc/item/1rz5450j>

Author

Leng, Weiyu

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Envelope Tracking Supply Modulator with Trellis Search-Based Switching and 160 MHz
Capability

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical & Computer Engineering

by

Weiyu Leng

2020

© Copyright by
Weiyu Leng
2020

ABSTRACT OF THE DISSERTATION

Envelope Tracking Supply Modulator with Trellis Search-Based Switching and 160 MHz Capability

by

Weiyu Leng

Doctor of Philosophy in Electrical & Computer Engineering

University of California, Los Angeles, 2020

Professor Asad A. Abidi, Chair

Envelope tracking is widely used to raise the efficiency of PAs. An envelope tracking supply modulator (ETSM) modulates PA's supply voltage to tracks the RF waveform's envelope, so that the PA will operate in saturation all the time. A hybrid amplifier is commonly used to realize the ETSM, which, in effect, partitions the envelope bandwidth into a low and high subband. An efficient switching buck converter tracks the low band. In parallel with it, an op-amp supplies the current in the high band. In prior arts, the hybrid amplifier is realized with feedback using a hysteresis comparator, whose output actuates the buck converter to respond to the changing envelope; the continuous-time op-amp makes up for the error. But a comparator-driven buck converter produces a slew-rate limited current that always lags the envelope waveform. This forces the op-amp to produce a larger current to correct the error, and the arrangement cannot guarantee that the buck converter switches no often than is absolutely necessary. We replace the hysteresis comparator with a novel trellis-search that, first, finds the optimal sequence to switch the buck converter to minimize the RMS current that the op-amp must deliver; second, to lower the loss from switching the capacitance of FETs, it penalizes a large number of switching events in the buck converter. Meanwhile, with a conventional on-chip hysteresis comparator, we can demonstrate ETSM operation up to 160 MHz modulation bandwidth. This is the widest bandwidth reported so far for any ETSM.

The dissertation of Weiyu Leng is approved.

Sudhakar Pamarti

Danijela Čabrić

Michael Stenstrom

Asad A. Abidi, Committee Chair

University of California, Los Angeles

2020

To my parents

TABLE OF CONTENTS

Acknowledgments	ix
Curriculum Vitae	x
1 Introduction	1
1.1 Motivation	1
1.2 Summary of Efficiency Enhancement Techniques	2
1.2.1 Dynamic Biasing (Class-B)	3
1.2.2 Load Modulation	3
1.2.3 Drain Bias Modulation	4
1.3 Envelope Tracking System Overview and Challenges	5
1.4 Dissertation Organization	6
1.5 Supply Modulator Architectures	7
1.5.1 Baseline Continuous-Time Hybrid Amplifier	7
1.6 Conventional Ways to Generate v_{sw}	10
1.7 Prior Arts to Reduce P_e	13
1.7.1 Method 1: Voltage Reduction	13
1.7.2 Method 2: Go Multibit	15
1.7.3 Method 3: Better Control	16
2 Novel DSP Algorithm for Envelope Tracking	21
2.1 Problem Formulation	21
2.2 Viterbi-Like Trellis Search	23
2.3 DSP Core Design	26

2.3.1	Pre-processor	27
2.3.2	BMU	28
2.3.3	TMU	29
2.3.4	TBU	32
2.4	FPGA Implementation	35
2.5	Systematic Study of the HA	36
3	Circuit Design	42
3.1	Design of the Buck Converter	42
3.2	Design of the Linear Amplifier	45
3.2.1	Op-Amp Requirements	45
3.2.2	Op-Amp Design Considerations	46
3.2.3	Op-Amp Simulation Results	51
3.3	DC-coupled Programmable Hysteresis Comparator	52
3.4	Chip Implementation	54
3.5	PCB Implementation	58
4	Lab Measurements	62
4.1	ETSM-Only Measurements	62
4.2	PA Characterization and Calibration for ET TX Measurements	65
4.3	ET TX Measurement Results	69
5	Supplemental Extension of the Author’s M.S. Thesis: Approximate Equivalent Circuits to Understand On-Chip Inductors	75
5.1	Introduction	75
5.2	Background of Extension	76

5.3	Improved Study on Inductor’s Conductor Losses	77
5.3.1	Skin Effect	77
5.3.2	Proximity Effect	82
5.3.3	Foster and Cauer Network Synthesis	88
5.4	Substrate Equivalent Circuit and Distribution Factor	91
5.5	Inter-winding Capacitance	93
5.6	Substrate Eddy Current	94
5.7	Definitions of Quality Factor	95
5.7.1	Physically Correct Equivalent Circuits	95
5.7.2	Apparent Q	97
5.7.3	True Q	99
5.7.4	Measuring Q_{tru} of On-Chip Inductor	105
5.8	Extra Case studies	106
5.8.1	Inductor Developed at a Semiconductor Company	107
5.8.2	Square Inductor and Design Space Exploration	109
5.8.3	Shielded Inductor	112
5.8.4	Tapered Inductor	115
6	Conclusions	122
6.1	Conclusion and Future Work of the Envelope-Tracking Supply Modulator Project	122
6.2	Conclusion and Future Work of the On-Chip Inductor Modeling Project	123
A	Important Verilog Codes	125
A.1	Verilog Codes for BMU and TMU	125
A.2	Verilog Codes for TBU	129

A.3	Verilog Codes for Delay Lines between TMU and TBU	131
A.4	Verilog Codes for VA Wrapper	132
B	Derivation from Proximity Effect's Universal Curves to the Equivalent Circuits . .	135
C	Calculation of Sec. 5.8.1's Reference Equivalent Circuit	136
	References	139

ACKNOWLEDGMENTS

In Fall 2010, I took EE10 (Circuit Theory I) with Professor Abidi, and embarked on the journey of circuit engineering. I cannot thank him enough for his mentorship during the past ten years, which not only provided tremendous help and guidance during my PhD studies but also will benefit me tremendously in my future endeavors.

I would also like to thank Professor Pamarti for his valuable input in helping me defining my research topic, which paved the path for my subsequent researches.

In addition, I would like to express gratitude to the rest of my committee: Professor Čabrić and Professor Stenstrom.

The following students offered me technical help during my PhD studies: Linqi Song, Sida Li, Wenlong Jiang, Dihang Yang, Kejian Shi, Alvin Chen, Hao Xu, Yan Zhang; I would like to thank all of them. Dr. Song brought up the idea of *Dynamic Programming* during our discussion, which prompted me to dive deeper on this topic and ultimately established Viterbi Algorithm as the key component of my research. Sida Li helped me construct and debug the Verilog codes for the SERDES link between FPGA and DAC. Dr. Zhang guided me through Verilog coding for FPGA.

I also want to thank the following people from Broadcom Inc.: Hooman Darabi, Sraavan Mundlapudi, Edward Roth, Guy Geshvindman, Ken Wong, Debopriyo Chowdhury, Ali Afsahi, Saeed Chehrazi, Mike DeGennaro, Richard Chen. Dr. Darabi and Dr. Afhasi offered me a much-appreciated chance of tapeout of my test chip. Sraavan Mundlapudi kindly shared his experience in ET with me during my time at Broadcom.

In the end, I would like to thank my parents for their understanding and unwavering support throughout this long march.

CURRICULUM VITAE

- 2009 – 2013 B.S. in Electrical Engineering, University of California Los Angeles (UCLA), Los Angeles, USA.
- 2013 – 2015 M.S. in Electrical Engineering, University of California Los Angeles (UCLA), Los Angeles, USA.
- 2015 – Present Ph.D. student in Electrical & Computer Engineering, University of California Los Angeles (UCLA), Los Angeles, USA.

CHAPTER 1

Introduction

1.1 Motivation

A modern wireless communication link consists of a transmitter (TX), a receiver (RX) and the channel. The TX generates a complex baseband signal, up-converts it to RF, and amplifies it in power. A power amplifier (PA) drives the final antenna with high power and is typically the most power-hungry block in the transceiver system. PA's efficiency largely determines the battery life of the mobile devices.

In the current and emerging mainstream wideband wireless communication standards, such as 802.11ax, 4G LTE and 5G NR, highly spectrum-efficient digital modulation schemes involving Orthogonal Frequency-Division Multiplexing (OFDM) are universally adopted. However, OFDM signals pose significant challenges for the PA. The signal's orthogonal sub-carriers add constructively or destructively, rendering a large peak-to-average power ratio (PAPR). Although many crest-factor-reduction (CFR) methods have been developed, they do not stop PAPR's increasing trend. Meanwhile, OFDM signals demand very high linearity. Since the RX uses FFT to demodulate the data, small distortion can lead to the failure of an entire OFDM symbol, rather than merely a few points on the periphery of the constellation. To ensure linearity for the high-PAPR signal, the PA must operate in deep back-off from saturation, which renders a low efficiency.

Envelope tracking (ET) is in wide use to raise the efficiency of PAs for high-PAPR signals. It operates on the principle that if PA's supply voltage tracks the RF waveform's envelope, the PA will operate close to saturation all the time, and therefore at its peak efficiency. A practical realization brings up many problems, chief being the realization of an envelope tracking supply modulator

(ETSM), whose own efficiency must be as high as possible. A hybrid amplifier (HA) is commonly used, which, in effect, partitions the envelope bandwidth into a low and high subband. An efficient switching buck converter tracks the low band, which contains most of the load current. In parallel with it, an op-amp with a Class-AB output stage supplies the current in the high band. Feedback servos the total current to the envelope waveform extracted from the baseband modulator.

In prior art the HA is realized with feedback using a hysteresis comparator, whose output actuates the two switches of the buck converter to respond to the changing envelope; the continuous-time op-amp makes up for the error. But a buck converter driven by a comparator that reacts to the instantaneous waveform produces a slew-rate limited current that, like a slewing op-amp, always lags the envelope waveform. This forces the op-amp to produce a larger current to correct the error, and the arrangement cannot guarantee that the buck converter switches no often than is absolutely necessary.

Many prior arts have tried to improve the ETSM at the circuit level by diversifying the forms of energy storage with more off-chip passives. However, the module space is very stringent in modern RF front-ends, so more passives may not always be available. On the other hand, the performance of CMOS digital signal processing (DSP) circuits has been improved a lot recently, so we seek to make full use of the DSP advancements to find an optimal control method for the buck converter with the minimum 1 off-chip inductor, such that we can leverage the digital process advancements over the module space. We want to find and achieve the theoretical upper limit of the minimum 1-inductor architecture.

1.2 Summary of Efficiency Enhancement Techniques

In a Class-A PA, when the output signal backs off from the peak swing, the bias current and the drain voltage of the PA stay the same, so the input power does not decrease (Fig. 1.1(a, b)). This leads to a quadratic back-off on efficiency with respect to the output voltage swing. How can we reduce the input power together with the output power? There are two dimensions we can explore: current and voltage.

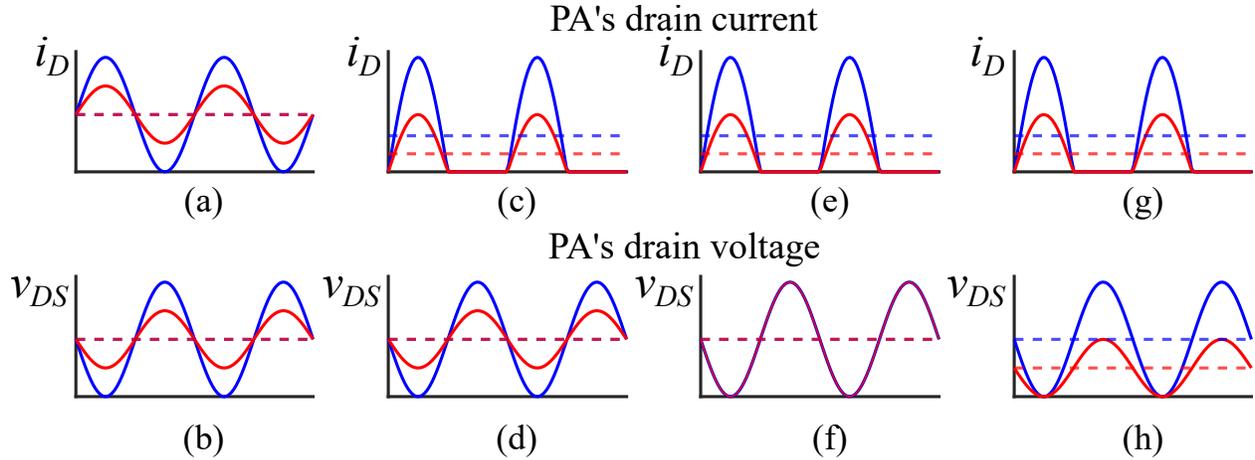


Figure 1.1: PA's drain current waveforms and drain voltage waveforms with their respective mean values of (a, b) Class-A PA; (c, d) Class-B PA; (e, f) Load modulation PA; (g, h) Envelope tracking PA

1.2.1 Dynamic Biasing (Class-B)

Dynamic biasing explores the dimension of drain current to improve the back-off efficiency of Class-A PA. If the average drain current of the PA can be reduced together with the signal current swing, we can achieve a linear back-off on the input power when output power backs off quadratically, as the drain bias voltage stays constant (Fig. 1.1(d)). This renders a linear back-off in efficiency. In practice, dynamic biasing can be achieved using the intrinsic rectifying characteristic from the input voltage to the transistor current when the transistor is biased on 50% duty cycle (Fig. 1.1(c)). This is the well-known Class-B PA.

1.2.2 Load Modulation

Dynamic load modulation explores the dimension of drain voltage to improve the back-off efficiency on top of the Class-B biasing. Conceptually, the load modulation PA, such as the Doherty PA [1] or the Chireix outphasing PA [2], increases the apparent load resistance when the signal current swing drops, keeping the drain voltage swing at maximum (Fig. 1.1(f)). Then, the PA's output power backs off linearly, rather than quadratically, when the output signal current swing backs off linearly. Meanwhile, the input power backs off linearly, so ideally the load modulation PA with Class-B bias maintains its peak efficiency all the time. In real implementation, modulation

on the apparent driving-point impedance is realized by RF signal combination [3].

1.2.3 Drain Bias Modulation

The other way to explore the dimension of drain voltage is directly reducing the bias voltage (DC average of v_{DS}) together with the output voltage swing. Fig. 1.1(g, h) show that when both the drain current swing and drain voltage swing are kept at maximum, the maximum efficiency can be maintained. In another perspective, the input power backs off quadratically when the output power backs off quadratically, so ideally PA also keeps its peak efficiency all the time. This is called Envelope Tracking (ET), a more practical implementation of Kahn's envelope-elimination-and-restoration (EER) architecture in [4].

The EER architecture is first proposed in [4] that a switching PA only amplifies the PM part of the RF signal, while a supply modulator modulates the drain of the switching PA to overlay the AM part on the PM signal at PA's output. EER is rarely adopted in commercial electronics, because of several fundamental challenges:

1. IQ-PM transformation extends the bandwidth even more than IQ-AM transformation does (Fig. 1.3);
2. The supply modulator needs to be extremely linear and wideband, because the PA is in complete saturation, and drain bias modulation is the only way to faithfully reproduce the entire AM information;
3. AM-PM distortion characteristic changes with respect to supply voltages;
4. AM and PM signals must be accurately aligned in time.

In ET, PA still operates close to saturation for different supply voltages, so 2)-4) of the aforementioned challenges of EER still exist. However, since the PA is not driven *deep* into saturation (switching mode), the requirements are much more relaxed. In this work, we focus on ET, which is marked as the most promising efficiency enhancement architecture in [5, Fig. 9.21].

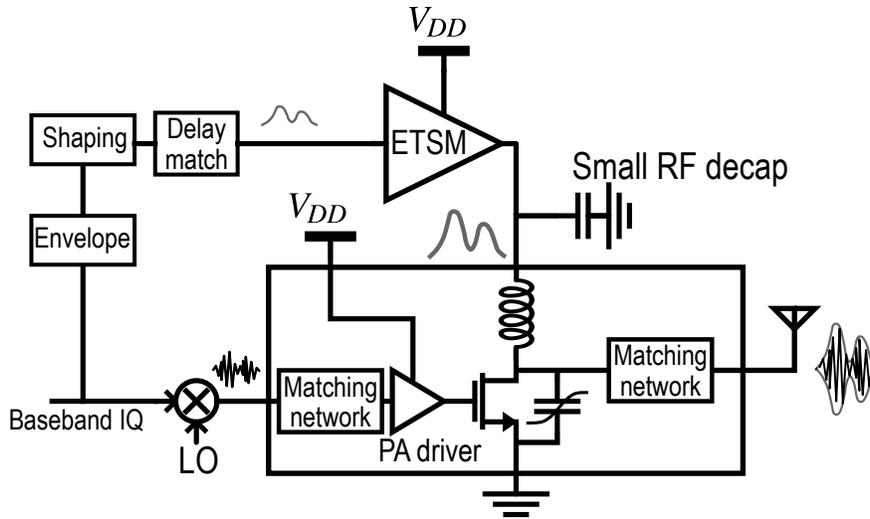


Figure 1.2: Block diagram overview of an ET system

1.3 Envelope Tracking System Overview and Challenges

As discussed in Sec. 1.2.3, ET is a method to improve the efficiency of a linear PA dealing with high PAPR signals. An envelope tracking supply modulator (ETSM) modulates the PA's supply to keep PA close to saturation all the time. This will raise PA's back-off efficiency (Fig. 1.2). Ideally, the PA should operate in Class-B mode for maximum efficiency (Fig. 1.1(g)). But practically, a Class-AB bias is often used for linearity concerns, such as reducing AM-PM distortion. Also, when the PA operates in linear mode as a waveform-engineered RF current source, its bias current, or its gain in effect, is theoretically not affected by supply variations. But in reality, modulating PA's supply voltage introduces small bias variations that may change the gain by as large as 2dB. Also, changes in PA's drain capacitance causes AM-PM distortion.

These are the main challenges of the ETSM in Fig. 1.2:

1. The bandwidth of the envelope signal extends the signal's baseband bandwidth as shown in Fig. 1.3. So the bandwidth of the ETSM should at least be $3 \times$ IQ signal's baseband bandwidth [6].
2. The ETSM must supply the entire drain current of the PA, so its own efficiency must be as high as possible. Otherwise, the overall efficiency, which is equal to the *product* of the theo-

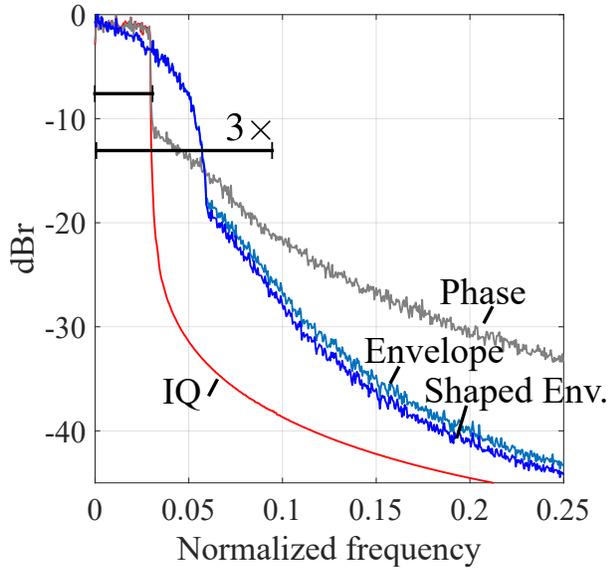


Figure 1.3: Bandwidth extension of IQ-polar transformation. Envelope shaping will not significantly reduce the bandwidth extension.

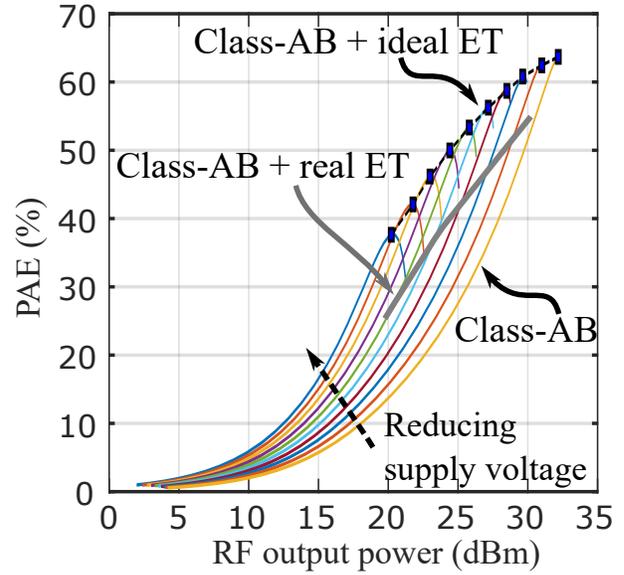


Figure 1.4: Efficiency enhancement of ET on practical Class-AB PA

retically enhanced efficiency of the PA and the efficiency of the ETSM, will be no different than the original efficiency of the PA (Fig. 1.4).

3. The ETSM must act as a voltage source with low output impedance. Equivalently, its output must have enough SNDR. Otherwise, the distortion on the supply waveform will affect the RF output. For a static supply, low impedance is generated with a large bypass capacitor on PA's supply, but large capacitor cannot be modulated by wideband signals, so the ETSM must provide low output impedance actively.

Due to these requirements, the design of the ETSM remains the major challenge for any ET system and is the focus of our work.

1.4 Dissertation Organization

This dissertation focuses on exploring techniques to improve the performance of the CMOS ETSM. It is organized as follow:

In the remaining of Chapter 1, existing ETSM architectures are reviewed, which leads to the target of our research.

Chapter 2 presents an innovative Viterbi-like trellis-search DSP algorithm to control the ETSM. This algorithm can push the existing ETSM architecture to its theoretical limit.

Chapter 3 includes the circuit design steps and choices of the ETSM for the 160MHz WiFi application. The ETSM chip is also compatible with the DSP method in Ch. 2.

Chapter 4 discusses the measurement setup and the measurement results of the ETSM chip under different operating modes.

Lastly, Chapter 5 attaches a side project extended from the author's Master Thesis on inductor modeling and optimization. The novel contents during the author's Ph.D. study are included.

1.5 Supply Modulator Architectures

1.5.1 Baseline Continuous-Time Hybrid Amplifier

For a buck converter, if the switching signal v_{sw} is an oversampled PWM signal, the input signal's spectrum will be copied to harmonics of f_s , where f_s is frequency of the PWM sampling waveform. Quantization noise level can be reduced by increasing the oversampling ratio. Then a 2nd-order low-pass filter (LPF) will pass the baseband spectrum, reject the harmonics, and also reduce the integrated quantization noise, rendering the output waveform in Fig. 1.5(a). However, if the input signal's bandwidth grows, or if f_s is limited by the buck converter's implementation, the spectra overlap, and the output waveform is distorted (Fig. 1.5(b)). This often happens for ET, because of the bandwidth extension in IQ-AM transformation. Also, we cannot choose f_s exactly as Nyquist rate, because a 2nd-order filter is not sharp enough to cut off right at $f_s/2$. To solve this problem, a hybrid architecture (Fig. 1.5(c)) replaces the load capacitor with a voltage source, but use it, with proper feedback, like a capacitor, which rejects low-frequency current, and compensates high-frequency current. In this case, v_{sw} does not need to be an oversampled PWM signal, and the output voltage is not distorted.

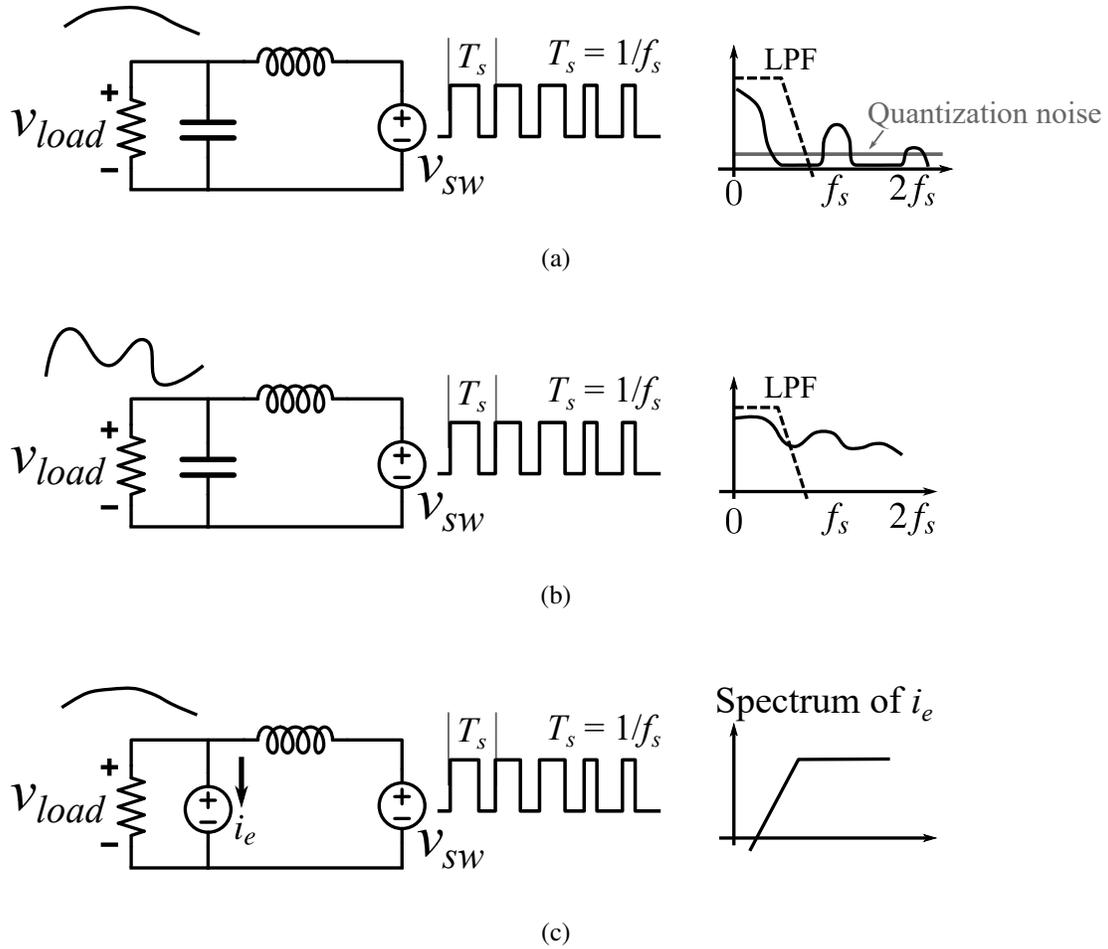


Figure 1.5: Illustration of buck converter's operation in (a) oversampling mode, (b) aliasing mode, and (c) hybrid mode

Fig. 1.6 shows the baseline realization of Fig. 1.5(c): a hybrid amplifier (HA) consisting of an op-amp in voltage feedback and a buck converter in current feedback. The op-amp dictates the supply voltage for PA, ideally with zero current (i_e). The buck converter supplies PA's bias current that varies with the envelope (Sec. 1.2.1), ideally with zero voltage error. This way, we get the bandwidth and linearity of the op-amp, and the efficiency of the buck converter at the same time. However, if the switching frequency of the buck converter $f_{sw} < \infty$, $i_e \neq 0$, and hence the power loss P_e in supplying i_e is not equal to 0.

Fig. 1.7 shows a simplified equivalent circuit to analyze the efficiency of the HA. For simplicity, we assume that the quiescent current of the op-amp is 0, such that the op-amp operates in ideal

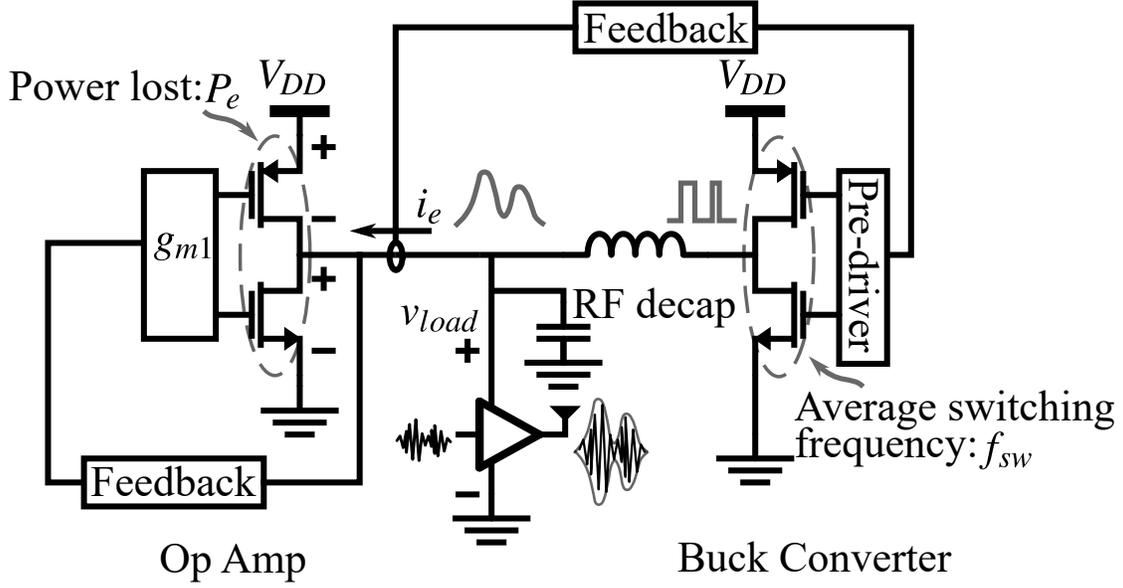


Figure 1.6: Baseline architecture of a hybrid amplifier (HA)

push-pull mode. The op-amp has infinite loop gain, so with the load conductance defined as G_{load} :

$$\text{Load voltage: } v_{load} = \alpha V_{DD} \quad (1.1)$$

$$\text{Load current: } i_{load} = \alpha G_{load} V_{DD}, \text{ where } \alpha \leq 1 \quad (1.2)$$

The inductor is lossless and carries a DC (for now) current of

$$i_{sw} = \alpha_0 G_{load} V_{DD}, \text{ where } \alpha_0 \leq 1 \quad (1.3)$$

α and α_0 are the back-off factors. To satisfy the circuit theory, the inductor is assumed to be in volt-second balance. Then, the theoretical efficiency of the HA can be derived as a piecewise function:

$$\eta = \begin{cases} \frac{\alpha}{\alpha_0}, & \text{if } \alpha \leq \alpha_0 \\ \frac{\alpha^2}{\alpha - \alpha_0 + \alpha \alpha_0}, & \text{if } \alpha_0 < \alpha \leq 1 \end{cases} \quad (1.4)$$

Fig. 1.8 overlays the efficiency curves for different α_0 's with respect to α on top of the distribution of an OFDM waveform's normalized envelope. Now we can see that if L is extremely large such that i_{sw} is very close to a DC current, the efficiency will back off on two sides from the peak value. But we can still get a reasonably high average efficiency if we set α_0 to be the most probable α .

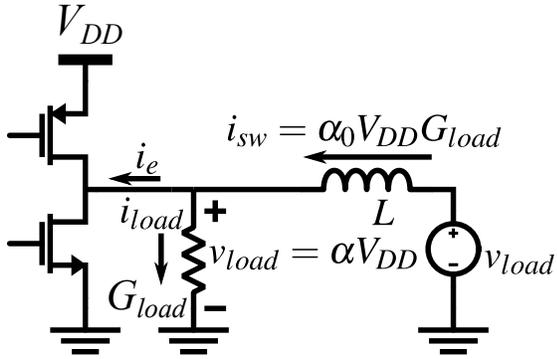


Figure 1.7: Simplified equivalent circuit to analyze the efficiency of the hybrid amplifier

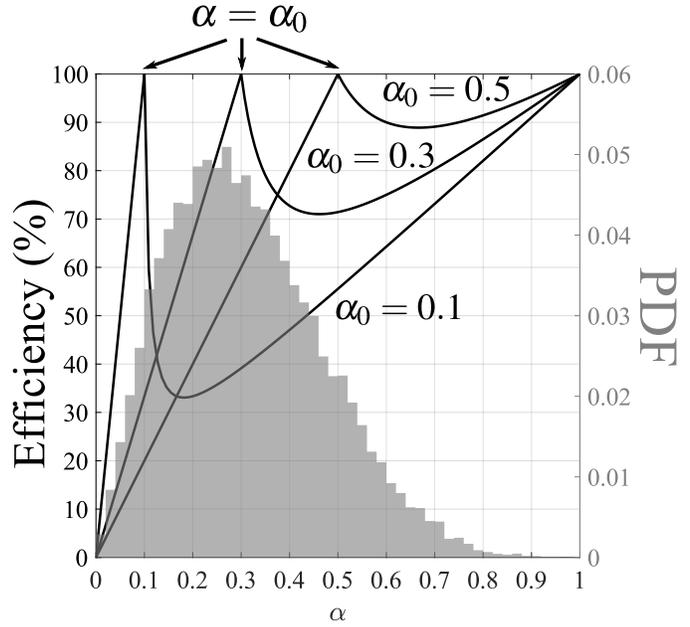


Figure 1.8: Efficiency curves for different inductor currents overlaid on top of probability distribution of a typical OFDM signal

On top of that, we can achieve even higher efficiency if L is reduced such that α_0 can be dynamically varied, with feedback, to track the instantaneous α . In other words, since $v_{load} = \alpha V_{DD}$ is pre-determined, we need to use feedback to find a switching signal v_{sw} such that $i_e \rightarrow 0$ as much as possible. Then P_e can be minimized.

1.6 Conventional Ways to Generate v_{sw}

The control signal v_{sw} for the buck converter can be generated by two kinds of feedbacks:

1. PWM control [7, 8] as shown in Fig. 1.9(a): filtered the mirrored i_e is compared with a sawtooth waveform of fixed frequency f_s to generate v_{sw} ;
2. Hysteresis control [9, 10] as shown in Fig. 1.9(b): scaled copy of the op-amp's compensating current i_e is fed into a hysteresis comparator directly to generate v_{sw} .

[11, Table I] shows a comparative study on these two control methods. It can be summarized that

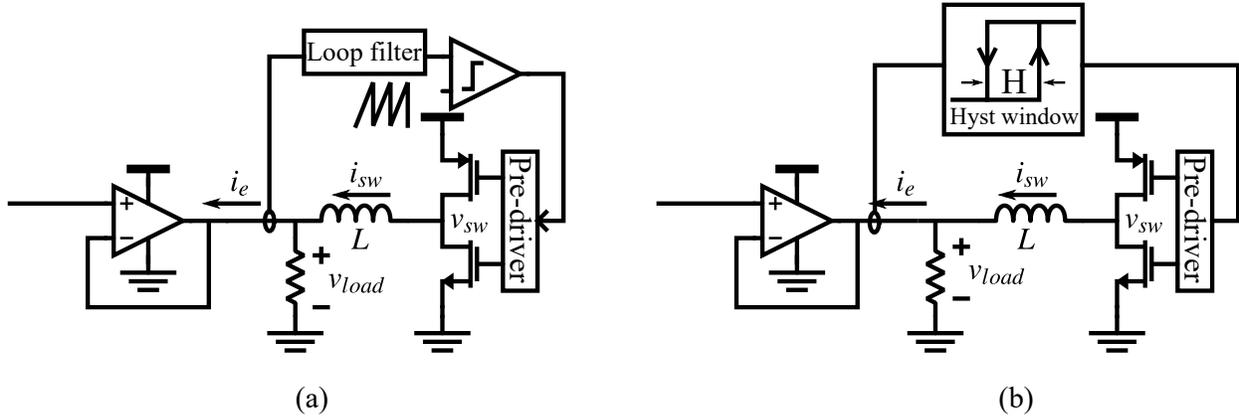


Figure 1.9: i_e is minimized by controlling buck converter's v_{sw} with (a) PWM or (b) hysteresis logic

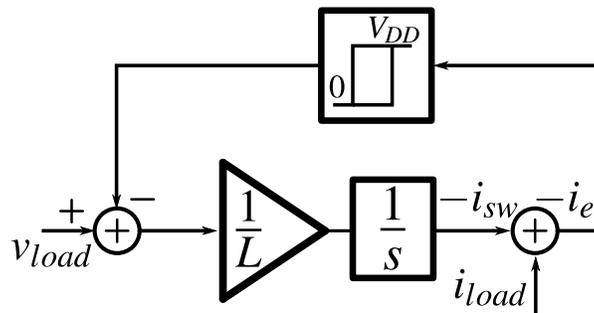


Figure 1.10: HA controlled by a hysteresis comparator is essentially an asynchronous $\Sigma\Delta$ -modulator

hysteresis control generally offers a wider tracking bandwidth than the PWM control does. PWM control with a sawtooth sampling signal is naturally an over-sampling system. The bandwidth of the loop filter must be a small fraction of the sampling frequency f_s to ensure stability. But in this case f_s is also the switching frequency of the buck converter, which cannot be high naturally. On the other hand, Fig. 1.10 shows that the hysteresis comparator and the inductor form a continuous-time asynchronous $\Sigma\Delta$ -modulator for v_{load} , and a Δ -modulator for i_{load} . The average switching frequency f_{sw} of the control loop is set by the hysteresis window H and the inductance L . This loop is robust and unconditionally stable, so the hysteresis control method is more popular in recent state-of-the-art designs. The other way to look at the hysteresis loop is that it is a self-running relaxation oscillator, whose oscillation frequency is determined by the hysteresis window H and the integrator's scaling factor $1/L$.

Fig. 1.11 shows the limitations of the conventional hysteresis method. The slew rates of i_{sw}

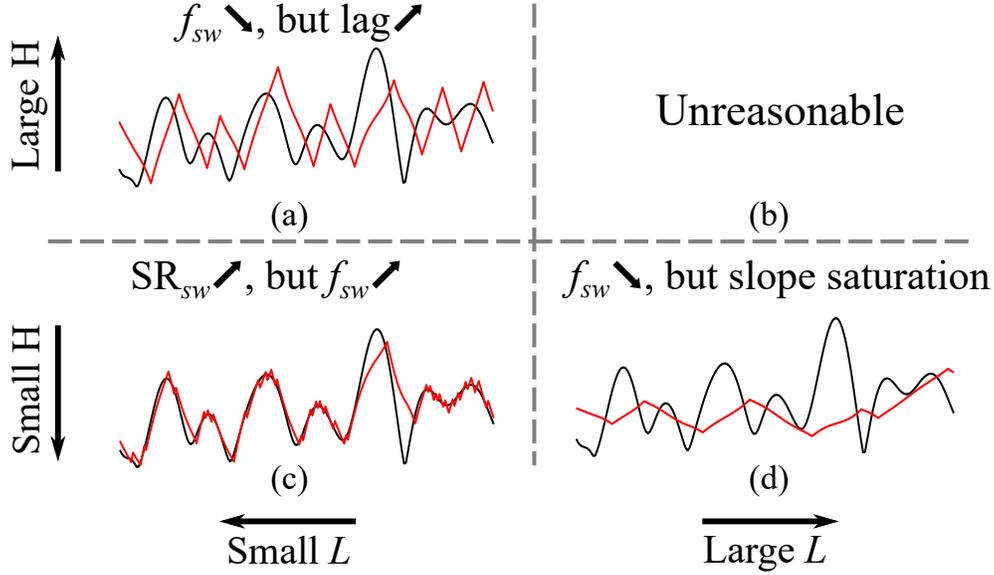


Figure 1.11: 4 modes of operation for a hysteresis controlled HA: (a) Delay-limited; (b) Unreasonable; (c) Oversampling; (d) Slope-saturated

and i_{load} are defined as SR_{sw} and SR_{load} respectively. SR_{sw} is limited to $((V_{DD}, 0) - v_{load})/L$. If $SR_{sw} \gg SR_{load}$, and H is chosen to be small as shown in Fig. 1.11(b), i_{sw} can track i_{load} very well with small i_e , by oscillating in a small window H around i_{load} . If the buck converter is ideal, and we solely want to minimize i_e , this mode should be chosen. However, this mode demands an f_{sw} much higher than the envelope's bandwidth. For a practical buck converter, large f_{sw} reduces its efficiency, so the reduction in P_e may be surpassed by the increased loss in the buck converter.

How should we limit f_{sw} to ease the design of the buck converter? Firstly, if we keep L and reduce H , f_{sw} is also reduced as shown in Fig. 1.11(a), and SR_{load} is not sacrificed. But a larger H results in a lagging response to the instantaneous i_{load} variation. Also, at the occasions when $SR_{sw} \gg SR_{load}$, the swing of i_e will be larger, due to the wider limiting window. These effects increase P_e . Secondly, if we keep H and increase L such that $SR_{sw} \ll SR_{load}$, f_{sw} will be reduced as shown in Fig. 1.11(d), but the Δ -modulator for i_{load} now operates completely in slope-saturation mode, and i_{sw} is like a quasi-constant current source. Then, i_e will be larger, so does P_e , because α_0 in Fig. 1.8 loses track on the instantaneous α . To summarize, since the loop is self-oscillating, there is no guarantee that the buck converter switches no often than is absolutely necessary to achieve a desired P_e .

1.7 Prior Arts to Reduce P_e

To break the limitations of the baseline hysteresis controlled HA described in Sec. 1.6, prior arts have developed 3 main solutions. As shown in Fig. 1.6, P_e is determined by two factors: the error current i_e and the voltage drop across the op-amp's output transistors v_{ds} . Thus, without increasing f_{sw} , we can work on 3 dimensions:

1. Reduce v_{ds} of op-amp's output transistors (Fig. 1.12(a));
2. Reduce i_e with multiple quantization bits on v_{sw} (Fig. 1.13(a));
3. Find a better control method to generate v_{sw} other than the two described in Sec. 1.6.

Now, we discuss the advantages and disadvantages of these 3 methods with examples from the literature.

1.7.1 Method 1: Voltage Reduction

[10] proposes a cascade HA as shown in Fig. 1.12(b). A slowly switching auxiliary buck converter is added to the supply of the op-amp and tracks the “envelope” (denoted as v_{supply}) of the envelope of the RF signal. Since v_{supply} has a much lower bandwidth than that of v_{load} [10, Fig. 8, 23], the auxiliary buck converter does not need an additional op-amp to servo v_{supply} . Or equivalently, the auxiliary buck converter operates in oversampling PWM mode. [10, Fig. 28] shows that this approach can increase ETSM's efficiency by 4%.

[12] recognizes the fact that the knee voltage of its CMOS PA load is high such that the practical v_{load} for ET does not swing down to GND. As shown in Fig. 1.12(c), a buck converter is inserted between V_{DD} and V_{SS} of the op-amp to create another low-voltage rail. When the op-amp is sinking current ($i_e > 0$), the instantaneous P_e is reduced. [12, Fig. 14] shows that the ETSM's efficiency can be improved by 5% at 6dB back-off.

[13] takes another route by reducing the supply voltage of the op-amp with a buck-boost converter (instead of buck to deal with real batteries). But this will limit the peak swing of the op-amp's

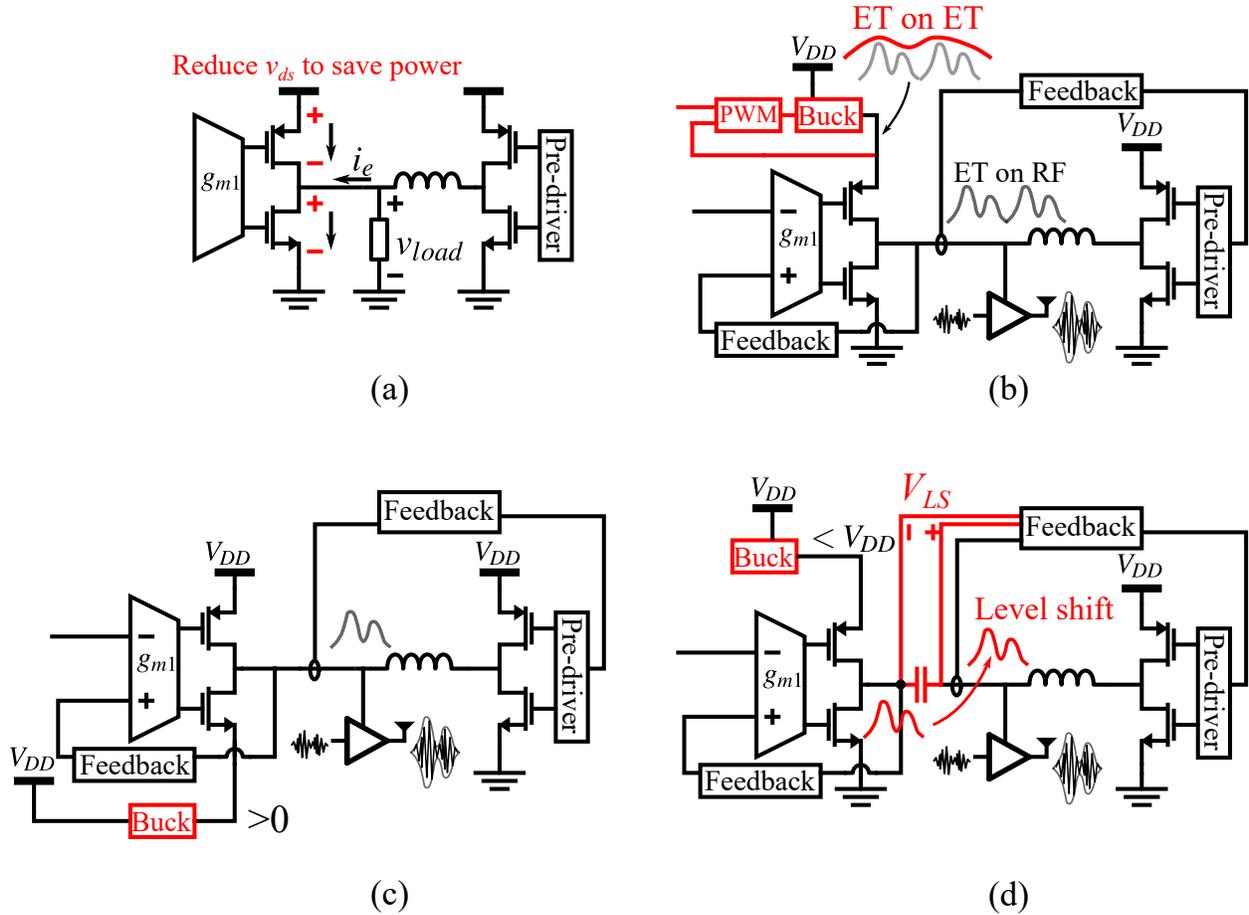


Figure 1.12: (a) When i_e cannot be reduced further without increasing f_{sw} , v_{ds} of the op-amp is reduced to save P_e by (b) Modulating op-amp's supply, (c) Raising V_{ss} of the op-amp, and (d) Reducing the supply of the op-amp and AC couple the op-amp's output to PA's supply.

output. As [12] has proven, the actual v_{load} may have a limited swing not to disturb the linearity of the original PA, so the op-amp's output with reduced swing can be level-shifted up to PA's supply with an AC coupling capacitor (Fig. 1.12(d)). The control logic to generate v_{sw} needs to be modified accordingly, because the logic needs to control both the i_e and the level-shifting voltage V_{LS} . A common problem of [10, 12, 13] is that they all need additional off-chip inductors to build the auxiliary buck converters, static or modulated. Typically, the inductance needed in the auxiliary buck converter is even larger than L of the main buck converter.

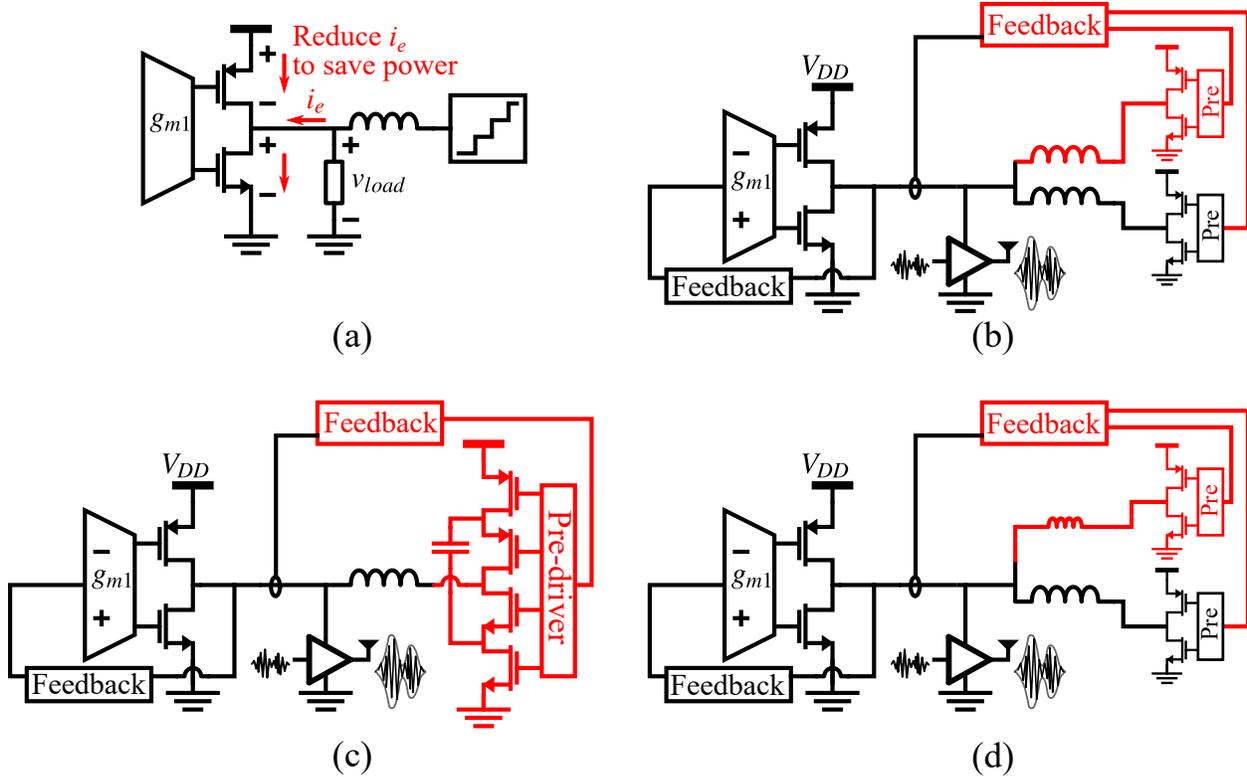


Figure 1.13: (a) Increase effective quantization bits on v_{sw} by (b) Multiphase switching, (c) Multilevel switching, and (d) Multiple subbands.

1.7.2 Method 2: Go Multibit

The baseline HA shown in Fig. 1.6 uses a binary quantized v_{sw} , which gives 2 possible SR_{sw} 's, $((V_{DD}, 0) - v_{load})/L$. If v_{sw} can be quantized to more levels, more SR_{sw} 's become available, enabling a better interpolation on i_{load} . In [8], multibit is achieved by using 2 inductors of inductance $2L$ (Fig. 1.13(b)). This way, SR_{sw} can be $((V_{DD}, V_{DD}/2, 0) - v_{load})/L$. But f_{sw} of each buck converter does not need to be higher, because the middle level $V_{DD}/2$ is generated by the intrinsic XOR function on two outphased binary signals. [14] shows another variation of multiphase switching: multiple subbands, as shown conceptually in Fig. 1.13(d). Two inductors of drastically different inductance are jointly controlled. The buck converter with the larger inductor supplies of the low subband of i_{load} , including DC, and switches slowly. The buck converter with the smaller inductor supplies of the high subband of i_{load} and switches fast. The two buck converters can be optimized for specific switching frequencies. The op-amp supplies any residual highest subband of i_{load} and

absorbs the ripple. Instead of quantizing directly with 3 levels (1.5 bits), this architecture essentially assigns 1 bit for i_{load} 's low subband and 1 bit for i_{load} 's high subband, so it is an intrinsically multibit solution.

Multiphase and multiple subbands both require additional inductors. To save the number of inductors, [15] uses an on-chip switched-capacitor multilevel converter to directly create more quantization levels (Fig. 1.13(c)). However, this topology needs a 12 nF on-chip flying capacitor and sophisticated feedback control to regulate the capacitor's voltage to $V_{DD}/2$. The apparent f_{sw} is increased for a multilevel converter, but the switching voltage difference is reduced, so effectively, we can say that i_e is reduced while maintaining f_{sw} .

The 3 realizations in Fig. 1.13 can be combined arbitrarily for cumulative improvement. For example, [16] uses a 3-phase 3-level buck converter for the envelope of 20 MHz LTE signal, and [17] has proven that a 2-phase 3-level buck converter with filtering is good enough to track the envelope of 20 MHz LTE signal, even without a compensating op-amp (Fig. 1.5(a)). The methods in Sec. 1.7.2 can also be combined with methods in Sec. 1.7.1, as in [6].

We can notice that the methods described in Sec. 1.7.1 and Sec. 1.7.2 share a common feature: they need extra energy-storage passives, i.e. large inductors or capacitors, to partition and condition the power drained from the supply. An inductor whose current is regulated by volt-second balance behaves like a lossless current source, and a capacitor whose voltage is regulated by amp-second balance behaves like a lossless voltage source. More energy-storage passives provide more diverse ideally lossless energy sources for the load, so that the op-amp does not have to compensate the extra energy in a lossy way. But there are costs: complicated nested loops are necessary for multiple passives contributing to the total current to coordinate timing perfectly, which can lead to distortion in the load current; and off-chip passives, especially inductors, add to the physical volume, which is at a premium in miniaturized modules.

1.7.3 Method 3: Better Control

The last way to reduce P_e without raising f_{sw} is designing a better control method for the buck converter such that it switches at more appropriate time instances. [7, 8] realize retiming by “ex-

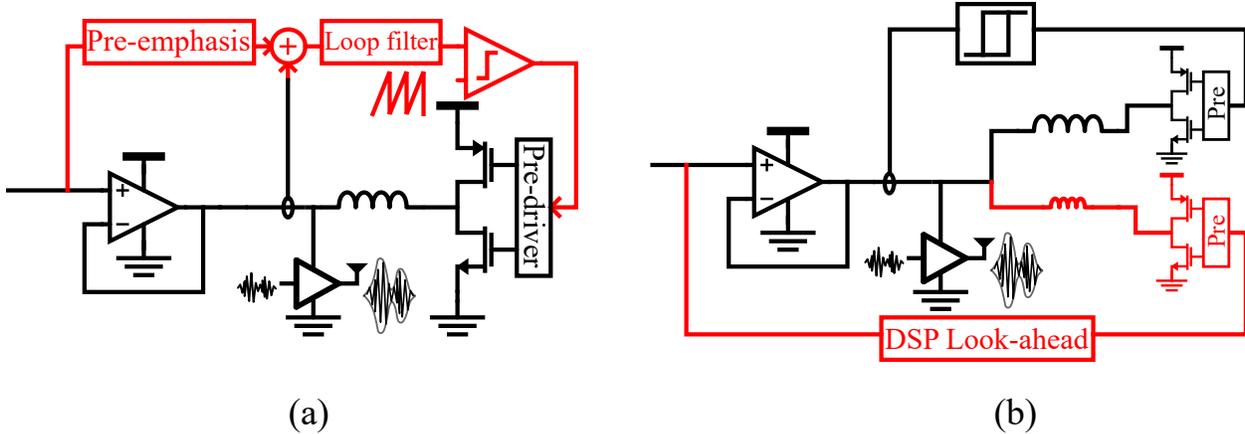


Figure 1.14: (a) Pre-emphasis before PWM to “extend” the buck converter’s bandwidth. The change of dimension has been ignored for simplicity. (b) DSP-assisted v_{sw} generation

tending” the bandwidth of the buck converter through a feed-forward path. A pre-emphasized version of the input envelope signal is added to the sensed i_e prior to the PWM controller, as shown in Fig. 1.14(a). However, the effectiveness of the pre-emphasis remain doubtful, because [7, 8] use multiple inductors or capacitors at the same time. A common design challenge of this method is that the loop’s stability depends on matching between the gains of two paths at the summation point in Fig. 1.14(a), which may vary with respect to different loadings.

In contrast, [18] provides a more direct approach to solve the lag problem of the conventional hysteresis comparator described in Sec. 1.6. A look-ahead hysteresis logic is proposed in DSP where the k -th i_{load} sample ahead of the current sample is used to generate v_{sw} :

$$\begin{aligned}
 &\text{If } i_{sw}[n] < i_{load}[n+k] - H, \\
 &\quad v_{sw} = 1 \text{ (pull up);} \\
 &\text{else if } i_{sw}[n] > i_{load}[n+k] + H, \\
 &\quad v_{sw} = 0 \text{ (pull down);} \\
 &\text{else } v_{sw}[n] = v_{sw}[n-1] \text{ (no change)} \tag{1.5}
 \end{aligned}$$

If (1.5)’s look-ahead index k is set to 0, it is no different than a conventional hysteresis comparator. [18]’s logic is very similar to the dynamic threshold hysteresis comparator described in [19, Fig. 3], where the hysteresis window for partial-response maximum-likelihood detection is moved up or

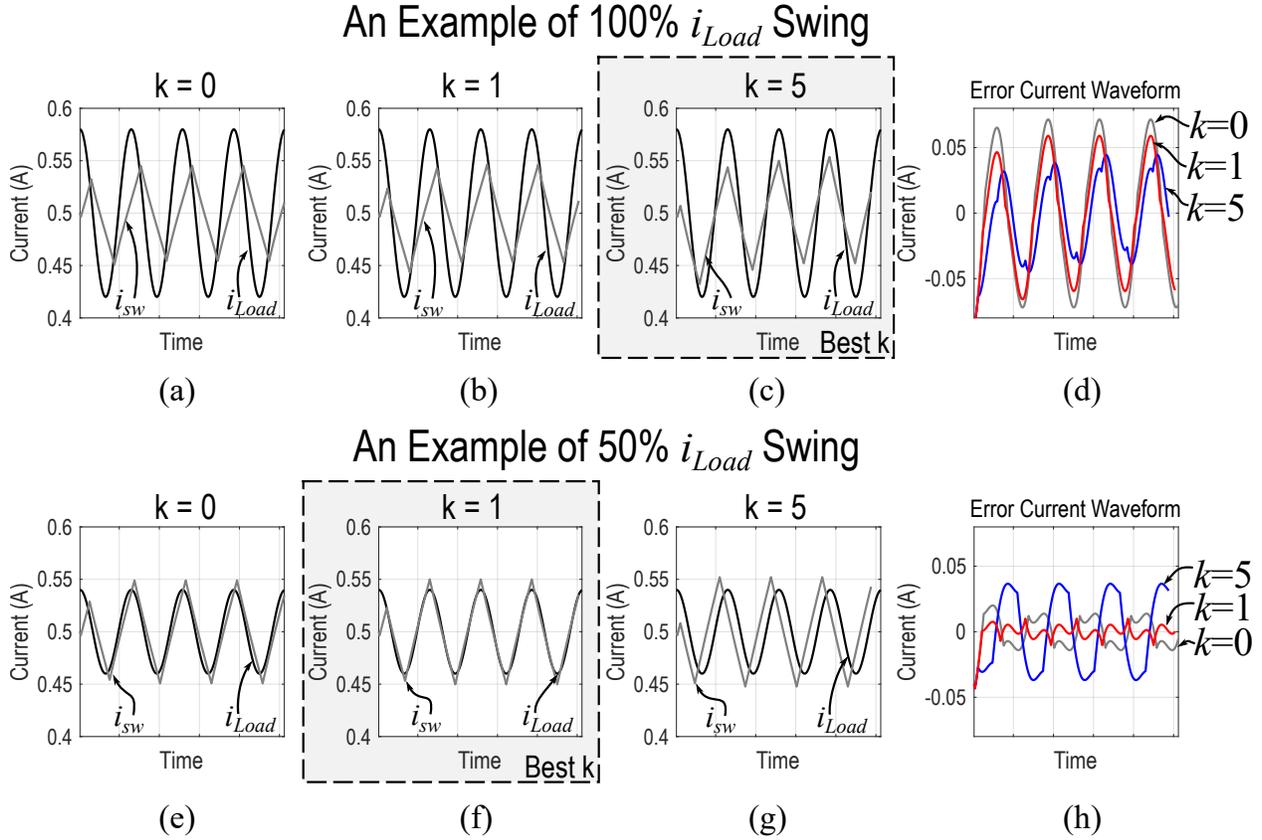


Figure 1.15: Experimental results that illustrate how optimal k can be different for different i_{load} swings on waveform level: (a-d) 100% swing (e-h) 50% swing

down depending on the immediate decision made. (1.5) modulates the hysteresis window's position by $\Delta i_{load} = i_{load}[n+k] - i_{load}[n]$, and v_{sw} toggles earlier when there is an incoming peak or trough. For example, conventionally when v_{sw} needs to toggle from 0 to 1, i_e has to hit $-H/2$. But in (1.5)'s look-ahead, if i_{load} is about to rise ($\Delta i_{load} > 0$), the lower threshold is shifted up to $-H/2 + \Delta i_{load}$, causing v_{sw} to flip high earlier, such that i_{sw} better catches i_{load} 's arriving peak.

As shown in Fig. 1.14(b), [18] uses the look-ahead algorithm on a buck converter with a smaller inductor that covers the highband. Since (1.5) generates v_{sw} in an open-loop fashion, it is impossible to achieve volt-second balance across the small inductor because of DC imbalance. Thus, a slow buck converter controlled by a hysteresis comparator in feedback is added to source a nearly constant current, and the fast buck converter driven by (1.5) supplies most of i_{load} 's AC part, with the rest compensated by the op-amp. Due to this design choice, it is a bit hard to judge whether

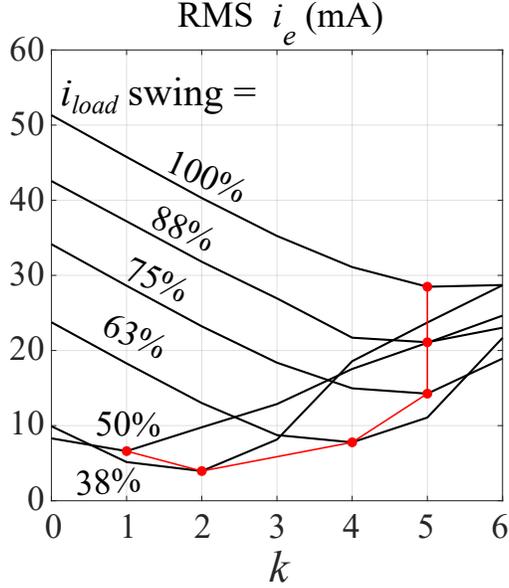


Figure 1.16: Experimental results that illustrate how minimum RMS i_e 's are achieved with different optimal k 's for different i_{load} swings

the 6% improvement shown in [18, Fig. 17] is caused by the extra inductor or the look-ahead algorithm.

Our simulations have shown that, if we avoid DC imbalance by inserting a large DC block capacitor at the op-amp's output, with a single inductor and an optimal k , the look-ahead hysteresis logic can already reduce P_e by roughly 25%, compared to P_e in the conventional hysteresis loop at the same f_{sw} . This observation raises our curiosity, because (1.5) does not find the global optimum of v_{sw} .

We illustrate the reasoning for the sub-optimum solution of v_{sw} from (1.5). An experiment is run with a sinusoidal i_{load} waveform, upsampled by 16. Two cases are compared: 100% i_{load} amplitude where $SR_{sw} < SR_{load}$ and 50% i_{load} amplitude where $SR_{sw} \approx SR_{load}$. Fig. 1.15(a, e) show i_{load} and i_{sw} of the two cases from conventional hysteresis comparator ($k = 0$). Then, Fig. 1.15(b, c, f, g) show that the optimal k for 100% swing is 5, while the optimal k for 50% swing is 1. The calculated i_e waveforms are included in Fig. 1.15(d, h) for visualization. The optimal k that results in the lowest i_e (or P_e equivalently) is a function of the specific i_{load} 's swing for each excursion. In a random waveform, we can only choose a best k over the entire packet on average as shown in Fig. 1.16, but if we can find a way to optimize v_{sw} for every i_{load} excursion, we can achieve the

globally optimized switching and fully explore the capability of the single-inductor HA. Then, we can go one step further than the prior art. At least, we will be confident that we are indeed at the optimum.

CHAPTER 2

Novel DSP Algorithm for Envelope Tracking

2.1 Problem Formulation

As discussed in Sec. 1.7.3, [18]’s look-ahead DSP algorithm to optimize v_{sw} is ad-hoc and proven suboptimal. Thus, our objective is finding a DSP tool that dynamically optimizes the switching instance of v_{sw} for every excursion of i_{load}

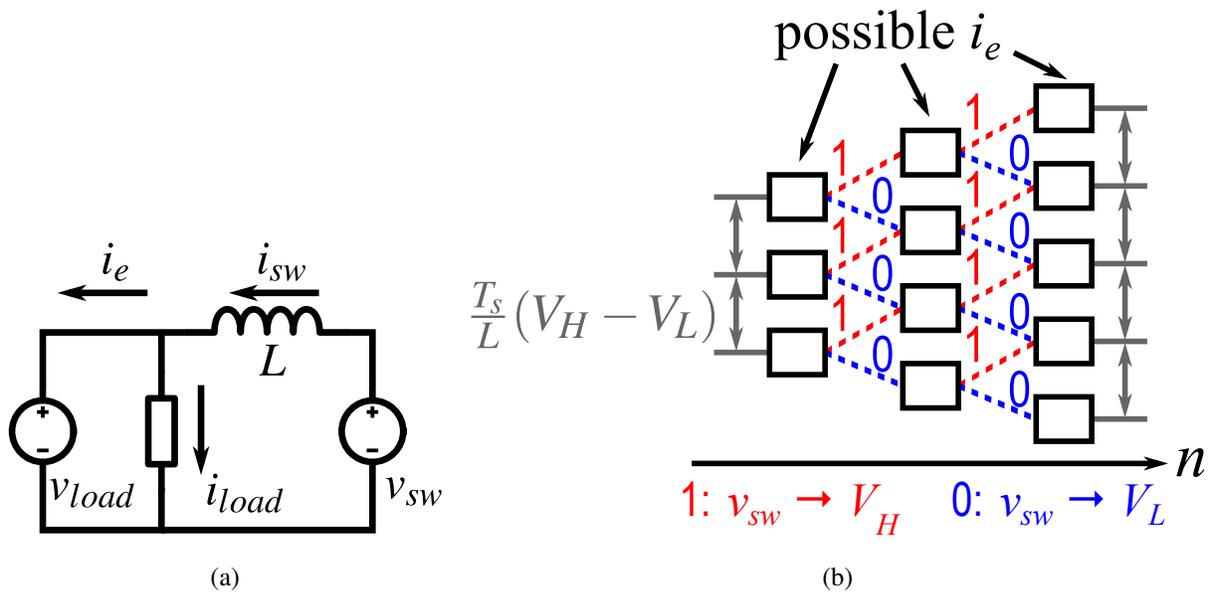


Figure 2.1: (a) Simplified schematic of the HA for DSP formulation; (b) Possible values of i_e are equally spaced, and the total count adds 1 for each time sample

Fig. 2.1(a) shows the simplified schematic of the hybrid amplifier (HA). Its governing KCL is:

$$i_e(t) = i_{sw}(t) - i_{load}(t) \quad (2.1)$$

$$= \frac{1}{L} \int_0^t (v_{sw}(\tau) - v_{load}(\tau)) d\tau - i_{load}(t) \quad (2.2)$$

(2.2) can be re-written in discrete-time form:

$$i_e[n] = i_{sw}[n] - i_{load}[n] \quad (2.3)$$

$$= \underbrace{\frac{T_s}{L} \sum_{i=0}^{n-1} v_{sw}[i]}_{\text{unknown}} - \underbrace{\left(\frac{T_s}{L} \sum_{i=0}^{n-1} v_{load}[i] + i_{load}[n] \right)}_{\text{known}}, \quad (2.4)$$

where T_s is the sample time, and $v_{sw} = V_H$ or V_L

The known part of (2.4) can be created easily as an input to the DSP core. V_H and V_L represent the two voltage levels of a buck converter. If the inductor and the switches are ideal, $V_H = V_{DD}$ and $V_L = 0$. If the load is resistive, $i_{load} = G_{load} \times v_{load}$. For a real PA, the relationship between i_{load} and v_{load} may not be linear, so a look-up table (LUT) is necessary to calculate i_{load} .

Since v_{sw} is quantized (in this case binary), at any time sample n , i_e can only be at discrete values that are equally spaced by $T_s/L \cdot (V_H - V_L)$, due to the accumulation operation in the unknown part of (2.4). Moreover, the number of possible values of i_e at time sample $(n+1)$ will be more than the number at time sample n by 1. For example, if we assume a zero initial condition at $n=0$, then at $n=n_0$, there will be (n_0+1) equally spaced possible i_e values, the smallest i_e will be:

$$\min i_e[n_0] = (n_0 - 1)V_L - \left(\frac{T_s}{L} \sum_{n=0}^{n_0-1} v_{load}[n] + i_{load}[n_0] \right), \quad (2.5)$$

and the largest i_e will be:

$$\max i_e[n_0] = (n_0 - 1)V_H - \left(\frac{T_s}{L} \sum_{n=0}^{n_0-1} v_{load}[n] + i_{load}[n_0] \right). \quad (2.6)$$

Fig. 2.1(b) visualizes the expansion of i_e states over time. Now, any v_{sw} sequence can be represented by a trellis in Fig. 2.1(b) that links branches between time samples. If we associate a cost function for each branch, finding the optimal v_{sw} sequence is then equivalent to finding a trellis with the least cumulative cost.

We now formulate the cost function. The true objective function that we want to minimize is the error energy lost in the op-amp W_e during a period of time. But this does not pose any

restriction on f_{sw} , so v_{sw} will switch as much as possible to minimize W_e . This problem can be solved by including a switching penalty constant P_{pen} , whenever the trellis wants to change its direction. P_{pen} has the dimension of power and can be set to the switching power loss in the buck converter at the desired f_{sw} (Sec. 3.1):

$$\text{OBJ}_{2\text{-norm}} = W_e = \sum_{i=0}^{n-1} (P_e[i] + \sigma_2[i]), \text{ where } P_e[i] = \begin{cases} i_e[i]v_{load}[i], & \text{if } i_e[i] \geq 0 \\ i_e[i](v_{load}[i] - V_{DD}), & \text{if } i_e[i] < 0 \end{cases} \quad (2.7)$$

$$\sigma_2 \text{ has unit of power, } \sigma_2[i] = \begin{cases} 0, & \text{if } v_{sw}[i-1] = v_{sw}[i] \\ P_{pen}, & \text{if } v_{sw}[i-1] \neq v_{sw}[i] \end{cases} \quad (2.8)$$

(2.7), which is a 2-norm cost function, involves a full multiplier and is computationally expensive. Moreover, Fig. 1.8 implies that minimizing $|\alpha - \alpha_0|$, or $|i_e|$ equivalently, can already lead to the minimum power loss, and we have no control on the voltage dimension in P_e anyway. Minimizing i_e should be our most tangible objective. We then simplify the cost function to 1-norm, which accumulates i_e 's absolute value, to significantly reduce the computational complexity:

$$\text{OBJ}_{1\text{-norm}} = \sum_{i=0}^{n-1} (|i_e[i]| + \sigma_1) \quad (2.9)$$

$$\sigma_1 \text{ has unit of current, } \sigma_1[i] = \begin{cases} 0, & \text{if } v_{sw}[i-1] = v_{sw}[i] \\ I_{pen}, & \text{if } v_{sw}[i-1] \neq v_{sw}[i] \end{cases} \quad (2.10)$$

Till now, the problem formulation has clearly reminded us of the Viterbi Algorithm, which was developed to efficiently find the least costly path in a trellis diagram. Fig. 2.1(b) is similar to the trellis diagram in [20, Fig. 8(a)], whose branch lengths are equivalent to our (2.9). Finding an optimal v_{sw} sequence is like the processes in [20, Fig. 8(b)]. In this work, we modify the well-known Viterbi Algorithm for our ET application.

2.2 Viterbi-Like Trellis Search

The Viterbi algorithm (VA) [20, 21] has been widely adopted to decode the convolutional codes used in 3G and legacy WLANs. Recently, it also finds its application in speech recognition and

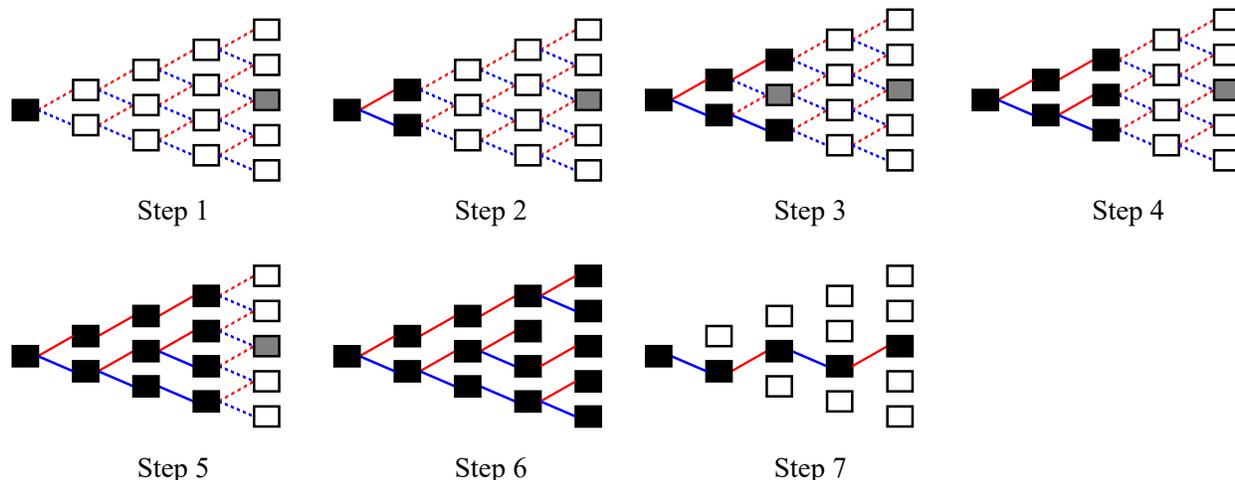


Figure 2.2: Visualized ACS processing steps of Alg. 1 to find the trace from the start (step 1) to the end (step 7) with the least cost.

bioinformatics. VA is a dynamic programming algorithm to find the most likely sequence of hidden states—called the Viterbi path—that results in a sequence of observed events in the context of the hidden Markov model (HMM). Our problem, formulated in Sec. 2.1, fits HMM very well. The known part in (2.4) can be treated as the observed sequence of events, and the unknown part in (2.4) is considered the “hidden cause” of the observation. The Viterbi algorithm finds the most likely “hidden cause” that minimizes the 1 -norm distance to the observation (2.9) (2.10).

To follow the commonly used nomenclature of the VA, we denote (2.9) (2.10) as the trace metric (TM). The core of the VA is an add-compare-select (ACS) logic. Firstly, for each state, the accumulator forms new TM candidates by *adding* branch costs to the existing TMs of the previous states, to which the branches are connected. Then for each state, the candidates are *compared*, and the smallest one is *selected* as the new TM. The selected branch is called the *survivor*.

For our problem formulation shown in Fig. 2.1(b), the candidate count for each state will be 2, so only a comparator and a MUX are needed. Assuming at any sample instant, the total number of states is N , we can develop the rudimentary ACS logic specifically for ET in Alg. 1. TM0 and TM1 stand for the candidates following a 0 branch and an 1 branch, respectively. BS stands for the branch selection (0 or 1) for each state. Note that the bottommost and the topmost states need special treatments. Fig. 2.2 visualizes the execution of Alg. 1. Step 2 is straightforward because

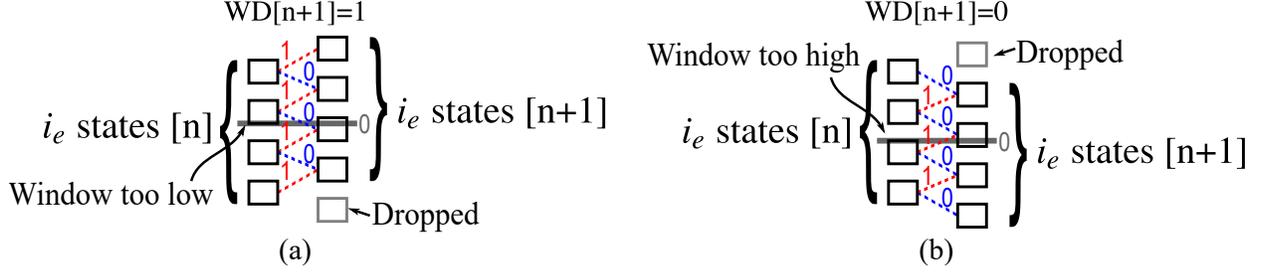


Figure 2.3: Bound a window of i_e states around 0: (a) The window is too low, so the bottommost state is dropped, and the window is moved up ($WD = 1$); (b) The window is too high, so the topmost state is dropped, and the window is moved down ($WD = 0$).

there is no middle i_e state with two incoming branches. Step 3 shows that the middle state (marked in gray) needs to execute the ACS logic in Alg. 1 and select from two TM candidates, and the decision is illustrated in step 4, where only 1 branch is kept as the survivor. Step 5-6 repeat the execution until the target is hit. Finally, the method in [22, Fig. 3] is used to retrieve the optimal path from the start to the end, which should be $v_{sw} = 0101$.

Now, we need to solve a final problem before hardware implementation: how to keep the number of states constant? From Fig. 1.8 and Fig. 1.15(d, h), we see that if the algorithm tries to minimize the 1-norm error between the unknown “hidden cause” and the known “observation” in (2.4), i_e states must be placed around 0. Otherwise, i_e will drift away indefinitely, and $|i_e|$ will be very large. Thus, if the state with the largest $|i_e|$ (most deviation from 0) is dropped every cycle, a finite number of i_e states can be maintained, and a search window of i_e will be bounded around 0. Or equivalently, according to (2.3), a search window of i_{sw} is placed around i_{load} . This is achieved by initializing the N i_e states as:

$$-\frac{N}{2} \times \frac{T_s}{L}(V_H - V_L) + (i - 1) \times \frac{T_s}{L}(V_H - V_L), \quad i \in [1, N] \quad (2.11)$$

Then for every cycle, the absolute values of the bottommost and topmost states are compared, and the larger one is dropped. As a result, N i_e states will be locked around 0 in a limit cycle as

illustrated in Fig. 2.3. The choice of window's direction (WD) is saved for subsequent processing.

Algorithm 1: Rudimentary ACS logic to pick survivor for every state

```

/* Add                                                    */
// Bottommost state, no incoming 1 branch
TM01 ← TM1 + |ie|1 + (BS1 = 0 ? 0 : Ipen);
TM11 ← ∞;
// Topmost state, no incoming 0 branch
TM0N ← ∞;
TM1N ← TMN-1 + |ie|N + (BSN-1 = 1 ? 0 : Ipen);
// Middle states
for i = 2 to N - 1 do
    | TM0i ← TMi + |ie|i + (BSi = 0 ? 0 : Ipen);
    | TM1i ← TMi-1 + |ie|i + (BSi-1 = 1 ? 0 : Ipen);
/* Compare and select                                    */
for i = 1 to N do
    | TMi ← Min(TM0i, TM1i);

```

2.3 DSP Core Design

A hardware Viterbi decoder consists of 3 main blocks [23, Fig. 6]:

1. **Branch metric unit** (BMU) calculates branch metrics, which are the costs of jumping to the next state through candidate branches, as discussed in Sec. 2.1. Our branch metric is $|i_e|$ plus a proper switching penalty. In our case, BMU also needs to decide the direction of the window.
2. **Trace metric unit** (TMU), also known as path metric unit, accumulates the branch metrics to get 2 candidate trace metrics for each state, one of which can eventually be chosen as the surviving optimum. Every clock TMU makes N decisions, throwing off wittingly non-optimal paths. The results of these decisions are saved for traceback.

3. **Traceback unit** (TBU) restores a least costly trace from the decisions made by TMU. It does it in the reverse direction, so some tricks are needed to reconstruct a correct order.

Our specific problem formulation requires an additional pre-processor in front of the BMU such that the BMU can correctly produce (2.4) with minimum calculation. Also, we need to process the input envelope data at full throughput with finite latency. One input sample comes in, while one v_{sw} should be produced by the DSP core.

2.3.1 Pre-processor

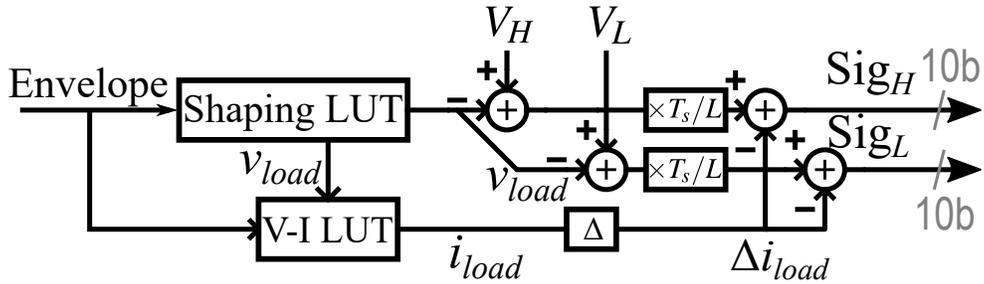


Figure 2.4: Pre-processor

The pre-processor is shown in Fig. 2.4. The envelope signal (in dBm) is first fed into a calibrated shaping LUT to calculate the desired v_{load} as the supply to the PA. Then, the desired v_{load} and the input envelope signal are fed into another 2-D LUT that produces the load current signal i_{load} . For a simple resistor load, this 2-D LUT degenerates to $i_{load} = G_{load}v_{load}$. Since in (2.4), i_{load} is outside of the accumulator, it would take an extra adder if (2.4) were executed directly. So instead, we take the derivative (difference) of i_{load} , such that it will be integrated back to i_{load} in TMU. The pre-processor provides two signals Sig_H and Sig_L , representing the *change* of i_e when v_{sw} is evaluated as V_H or V_L :

$$Sig_H[n] = \frac{T_s}{L} V_H - \left(\frac{T_s}{L} v_{load}[n] + i_{load}[n] - i_{load}[n-1] \right) \quad (2.12)$$

$$Sig_L[n] = \frac{T_s}{L} V_L - \left(\frac{T_s}{L} v_{load}[n] + i_{load}[n] - i_{load}[n-1] \right) \quad (2.13)$$

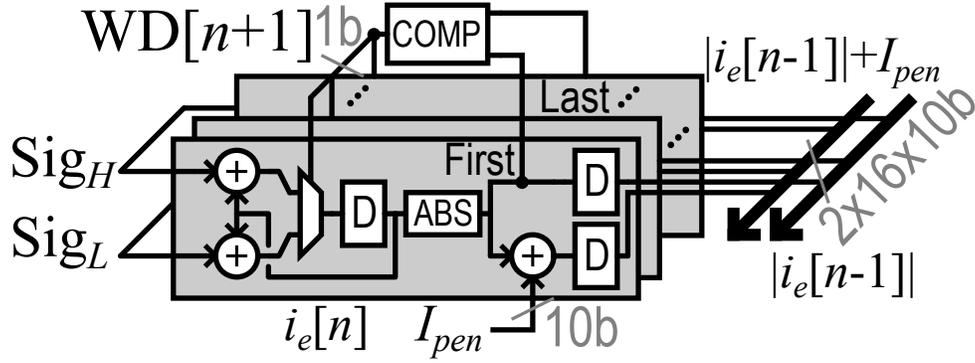


Figure 2.5: Branch metric unit (BMU)

2.3.2 BMU

Fig. 2.5 shows the hardware implementation of BMU. The first set of N registers stores the error current samples in the window i_{ei} . In each clock cycle, the window is bounded around 0 with:

Algorithm 2: BMU's logic to regulate search window's position

```

// Skewed absolute values:
top =  $i_e[n]_N < 0 ? \text{BitwiseInvert}(i_e[n]_N) : i_e[n]_N$ ;
bot =  $i_e[n]_1 < 0 ? \text{BitwiseInvert}(i_e[n]_1) : i_e[n]_1$ ;
for  $i = 1$  to  $N$  do
    if  $\text{top} > \text{bot}$  then           // Window is too high, so everybody moves down
         $i_e[n+1]_i \leftarrow i_e[n]_i + \text{Sig}_L$ ;
         $\text{WD}[n+1] \leftarrow 0$ ;
    else                           // Window is too low, so everybody moves up
         $i_e[n+1]_i \leftarrow i_e[n]_i + \text{Sig}_H$ ;
         $\text{WD}[n+1] \leftarrow 1$ ;

```

Since the regulation of the window's position does not need accurate values, to reduce delay and hardware, Alg. 2 uses *bitwise inverse* to quickly calculate the skewed absolute values of the 2's compliment representations of $i_e[n]_1$ and $i_e[n]_N$. To ease the loop timing for the subsequent TMU, two sets of the delayed signals $|i_e[n-1]_i|$ and $|i_e[n-1]_i| + I_{pen}$ are prepared as shown at the output ports of Fig. 2.5. TMU will select which signal to use for each state based on its branch selection

memory (Alg. 1). In this work, we use 16 i_e states and 10 bits for each i_e state, such that the i_e window encloses 0 (or equivalently, the i_{sw} window encloses any kind of i_{load}), for most of the time (Fig. 2.6). At the end, 32×10 bits are passed to TMU.

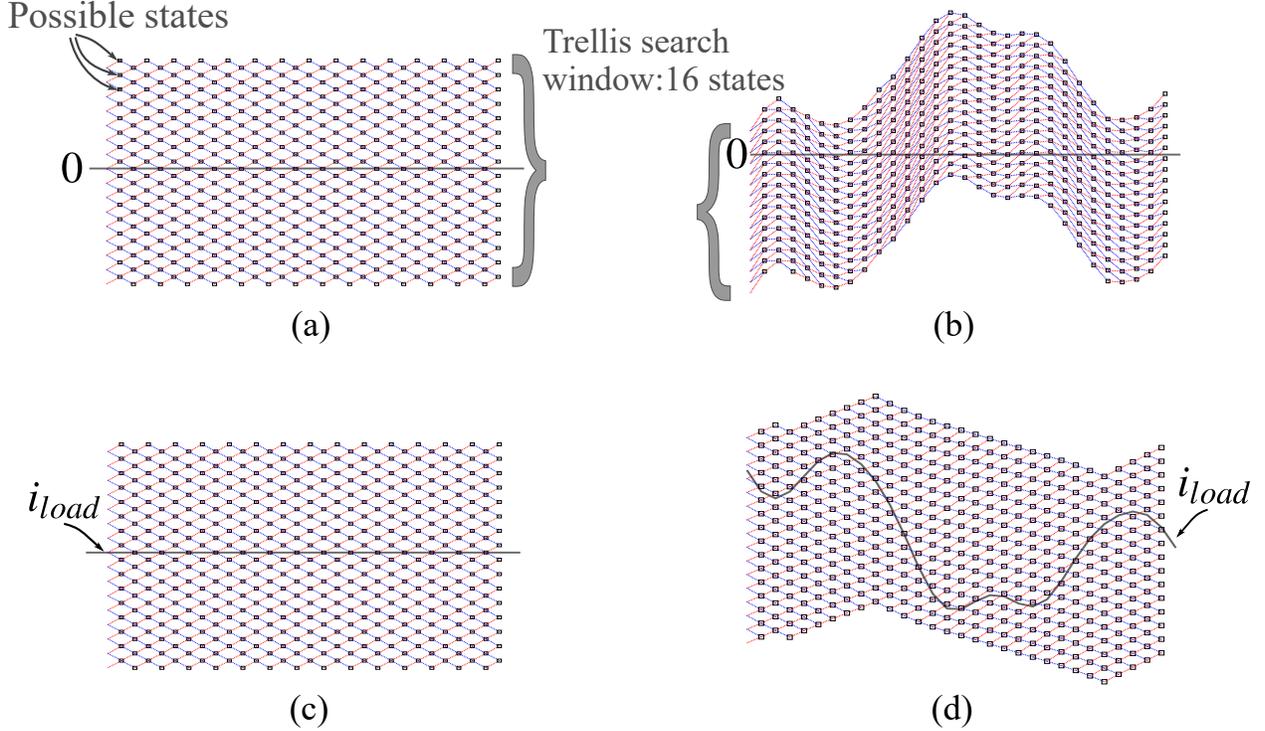


Figure 2.6: Actual window of the 16 i_e states in BMU with (a) DC and (b) random waveform; equivalent window of the 16 i_{sw} states covering i_{load} for (c) DC and (d) random waveform

2.3.3 TMU

To incorporate the WD signal from BMU, Alg. 1 has to be modified. Depending on the WD signal, each state's candidate branches may come from 2 of the 3 directions (up, flat, or down). This is explicitly included with pseudo codes in Alg. 3. For readability, we denote $WD[n-1]$ as WD_{in} , the i -th **U**npenalized **B**ranch metrics $|i_e[n-1]_i|$ as UB_i , and the i -th **P**enalized **B**ranch metrics $|i_e[n-1]_i| + I_{pen}$ as PB_i .

Since the branch metrics are always non-negative, there must be an additional circuit preventing TM registers from overflow during the accumulation of non-negative inputs. This is done by subtracting a common **Renorm** from all TM registers before register update, as shown at the end

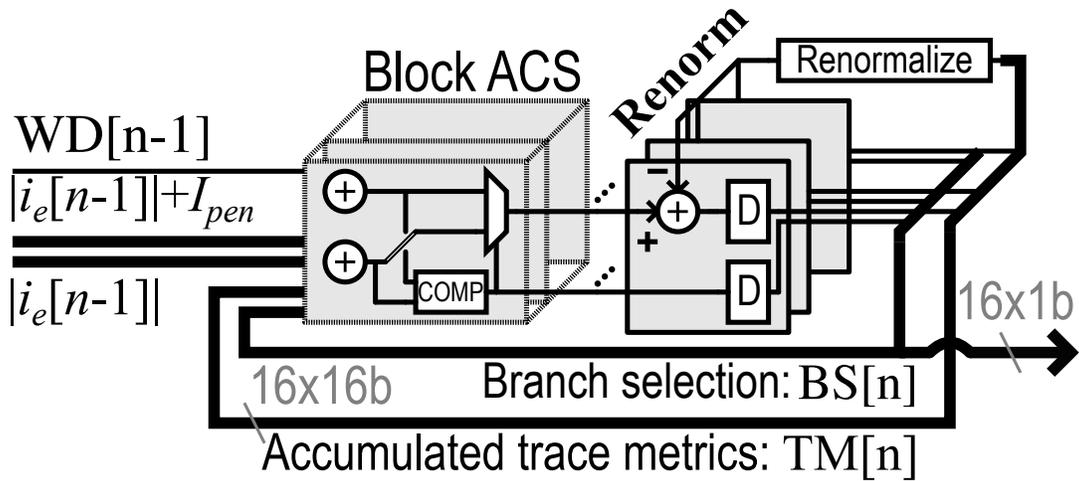


Figure 2.7: Trace metric unit (TMU)

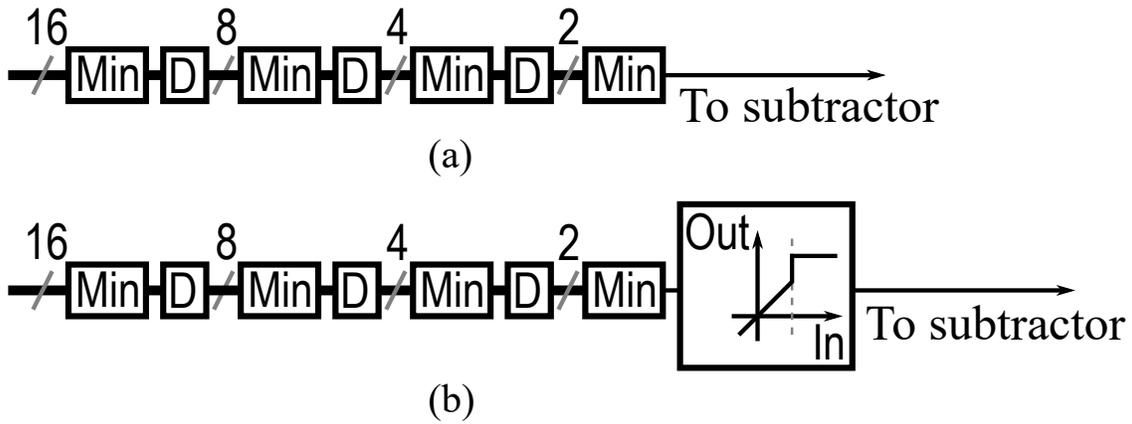


Figure 2.8: (a) Linear pipelined renormalization block that will lead to instability; (b) Pipeline and clip renormalization block to stabilize the renormalization loop.

of Alg. 3 and Fig. 2.7.

The simplest **Renorm** would be the minimum of all TM registers every cycle, but it is impossible to find the minimum value of 16 16-bit values in 1 DSP cycle. If the pipeline registers are inserted into the radix-2 minimizer (Fig. 2.8(a)), the delayed minimum will lead to diverging values in TM registers after closing the renormalization loop (Fig. 2.9(a)). To solve this issue, a nonlinear function is added at the end of the pipeline to generate **Renorm** (Fig. 2.8(a)). If the output of the minimizer is smaller than a threshold, **Renorm** will be the pipeline's output. Otherwise, **Renorm** will be clipped to a fixed value. The specific choices of the threshold and the clipping level depend

Algorithm 3: Implemented ACS logic incorporating WD signal and renormalization

```
/* Add                                                                 */
if  $WD_{in}$  is 1 then                                                    // Window moved up
┌   for  $i = 1$  to  $N - 1$  do      // Window moved up, so 0 branches are downward
├    $TM0_i = TM[n]_{i+1} + (BS[n]_{i+1} = 0 ? UB_i : PB_i)$ ;
├    $TM0_N = MAX$ ;                // Topmost state's special treatment
├   for  $i = 1$  to  $N$  do          // Window moved up, so 1 branches are flat
├    $TM1_i = TM[n]_i + (BS[n]_i = 1 ? UB_i : PB_i)$ ;
└   else                                                                // Window moved down
├   for  $i = 1$  to  $N$  do          // Window moved down, so 0 branches are flat
├    $TM0_i = TM[n]_i + (BS[n]_i = 0 ? UB_i : PB_i)$ ;
├    $TM1_0 = MAX$ ;                // Bottommost state's special treatment
├   for  $i = 2$  to  $N$  do          // Window moved down, so 1 branches are upward
├    $TM1_i = TM[n]_{i-1} + (BS[n]_{i-1} = 1 ? UB_i : PB_i)$ ;
└   /* Compare and select                                            */
for  $i = 1$  to  $N$  do
┌    $TM[n+1]_i \leftarrow (TM0_i < TM1_i ? TM0_i : TM1_i) - Renorm$ ;
├    $BS[n+1]_i \leftarrow TM0_i < TM1_i ? 0 : 1$ ;
```

on the amplitude of the waveform, and simulations are necessary to pick the most suitable values. In our work, the threshold is chosen to be 0, and the clipping level varies with the envelope's distribution and amplitude. Fig. 2.9(b) shows the stabilized TM registers, which do not exceed the upper limit 0x7FFF. The ACS loop in TMU is the bottleneck of the DSP core, containing 3 adder delays and 1 MUX delay.

At the end, the branch selection for each state $BS[n]_i$ and the delayed $WD[n-2]$ are passed to the TBU for further processing.

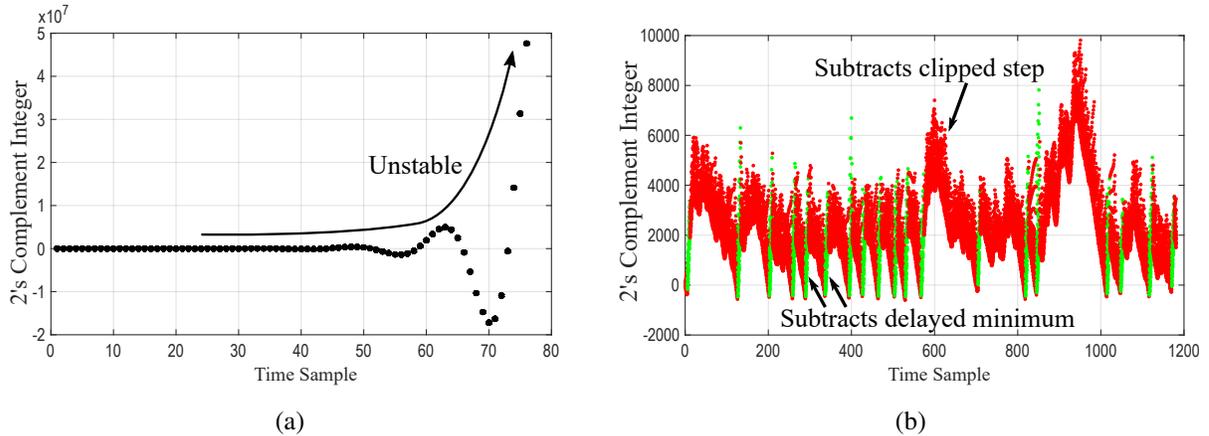


Figure 2.9: (a) TMU registers' digital waveforms from (a) Pipeline-subtraction; (b) Pipeline-clip-subtraction

2.3.4 TBU

At first glance, we would think that the trace back procedure can only start after $BS[n]$ and $WD[n]$ are calculated for the entire signal in BMU and TMU. The trace back should start from the state with the smallest final TM value. But there are two problems here:

1. As discussed in Sec. 2.3.3, finding the minimum value from an array of registers cannot be done on the fly;
2. This will not allow seamless data transmission, as forward transmission needs to be interrupted for reverse processing.

How can we bypass these two problems and reconstruct a unique v_{sw} out of N possible traces at the same throughput as BS and WD are produced? We have to first explore the merging property of survivor paths in VA. It has been shown in [24, Fig. 2b] and [25, Fig. 6.5] that the survivor paths traced back from any terminal state always merge after a delay. The number of time steps that have to be traced back for the paths to merge with very high probability is called the survivor depth D . The latency of the decoding should be no smaller than D . Fig. 2.10 illustrates the merging phenomenon in our context. From simulations on our envelope waveforms, we decide to use 50 time steps ($> 3 \times 16$ states) to guarantee merging. Now, we have solved the problem in twofold:

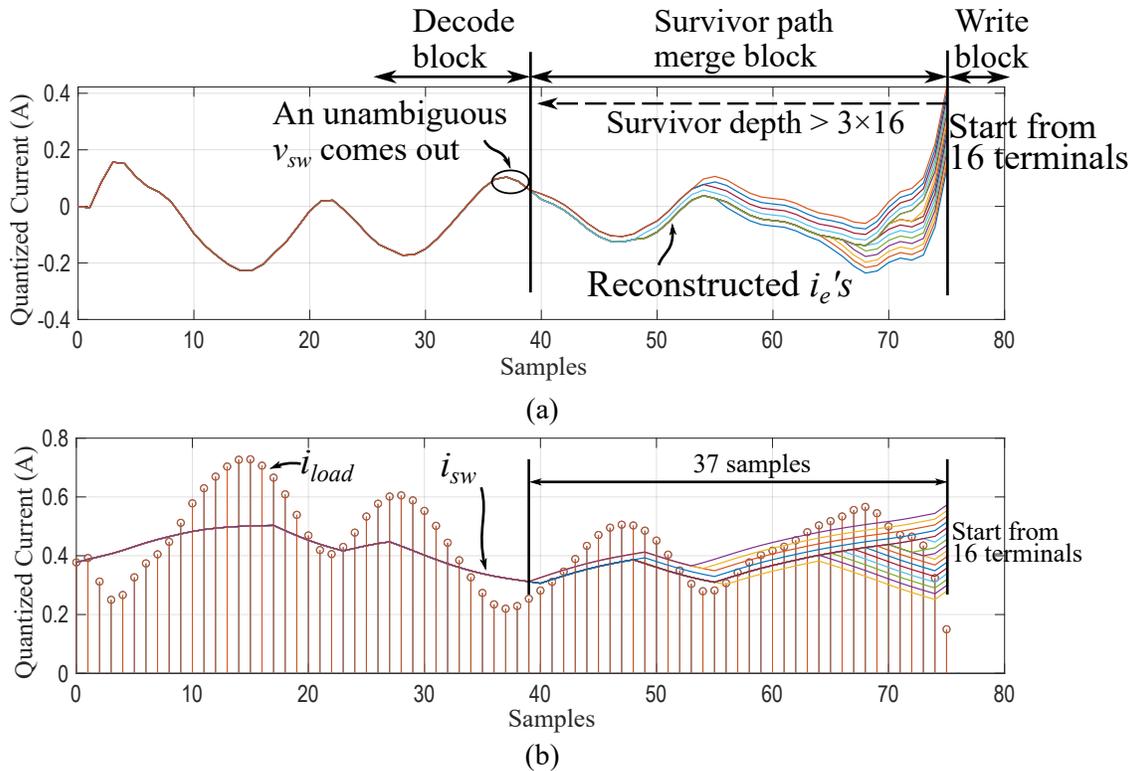


Figure 2.10: Reconstructed (a) i_e and (b) equivalent i_{sw} waveforms from v_{sw} traced back from different terminals

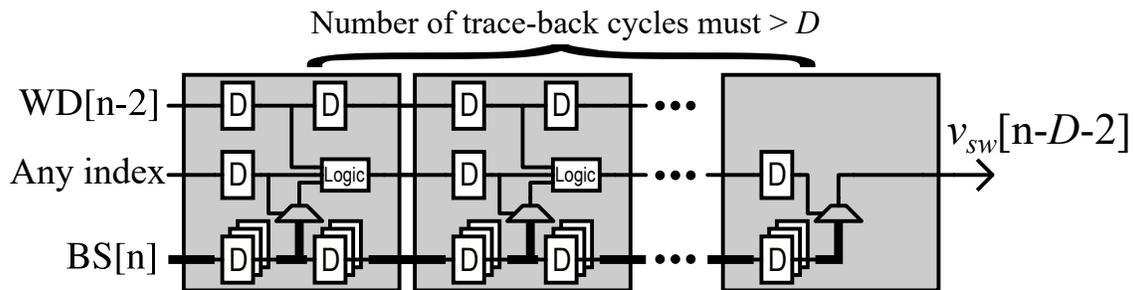


Figure 2.11: Trace back unit (TBU)

1. Any state can be the starting state because the survivor path will merge eventually;
2. Only 1 unambiguous v_{sw} at D time steps earlier needs to be retrieved.

The actual hardware uses a pipeline of 50 stages as shown in Fig. 2.11. The address index can start from any number $\in [1, 16]$. Each stage reads a new address index from the FIFO memories (WD and BS), and calculates a new address index on the fly for the next stage with the logic in

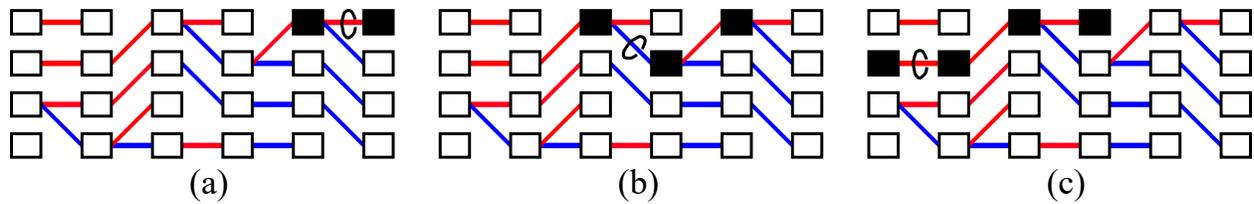


Figure 2.12: Illustration of TBU's pipeline operation. The circled branch are being traced through simultaneously (a) in 1st stage, (b) in 2nd stage, and (c) in 3rd stage. All of the circled evaluation started from the topmost state.

Algorithm 4: TBU's address update logic from the pointers read out

```

/* Add                                                    */
if WD memory read out is 1 then                        // Window moved up
|
|   if BS memory <address in > = 1 then
|   |   address out = address in;
|   else
|   |   address out = address in + 1;
|
else                                                    // Window moved down
|
|   if BS memory <address in > = 1 then
|   |   address out = address in - 1;
|   else
|   |   address out = address in;
|

```

Alg. 4. The reading indexes flow twice as fast as the memories do, because as new memories come in, the reading addresses need to chase the shifting memories (Fig. 2.12). This reverse timeline is created without interrupting the forward timeline. The last stage only needs to read out a final output v_{sw} .

2.4 FPGA Implementation

As shown in Fig. 2.13(a), the DSP core is realized on Xilinx KC705 and passes the timing constraints at a maximum clock of 230 MHz. Due to the SERDES limitation, we select 200 MHz as the DSP clock. A clock cycle defines one step in the trellis search and limits the operation to modulation bandwidths of 20 MHz. RTL codes for BMU and TMU are written jointly in Appx. A.1. RTL codes for TBU are included in Appx. A.2. They are connected through delay lines in Appx. A.3 for better separation between the two blocks on the fabric. Appx. A.4 shows the codes for the top-level wrapper of the DSP core. RTL codes of the DSP core are simulated and matched to an equivalent MATLAB script with 2's complement integers. Fig. 2.13(b) shows that BMU and TMU are the most complicated blocks. Although TBU has 50 stages, its simple logic does not require a lot of slices on FPGA's fabric. v_{ET} (16-bit), Sig_H (10-bit), and Sig_L (10-bit) are pre-calculated with Fig. 2.4 and stored in FPGA's BRAMs. The scaled v_{ET} is used as the desired v_{load} in the DSP core. v_{ET} is delayed to match the latency of the DSP core. Then all data are sent to a JESD204b SERDES IP, which drives TI DAC39J84 through the high-speed FMC connector. An external 400 MHz clock is fed to the clock distributor in LMK04828, which provides the high-speed clock for SERDES. The differential 400 MHz clock is buffered on FPGA to a single-ended 200 MHz clock, which drives the entire DSP core. The unilateral data flow of the DSP core relaxes the clock skew constraint significantly. Table 2.1 shows that the DSP's computational part only takes a small fraction of the FPGA fabric. A virtual IO block is added to program I_{pen} , V_H , V_L , the renormalization block's threshold voltage, and the clipping level conveniently. The lab connection is shown in Fig. 2.14.

Fig. 2.15(a) illustrates, with DC for readability, that a lower I_{pen} will increase f_{sw} . Fig. 2.15(b,c) show that the DSP core retimes switching instances and can generate v_{sw} at full throughput even with continuous packages.

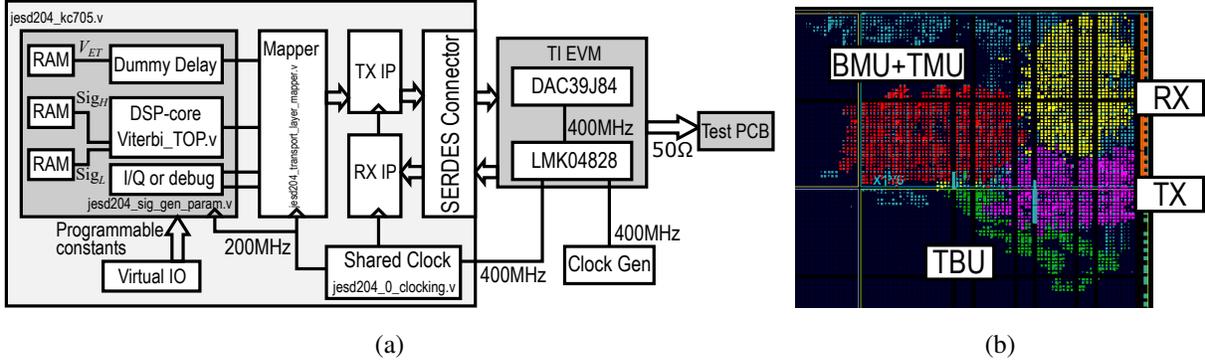


Figure 2.13: (a) Complete TX DSP system built on FPGA including interface with DAC; (b) FPGA's floor plan without showing BRAM

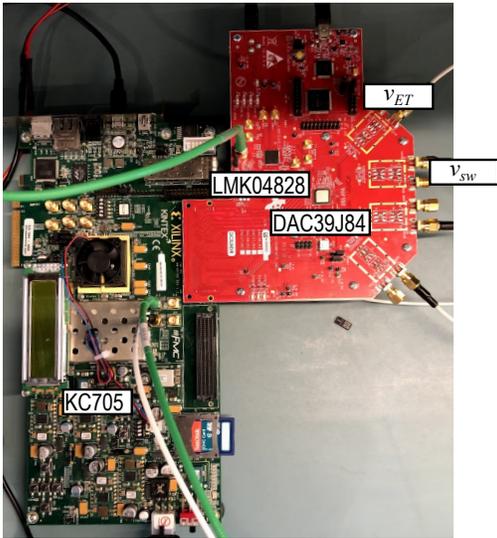


Figure 2.14: FPGA-DAC setup

Table 2.1: FPGA's usage

	Usage	%
LUT	10141	5
Flip-flop	10333	2.5
BRAM	294	66
I/O	20	4
High-speed TX/RX	4	25

2.5 Systematic Study of the HA

With the DSP algorithm and the buck converter's model in Sec. 3.1, we can explore the entire design space of the HA with (2.14). For a given v_{load} and i_{load} , the hysteresis comparator (HC) operation mode only has 2 design variables: inductance L and hysteresis window H . Our trellis-search (TS) operation mode also has 2 design variables, L and switching penalty I_{pen} . For each design point, two different v_{sw} 's can be calculated by running the TS algorithm or the hysteresis logic. Then from f_{sw} we can find the theoretical minimum buck converter's loss P_{buck} with (3.4),

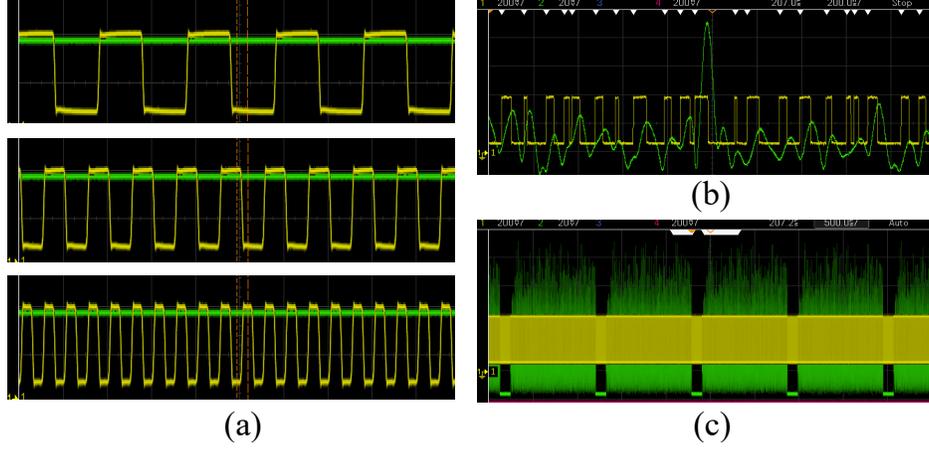


Figure 2.15: (a) Generated v_{sw} for a DC v_{load} with I_{pen} decreased to 0 from top to bottom; (b) Zoom-in view showing aligned v_{load} and v_{sw} ; (c) v_{sw} can be generated continuously over packets.

assuming $(1 + \gamma)R_{sw}C_{sw} \approx 70$ ps for our process. From i_e waveform, we can find P_e with op-amp's Class-AB push-pull loss expression in (2.15). To make the model more realistic, we assume a nonzero quiescent bias current I_{Q2} in the op-amp's output stage. And we also include the inductor's series resistance (0.15Ω) and a nonzero constant bias current of op-amp's input stage (7.8 mA). But the dominant trade-off is still between P_{buck} and P_e (on average) in (2.14). If we can achieve a smaller P_e with the same P_{buck} , we can raise ETSM's overall efficiency η_{ETSM} .

$$\eta_{ETSM} \approx \frac{P_{load}}{\underbrace{P_{buck} + P_{e,avg}}_{\text{trade-off}} + \underbrace{i_{sw,avg}^2 R_L}_{\text{inductor}} + \underbrace{V_{DD} I_{Q1}}_{g_{m1}} + P_{load}} \quad (2.14)$$

$$P_e = \begin{cases} i_e v_{load}, & \text{if } i_e \geq I_{Q2} \\ i_e (v_{load} - V_{DD}), & \text{if } i_e \leq I_{Q2} \\ (I_{Q2} + 0.5i_e)v_{load} \\ + (I_{Q2} - 0.5i_e)(V_{DD} - v_{load}), & \text{otherwise} \end{cases} \quad (2.15)$$

We first test a shaped envelope waveform of 6dB PAPR on a resistor load of 5Ω . The load power is 550 mW. We assume the op-amp has $I_{Q2} = 10$ mA on its output stage for the 20 MHz case. The color maps of ETSM's efficiencies for the HC and TS cases are plotted with respect to their own design variables in Fig. 2.16. We also overlay the switching frequency contours on top

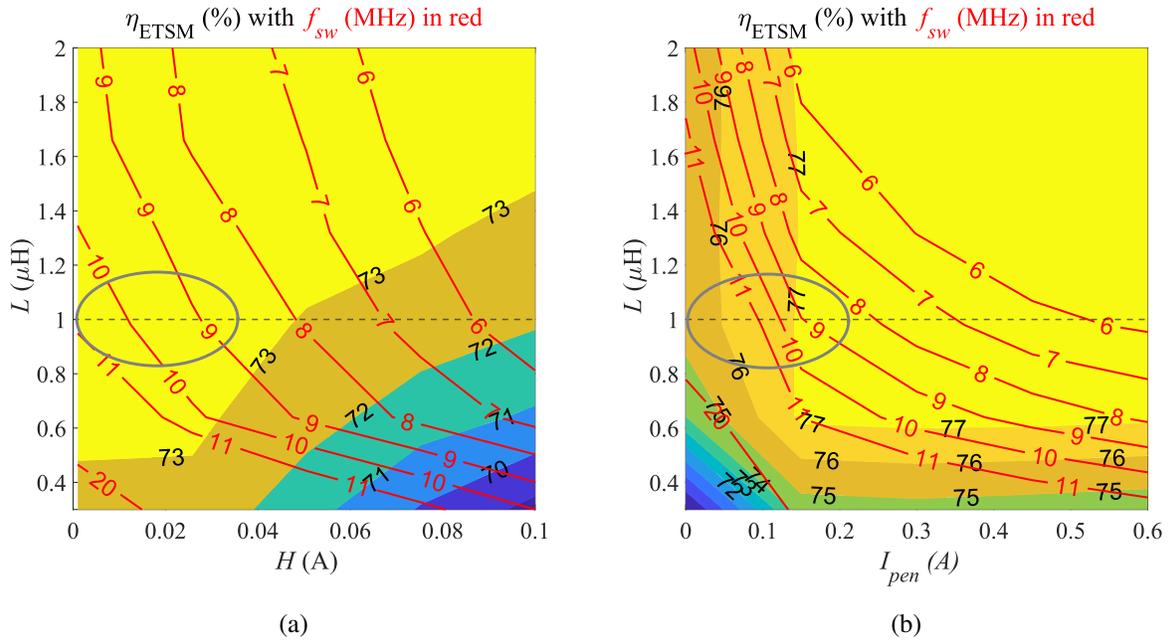


Figure 2.16: Contours of simulated ETSM’s overall efficiency η_{ETSM} and switching frequency f_{sw} with (a) HC and (b) TS for the 20 MHz mode

of the color maps for comparison. Firstly, Fig. 2.16(a) verifies that larger L and larger H indeed limit f_{sw} for the HC case. Fig. 2.16(b) shows that larger switching penalty does reduce f_{sw} for the TS case. We see that the peak of ETSM’s efficiency increases 4% from 73% to 77%, where our inductor pick of 1 μH cuts through, for $f_{sw} \approx 10\text{MHz}$. If we overlay the switching frequency contours on top of the P_e contours in Fig. 2.17, we can see that at f_{sw} around 10 MHz, TS saves P_e by 40%, from roughly 100 mW to roughly 60 mW, while P_{buck} is still kept around 65 mW. It is this save that raises the overall efficiency of the ETSM.

Fig. 2.18 illustrates the effect of TS on the waveform level. The conventional HC actuates a generally lagging i_{sw} with respect to i_{load} . Instead, our TS algorithm expands a search window around i_{load} , and finds an optimal trellis by customizing the switching instances for different wave excursions. As a result, v_{sw} is retimed, and i_e can be minimized to its theoretical limit for the target L and f_{sw} .

Although our FPGA cannot handle the upsampled clock speed of 1.6 GHz for the 160 MHz mode, we still want to analyze whether TS can offer any theoretical improvement for the 160 MHz

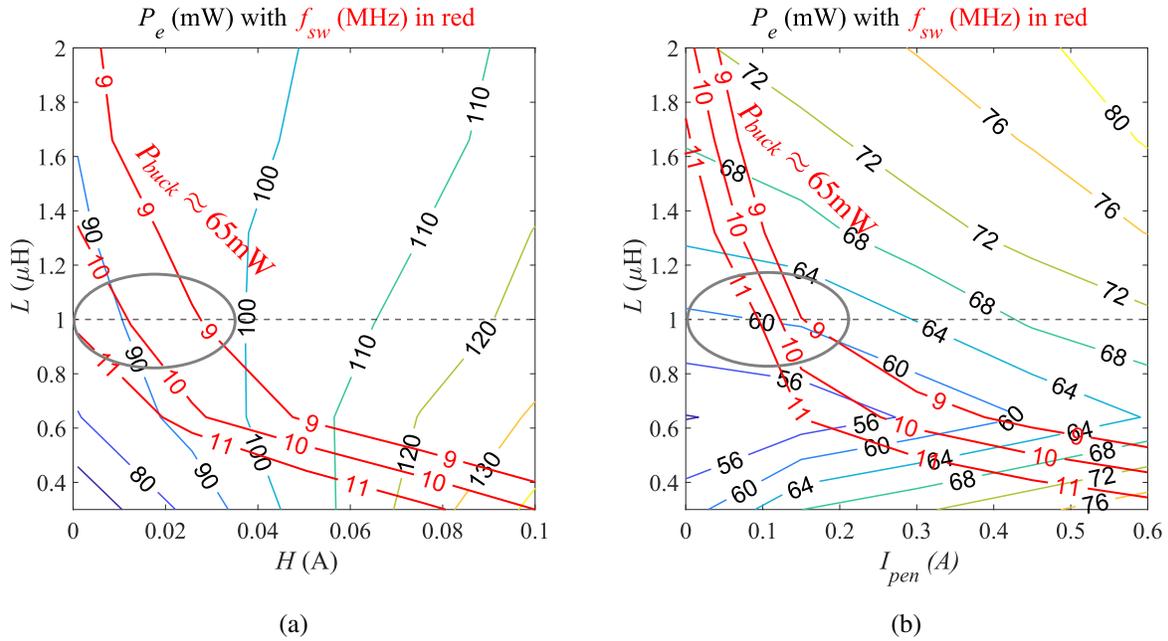


Figure 2.17: Contours of error power P_e and switching frequency f_{sw} with (a) HC and (b) TS for the 20 MHz mode

mode out of our curiosity. We assume the op-amp has $I_{Q2} = 20\text{mA}$ on its output stage for the 160 MHz case for wider bandwidth. The analytical results are included in Fig. 2.19. We see that the optimum efficiency intersects switching frequencies that are much lower than 160 MHz. And TS does not seem to provide any noticeable improvement in ETSM's efficiency, merely from 70% to 71%. It is explained as follow. Our original target is reducing the gap between i_{sw} and i_{load} with a reasonable switching frequency. This target can be achieved for the 20 MHz mode, with a switching frequency of around 10 MHz where the buck converter is still very efficient. Then, TS retiming clearly reveals its effect on a better alignment between i_{sw} and i_{load} , as shown in Fig. 2.20(a). But for the 160 MHz mode, if we just compress the time axis by a factor of 8, we should get the same relationship between i_{load} and i_{sw} at $8 \times f_{sw}$. The increased f_{sw} requires at least $2.8 \times$ the original P_{buck} according to (2.14), which is 120 mW extra loss. This is much more than the increase in P_e , if we just keep f_{sw} low and let the HA operate in complete slope-saturation mode, as described in Fig. 1.11(d) of Sec. 1.6. The other perspective is that P_e will eventually saturate to a maximum value when we switch slowly with a large inductor, and all the AC power

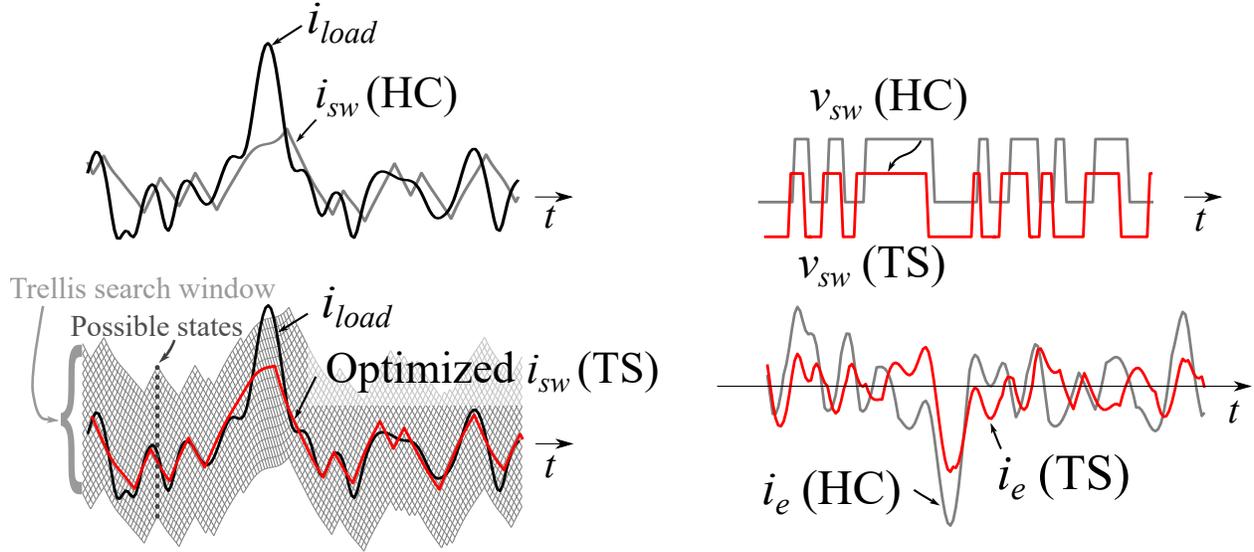


Figure 2.18: Illustration of how TS reduces i_e compared to HC on waveform level.

of i_{load} is supplied by the op-amp. But P_{buck} will increase $\propto \sqrt{f_{sw}}$ indefinitely, if f_{sw} keeps rising. Therefore, the complete design space exploration predicts a slope-saturation mode for the 160 MHz mode, if we restrict ourselves to a single inductor. In slope-saturation mode, retimed switching will not help anymore, because the op-amp must supply the entire AC part of i_{load} , despite the switching instance. Thus, in this work we only apply TS to ET for 20 MHz bandwidth. And for the 160 MHz mode, all burden falls on the op-amp, which needs to be designed accordingly in Sec. 3.2.2.

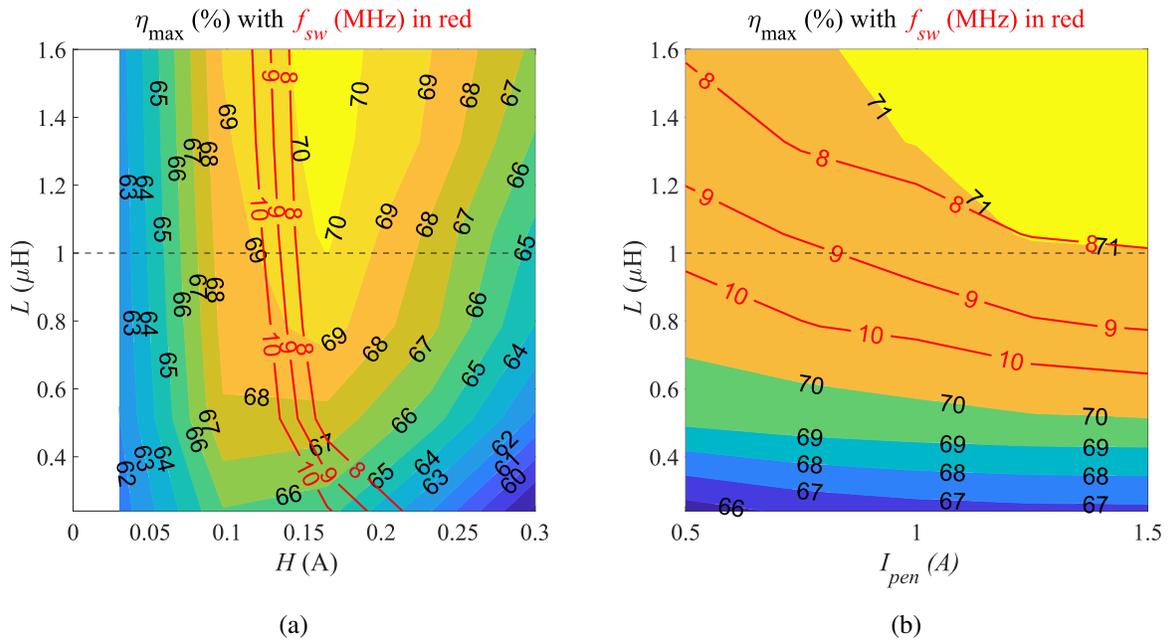


Figure 2.19: Contours of simulated ETSM's overall efficiency η_{ETSM} and switching frequency f_{sw} with (a) HC and (b) TS for the 160 MHz mode

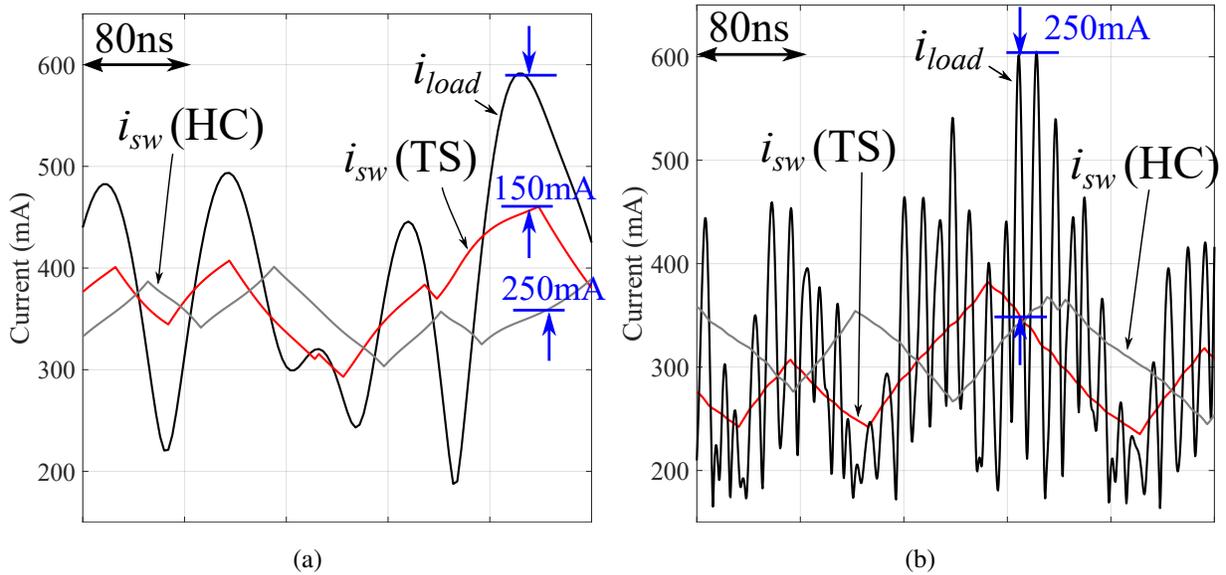


Figure 2.20: (a) Retimed switching reduces the gap between i_{sw} and i_{load} with reasonable f_{sw} ; (b) Retimed switching does not significantly affect P_e in slope-saturation mode.

CHAPTER 3

Circuit Design

3.1 Design of the Buck Converter

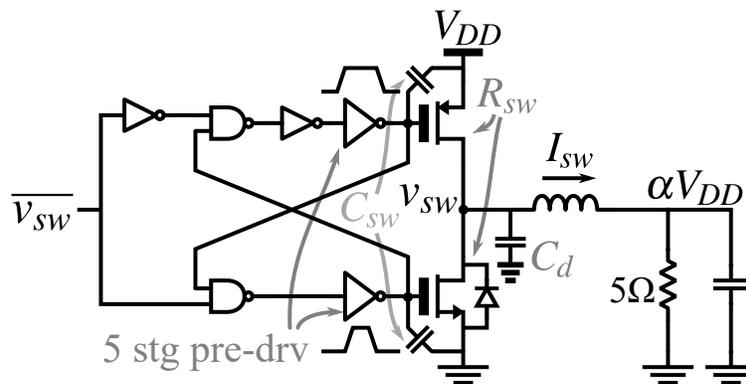


Figure 3.1: General design of a buck converter, including the non-overlapping pre-driver

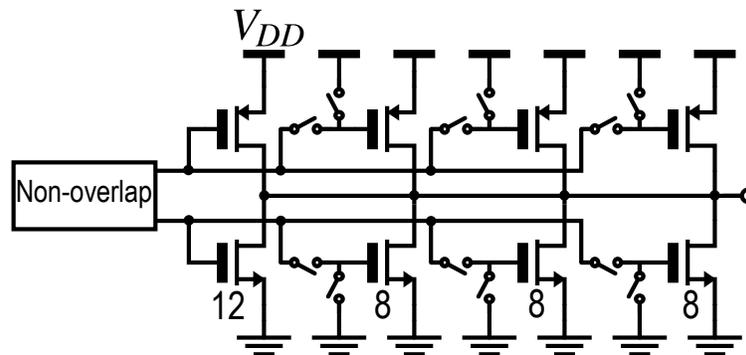


Figure 3.2: Sliced output stage of the buck converter

The general design of a buck converter is shown in Fig. 3.1. The sizing of the output FETs and the last pre-driver stage is the most important design choice. The output stage must also stand full V_{DD} . We start our analysis from the simple inverter design of the output stage.

The dominant power losses of a buck converter consist of the following:

1. Conduction loss of power FETs' on-resistance R_{sw} (modeled on average for PMOS and NMOS);
2. Switching loss of power FETs' gate capacitance C_{sw} (modeled on average for PMOS and NMOS);
3. Switching loss of drain capacitance C_d ;
4. Crowbar current loss when both FETs are on during switching;
5. Free-wheeling NMOS's body diode loss when both FETs are off;
6. Switching loss of pre-driver stages.

The pre-drivers in Fig. 3.1 are connected in a loop such that a non-overlapping version of the input v_{sw} can be generated for the PMOS and NMOS. This can eliminate the crowbar current, and roughly 50% of C_d 's switching loss, because i_{sw} stored in the inductor discharges C_d adiabatically at the falling edge, at the cost of a slightly higher body diode loss. At the rising edge, parasitic C_d still needs to be switched hard. Detailed simulations are needed to pick an optimal non-overlapping duration such that the body diode loss becomes negligible. In this work, we use roughly 1.5 ns on both edges.

The FET's conduction loss and switching loss remain the dominant trade-off. They have a clear trade-off, as the output transistor's width is the only design parameter that we can optimize [26]:

$$P_{buck} = P_{sw} + P_{sw,pre} + P_{\Omega} \quad (3.1)$$

$$= C_{sw}V_{DD}^2f_{sw} + \gamma C_{sw}V_{DD}^2f_{sw} + I_{sw}^2R_{sw} \quad (3.2)$$

$$= (1 + \gamma)C_{sw}V_{DD}^2f_{sw} + \frac{P_{load}^2R_{sw}}{\alpha^2V_{DD}^2} \quad (3.3)$$

$$\geq \frac{2}{\alpha} \sqrt{(1 + \gamma) \underbrace{R_{sw}C_{sw}}_{\text{Technology}} f_{sw} \cdot P_{load}} \triangleq \varepsilon P_{load}. \quad (3.4)$$

$$\text{where } v_{load} = \alpha V_{DD},$$

$$\varepsilon = \eta_{max}^{-1} - 1. \quad (3.5)$$

Since the pre-driver's capacitance scales with the capacitance of the output stage, and they do not have conduction loss, $P_{sw,pre}$ can be expressed as γP_{sw} , where γ could be around 0.6 ([26, (6)]). $R_{sw}C_{sw}$ in (3.4) is a technology constant depending on the FET's V_{DD} and the channel length, and γ simply enlarges that by a constant, so $(1 + \gamma)R_{sw}C_{sw}$ can be referred to as τ_{sw} , which is simulated as 70 ps in our process. (3.4) and (3.5) imply that the maximum efficiency η_{max} of a buck converter only depends on the back-off voltage ratio α and f_{sw} . FETs' widths should be selected such that $P_{sw} + P_{sw,pre} = P_{\Omega}$ at a given f_{sw} .

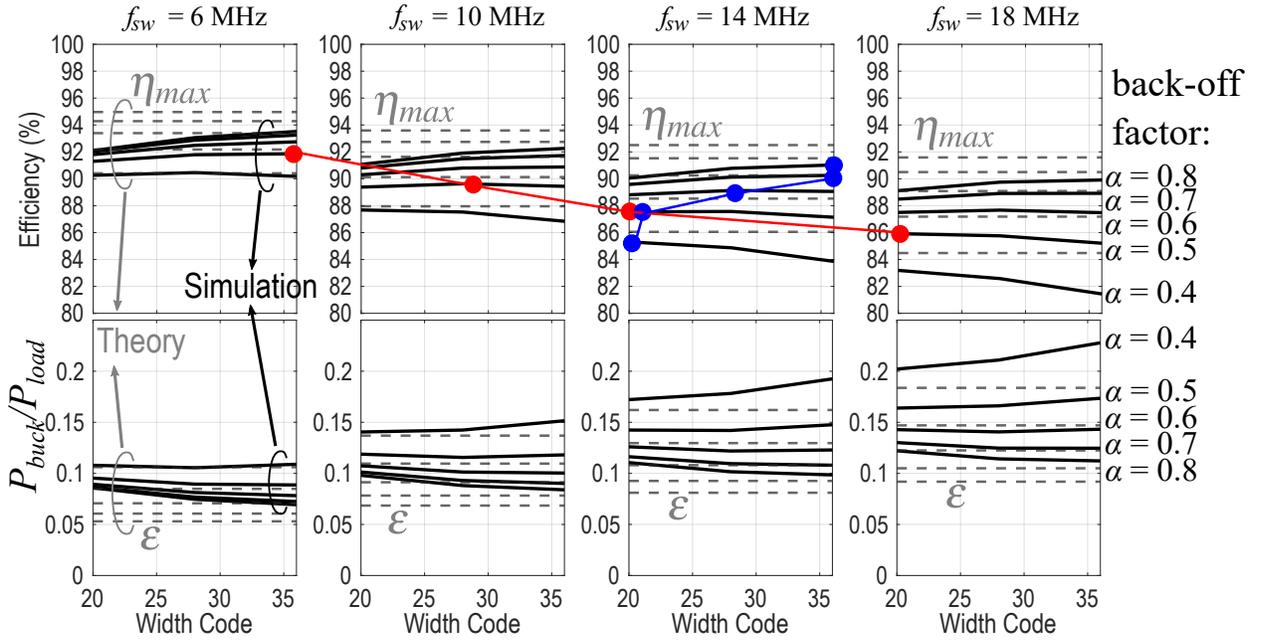


Figure 3.3: Simulated efficiency η (top) and $(\eta^{-1} - 1)$ versus theoretical η_{max} and theoretical ϵ for various f_{sw} and voltage back-off α ((3.4) (3.5)) with $(1 + \gamma)R_{sw}C_{sw} = 70$ ps.

The actual buck converter is sliced such that FETs' width can be programmed to the optimum values for different switching frequencies and output voltage levels (Fig. 3.2). For example, at a given switching frequency, as the output voltage level rises, a higher maximum efficiency can be approached with more slices. Or at a given output voltage level, as the switching frequency decreases, a higher maximum efficiency can be approached with more slices. The simulated maximum efficiencies are close to the modeled maximum efficiencies with the optimum width codes, as shown in Fig. 3.3.

3.2 Design of the Linear Amplifier

3.2.1 Op-Amp Requirements

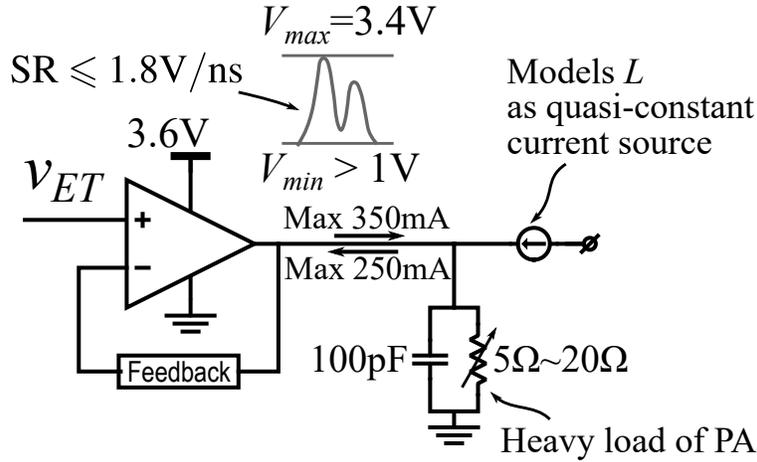


Figure 3.4: Design requirements of the op-amp for 160 MHz mode

For ETSM's toughest 160 MHz mode, there are several design requirements for the op-amp, since Sec. 2.5 has discussed that the op-amp essentially takes all the burden of i_{load} 's AC part (Fig. 3.4). In this case, the inductor can be approximated as a current source, and the op-amp drives the full apparent load of the PA. It includes a RF bypass capacitor of 100 pF integrated in the PA module. The resistive apparent load at PA can be calculated by differentiating the measured V-I realization at the PA's supply port as $\partial V_{DD}/\partial I_{DD}$. It varies between 5 Ω and 20 Ω for our PA. The op-amp's bandwidth must be at least $3\times$ half of the channel bandwidth, due to IQ-to-envelope bandwidth extension. The op-amp needs to sustain a full battery voltage of 3.6 V and handle a high swing of 1 V \sim 3.4 V. The low-side headroom is more relaxed, because the minimum supply voltage of the PA is typically above ground for a low EVM. The high-side headroom is more stringent, as the maximum supply voltage of the PA needs to be as high as possible to raise PA's saturation power. From simulation, the op-amp's output stage needs to source a maximum current of 350 mA, and sink a maximum current of 250 mA. For the 160 MHz mode, the slew rate of the waveform v_{load} after shaping is simulated to be as large as 1.8 V/ns.

3.2.2 Op-Amp Design Considerations

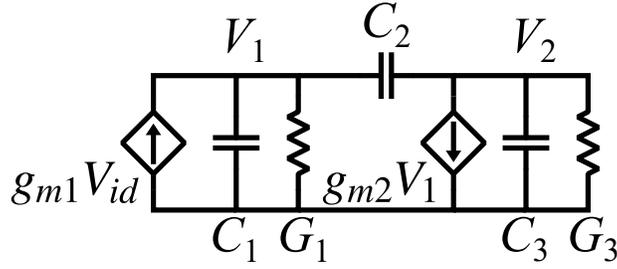


Figure 3.5: Canonical- π equivalent circuit of a two-stage amplifier

The design reasoning of the op-amp goes as such. Heavy resistive load means that we need at least a 2-stage amplifier for enough loop gain and low output impedance. Otherwise, the close-loop output equivalent circuit will be heavily loaded by the small load resistance, leaving high voltage distortion with high current drawn from the op-amp.

Fig. 3.5 shows the canonical- π equivalent circuit of a two-stage amplifier. C_3 and G_3 are the loads. C_2 is the miller capacitor, and C_1 is primarily the output transistor's gate capacitor. g_{m1} models the input stage.

When the feedback factor is β . We have the loop gain as:

$$T(s) = A_0 \frac{1 - s/\omega_z}{(1 + s/\omega_1)(1 + s/\omega_2)} \quad (3.6)$$

$$= \frac{\omega_u}{s} \frac{1 - s/\omega_z}{(1 + \omega_1/s)(1 + s/\omega_2)} \quad (3.7)$$

$$= \frac{R_1 R_3 \beta g_{m1} g_{m2} \left(1 - s \frac{C_2}{g_{m2}}\right)}{1 + s \left((1 + g_{m2} R_3) R_1 C_2 + R_1 C_1 + R_3 C_3 + R_3 C_2 \right) + s^2 R_1 R_3 \sum C_i C_j}, \quad (3.8)$$

$$\text{where } \sum C_i C_j \triangleq C_1 C_2 + C_2 C_3 + C_1 C_3 \quad (3.9)$$

$$(3.10)$$

$$\text{Unity-gain frequency: } \omega_u \triangleq 2\pi f_u \approx \left[\frac{\beta g_{m1}}{C_2} (1 \parallel g_{m2} R_3) \right] \parallel \left[\frac{g_{m2}}{C_3} \cdot \frac{\beta g_{m1}}{G_1} \right] \quad (3.11)$$

$$\approx \frac{\beta g_{m1}}{C_2} (1 \parallel g_{m2} R_3), \text{ if } \frac{\beta g_{m1}}{G_1} \rightarrow \infty \quad (3.12)$$

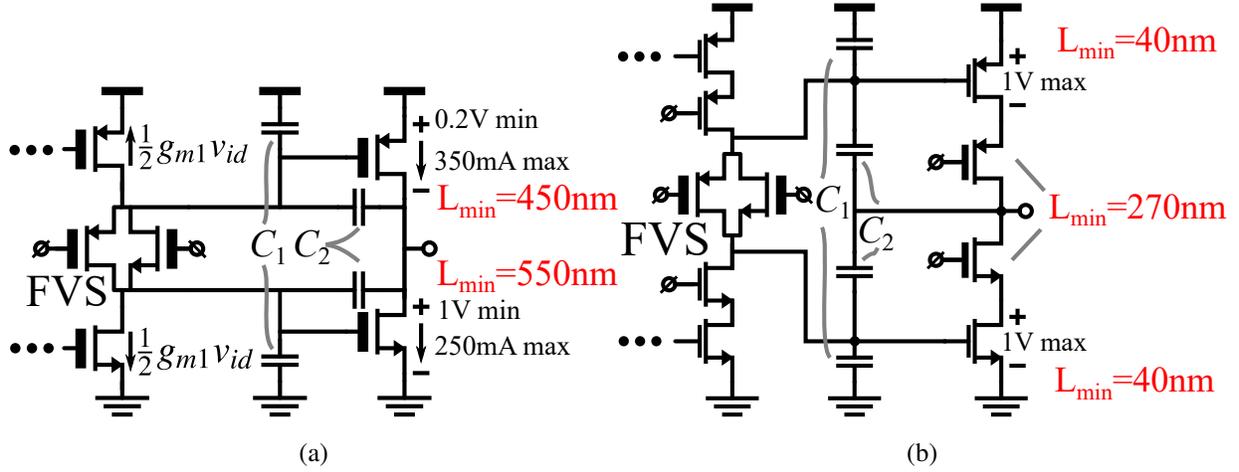


Figure 3.6: Op-amp's output stage with (a) HV IO devices; (b) Stacking of core devices and regular IO devices.

$$\text{Second pole frequency: } \omega_2 \triangleq 2\pi f_2 \approx \frac{g_{m2} \frac{C_2}{C_1+C_2} + G_3}{C_3} \quad (3.13)$$

$$\text{RHP zero frequency: } \omega_z \triangleq 2\pi f_z = \frac{g_{m2}}{C_2} \quad (3.14)$$

To meet the close loop bandwidth requirement with some margin, the unity gain frequency f_u is selected as 300 MHz. For a flat close-loop response, we need to push the f_2 away to twice f_u (3.12), by having a certain g_{m2} , and a reasonable miller factor $C_2/(C_1 + C_2)$, because the pole formed by $G_3 = 20\Omega$ and $C_3 = 100\text{pF}$ is only at 80 MHz (3.13). There is a maximum available g_m out of a certain bias current, so the requirement on g_{m2} largely determines the quiescent bias current of the output stage.

To design a large f_u , we need a smaller C_2 without penalizing βg_{m1} . A smaller C_2 also pushes away RHP zero such that $f_z \gg f_2$ in (3.14). But if we reduce C_2 , C_1 has to be reduced as well, because we need a reasonable miller factor on g_{m2} for its effect in (3.13). So basically we need a faster output transistor to save βg_{m1} . If we use the HV IO devices to handle the 3.6 V supply, the requirement of high current handling capability demands a very high W/L (Fig. 3.6(a)). Which leads to undesirably large C_1 and C_2 . One way to solve this problem is adding a buffer stage between the input stage and the output stage [6]. But this method poses new problems in our particular technology:

1. The input capacitance of the HV IO devices is still on the order of 100 pF, which needs a high current to drive with a bandwidth much larger than f_u ;
2. The buffer still needs to sustain 3.6 V full battery voltage, such that the same problem of the original op-amp falls onto the new internal buffer;
3. If we reduce the gate swing of the output transistor and use core devices, we need extra voltage rails (1 V and 2.6 V). And the slow HV IO devices becomes even slower, because of the smaller over-drive voltage.

Instead, we stack core devices (1 V) and regular IO devices (2.5 V) for speed and voltage tolerance at the same time (Fig. 3.6(b)). This way, the output current itself pushes the cascode pole away from f_2 , so additional bias currents are not longer necessary. Gates of the IO devices are biased to limit V_{ds} of the core devices to 1 V. At the extremes, the IO devices can enter triode region but the core devices remain in saturation. As a result, the input stage enjoys driving fast output transistors and gives an f_u with smaller current. The input stage can also be changed to the cascode structure to raise its output impedance for extra loop gain, thanks to the speed of the core devices. The Class-AB bias of the output stage is realized with a Monticelli floating voltage source. The coupled complimentary source followers can push the gate biases of the output transistors independently towards the rails and control the quiescent bias current I_{Q2} .

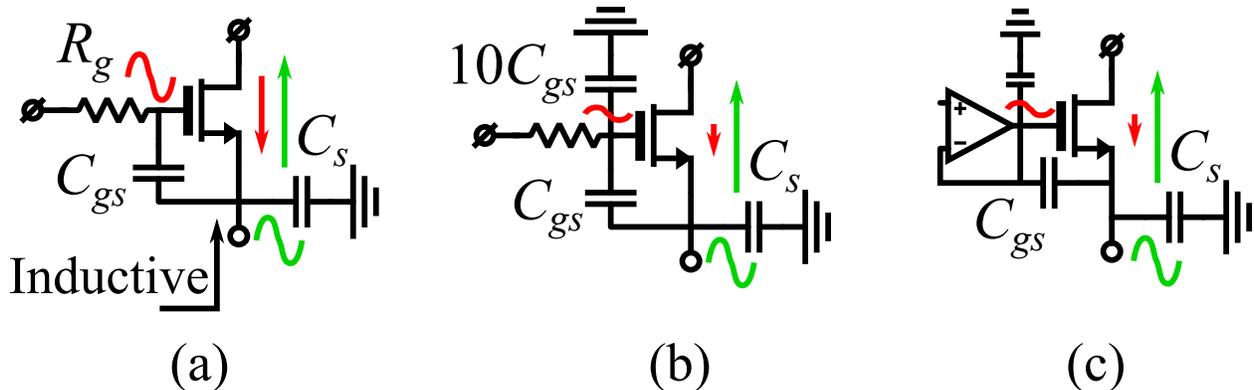


Figure 3.7: Cascode biases

Cascode biases need special attention for this op-amp. In a common gate device with a large gate resistance R_g and a large C_{gs} , the input impedance looking into the source terminal becomes

inductive early on, because of the feed-forward voltage through the high-pass filter generates a current cancelling the transconductance current from the source voltage (Fig. 3.7(a)):

$$Z_s = \frac{1}{ng_m} \underbrace{\frac{1 + sR_g C_{gs}}{1 + sR_g C_{gs} \frac{n-1}{n}}}_{\text{inductive}} \parallel \left(\frac{1}{sC_{sb}} + R_g \right) \quad (3.15)$$

This is undesirable for our ET application, as the current of op-amp's output transistor will change dramatically from sourcing to sinking.

To solve this, we can insert a bypass capacitor that is much larger than C_{gs} to create a low-impedance bias on the gate (Fig. 3.7(b)). Then the input impedance will be resistive ($1/ng_m$) with a bandwidth of half of transistor's f_T , assuming $C_{sb} \approx C_{gs}$. But even if the regular IO devices have shorter L_{\min} than HV IO devices do, to handle the high output current, C_{gs} of the IO devices are still on the order of tens of pico farad in our case. Hence, inserting bypass capacitors on the order of hundreds of pico farad for the two output cascode transistors burdens chip's floor plan. Instead, we choose to use auxiliary op-amps to create the low impedance bias actively, with a total penalty of 4 mA extra bias current (Fig. 3.7(c)). In the end, we have a cascode pole on P-side at 2 GHz, and a cascode pole on N-side at 8 GHz, which are both larger than our f_2 at roughly 600 MHz.

The op-amp is then sized following these steps:

1. From voltage headroom, maximum current capacity and $L = L_{\min} = 40$ nm, the output stage's FETs can be sized. Then, C_1 can be roughly estimated in total as about 2.5 pF;
2. Assuming a reasonable miller factor $C_2/(C_1 + C_2)$ of 0.7, C_2 can be estimated as roughly 6 pF;
3. From ω_2 , we can calculate g_{m2} as roughly 500 mS at quiescent bias point using (3.13). This leads to $I_{Q2} \approx 17$ mA with weak inversion assumption. Then, we can design our Monticelli bias voltages;
4. From C_2 and ω_u , βg_{m1} can be calculated as around 15 mS by *successive approximation* with (3.11) and (3.12), since we do not have a priori knowledge of input stage's gain g_{m1}/G_1 ;

C_2 are designed to be programmable for the easier 20 MHz mode.

3.2.3 Op-Amp Simulation Results

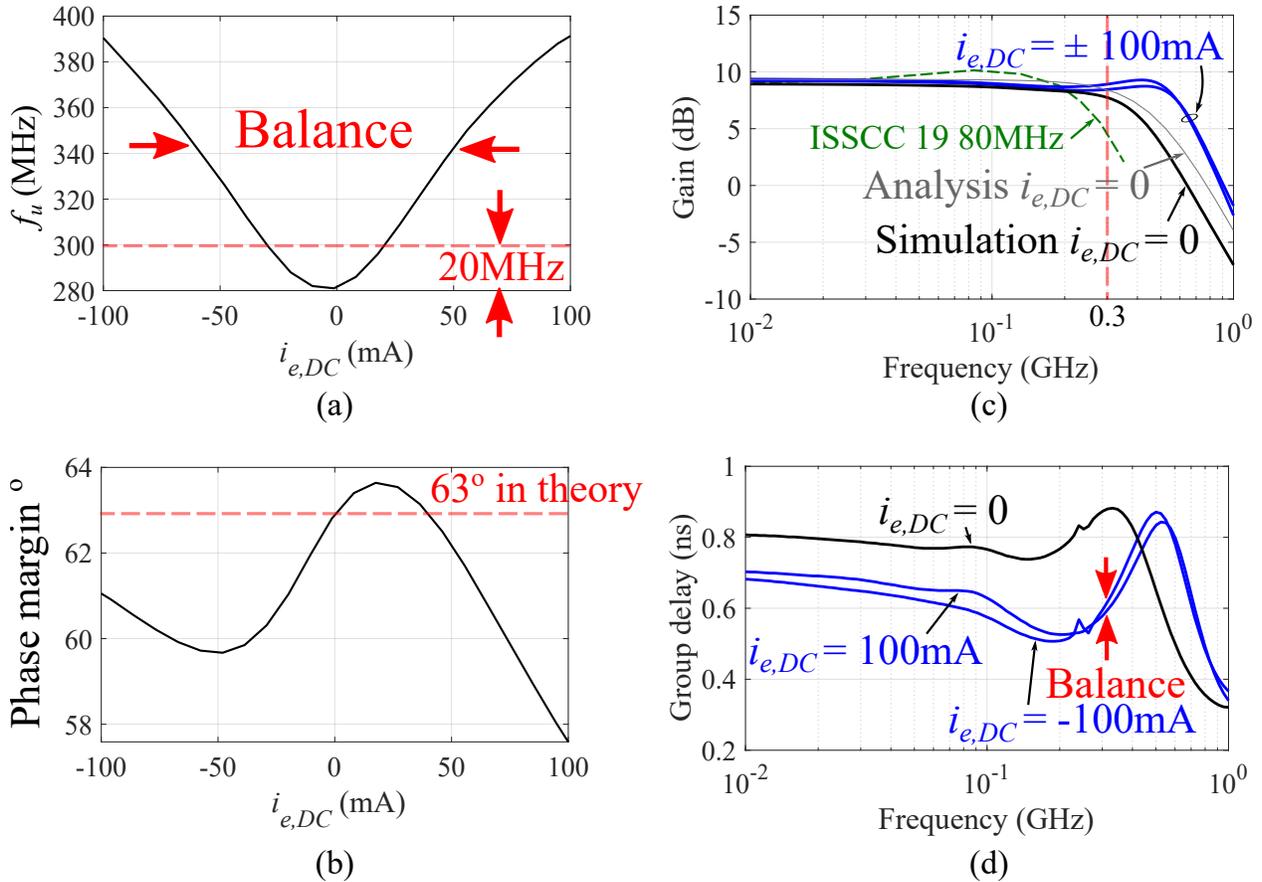


Figure 3.9: Simulation results of the complete op-amp: (a) Loop-gain's unity gain frequency f_u and (b) Loop-gain's phase margin as functions of the op-amp's output current (+: sourcing / -: sinking); (c) Close-loop transfer function; (d) Close-loop group delay

Simulation shows that the close-loop response meets the design target at quiescent in Fig. 3.9(c). When sourcing and sinking different currents, the op-amp's close-loop transfer function will inevitably change, so we need to fine-tune the biases and sizing in the simulator to balance the close-loop transfer function when sourcing and sinking the same amount of current, as shown in Fig. 3.9(a-c). Compared to the published simulation results in [15, Fig. 27.5.3] for 80 MHz ET, our op-amp does have enough bandwidth for the 160 MHz target. The group delay variation at

300 MHz is 0.25 ns as shown in Fig. 3.9(d). This variation meets our specification, because it is on the order of the 0.1 ns calibration resolution of the delay between the RF signal and the ET signal for the 160 MHz mode, as allowed by the measurement instruments in the lab.

3.3 DC-coupled Programmable Hysteresis Comparator

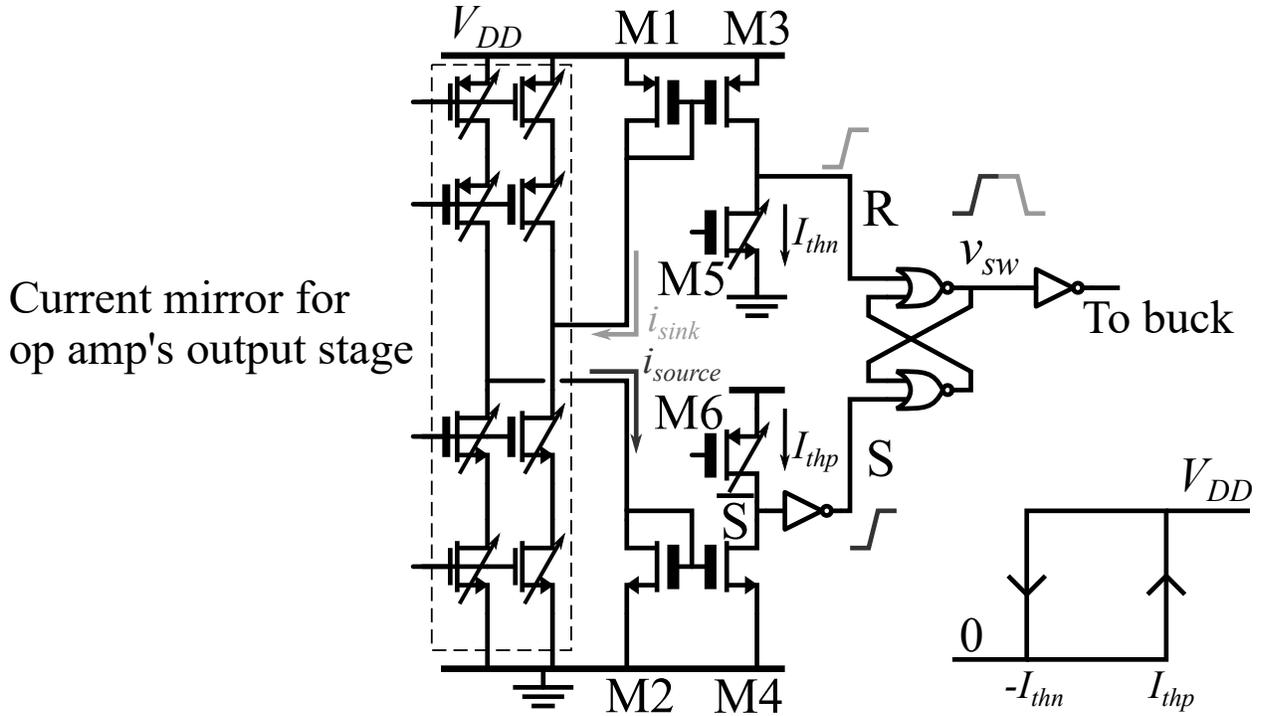


Figure 3.10: Schematic of the programmable DC-coupled hysteresis comparator

The DC-coupled hysteresis comparator adopts the most straightforward architecture (Fig. 3.10). Two scaled copies of the op-amp's output stage mirror the op-amp's sinking current (i_{sink}) and sourcing current (i_{source}). They are both injected to the current mirrors built in HV IO devices (M1-M4). The pull-down M5 and the pull-up M6 are biased by DC currents of I_{thn} and I_{thp} respectively. Then, node \bar{S} and node R become high-impedance nodes that act like current comparators for i_{source} versus I_{thp} , and i_{sink} versus I_{thn} . An SR latch will generate the hysteresis operation as follow:

1. When $i_{source} < I_{thp}$ and $i_{sink} < I_{thn}$ ($i_e \approx 0$), \bar{S} is pulled up and R is pulled down, so v_{sw}

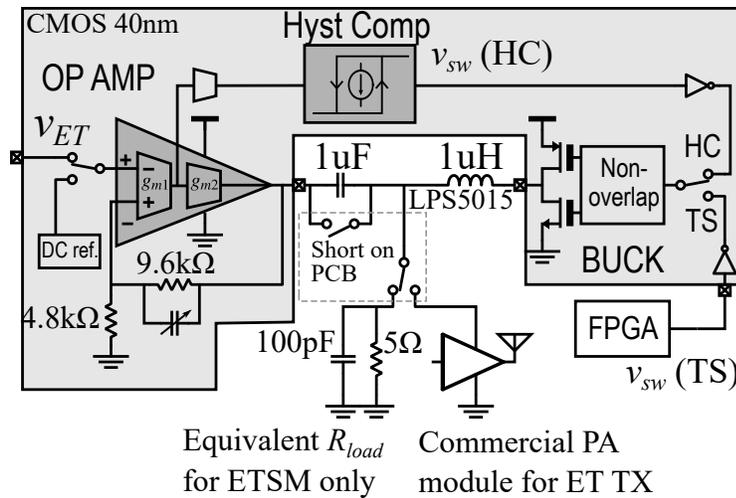


Figure 3.11: Implemented chip's schematic

maintains its value;

2. When the op-amp starts to source current to compensate the deficient i_{sw} ($i_{source} > I_{thp}$), \bar{S} is pulled down, so v_{sw} is set to V_{DD} ;
3. When the op-amp starts to sink current to drain excessive i_{sw} ($i_{sink} > I_{thn}$), R is pulled up, so v_{sw} is reset to 0.

This forms a hysteresis window between $-I_{thn}$ and I_{thp} for the scaled version of i_e .

The advantage of this topology is that it is the simplest possible hysteresis comparator that merges current sensing and comparing. It has a very low delay (< 1 ns) and does not require any static bias current. The disadvantage is that due to the large mirroring ratio and the mismatch in voltages between the current mirror nodes and the op-amp's actual output. The effective I_{thp} and I_{thn} may vary across PVT and operating conditions. To solve this issue, both the current mirrors' scaling ratios and the reference currents in M5-M6 are designed to be highly programmable to relax the need for accurate matching.

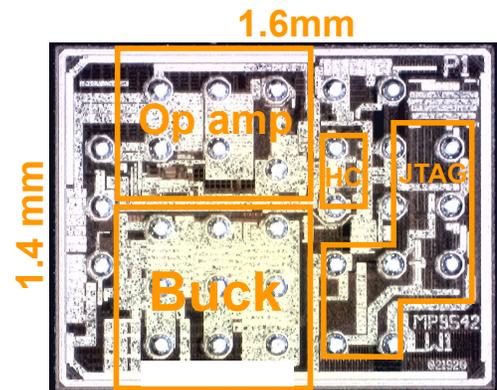


Figure 3.12: Die photograph

3.4 Chip Implementation

The overall system is taped out in TSMC 40nm process as shown in Fig. 3.11. The buck converter takes v_{sw} from either an external port connected to the DSP core on the FPGA, or from the conventional current-mode hysteresis comparator on the chip. The op-amp and the buck converter take about 60% of the total area, while the rest 40% contains the JTAG's digital blocks and on-chip bypass capacitors. Fig. 3.12 shows the chip's photograph. v_{ET} and v_{sw} input bumps are on the left side. The buck converter and the op-amp's supplies, grounds, and outputs use multiple bumps in parallel to satisfy the requirement on the bump's current density. The system can be programmed via a JTAG interface on the right side of the chip.

Chip layout is a key aspect in this project, due to the high currents in the on-chip metal connects. Layout of the op-amp's output stage and the buck converter's output stage is particularly difficult. Thus, we include some special considerations for chip layout, in the hope of guiding any future researchers.

The power transistors of the op-amp's output stage and buck converter's output stage need to be laid out hierarchically. The technique is useful for any high current CMOS layout. Fig. 3.13(a) summarizes the 1P6M metal stack in this process. M6 is a special ultra-thick (3.5 μm) copper layer, with extremely low sheet resistance. The top is a thinner aluminum layer used to connect the WLCSP bumps. M2 through M5 have the same intermediate sheet resistance. M1 has the highest sheet resistance. Thus we allocate the metal layers as follow:

- M6: Macro distribution fingers to carry high currents;
- M2: Gate metal connection;
- M4 ~ M5: Stacked layers with as many vias as possible that serve as the fine horizontal distributor of the current from M6;
- M3: Local source and drain fingers spamming 100% width with as many via as possible down to N+ and P+;
- M1: Body bias connection.

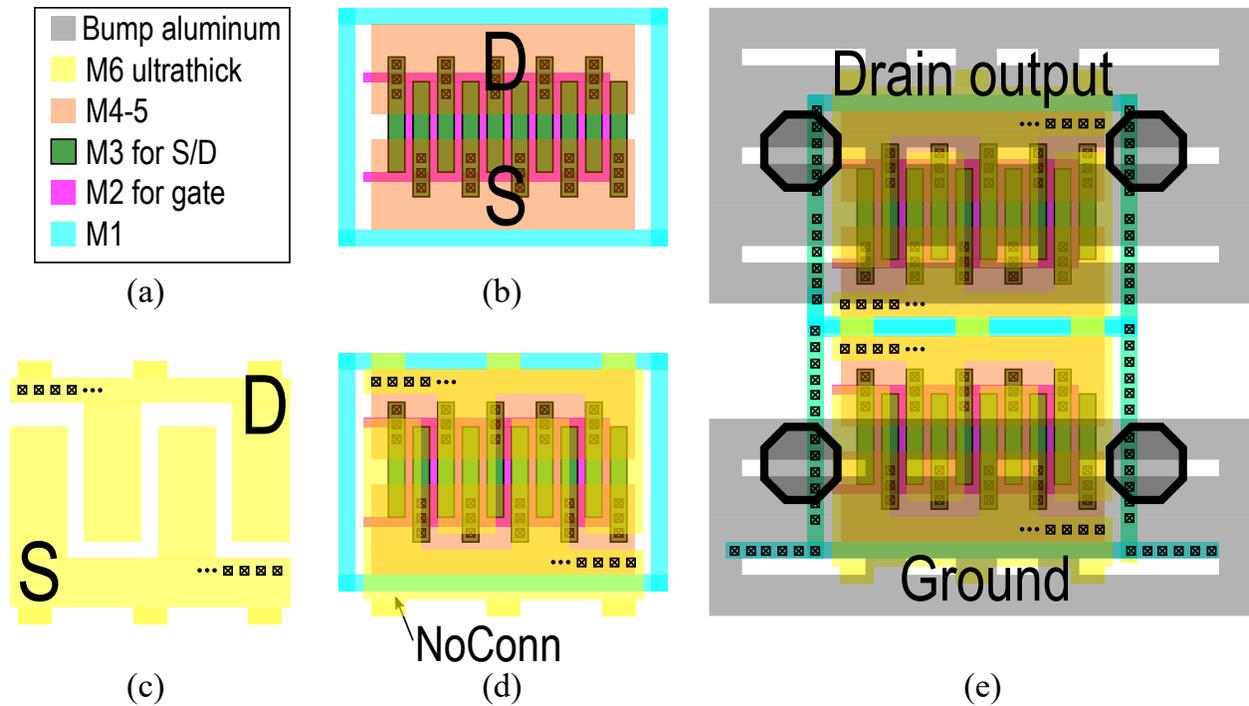


Figure 3.13: Important layout steps of the high-power transistors: (a) Chip’s metal allocation; (b) Layout of the lower metal layers as fine current distribution fingers; (c) Layout of the ultra-thick metal layer as coarse current distribution fingers; (d) One complete unit cell of the power transistor which can overlap with copies of itself to form larger transistors. The body is intentionally not connected; (e) After the unit cell is copied and reused, a large vertical distribution network is formed on M6, which interfaces with the bump aluminum layer. The body is connected to ground bump on this layer.

We first customize the unit cell with multiple fingers in one row as shown in Fig. 3.13(b). The gate is connected through M2 on both sides to minimize the gate resistance, because charging and discharging the buck converter’s output transistors’ gates demand substantial current spikes. Then M3 through M5 are stacked with the densest vias. They form the local fine current distribution fingers. In the meantime, since M6 has a larger minimum pitch than M1 to M5 do, we build local coarse current distribution fingers on M6 with wider traces. The coarse fingers are then joined on top and bottom with a horizontal bar to form high current paths for source (S) and drain (D) (Fig. 3.13(c)). Note that we intentionally thin the horizontal joint on the D side, because the width will be doubled after cell reuse. The small teeth, which will overlap each other, are added to reduce the density for design rule check (DRC) after cell reuse. The transistor must be surrounded by a

guard ring for its body bias, but there is another special consideration. Since M1 of this process is very thin and has a low electromigration limit, we must not connect the guard ring to the source (or ground) locally. Otherwise, even a small portion of the huge D-S current flowing through M1 will cause electromigration failure. During layout of the unit cell, there will be thousands of electrical rule check (ERC) errors due to the unconnected body, but they will be fixed in batch once we combine all the unit cells on the top level. The complete unit cell is illustrated in Fig. 3.13(d).

Fig. 3.13(e) shows how the cells are reused and overlapped to form bigger transistors, which eventually interface with the aluminum top. Note that all the M1 guard rings now form a guard mesh at the bottom. We just need to use a few vertical M6 bars on the sides to connect the guard mesh directly to the bump. This will bias the body and pass ERC, but the metal does not carry any current. All high currents are forced onto the carefully designed distribution fingers. This is called the “star” connection, as opposite to the “delta” connection. Our whole chip electromigration simulation shows that the core devices of the op-amp’s output stage have the highest current density on their D/S fingers.

The whole chip floor plan needs some special considerations as well (Fig. 3.14). The supply and ground of the buck converter are expected to be very noisy due to the switching dynamics of large power transistors, so we isolate the buck converter’s supply and ground properly on one side of the chip. The buck converter’s output is sandwiched by two rails, which forms a GSG port. The analog ground is designed as a very wide loop from M6 down to the substrate, which shield the op-amp from the switching noise. Furthermore, A stripe of high resistivity native silicon well in the substrate is also added around the buck converter to create a partially isolated p-type bulk for the buck converter. 4 bumps are allocated for the analog ground as shown in Fig. 3.14, the right-side 3 ground bumps sink high currents in parallel, and the left-side bump provides a low inductance return path for the op-amp’s input stage. There is only one signal that needs to jump between the two grounds: v_{sw} generated by the DC coupled HC. This signal is buffered locally with the buck converter’s rails after it crosses the boundary. All analog bias currents are generated from a bank of current mirrors that takes a 26 μA reference current from outside. The DC bias currents are distributed by buses at the center of the chip as marked on Fig. 3.14. They are safely shielded by

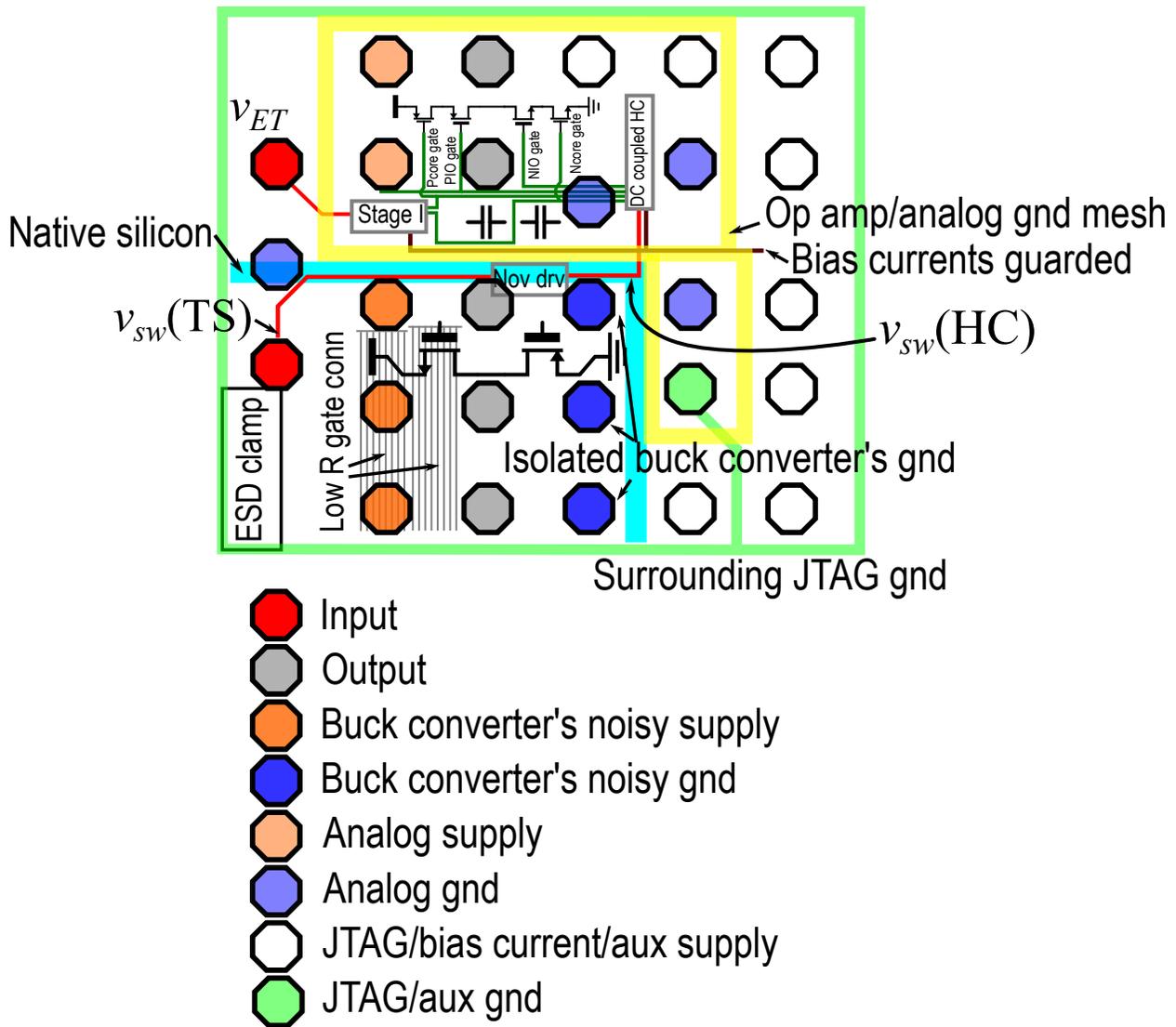


Figure 3.14: Conceptual illustration of chip's floor plan

the analog ground on top and bottom.

Here are some notes on the routing of several important voltages. The long extension of the buck converter's output transistors' gates renders high resistance, so we use wider M2 traces and a double-side connection as shown in Fig. 3.13(b). The op-amp's input stage generates gate signals of the core devices on the left side, which are fed to the big output stage and the DC coupled HC with properly shielded buses in the middle of the chip. The close-by metal fills around these two long routes are completely removed to minimize the parasitic capacitance that would lower the core transistor's f_T in effect. Two auxiliary op-amps that bias the gates of the cascode IO devices

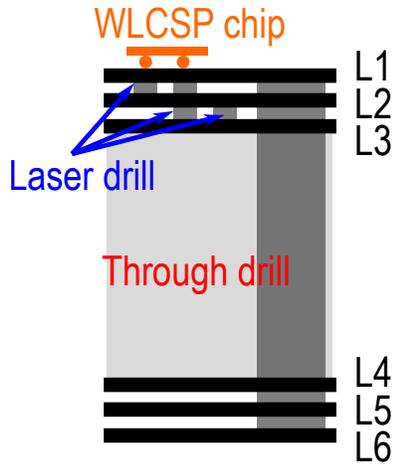


Figure 3.16: Metal stack of the PCB that is compatible with the WLCSP package

). The core of the PCB is a 21 mil thick layer that only allows mechanical through drills that have high via inductances.

In this project, the high current swing of the op-amp's Class-AB output stage and the pulsating current of the buck converter demand very low supply inductance. In addition, the large load capacitance (100 pF) will resonate with merely 1 nH inductance at 500 MHz in a series LCR network at the op-amp's output. This will erode the phase margin, as our f_u is right around 300 MHz. To get the minimum interconnect inductance, we use the top 3 layers with laser drills for routing in close proximity to the chip. L3 is used as the local PCB ground of the chip (Fig. 3.18(c)), and L2 is used as the local signal and supply routing for the chip (Fig. 3.18(b)). The rest of L2 is filled with the ground. Note that since the distance between L1 and L2 is very short, we have to chop L2 and use L3 as the ground of the transmission lines such that the line width can be larger for lower loss. L5 is designated as the solid ground for the entire PCB. L4 and L6 are used to route the supplies for the buffers that drive the chip's inputs as shown in Fig. 4.1. All through drills for L4-L6 are kept away from the chip's stencil. The purpose is not to disturb the chip's signal lines and ground as much as possible.

We co-designed the PCB with the chip's layout, such that we can verify the effects of PCB's parasitics. Thanks to the proper assignment of the chip's bumps, we can put an array of local bypass

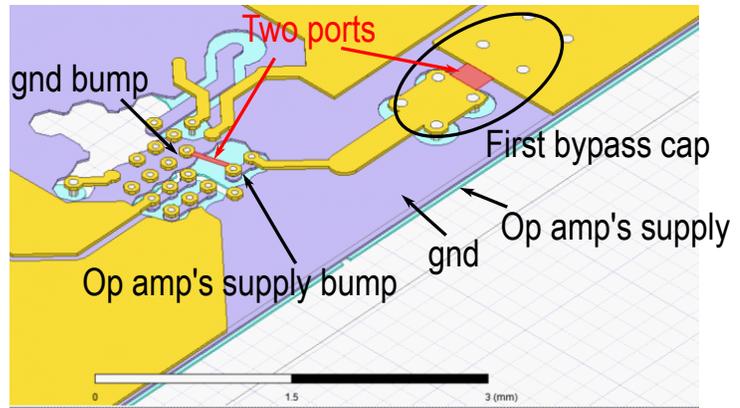


Figure 3.17: One example of the HFSS PCB interconnect simulation setup for co-design with chip's layout. The inductance of the PCB via together with the chip's bump is roughly 50 pH.

capacitors for the buck converter's supply and the op-amp's supply right next to the chip. Both the chip's input and output signals can then be sandwiched by rails. All important ports are partially cut from the whole PCB with ANSYS SIwave and simulated in HFSS. One example is shown in Fig. 3.17. Then, two port S-parameters are placed in EMX to fit to a 10-segment transmission line model. As expected, the characteristic impedance is extremely low ($< 10\Omega$), and the line looks very capacitive. But the co-design and co-simulation build up our confidence for a successful chip fabrication.

The layout of the differential-to-single-end buffers that amplify the weak signals from the DAC are customized to be narrow to handle the fan-in from the SMA connectors (Fig. 3.18(d)(e)). The buffers' supplies are routed through L4 and L6 aside from the chip.

Since the buck converter's inductor is large, it is difficult to place it close to the chip, while keeping the op-amp's output close to the load. There are two options:

1. Place the inductor on L1, but leave PA a bit further away from the op-amp's output;
2. Place it on L6 and use a lot of mechanical through drills from L1 to L6.

We choose method 1, because the layout of PA's RF ports dictates the PA's position, and, more importantly, too many mechanical drills disrupt the valuable ground planes on L2 and L3 that we strive to build. The layout of the PA, the inductor, and the chip's sitting pad is shown in Fig. 3.18(f).

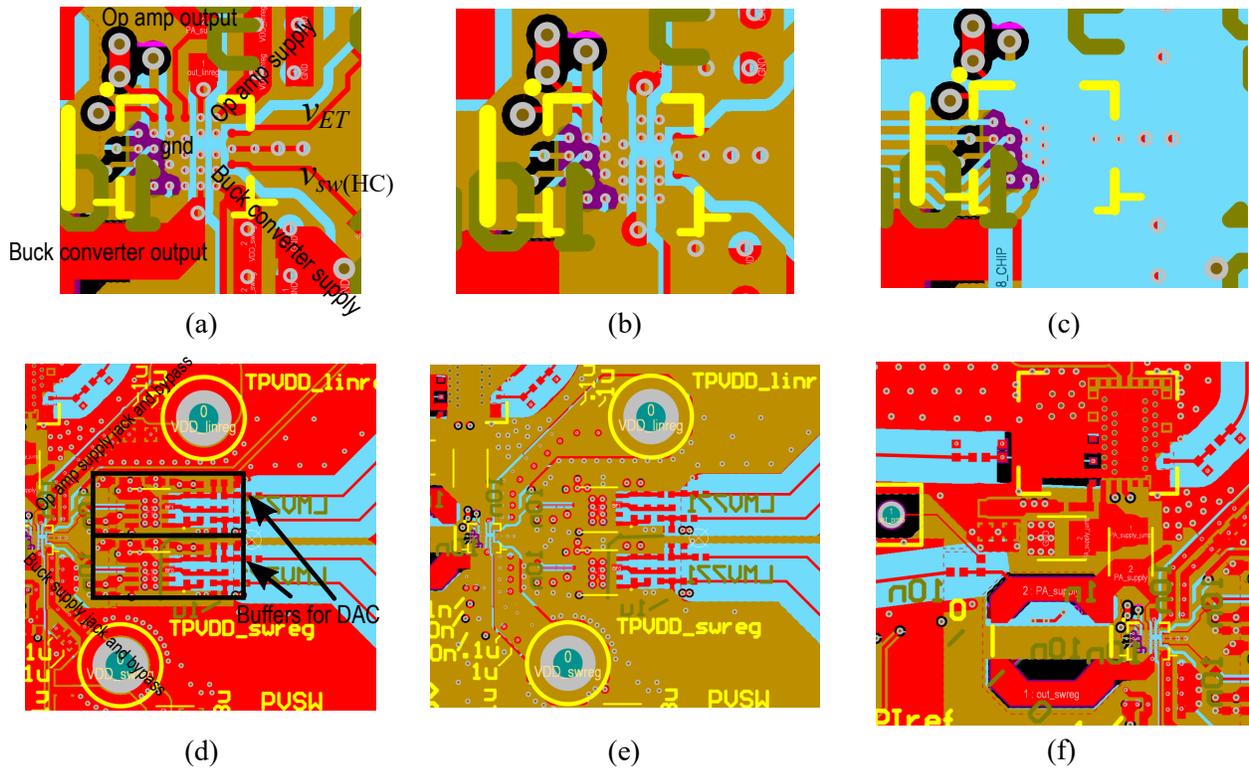


Figure 3.18: Layout around the chip (a) L1, (b) L2, (c) L3. Layout of DAC's buffer as in Fig. 4.1 (d) L1, (e) L2. (f) Placement of buck converter's inductor and the PA.

CHAPTER 4

Lab Measurements

4.1 ETSM-Only Measurements

We start from characterizing the ETSM only. A $5\ \Omega$ resistor in parallel with a $100\ \text{pF}$ capacitor are attached to the output of the ETSM to give a P_{sat} of roughly $2\ \text{W}$. For the conventional DC-coupled case, the measurement setup is shown in Fig. 4.1. For the $160\ \text{MHz}$ mode, we have to use Keysight M8190A to play the envelope of the WiFi packet. For the $20\ \text{MHz}$ mode, Kintex 7 FPGA drives Texas Instrument DAC39J84 to play both the envelope waveform and the DSP-calculated switching sequence. The differential-to-single-end buffer sets the proper swing and DC bias of the input signals (v_{ET} and v_{sw}) to the chip's bumps. In both setups, the output signals are picked up

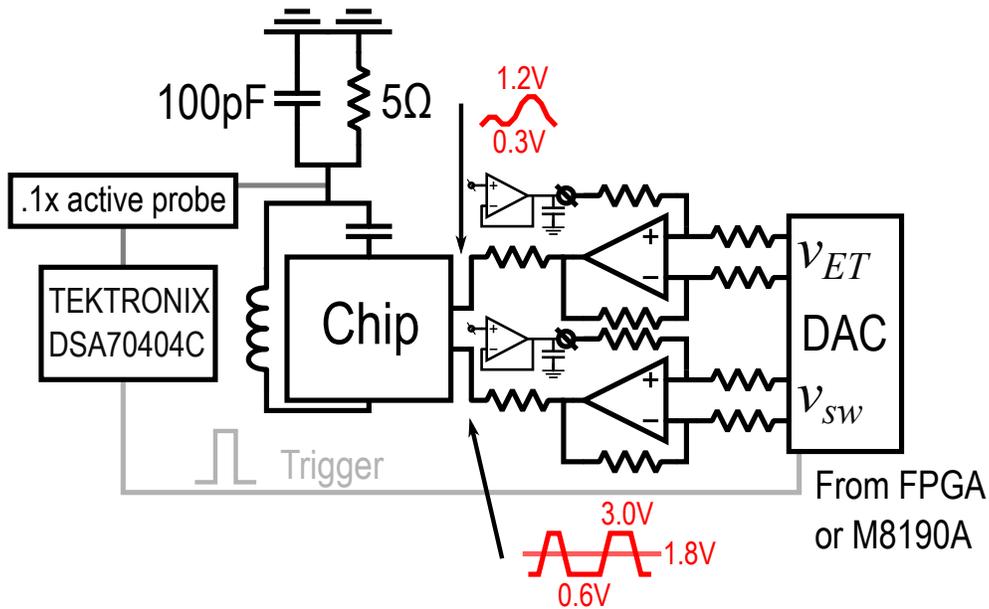


Figure 4.1: Measurement setup of ETSM only with RC load

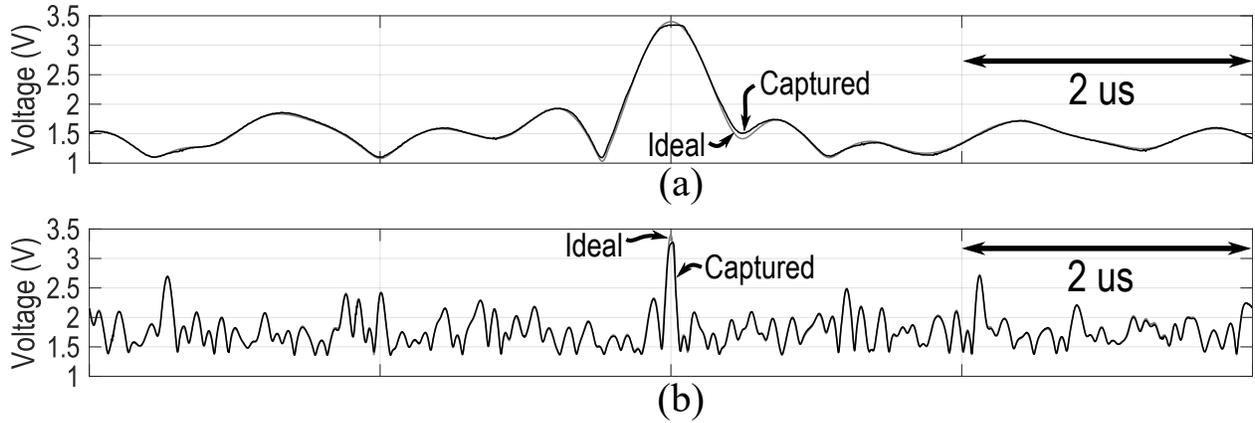


Figure 4.2: Measured ETSM's waveforms with respect to the references on 5Ω resistor load: (a) 20 MHz captured waveform with trellis-search; (b) 160 MHz captured waveform with hysteresis comparator

by a $0.1\times$ active probe and digitized using Tektronix DSA 70404C. The capturing and processing steps are as follow:

1. The waveform with the duration of one WiFi packet is captured, triggered by one of the DAC's output as shown in Fig. 4.1. Then the 8-bit sampled sequence from the scope is smoothened by a Savitzky-Golay filter.
2. The AC part is calculated by subtracting the signal by its average over the entire packet. An coarse gain is estimated from the AC parts of the reference signal and the smoothened signal.
3. An integer delay between the reference signal and the captured signal is found by comparing the AC part of the smoothened signal scaled by the rough gain to the AC part of the reference signal with the MATLAB *finddelay* function.
4. Rough DC offset is then found by averaging the difference between coarsely adjusted captured signal in gain and delay, and the reference signal.
5. With the coarse gain, integer delay, and coarse offset known, a 3-layer successive loop is then instated to search the fine gain and fine offset for minimum error defined as (4.1). Only integer delays are used here, as the sampling rates of the scope can be as high as 25GHz.

$$\text{mean} \left[\left| \frac{v_{\text{capture}} - v_{\text{ref}}}{v_{\text{ref}}} \right| \right] \quad (4.1)$$

Depending on the required resolution, the last step can be iterated multiple times with smaller steps. In our experiments, we find that a gain resolution of 0.1 and an offset resolution of 10 mV will be enough for the 160 MHz mode. Examples of the gain-delay-offset adjusted waveforms are plotted with respect to the references in Fig. 4.2.

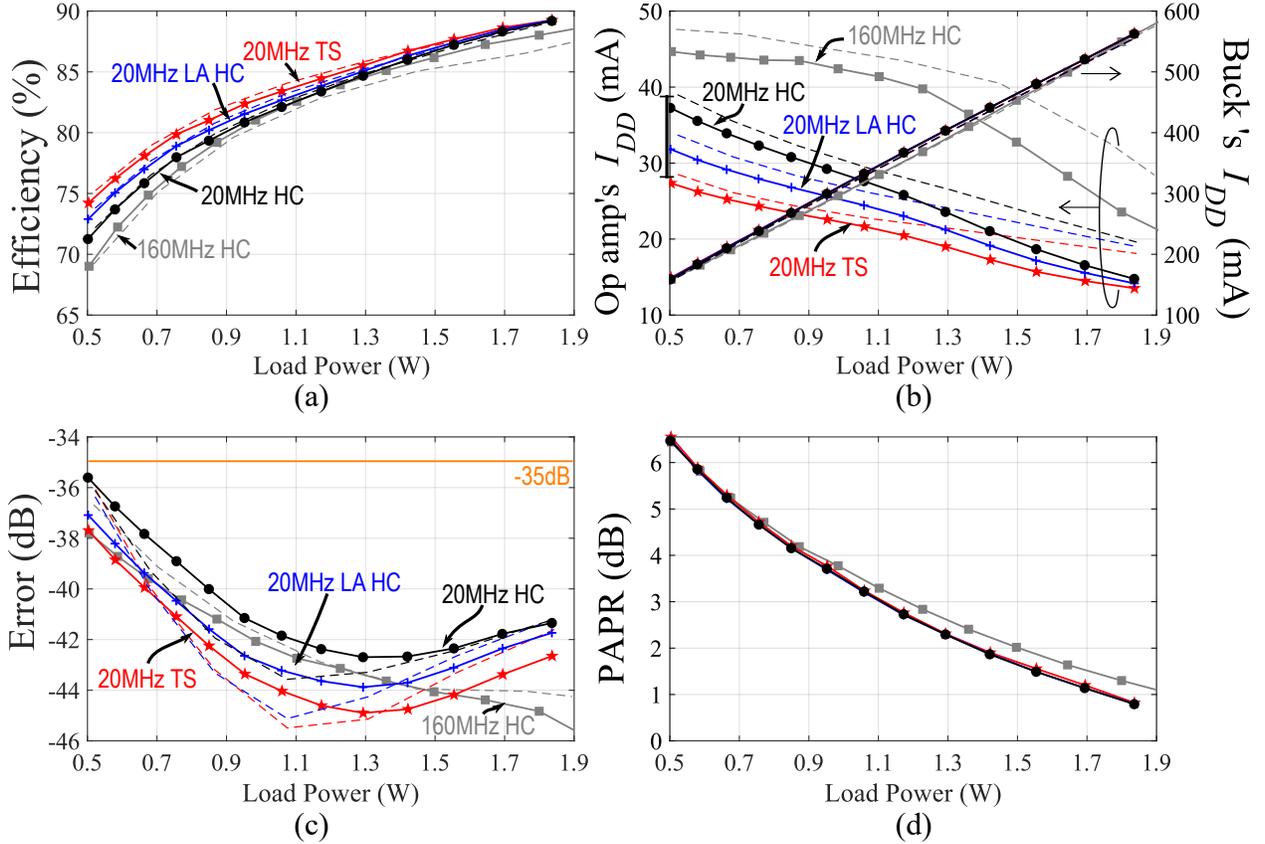


Figure 4.3: Measured ETSM's metrics for the conventional HC, look-ahead (LA) HC, and the trellis-search (TS): (a) ETSM's efficiencies with $5\ \Omega$ load resistor; (b) Supply currents of the op-amp and the buck converter; (c) Waveform errors; (d) Measured waveforms' error ($= \text{SNDR}^{-1}$)

Fig. 4.3 shows the measured ETSM's metrics with a $5\ \Omega$ resistor and a $100\ \text{pF}$ capacitor. Output power is swept with an experimental shaping function on the normalized envelope, which sets the maximum output voltage and the minimum output voltage:

$$v_{load} = V_{max} \left[\left(\left(1 - \frac{V_{min}}{V_{max}} \right) \times \text{normalized envelope} \right)^{1.1} + \frac{V_{min}}{V_{max}} \right] \quad (4.2)$$

Normalization means that the maximum envelope of the entire packet is normalized to *unity*. (4.2) is not used in Sec. 4.2 for the PA load, as the PA needs more sophisticated calibration on v_{load} .

Power is increased by reducing the PAPR with a raised V_{min} , while maintaining V_{max} Fig. 4.3(d). The ETSM's peak efficiency is about 88% when the output voltage is pushed to complete saturation (Fig. 4.3(a)). For 20 MHz we compare the efficiency with the same waveform on 3 cases:

1. Conventional hysteresis comparator (HC) on our chip;
2. Look-ahead (LA) hysteresis with a k picked on average;
3. Our trellis-search (TS) method.

The last two methods' v_{sw} signals are fed from outside via the bump.

At the toughest 6 dB back-off, TS improves the efficiency of the complete HA by 3%, from 71.2% to 74.2%, compared to the HC method. This is pretty close to our estimated 4% from the color map in Fig. 2.17. We also do slightly better than the LA hysteresis method, where the k is picked on average. As PAPR gets lower, all cases merge, because the signal looks more like a DC, where the switching optimization does not matter anymore.

TS unburdens the op-amp, as the measured op-amp supply current is lowered by almost 10 mA at high PAPR, when buck converter's currents are matched (Fig. 4.3(b)). The saving from TS is more significant than that from the LA hysteresis. For the 160 MHz mode, we have a lower efficiency due to higher bias currents and the slope-saturation mode as expected. We make sure that the waveform error is below -35 dB (4.1) for all cases, so that we are confident before moving to the PA measurement (Fig. 4.3(c)).

4.2 PA Characterization and Calibration for ET TX Measurements

To measure the ETSM in a complete TX system, we use a commercial PA that is originally developed for 80 MHz ET, but it can be pushed to operate for a signal bandwidth up to 160 MHz. The PA pre-driver's V_{DD} is held constant at 3.4 V for better linearity. The PA output stage's V_{DD} is modulated by the ETSM chip. To quickly demonstrate the complete ET system, we choose to fairly compare ET and constant supply without DPD. Then, the shaping between RF power and PA's V_{DD} is our only knob for a good efficiency-EVM trade-off (Fig. 1.2).

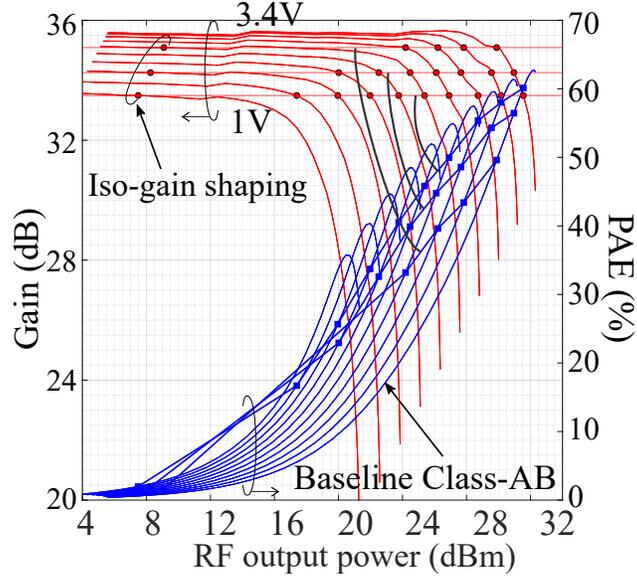


Figure 4.4: Waterfall curves to generate Coarse LUT for ET

Fig. 4.4 shows the waterfall curves that are used to find a good look-up table (LUT) between RF power and PA's V_{DD} . PA's gains and efficiencies are plotted with respect to the actual RF output powers over different V_{DD} 's ranging from 1 V to 3.4 V. The waterfall curves are measured by feeding the supply ports of the PA directly with well-calibrated DC supplies, so that the supply currents can be measured. We can see that the variation on PA's V_{DD} still affects its small signal gain by as much as 2 dB, so the first goal of our LUT is equalizing PA's gain across PA's different V_{DD} 's. If we choose to do an iso-gain shaping interpolation starting from a low V_{DD} (e.g. 1 V) as shown in Fig. 4.4, a coarse $LUT_{V_{DD}}$ between v_{load} and RF_{out} can be generated in Fig. 4.5(a). Then, the highest theoretical efficiency enhancement after ET can be achieved. But the iso-gain level hits a very compressed gain for the maximum V_{DD} , where significant AM-PM distortion that varies with V_{DD} also sets in, even if we have a flat gain with respect to the RF output power. Since we do not have AM-PM DPD at the PA's input port, EVM will be hurt. Instead, if we choose to run the iso-gain interpolation from a high V_{DD} (e.g. 1.8 V), then the gains are less compressed, but the theoretical efficiency enhancement after ET will be lower as shown in Fig. 4.4. In our measurements, a V_{min} of 1.2 V is eventually chosen for the 20 MHz mode, and a V_{min} of 1.3 V is set for the 160 MHz mode. These choices render satisfying EVMs as shown in Sec. 4.3 while maximizing the theoretical efficiency enhancement.

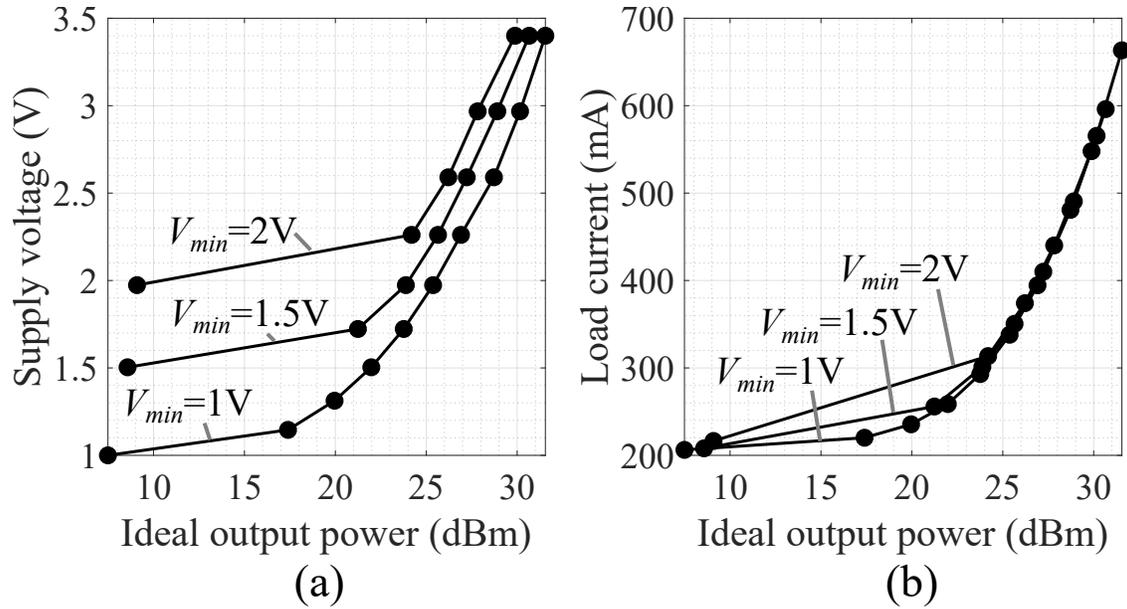


Figure 4.5: Coarse LUTs generated from different iso-gain interpolations in Fig. 4.4 for PA: (a) $LUT_{VDD}: v_{load} - RF_{out}$; (a) $LUT_{current}: i_{load} - RF_{out}$

Before running real waveforms on the ET system, we must fine calibrate the coarse LUT_{VDD} in Fig. 4.5(a) generated from the iso-gain interpolation on waterfall curves, because the gain error and offset of the ETSM will introduce finite errors to the actual LUT_{VDD} that the PA sees. For 802.11ax packets, PA's gain flatness has to be ≤ 0.1 dB not to degrade EVM after applying ET, so fine calibration is essential. The fine calibration steps are as follow:

1. The original coarse LUT_{VDD} is used to calibrate the gain error and offset of the ETSM. They can be corrected by adjusting the digital codes to the DAC, or the analog gain and offset of the DAC. The biasing op-amp for the differential-to-single-end buffer in Fig. 4.1 can also be used to correct the offset in an analog fashion. This step will give us a rough matching between the ideal v_{load} and actual measured v_{load} as shown in Fig. 4.7(b).
2. The average again from the coarse LUT_{VDD} is used as the fine calibration's target gain. Then, for each given RF_{out} level, the corresponding v_{load} is poked by several offsets (~ 10 mV) above and below the original v_{load} as shown in Fig. 4.6. The one offset that renders the closest gain to the target gain is updated as the new v_{load} in LUT_{VDD} . This step is repeated several times until a flat gain within 0.1 dB error is reached. In another perspective, this is a

linearization by manual feedback.

In the end, we should get our final LUT_{VDD} as shown in Fig. 4.7(b) with the verified flat gain and a static reference efficiency including the ETSM, as shown in Fig. 4.7(a).

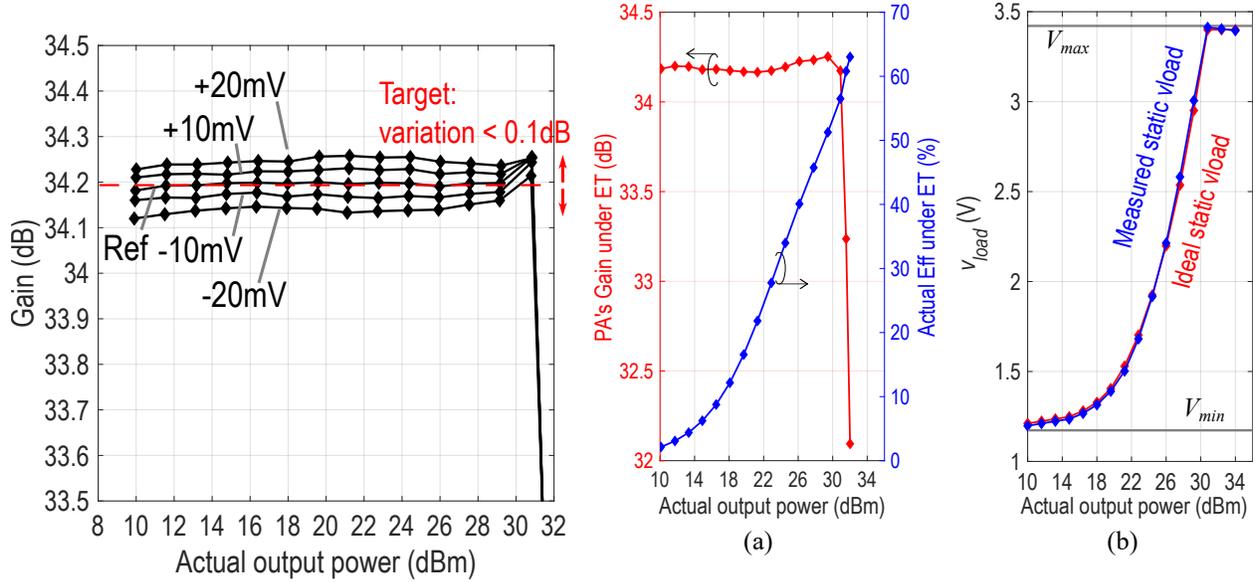


Figure 4.6: Manual feedback calibration to flatten the gain with variation ≤ 0.1 dB

Figure 4.7: (a) PA's gain and overall static efficiency including the ETSM after fine calibration; (b) Matching of measured static v_{load} and ideal v_{load} from LUT_{VDD}

There are two more considerations necessary for the ET measurements in the TS mode (open-loop DSP) with the PA:

Firstly, due to the series loss in buck converter's switches and the inductor, the actual V_H and V_L used in the TS algorithm should be slightly lower than than V_{DD} and 0 respectively in digital representation. Otherwise, the average of the actual v_{load} will be lower than the target v_{load} , i.e. the DC block capacitor will have a non-zero DC offset voltage. A good approximation for V_H and V_L is:

$$(V_H, V_L) = (V_{DD}, 0) - \langle i_{load} \rangle R_s, \text{ where } i_{load}'\text{s averaging is over entire packet} \quad (4.3)$$

R_s is the total series resistance of the buck converter, including the inductor's resistance, the switches' resistance and the PCB traces' resistance. $\langle i_{load} \rangle R_s$ is the offset voltage that changes

with the output power level. It is at most a few LSBs in digital representation and requires manual adjustments to the DSP core during the power sweep. A possible alternative to improve the accuracy of this calibration step is postulated in Sec. 6.1. This additional calibration step should be done together with the first step of the conventional calibration, and the objective is a match between target v_{load} and measured v_{load} as shown in Fig. 4.7(b).

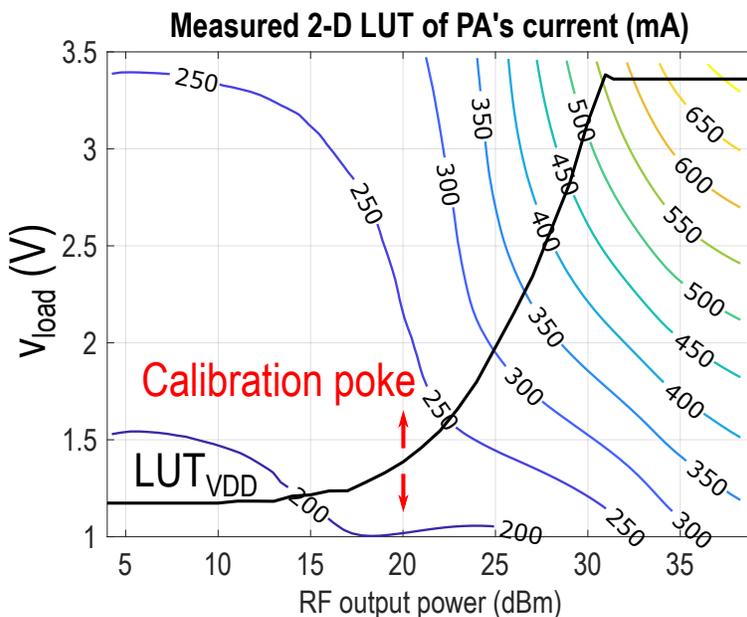


Figure 4.8: 2-D LUT for calibration in TS mode

Secondly, during the fine iso-gain calibration, instead of only calibrating the 1-D LUT_{VDD} table, we have to include the change in i_{load} when we poke the target v_{load} at a given power. Fig. 4.8 visualizes this. A 2-D map of i_{load} (PA's current) with respect to RF output power and v_{DD} 's must be prepared in high resolution for 2-D interpolation. Then, fine calibration of LUT_{VDD} is equivalent to moving the curve along the v_{load} axis in the current map as shown in Fig. 4.8. The final target is still a flat gain as illustrated in Fig. 4.7(a), despite the TS drive.

4.3 ET TX Measurement Results

The complete ET is tested with a two-stage Class-AB PA transmitting 802.11ax packets on a 2.5 GHz carrier. Each channel has 256-QAM. The EVM testing step is shown in Fig. 4.9(a, b). The

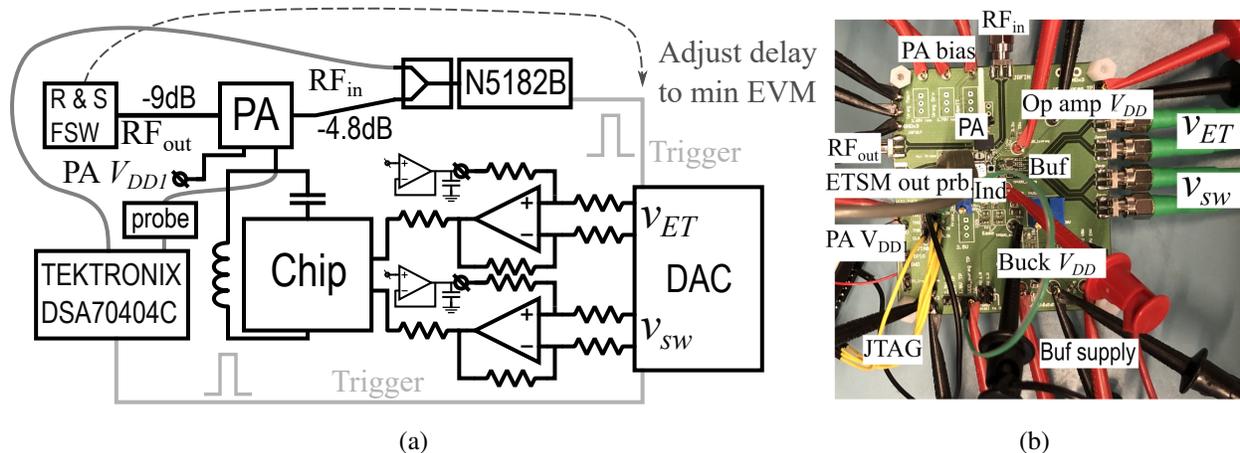


Figure 4.9: Measurement setup of ET TX system: (a) Schematic; (b) Lab photo.

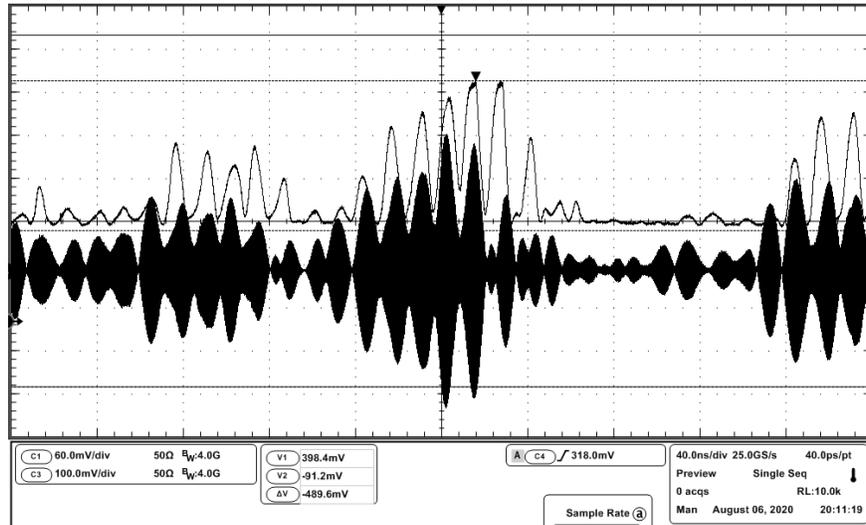
packet signal is upconverted with a signal generator and fed into the PA. The RF signal generator triggers the DAC that plays v_{ET} and v_{sw} in synchronous with RF signal generator's modulation. The DAC's outputs are buffered, level shifted and drive the chip's bump as they did for the resistive load.

The PA stage's V_{DD} is modulated by the ETSM chip, which is monitored by a scope with respect to the RF waveform for sanity check (Fig. 4.10(a)). The RF output is measured by a signal analyzer to record the spectrum and the constellation as shown in Fig. 4.10(b). 256-QAM is used in each OFDM subchannel, which has a tough EVM limit of -32 dB. Timing wise, we need to first adjust the delay through the trigger signal between the RF signal, and v_{ET} and v_{sw} to minimize EVM. Delay matching has to be within ~ 0.1 ns for the 160 MHz mode. Then, we need to slightly adjust the delay between v_{ET} and v_{sw} across two channels of the DAC for maximum efficiency, while maintaining the delay between v_{ET} and the RF signal.

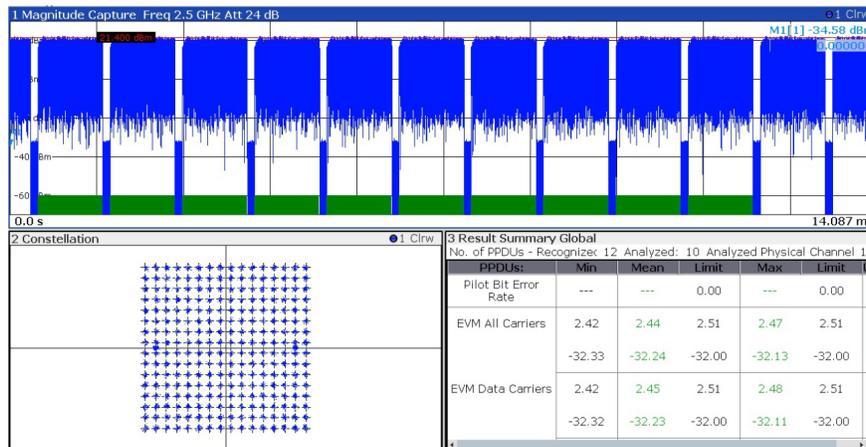
Compared to when the PA operates at a constant V_{DD} , ET using TS on a 20 MHz-wide signal improves the total power-added efficiency of PA, including the supply modulator, by 10.3% (48.8% relative) at EVM of -32 dB. ET using HC improves it by 9.4% at EVM of -31 dB (Fig. 4.11(a)). Due to the multiplicative nature of efficiencies, 3% of ETSM's efficiency improvement only translates to 1% of overall efficiency improvement, compared to the HC. For the 160 MHz-wide channel, the comparator-based ET with a conservative shaping function improves total efficiency by 7.2% at an

EVM of -32.7 dB Fig. 4.11(a). We can see from the EVM-efficiency trade-off curves in Fig. 4.11(b) that ET indeed moves the curve to the right, meaning that we are not cheating higher efficiency at a price of lower EVM. PA's output spectra are plotted in Fig. 4.11(c-d) at powers where EVM hits the MCS9 limit. We can see that ET does not worsen the spectra, and all masks are satisfied. Especially for 20 MHz mode, ET even lowers ACPR.

Fig. 4.12 shows the comparison between our work and the prior arts, We do not use unrealistically low V_{DD} 's, like some ETSM papers do without including any PA measurement [31, 32]. We have built a full ET transmitter for the highest bandwidth, compared to all previous works. In terms of ETSM's own efficiency, at similar back-off power, we see that TS does a little better than conventional hysteresis with one inductor. But more off-chip passives still offer more significant improvement, at the cost of the module space. Specifically, 1 extra off-chip inductor's size ($5\text{ mm} \times 5\text{ mm}$) is much larger than an ETSM chip's size ($1.6\text{ mm} \times 1.4\text{ mm}$). We did a very thorough measurement with the PA as well. We have the highest RF PAPR, and all EVMs meeting the standards, We are able to achieve similar efficiency enhancement compared to other works.



(a)



(b)

Figure 4.10: (a) Example waveform measurements in 160 MHz mode; (b) Example EVM measurements at EVM's limit.

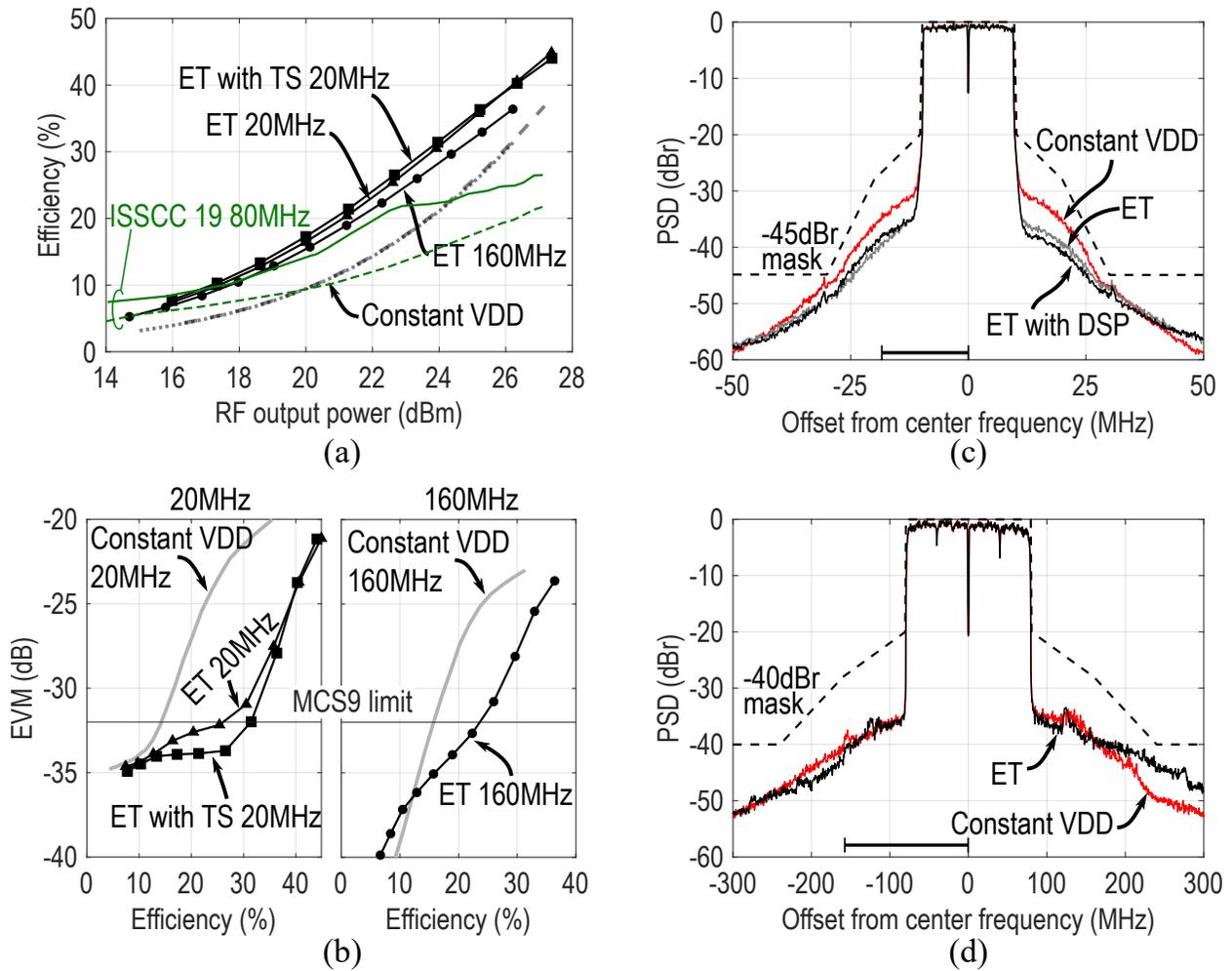


Figure 4.11: Measured ET TX's metrics: (a) Efficiency enhancements for various cases compared to a reference enhancement in [15]; (b) EVM-Efficiency trade-off curves for 20 MHz and 160 MHz mode; Spectra when EVM hits MCS9 EVM limit of -32 dB at (c) 24 dBm for 20 MHz mode and (d) 22.3 dBm for 160 MHz mode.

	This work		ISSCC 2019	ISSCC 2019	TPE 2019	ISSCC 2017	MTT 2017	JSSC 2010
Node (nm)	40		90	65	180	28	180	350
Supply (V)	3.6		5	2.4	3.6	3.6	4	3.3
Design	HC	TS	Subband + AC coupled	3-level	HC	HC	HC	2-phase
Signal BW (MHz)	160	20	100	80	20	40	60	N/A ²
Peak Eff (%)	88		88	91	90	95	80	89
Peak Power (W)	1.9		3	1	2.5		1.1	2
# Inductor	1		4	1 ¹	1	1	1	2
Back-off Eff (%)	74	77	81	68	70.6	75	67	74.2
Back-off (dB)	5	5	5	5	5		3.4	
Error (dB)	-39	-39		-40				
RF PAPR (dB)	12	10		7.84		9	11.2	N/A
RF Pout (dBm)	22.3	24	23	24		19	21.7	
EVM (dB)	-32.7	-32				-34	-28.9	
ACPR (dBc)	-35	-35.4	-38	-32.5	-35		-30	
ET+PA Eff (%)	22.3	31.4		23	38	30		
Relative Enhancement (%)	47.7	48.8		44	69	34		

¹12nF on-chip capacitor

²WCDMA signal

Figure 4.12: Performance summary and comparison to prior arts. The reference are listed in sequence: ISSCC 2019 [6], ISSCC 2019 [15], TPE 2019 [28], ISSCC 2017 [29], MTT 2017 [30], JSSC 2010 [8]

CHAPTER 5

Supplemental Extension of the Author's M.S. Thesis: Approximate Equivalent Circuits to Understand On-Chip Inductors

5.1 Introduction

The 3D electromagnetic fields surrounding a planar spiral inductor on a silicon substrate are complicated, and pose the main hurdle in reaching a simple analytical model for inductor design. The problem is solved by simulators such as PeakView™ and VeloceRF™, which formulate a two-port s -parameter model based on actual layout; they also optimize the geometry for a user-specified objective, such as the highest inductor quality factor (Q) at a given frequency. With these fast simulators the accurate design of on-chip inductors is now routine. On many occasions, though, circuit engineers want to know what trade-offs went into the optimal inductor geometry, and how sensitive the inductor's quality is to changes in geometry. It is to answer this that we present a simple, *approximate* equivalent circuit which helps to explain the electrical properties of the simulator's chosen geometry.

Before the advent of fast simulators, numerical methods such as Partial Element Equivalent Circuit were in use [33–35]. These are accurate but, in essence, they are little different than the aforementioned simulators. Frequency-dependent models such as [36–40] are non-physical and can only be used at one frequency. Compact frequency-independent models [41–46] offer the most insight because they are simple and they can approximate the underlying physics well. But when the model parameters are found by direct fitting to EM simulations by, for instance, choosing element values that minimize least-mean-square errors on the two-port parameters, it can stray far

from the true physics. This broken link between fitted circuit parameters and the physics has remained a shortcoming in that the model usually cannot reveal the major contributors reliably from among the various sources of loss, or the role of parasitic elements.

To reach a better understanding, we present an approximate equivalent circuit with essentially frequency-independent elements calculated directly from the inductor geometry. This equivalent circuit can reveal how simulators balance various losses to arrive at the best design. And since the optimum design is only as useful as the objective function that led to it, we summarize two definitions of quality factor and present new design-oriented expressions for them.

The equivalent circuit was constructed by extending or simplifying published methods and analysis to model the four basic effects in an on-chip inductor:

1. the skin effect in a conductor,
2. the proximity effect in adjacent conductors,
3. the substrate capacitance and loss, and
4. inter-winding capacitance.

To validate the approximate circuit, we will use it to model well-characterized inductors, explore their design spaces, and explain why their particular geometries gave the highest quality factor. These case studies are the main contribution of this paper. Although this model can never surpass the simulator in accuracy over a broad frequency range, it offers physical insights of practical value that are adequate, we believe, to inform the user why a particular geometry is optimum.

5.2 Background of Extension

The author's M.S. thesis [47] has briefly covered the substrate capacitance and loss, and inter-winding capacitance among the four topics listed in Sec. 5.1. Some modifications are made during the Ph.D. study. In Summer 2015, we presented the M.S. work at a semiconductor company. Based

on the audience’s feedback, we carefully extended the research to elaborate on two specific topics that raised significant discussion during the presentation:

1. modeling of tapered inductors, and
2. definition of quality factor in the context of inductor modeling.

Since tapering only optimizes the conductive loss in an inductor, its modeling requires a rigorous understanding of the frequency-dependent current redistribution effect in the winding. Sec. 5.3 presents a more comprehensive study on inductor’s series loss, compared to the rudimentary version in the author’s M.S. thesis [47].

Sec. 5.4 includes slightly improved substrate models, compared to [47].

Sec. 5.6 adds a quantitative justification of why substrate eddy current can be safely ignored in modern RF CMOS processes, with some important references.

Sec. 5.7 corrects the ambiguities of [47] and proposes a new design-oriented expression of the quality factor of an inductor, which facilitates easy decomposition of inductor’s loss contributions.

The additional case studies carried out during the author’s Ph.D. study are included in Sec. 5.8.3 and Sec. 5.8.4. They put the extended analytical tools to use. The improved model reveals valuable information on how the various losses in the iuctor are distributed.

5.3 Improved Study on Inductor’s Conductor Losses

5.3.1 Skin Effect

We use the volume-filament-method of [48] to model a family of rectangular conductors with different aspect ratios but constant area of cross-section. This method is, in effect, the solution to 2D magneto-static fields as in [49] but uses a distributed equivalent circuit that enables a ready visualization by circuit engineers. It divides the conductor into filaments such that up to the maximum frequency of interest, the current density in each filament is uniform; although because of finite boundaries, it will change from filament-to-filament.

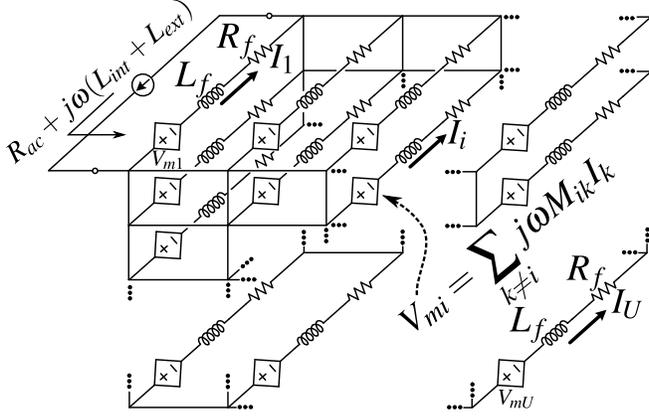


Figure 5.1: Equivalent circuit of a conductor based on the volume-filament-method in [48].

The filament i is modelled by a frequency-independent resistance R_f , self inductance L_f [50, Eqn. (20)], and mutual inductance M_{ik} to the other filaments at a distance d_{ik} [50, Eqn. (12)] and [51, Fig. 3]. The equivalent circuit in Fig. 5.1 illustrates these elements. The calculation to follow assumes that the filament is straight and infinitely long.

The circuit of Fig. 5.1 defines the net impedance $R_{ac} + j\omega(L_{int} + L_{ext})$ at the terminals, that is, at the driving point, of all the filaments in parallel that comprise a conductor. L_{ext} is frequency-independent and models the magnetic energy stored *outside* the conductor. L_{int} is $\propto 1/\sqrt{f}$ and models the magnetic energy stored *within* the conductor [49, Fig. 3(b)]. R_{ac} ($= R_{dc}$ at DC) models the loss in the conductor and increases $\propto \sqrt{f}$ [49, Fig. 3(a)]. The angle of the internal impedance $Z_{int} = R_{ac} + j\omega L_{int}$ starts at 0° and converges to a value $\leq 45^\circ$ depending on the aspect ratio of the conductor cross section¹. However, according to [53, Fig. 4], since $L_{int} < 50 \text{ nH/m} \ll L_{ext}$ at DC and L_{int} falls off at high frequencies, L_{ext} will dominate the total inductance of the conductor across a wide range of frequencies over which, therefore, the net inductance remains constant. Thus to model R_{ac} , we only need to synthesize an impedance with an asymptotic phase smaller than 45° and a value of R_{dc} at low frequencies. Any error in the phase, that is, in L_{int} , is lessened in its net impact because, to repeat, the frequency-independent L_{ext} in series will dominate.

¹If $R \propto \sqrt{f}$ and $X \propto f/\sqrt{f} = \sqrt{f}$, then $\arctan(X/R) \rightarrow 45^\circ$ because equal resistance and reactance \Rightarrow minimum impedance; see, for example, [52, Sec. 11.4] Current flows along the path of least impedance.

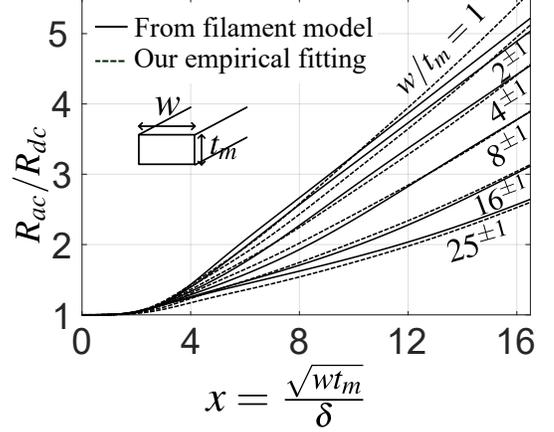


Figure 5.2: Skin effect normalized curves and our fitted results.

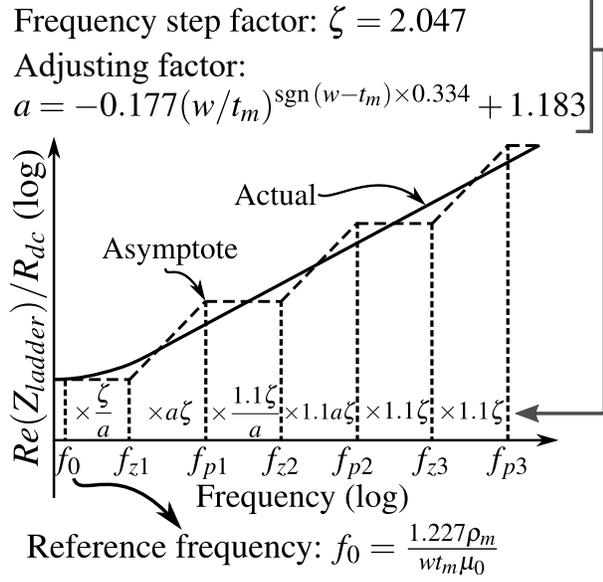


Figure 5.3: Ladder’s poles and zeros with their relative positions indicated by ratios.

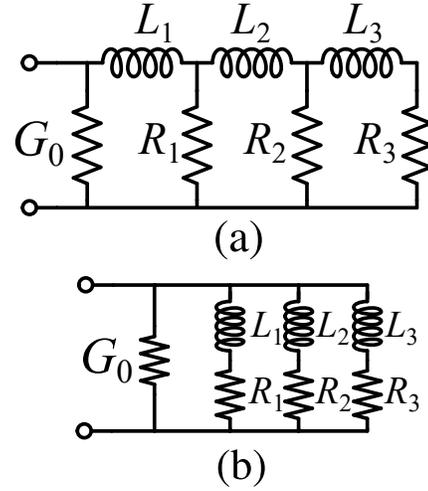


Figure 5.4: (a) Cauer-type LR ladder; (b) Foster-type LR ladder. They are equivalent.

The driving-point resistance R_{ac} of the network in Fig. 5.1 is plotted in Fig. 5.2 as a family of curves of R_{ac}/R_{dc} with respect to a normalized variable $x \triangleq \sqrt{wt_m}/\delta$, where δ is the skin depth². With proper scaling of the axis, it matches the family of normalized curves given in [54, Sec. 2-4, Fig. 3]. The thickness of the top layer metal on a chip rarely exceeds $4\ \mu\text{m}$, and the trace width in a typical on-chip inductor is usually smaller than $30\ \mu\text{m}$. To estimate practical bounds on x , assume the metal is pure copper, whose skin depth at 10 GHz is $0.66\ \mu\text{m}$; then an upper bound on this normalized variable is:

$$\max(x) = \frac{\sqrt{30 \times 4}}{0.66} \approx 16.5. \quad (5.1)$$

For conductors with smaller cross sectional area or at lower frequencies, $x \ll \max(x)$.

Solving field equations with a distributed network might be interesting to circuit designers, but it does not advance our search for a compact equivalent circuit. We will use the normalized curves of Fig. 5.2 from now on to synthesize a 6-component lumped network that approximates

²This distributed network was simulated by solving the circuit’s mesh matrix in MATLAB. The thickness of a filament is assumed to be $1/2$ the smallest skin depth, so determines the number of filaments needed for accurate results. In the case of $x_{\max} = 16.5$, we need $4x_{\max}^2 = 1089$. This leads to a 1089×1089 matrix, which MATLAB solves very quickly. Accounting for symmetry around the center line along the length of the conductor, the matrix can be shrunk $4 \times$ on each side.

an impedance which changes with a fractional power of frequency. We will show in Sec. 5.8 that this relatively simple network works well to model skin effect in many types of inductors. Following well-known approximation methods [55][56, Ch. 9], we consider a 3rd-order *LR* ladder circuit whose impedance is specified by negative real poles and zeros with certain ratios, interlaced in frequency:

$$Z_{ladder} = R_{dc} \frac{(1 + \frac{s}{\omega_{z1}})(1 + \frac{s}{\omega_{z2}})(1 + \frac{s}{\omega_{z3}})}{(1 + \frac{s}{\omega_{p1}})(1 + \frac{s}{\omega_{p2}})(1 + \frac{s}{\omega_{p3}})} \approx Z_{int}, \quad (5.2)$$

such that Z_{ladder} approximates the Z_{int} simulated with the volume-filament-method.

Fig. 5.3 gives approximate expressions for relative frequencies of poles and zeros in (5.2) as functions of the conductor geometry. The factor a adjusts for the unique frequency characteristic at the onset of skin effect in a rectangular conductor of aspect ratio $w/t_m > 1$; at higher frequencies $a \approx 1$.

(5.2) satisfies both the separability property and positive definiteness [56, Ch. 6] to be realized by a 3-section Cauer-type or Foster-type ladder with frequency-independent, positive L 's and R 's (Fig. 5.4). This synthesis is part of a MATLAB toolbox. The circuit elements in the ladder map uniquely to the variables in (5.2), including R_{dc} ; one or the other can be used for analysis, whichever is more convenient. Fig. 5.2 shows that for $x \leq 16.5$, the fitted frequency ratios using the equivalent circuit parameters of Fig. 5.3 match the resistance ratio curves generated by the volume-filament-method quite well. This approach to model the skin effect leads to the same, or a similar, equivalent circuit to previous work [57–59], but we are the first to synthesize it with poles and zeros for a frequency response that involves fractional powers of frequency (Fig. 5.2) arising from the physics of current flow.

Fig. 5.5 shows the comparison between the current distribution at various x 's calculated from the numerical Maxwell's equations in [49] and our volume-filament equivalent circuit. Note that in order to make the comparison, more filaments have been used in MATLAB to cover $\max(x)$ of 56 in Fig. 5.5. This is not needed for on-chip inductor modeling. Fig. 5.5 verifies that the current density around the center line parallel to the length of the conductor is an even 2D function across the conductor's cross section.

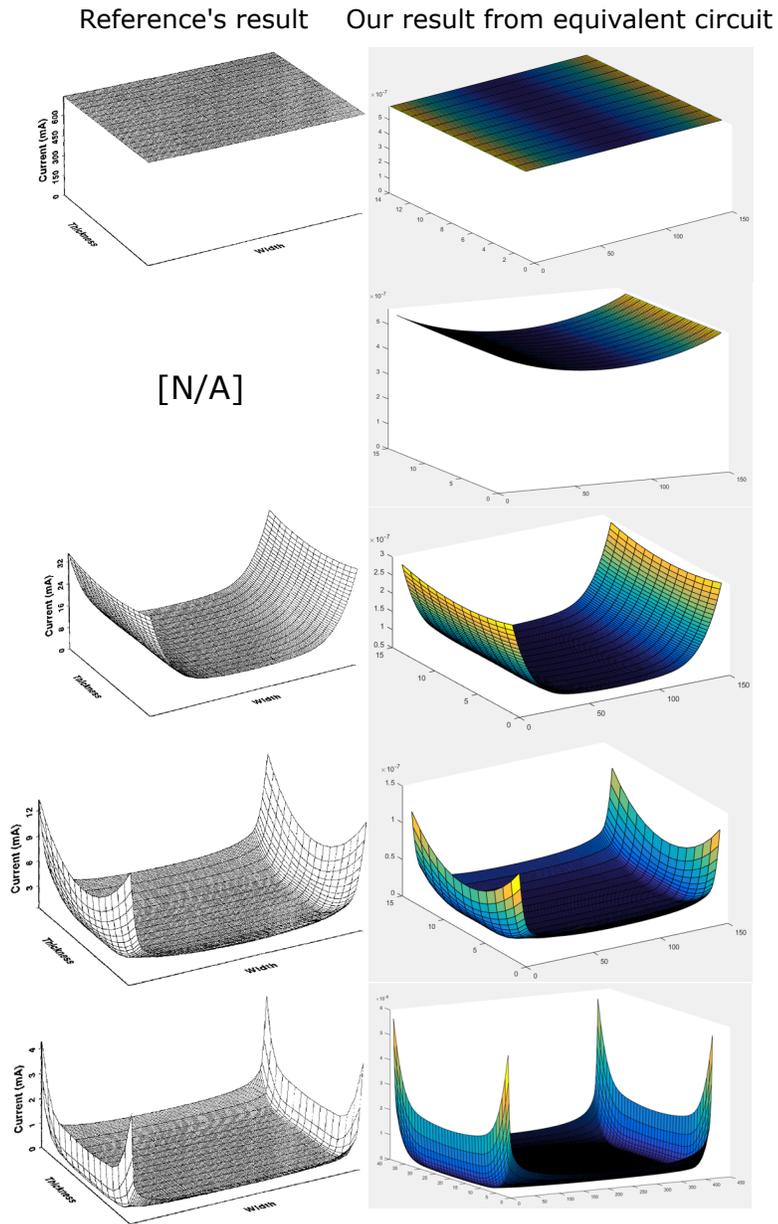


Figure 5.5: Comparison of the current distribution in the conductor of $w/t_m = 10.71$; Pictures on left column are results from [49]. Pictures on the right column are calculated from the volume-filament-method. From top to bottom: $x = 0.56$, $x = 1.77$, $x = 5.61$, $x = 17.74$, $x = 56.09$;

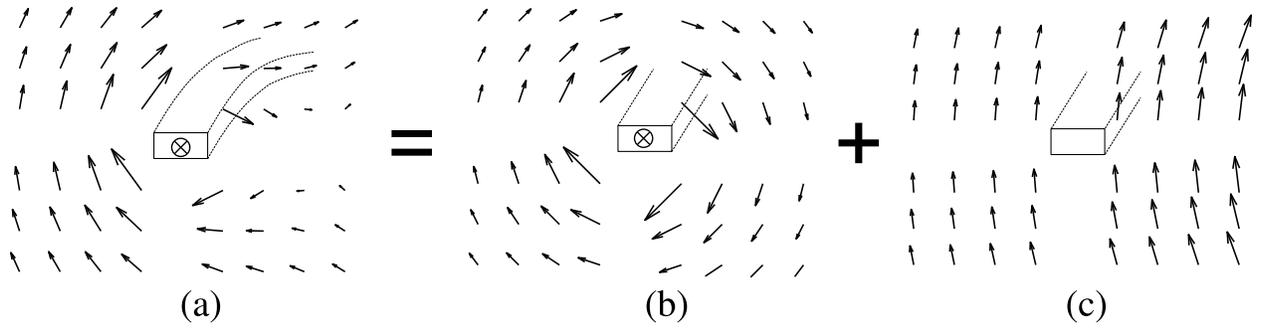


Figure 5.6: Illustration of the decomposition and untangling of skin effect and proximity effect for a 2-D case. (a) is the total magnetic field of a wound conductor; (b) is the purely odd magnetic field of a straight conductor; (c) is the external field due to winding, which can be further approximated as uniform (purely even)

5.3.2 Proximity Effect

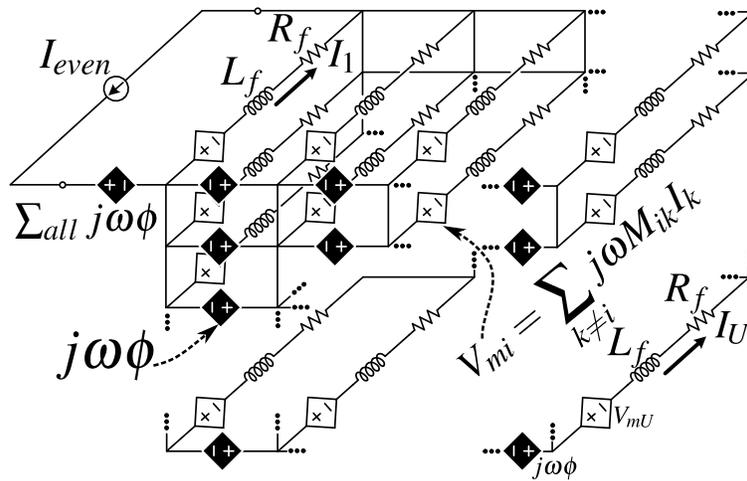


Figure 5.7: Equivalent circuit based on the volume-filament-method of a wound conductor subject to external magnetic field due to winding

Dissipation from the proximity effect (defined in [54, Sec.2-4]) is first studied for on-chip inductors by [60], and later by [61][39][43]. [60, Eqn. (9)] ignores the back EMF generated by the eddy current itself and concludes that the proximity effect loss increases $\propto f^2$. We will show that this is only partly correct. [61, Fig. 6] qualitatively mentions the influence on the magnetic field from the eddy current itself, but does not include it in [61, Eqn. (18,19)]. [39, Eqn. (31,32)] take into account the mutual induction between the impressed current and the eddy current, which is

proven correct by the comparison in [39, Fig. 9]. However, [39, Eqn. (31,32)] imply that the series loss of a conductor will saturate after a corner frequency f_0 defined in [39, Eqn. (30b)]. Ultimately [43] shows the most complete picture by supposing the distribution of the eddy current is analogous to the impressed current, concluding that the series loss of the conductor will never saturate. But vague definitions of M_{prox} and L_{prox} in [43, Eqn. (7,8)] for [43, Fig. 5(b)] exclude the case where the metal width is smaller than the metal thickness. Yet this is often true in modern RFCMOS or RFSOI processes with ultra-thick top metals. We do not question the accuracy and effectiveness of [43]’s model of the proximity effect, but in this paper we take a step further to re-derive everything on a more fundamental basis, along the same lines as the skin effect model; together, they give a satisfactory account of series loss in an inductor.

What is the proximity effect? It is an added source of loss from magnetic coupling when the return path of current is in proximity to the conductor in question. In the case of cylindrical conductors wound around a magnetic core, AC current density is only a function of horizontal position, dictated by the 1-D Helmholtz Equation [62, Eqn. (26,27)]. The general solution of this equation [62, Eqn. (28)] contains an even-mode cosh-term and an odd-mode sinh-term. The even-mode distribution is triggered by the odd-mode tangential magnetic boundary condition, and *vice-versa*.

[63, Eqn. (11)] associates the even-mode solution with skin effect, and the odd-mode solution with proximity effect. As [63, Sec. 3, Eqn. (4)] explains, the magnetic field around an isolated sheet conductor is an odd function; winding the sheet around a magnetic core to form a closed path creates a uniform, thus even, *external* field. This is consistent with [64, Fig. 13.28], where the magnetic field is offset by a constant across each conducting layer in a stack. The qualitative solution in [64, Fig. 13.25 (b)] shows that there is an even current distribution common to each layer, and an odd current distribution that increases along the layers. The consistent quantitative solutions in [62, Eqn. (10), Fig. 8, Fig. 9] lead to an equivalent resistance ratio R_{ac}/R_{dc} consisting of a term (M') independent of the layer index (skin effect) and a term that increases with the index (proximity effect).

In this work, we generalize the 1-D case of a cylindrical conductor around a magnetic core, to

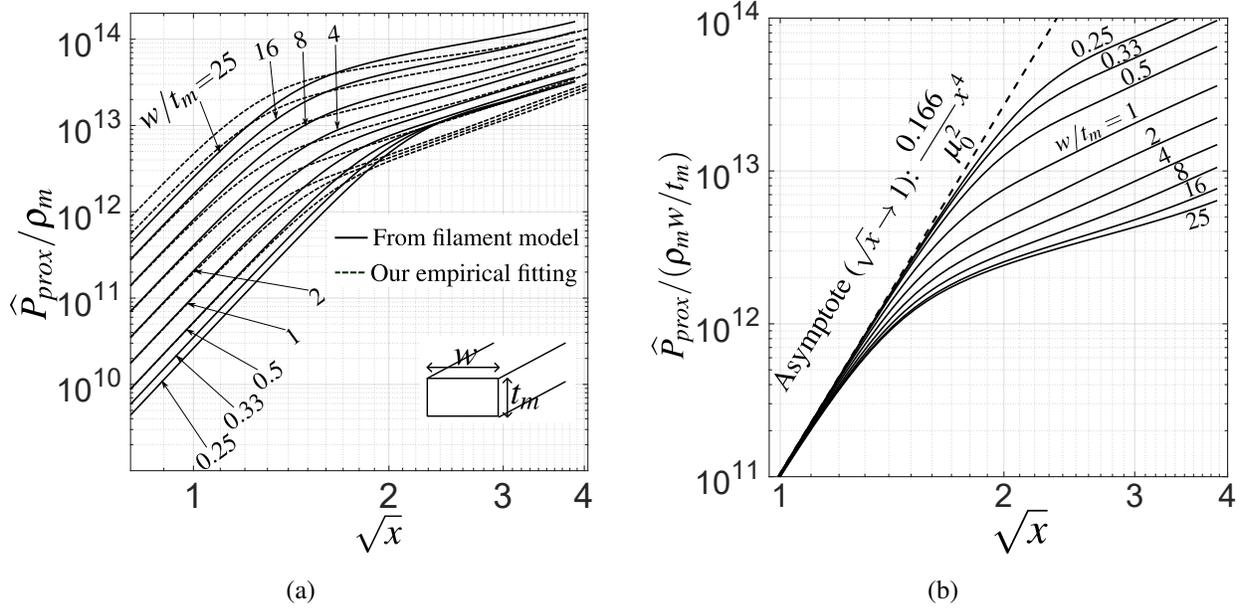


Figure 5.8: (a) Proximity effect normalized curve I compared with our fitted curves; (b) Proximity effect normalized curve II. $\sqrt{x} \triangleq (\sqrt{wt_m}/\delta)^{1/2}$.

the 2-D case for a rectangular conductor in a spiral air-core inductor. This is the same geometry we used in Sec. 5.3.1 to analyze skin effect. The curves in Fig. 5.2 resemble M' in [62, Fig. 9], but with an additional parameter w/t_m . We expect that the proximity effect loss curves for the 2-D case will resemble D' in [62, Fig. 9]. Fig. 5.6(a) shows the magnetic field of a loop conductor calculated using [65, Eqn. (11,12)]. Using the principle in [63, Sec. 3, (4)], we decompose it into a perfectly odd field for skin effect (Fig. 5.6(b)), calculated using Ampere's Law for a straight conductor ($H = I/2\pi r$), and a remaining field in Fig. 5.6(c).

We associate the remaining external field (Fig. 5.6(c)) with proximity effect. Unlike the 1-D cylindrical case, here the external field is not a purely even function, which means that skin effect and proximity effect are still entangled. But by assuming that the external field is uniform, as [61] does, the skin effect (odd-mode field and even-mode impressed current) can be decoupled from the proximity effect (even-mode field and odd-mode eddy current). This leaves us with solving a 2-D Helmholtz Equation for a straight rectangular conductor under uniform external field B_{ext} .

Again, to help us and other circuit engineers to visualize the fields being solved by the Helmholtz Equation we construct the distributed equivalent circuit in Fig. 5.7. New current-controlled voltage

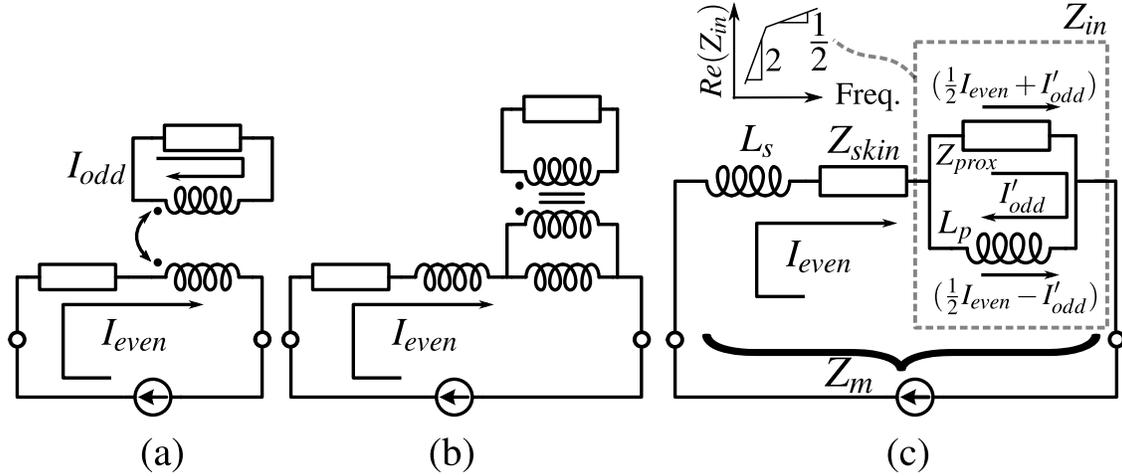


Figure 5.9: (a) Conventional equivalent circuit for skin effect and proximity effect: odd and even components of current defined. (b) Replacing coupled inductors with equivalent circuit involving ideal transformer [66, Ch. VI, Fig. 24c]. (c) Circuit of (b) simplified. This series L - R equivalent circuit for an inductor includes skin and proximity effects. Odd mode current is scaled by turns ratio of ideal transformer.

sources (CCVS's) (in black) are now inserted around each horizontal elementary mesh in Fig. 5.1, and one more CCVS (also in black) is added in series with the driven port to satisfy reciprocity. $j\omega\phi$ is current controlled, because $B_{ext} = \phi / (\text{Mesh area})$ is a function of the impressed current I_{even} and the geometry of the winding. From the source-shifting theorem of circuit theory [67, Sec. 3-3], all black CCVS's may be pushed in series with the sources V_{mi} so that the nodes on the front plane of Fig. 5.7 will merge into one. We assume that winding the conductor produces a uniform external magnetic flux B_{ext} .

By inspection, we see that two orthogonal modes may co-exist in the 2-D circuit in Fig. 5.7. The even-mode currents remain the same as those calculated in Sec. 5.3.1 and sum up to the impressed current I_{even} . The odd-mode currents are found by forcing $\sum I_i = 0$: these currents will circulate in the meshes formed by all the filaments in parallel, excluding the branch with the impressed current I_{even} .

Even and odd modes are, by definition, orthogonal [63, Eqn. (8~12)]. We can develop another sub-circuit which captures the new odd-mode current loss in series with the original equivalent circuit for skin effect to model the complete wound conductor. This sub-circuit for odd-mode proximity effect loss is developed as follows.

The odd-mode power dissipation per unit length normalized to $|B_{ext}|^2$ is:

$$\widehat{P}_{prox} = \sum_{i=1}^U \frac{1}{2} |I_i|^2 \rho_m w t_m \Big|_{|B_{ext}|=1}, \quad (5.3)$$

where ρ_m is the conductor's resistivity. By evaluating (5.3) for different w/t_m 's, we plot in Fig. 5.8(a) $\widehat{P}_{prox}/\rho_m$ as a function of the square root of the normalized variable $x = \sqrt{wt_m}/\delta \propto \sqrt{f}$; this is the same normalized variable used in Fig. 5.2 in connection with the skin effect. To reveal the underlying structure, we normalize \widehat{P}_{prox} in (5.3) over $\rho_m w/t_m$ so that a common asymptote appears at low \sqrt{x} in Fig. 5.8(b). \widehat{P}_{prox} is proportional to f^2 at low frequencies; but to \sqrt{f} at high frequencies. The transition depends on the aspect ratio of the conductor. These results are consistent with the derivations at low frequencies in [61][39] for f^2 -dependency, but they extend the results to higher frequencies as $\propto \sqrt{f}$. The normalized curves in Fig. 5.8(a) also match those in [68, Fig. 7], which were calculated by an FEM simulator. But we have obtained them by analyzing the distributed circuit of Fig. 5.7. Our analysis also matches [43]'s qualitative explanation that the eddy currents become skin-effect-like because as Fig. 5.8 shows, the proximity effect is proportional to \sqrt{f} at high frequencies. From here on, we will use these normalized curves to model proximity effect in inductors.

The equivalent circuit of Fig. 5.9(a) has been used [43] to capture skin effect and proximity effect simultaneously. I_{even} models the even-mode skin effect current and I_{odd} models the odd-mode proximity effect current. By proper choice of the coupled inductors' values and of the transformed impedance, the total loss of the circuit may be made to match the sum of skin and proximity effect losses that we have calculated. But Fig. 5.9(a) can be reduced to Fig. 5.9(c) through the intermediate step in Fig. 5.9(b). Then the left-over series inductance L_s is lossless, so the dissipated power vs. frequency in the shunt branch Z_{in} should match to the curves in Fig. 5.8(a). Here, Z_{prox} comprises a similar network to the ladder circuit (Fig. 5.4) used to model skin effect in Z_{skin} . The circulating odd mode current I'_{odd} is the actual odd-mode current in the volume filaments, scaled by the fictitious ideal transformer in the equivalent circuit (b). At low frequencies when $\omega \ll R/L_p$, $\text{Re}(Z_{in}) = \omega^2 L_p^2/R$: this is a consequence of the definition of impedance³. R is the DC resistance

³We remind readers of frequency dependence that appears in driving point immittance. A network with two acces-

of the ladder (Fig. 5.4) comprising Z_{prox} . If the ladder were replaced by a resistor R , then $\text{Re}(Z_{in})$ would level to R at radian frequencies beyond the corner R/L_p . However, by synthesizing a ladder with DC resistance of R , and a first zero close to the frequency R/L_p , with properly interlaced poles and zeros beyond, $\text{Re}(Z_{in})$ continues to rise $\propto \sqrt{\omega}$ for $\omega > R/L_p$. An inductor operating in the frequency band where $\text{Re}(Z_{in})$ rises gently with frequency will tend to maintain its quality factor better at high frequencies.

To fit the dissipation of Z_{in} to $P_{prox} = \widehat{P}_{prox} \cdot |B_{ext}|^2 \cdot l_{tt}$, where l_{tt} is the total length of the conductor, we follow these steps: At low frequency ($P_{prox} \propto f^2$),

$$P_{prox} = \frac{1}{2} |I_{even}|^2 \text{Re}(Z_{in}) = \frac{1}{2} |I_{even}|^2 \omega^2 \frac{L_p^2}{R} \quad (5.4)$$

$$= |B_{ext}|^2 l_{tt} \omega^2 \times 0.0415 \times \frac{w^3 t_m}{\rho_m} \quad (5.5)$$

The coefficient of 0.0415 is found empirically to give the best fit to the normalized asymptote in Fig. 5.8(b). To improve the fit to the quality factor, as well as to $\text{Im}(Z_{in})$ and the skin effect impedance, we intentionally double the loss at low frequencies, when the proximity effect loss is negligible compared to the loss due to skin effect and R_{dc} . The details are included in Appx. B.

$$\frac{L_p^2}{R} = 2 \times 2 \times 0.0415 \times \left| \frac{B_{ext}}{I_{even}} \right|^2 \frac{w^3 t_m l_{tt}}{\rho_m} \quad (5.6)$$

In a multi-turn inductor, B_{ext}/I_{even} of the i^{th} turn, now called B_i/I_{even} , is calculated using [65, Eqn. (11)]. This is where spacing between conductors, central to the proximity effect, is taken into account. Then,

$$\left| \frac{B_{ext}}{I_{even}} \right|^2 \times l_{tt} = \sum_{i=1}^N \left| \frac{B_i}{I_{even}} \right|^2 \times l_i \quad [\text{H}^2/\text{m}^3] \quad (5.7)$$

where l_i is the length of i^{th} turn.

Fig. 5.11 illustrates the current density as shaped by the proximity effect. By contrast to the distribution of current density from the skin effect (Fig. 5.5), this is an odd function along the conductor width around a center line running through the conductor along its length.

sible terminals consists of R in shunt with L_p , both parameters *independent* of frequency. The driving point impedance is defined at those two terminals. Thus $Z_{in}(j\omega) = \left(\frac{1}{R} + \frac{1}{j\omega L_p} \right)^{-1}$, so $R_{in}(\omega) = \text{Re}(Z_{in}(j\omega)) = \frac{\omega^2 L_p^2 / R}{1 + (\omega L_p / R)^2} \simeq \frac{\omega^2 L_p^2}{R}$ when $\omega < R/L_p$. We see that $R_{in}(\omega)$ *depends* on frequency. But $R_{in}(\omega)$ is a construct, *not* an element in the network.

$$f_0 = F_1 \times \frac{1.227\rho_m}{wt_m\mu_0} \quad F_1 = \begin{cases} 1 & \text{if } w/t_m > 1 \\ (t_m/w)^{\frac{2}{3}} & \text{if } w/t_m \leq 1 \end{cases}$$

$$\frac{R}{L_p} = F_2 \times 2\pi \times \sqrt{\frac{1}{2}} \times 2.51 \times \frac{\rho_m}{wt_m\mu_0} \quad F_2 = \begin{cases} \sqrt{0.72(w/t_m)^{-0.92} + 0.28} & \text{if } w/t_m > 1 \\ (t_m/w)^{\frac{2}{3}} & \text{if } w/t_m \leq 1 \end{cases}$$

$$a = \begin{cases} -0.177(w/t_m)^{0.334} + 1.183 & \text{if } w/t_m > 1 \\ 1 & \text{if } w/t_m \leq 1 \end{cases}$$

Figure 5.10: Expressions for transition frequency R/L , modified ladder's f_0 and modified adjusting factor a (relative to Fig. 5.3) to model proximity effect.

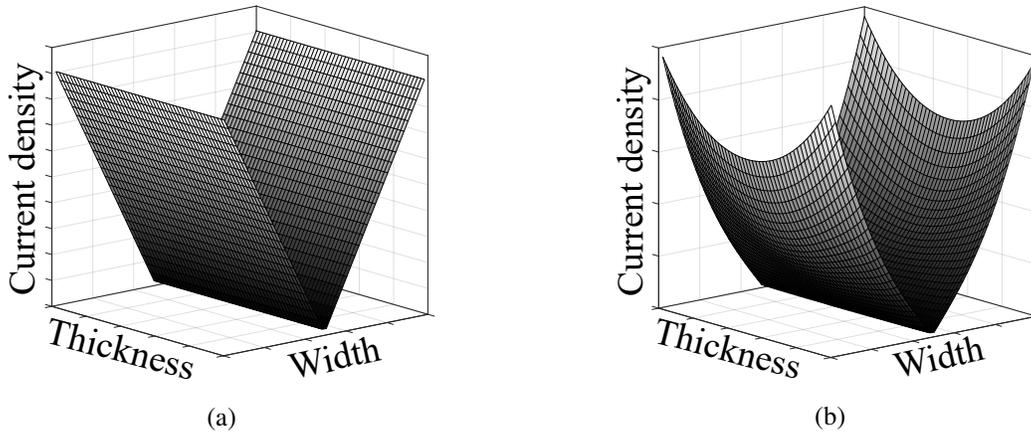


Figure 5.11: Proximity effect eddy current density (magnitude) in square conductor ($w = t_m$) (a) at low frequencies, where eddy current changes linearly from one side to the other; (b) at high frequencies where the eddy current redistributes in a skin-effect-like way, similar to Fig. 5.5 although not the same. Phase of the current density (not shown) has odd symmetry around the center line

5.3.3 Foster and Cauer Network Synthesis

When we model the inductor cases in Sec. 5.8, we do not actually calculate the values of L 's and R 's in Fig. 5.4, because we only need a valid expression of impedance to represent the equivalent circuit. Circuit theory guarantees that there is a one-to-one equivalence between a network of interconnected linear, time-invariant elements and a driving point function at the two terminals of that network, which is a ratio of polynomials in the complex Laplace frequency s subject to certain constraints. SPICE simulating and plotting the impedance of the network in Fig. 5.4 in AC

analysis, and MATLAB plotting the magnitude and phase of the driving point function in (5.2) after substituting $s = j\omega$ produce identical outputs. To clear the doubts, we will now illustrate the synthesis procedure briefly.

We start from the Foster network as shown in Fig. 5.4(b), as it is more straightforward. Using Partial Fraction Expansion, we can rearrange (5.2) in the form of:

$$\frac{1}{Z_{ladder}} = G_0 + \frac{1}{sL_1 + R_1} + \frac{1}{sL_2 + R_2} + \frac{1}{sL_3 + R_3} \quad (5.8)$$

Then, The component values in Fig. 5.4(b) are readily determined, because (5.8) represents the driving-point admittance of a parallel combination of 4 branches, with one branch of a single resistor. Partial Fraction Expansion is a built-in function in MATLAB's symbolic toolbox, and a few lines of scripts can transfer (5.2) to (5.8) accurately.

Next, Cauer form in Fig. 5.4(a) is a bit more involved as it requires a continued fraction form of Z_{ladder} :

$$Z_{ladder} = \frac{1}{G_0 + \frac{1}{sL_1 + \frac{1}{G_1 + \frac{1}{sL_2 + \frac{1}{G_2 + \frac{1}{sL_3 + \frac{1}{G_3}}}}}}} \quad (5.9)$$

This procedure is rather tedious by hand, so we recommend using [69] which transforms the form in (5.8) to the continued fraction form in (5.9).

We give an example here:

Listing 5.1: MATLAB code to generate Foster ladder element values from interlaced poles and zeros

```
syms s; % Declare Laplace
Imped = 3.6; % R_DC
```

```

% zero and pole staggered:
pir = 3.1415926;
wz1= 2*pir*3.693;wp1= 2*pir*7.535;wz2= 2*pir*17.03;
wp2= 2*pir*38.23;wz3= 2*pir*86.11;wp3= 2*pir*194;
% Impedance function:
Imped = Imped * (1+s/wz1) * (1+s/wz2) * (1+s/wz3) / (1+s/wp1) /
        (1+s/wp2) / (1+s/wp3);
% Partial fraction expansion on admittance
% decompose into N D form
[N,D] = numden(children(partfrac(1/Imped,s,'FactorMode','real')
    ));
% Print results
fprintf('Print foster form:\n')
temp1 = sym2poly(D(1) / N(1));
fprintf('L = %.4f nH, R = %.4f Ohm\n', [temp1(1) temp1(2)]);
temp2 = sym2poly(D(2) / N(2));
fprintf('L = %.4f nH, R = %.4f Ohm\n', [temp2(1) temp2(2)]);
temp3 = sym2poly(D(3) / N(3));
fprintf('L = %.4f nH, R = %.4f Ohm\n', [temp3(1) temp3(2)]);
R_first = double(D(4) / N(4));
fprintf('First resistor %.4f Ohm: \n', R_first);

```

This MATLAB code does partial fraction expansion and gives $L_1 = 82.9\text{pH}$, $R_1 = 44.9\Omega$, $L_2 = 257.4\text{pH}$, $R_2 = 6.0\Omega$, $L_3 = 152.9\text{pH}$, $R_3 = 16.4\Omega$, $G_0^{-1} = 37.1\Omega$ in (5.8) for Fig. 5.4(a). Or [69] transforms the Foster ladder to Cauer ladder as $L_1 = 44.5\text{pH}$, $R_1 = 14.5\Omega$, $L_2 = 86.2\text{pH}$, $R_2 = 9.8\Omega$, $L_3 = 345.6\text{pH}$, $R_3 = 12.6\Omega$, $G_0^{-1} = 37.1\Omega$ as (5.9) for Fig. 5.4(b). Fig. 5.4(a) and (b) are completely equivalent from the driving-point's perspective. We can use whichever is more convenient.

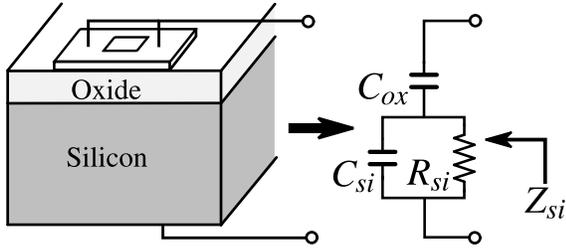


Figure 5.12: Development of Z_{si}

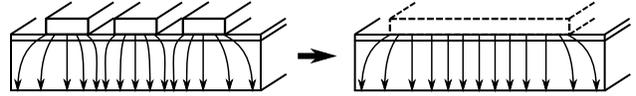


Figure 5.13: Replacing closely-coupled turns with one wide microstrip to approximate the substrate parasitic.

5.4 Substrate Equivalent Circuit and Distribution Factor

The silicon substrate is modeled by calculating, first, the impedance Z_{si} between the inductor and the substrate's backplane (Fig. 5.12). This impedance is best captured by a C - R - C sub-circuit. Then, depending on how the inductor is being driven, different multiples α of Z_{si} are used in the final complete equivalent circuit, in parallel with the series sub-circuit Z_m shown in Fig. 5.9(b).

We calculate Z_{si} with a simplified form of the method in [43]. R_{si} is the spreading resistance in the semiconducting substrate under the inductor's footprint, and $R_{si} \times C_{si}$ corresponds to the dielectric relaxation time constant of the semiconductor material. As shown in Fig. 5.13, the adjacent turns are typically closely coupled, so they can be approximated as a 1-turn circle of wide microstrip line with a width of $(d_{out} - d_{in})/2$ and a length of $\pi(d_{out} + d_{in})/2$, where d_{in} is the diameter of the inductor's inner hallow, and d_{out} is the diameter of the inductor's periphery. The substrate network per unit length is calculated with [70, Eqn. (13~17)], and the total Z_{si} is scaled to the length of the 1-turn microstrip. This simplification holds up well in the case studies of Sec. 5.8. With these assumptions the substrate C - R - C network is synthesized without sacrificing the essence of [43].

When d_{out} of a spiral is much smaller than the substrate thickness, the capacitance is simply given by $C_{si} = 2\epsilon_{si}d_{out}$, half the capacitance of an isolated disk of the same diameter situated in a medium with the permittivity of the silicon substrate [54, Sec. 2-31, Eqn. (127)]. It follows that $R_{si} = \rho_{si}/2d_{out}$.

We use a 2- π network [41, 43] to determine the multiples (α) of Z_{si} in parallel with the series sub-circuit Z_m taken from Fig. 5.9(b). Consider a solid metal bar placed above a ground plane,

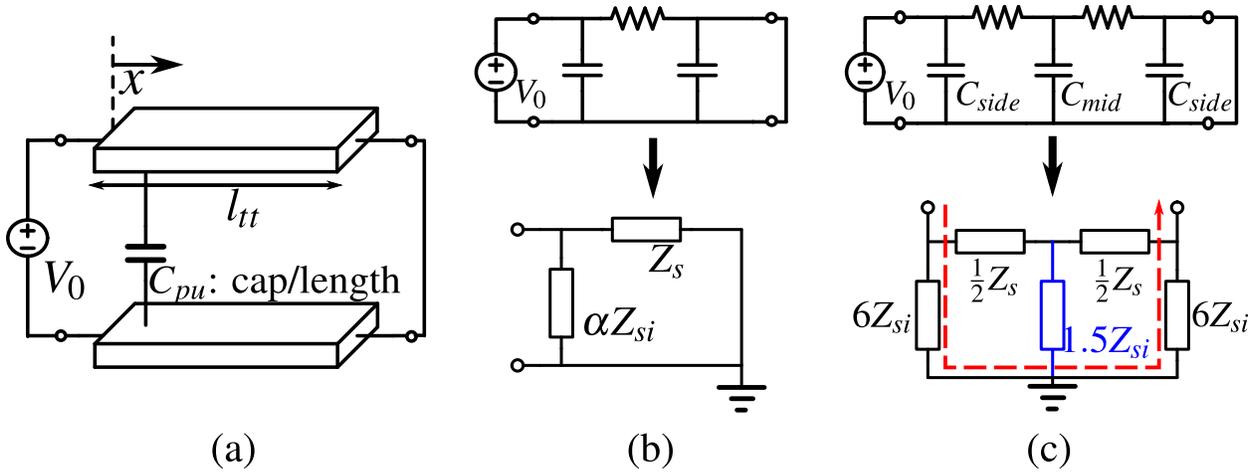


Figure 5.14: (a) Metal bar on a ground plane; (b) 1- π equivalent circuit generalized for inductor driven single-ended. $\alpha = 2$ or 3 depending on the symmetry and the driving point; (c) 2- π equivalent circuit generalized for differentially driven symmetrical inductor. Effective $\alpha = 12$.

with one port driven to V_0 and the other port grounded (Fig. 5.14(a)). C_{pu} is the capacitance to ground per unit length. At frequencies where the time of flight across the bar may be neglected, the voltage distribution will be linear, so the total energy stored is

$$W = \frac{1}{2} \int_0^{l_{tt}} C_{pu} (V_0 \frac{x}{l_{tt}})^2 dx = \frac{1}{2} \cdot \frac{1}{3} C_{pu} \cdot l_{tt} V_0^2 = \frac{1}{2} \cdot [\frac{1}{3} C_{total}] V_0^2 \quad (5.10)$$

where $C_{total} = C_{pu} \cdot l_{tt}$. This means that the effective lumped capacitance appearing at the driving point, across the terminals of the source V_0 , is $\frac{1}{3} C_{total}$. If both ports are driven by in-phase voltage sources, the apparent capacitance rises to C_{total} . A 1- π network (Fig. 5.14(b)) can model only one of the two situations correctly. But, a 2- π network (Fig. 5.14(c)) can model both, provided $C_{side} = \frac{1}{6} C_{total}$ and $C_{mid} = \frac{2}{3} C_{total}$. This is heuristically generalized to partition the substrate by replacing the ideal C_{pu} with a low-loss capacitive C-R-C sub-circuit per unit length. Then, α multiples of the total impedance Z_{si} of the C-R-C block appear in shunt, just as do different fractions of the total capacitor C_{total} (Fig. 5.14(b)). This derivation of distribution factor is close to the fitted distribution factor of a 2-T circuit given in [37].

The following cases are typical:

1. A symmetrical inductor is driven differentially at its two ports, when $\alpha = 12$. (Fig. 5.14(c))
2. A symmetrical inductor is driven single-endedly (seldom in practice; [40, Fig. 4, 5] show a

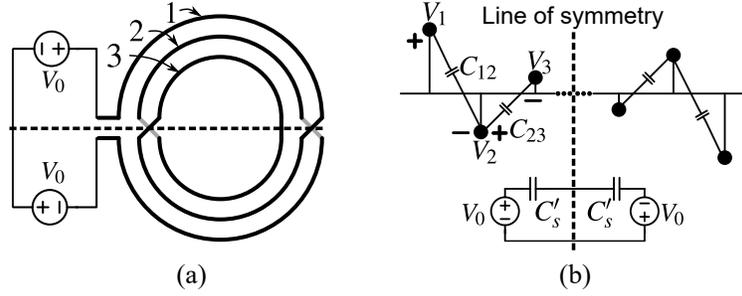


Figure 5.15: (a) Segments of a symmetrical inductor driven differentially; (b) Voltage profile of the inductor in (a) along the dashed line, from [71].

rare experimental example), when $\alpha = 3$. The 2- π network reverts to a 1- π network with a shunt $3Z_{si}$ at the driving point, which means that the capacitance is $\frac{1}{3}C_{total}$ as in (5.10).

3. An unsymmetrical inductor is driven single-endedly from its *outer port* with its inner port grounded, when $\alpha \approx 2$. The E-fields closer to the outer port spread out more in the substrate, whereas the E-fields associated with the inner port are relatively confined.

5.5 Inter-winding Capacitance

Sec. 5.4 shows how to model the vertical electric field from the inductor to the substrate as a lumped capacitance. In this section we turn to modelling the (mostly) horizontal electric field between the metal traces comprising the spiral inductor.

The inter-winding capacitance (C_s) captures the electric energy stored between inductor windings. We calculate C_s using the method in [71]. Assume a linear drop in potential between the turns of the spiral when it is driven by a voltage V_0 . The voltage on each turn in a symmetrical inductor is approximated by its average on that turn, as shown in Fig. 5.15. If the inductor is unsymmetrical, the voltage profile is as shown in [71, Fig. 2]. We use [72, Eqn. (9~18), Table II] to calculate the inter-wire capacitance between each pair of adjacent turns, e.g. C_{12} or C_{23} as shown in Fig. 5.15(b). The electric energy stored between adjacent turns sums up to W_s . Then, we lump all the distributed inter-winding capacitance into two series capacitors $C'_s = 2W_s/V_0^2$ across the input port (Fig. 5.15(b)), which amount to a single lumped capacitor $C_s = 0.5C'_s$.

The electric energy stored in the lossy capacitance of the substrate network (Sec. 5.4) resonates with the inductor’s magnetic energy at frequency (f_{sr}). It affects the fundamental quality factor (Q_{tru}), which we will define in Sec. 5.7. Although inter-winding capacitance C_s is also a key element of the equivalent circuit, we will show in Sec. 5.7.3 that below the inductor’s self-resonance frequency, it does not affect Q_{tru} .

5.6 Substrate Eddy Current

We also add a short study on the substrate eddy current to address the reviewer’s concern. This part is completely new to [47].

Like the proximity effect, the eddy current loss in the substrate is modelled by a transformed series resistance ($R_{s,ed}$) added to the L - R sub-circuit (Fig. 5.9(b)) [73].

To prevent latch-up, older CMOS processes used heavily-doped substrates with resistivity on the order of 0.01 to 0.05 Ω cm. In modern “RF CMOS” processes it is orders of magnitude higher, usually 10 Ω cm a move, as we will now show, mainly to enable realization of on-chip inductors with lower loss. The risk of latch-up is no worse.

[73] gives simulated values of $R_{s,ed}$ for inductors on heavily-doped substrates. [73, Table 1] shows the results of exploring different geometries of a 15 nH inductor comprising a stack of three stacked spirals in series, fabricated in the HP CMOS-14 process with a 0.05 Ω cm substrate which presents an effective series resistance $R_{s,ed}$ of a few ohms at 1.8 GHz. This scales down to 5 m Ω if that inductor is migrated to a modern process with a substrate resistivity of 10 Ω cm. A 10 nH inductor without stacking takes a larger area but according to [73, Fig. 7], it suffers from the largest $R_{s,ed}$; at 5 GHz, the simulated $R_{s,ed} = 15 \Omega$. This will scale down to 70 m Ω on a modern 10 Ω cm substrate. In our case studies, 70 m Ω amounts to roughly 1% of the typical series resistance at 5 GHz arising from skin effect and proximity effect.

Of the various sources of loss, how important then is the eddy current loss? For example, conductor loss at 5 GHz in the very wide traces of the single-turn 0.33 nH inductor described in [47, Sec. 5.3], is captured by a series resistance, $R_s \approx 0.3 \Omega$. A multi-turn 10 nH inductor on a single

layer of metal must occupy a larger area than this single-turn inductor, so it is safe to say that even then the $R_{s,ed}$ will remain negligible compared to R_s .

This trend is supported by [74, Fig. 8] which examines the impact of substrate eddy current by turning this loss component on or off in the simulator (solid versus dotted lines in [74, Fig. 8]). When the substrate's resistivity is larger than $1 \Omega\text{cm}$, the dotted lines and solid lines will converge, indicating that loss from substrate eddy currents is negligible.

We conclude that loss due to substrate eddy currents is not important in modern CMOS processes that use lightly-doped substrates. In extreme inductor geometries such as in [47, Sec. 5.3], this simplification causes some error. Otherwise in all common geometries of spiral inductors, Sec. 5.8 and [47, Ch. 5] show that predictions from our equivalent circuit, which neglects altogether loss from eddy currents in the substrate, are satisfactory compared to measurements.

Dissipation in the substrate is important when it arises from currents at high frequencies, capacitively coupled through the oxide and flowing to the ground plane.

5.7 Definitions of Quality Factor

5.7.1 Physically Correct Equivalent Circuits

A lumped equivalent circuit can capture electric and magnetic fields in a physical structure like a spiral inductor by, in effect, discretizing the fields in space. We assume that everything in the structure is linear. In general, the more the elements in the equivalent circuit connected in the right topology, the closer it approximates the effects of the actual fields as voltages and currents over a given span of frequencies [66, Ch. 1][75, Sec. 2]. At a minimum, it needs one inductor to capture stored magnetic energy, one capacitor to capture stored electric energy, and one resistor to model loss. Now the main point: if the topology of the equivalent circuit is correct and it contains a sufficient number of elements, then over the frequency span of interest these element values are frequency-independent [76, Fig. 2.4]. The reverse is also true: if over a frequency span of interest, an equivalent circuit comprising frequency-independent elements matches, at its terminals, the voltages and currents measured on a physical structure, then with each element is

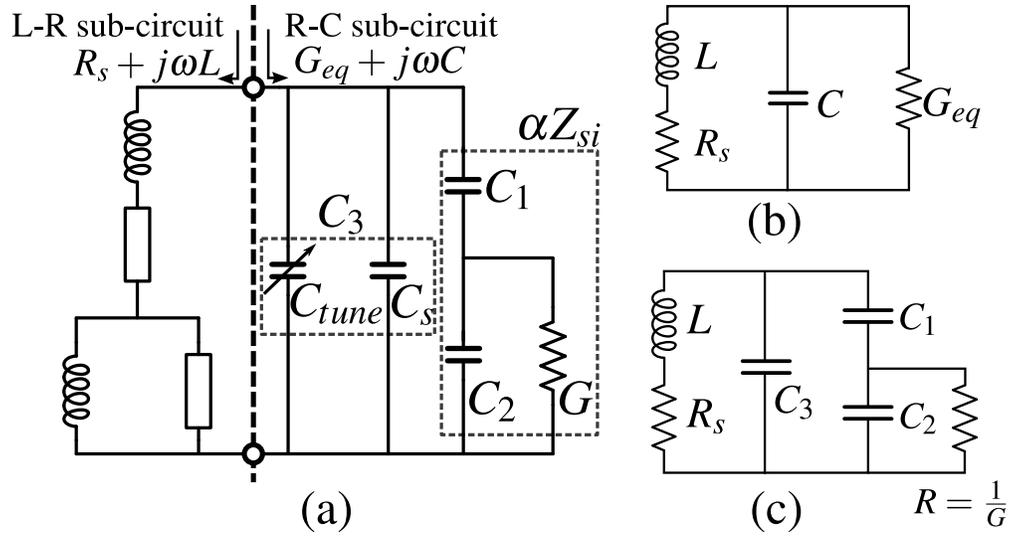


Figure 5.16: (a) Complete equivalent circuit of on-chip resonator. The L - R sub-circuit comes from Fig. 5.9(c). C_3 is the summation of a fictitious external tuning capacitor C_{tune} and C_s as the inductor's inter-winding capacitance. The scaled version of the substrate impedance αZ_{si} , depending on inductor's driving mode; (b) Simplified version of (a) for analysis of Q_{app} (c) Simplified version of (a) for analysis of Q_{tru}

associated physical meaning such as a form of energy storage or dissipation. There may be more than one equivalent circuit that can do this.

In RF circuits, a commonly used objective function to optimize inductor geometry is its *quality factor*, Q , at the operating frequency ω . After studying the literature, we find it necessary to clarify definitions of Q in the context of basic circuit theory with physically-correct equivalent circuits in mind.

The quality factor of an inductor L is originally defined when it is embedded in an LC resonator. Neglecting loss in the capacitor, Q determines the inductor's quality (or its merit, or ability) to produce clear resonance effects [66, Ch. 4, Sec. 28]. This Q also defines the quality factor of the complex conjugate poles of the resonator in the s -plane, which we discuss at greater length below.

5.7.2 Apparent Q

LCR meters and Q meters read out apparent Q , Q_{app} . It is called this because the meter is force-fitting its measurements to an equivalent circuit of two elements in series, an inductor and a resistor [77, Table 1-2]. But a physically-correct equivalent circuit such as Fig. 5.16(a) contains many more elements. The two will give the same driving point impedance $Z(j\omega)$ over frequency when the inductance and resistance in the simple series circuits are allowed to assume frequency dependence. Then, by definition

$$Q_{app}(\omega) \triangleq \frac{\text{Im}Z(j\omega)}{\text{Re}Z(j\omega)} \quad (5.11)$$

Suppose we measure Q_{app} of an inductor whose physically-correct equivalent circuit is Fig. 5.16(b). C lumps all inductor's self-capacitance, without any external C_{tune} . For now we will neglect the skin and proximity effects, although they will come back into play when calculating Q at a given frequency. The two independent reactances will lead to a quadratic expression in its impedance $Z(s)$, which we express in a standard format:

$$Z(s) = \frac{R_s}{1 + R_s G_{eq}} \frac{1 + \frac{s}{\omega_z}}{1 + \frac{s}{\omega_0 Q_0} + \frac{s^2}{\omega_0^2}}, \quad \omega_0^2 \triangleq \frac{1 + R_s G_{eq}}{LC}; \quad Q_0 \triangleq \left(\frac{\omega_0 L}{R_s} \parallel \frac{\omega_0 C}{G_{eq}} \right) = Q_{L0} \parallel Q_{C0} \quad (5.12)$$

Because ω_0 is a constant, so are Q_L and Q_C as defined above. And $Q_L, Q_C \geq Q$. We introduce the normalized frequency

$$\Omega \triangleq \frac{\omega}{\omega_0} \quad (5.13)$$

With straightforward manipulations it follows that

$$Q_{app}(\omega) = \frac{\Omega Q_{L0} (1 - \Omega^2) - \frac{\Omega}{Q_0}}{(1 - \Omega^2) + \Omega^2 \frac{Q_{L0}}{Q_0}} \quad (5.14)$$

$$= \frac{\Omega \left[\left(Q_{L0} - \frac{1}{Q_0} \right) - \Omega^2 Q_{L0} \right]}{1 + \Omega^2 \left(\frac{Q_{L0}}{Q_0} - 1 \right)} \quad (5.15)$$

In a useful inductor, $Q > 3$, so the numerator of this expression changes sign at $\Omega \simeq 1$. This implies that L , associated with the numerator from (5.11), crosses zero at a frequency close to ω_0 and becomes negative [77, Fig. 5-12(b)]: But physical inductance cannot be zero or negative: so this must be an artifact of a non-physical equivalent circuit which, in this case, omits to account

for the capacitance associated with the inductor that is causing an internal self-resonance at ω_0 . As a plot of measured inductance versus frequency in [77, Fig. 5-12(b)] shows, this two-element equivalent circuit is good enough to represent the actual physics of the inductor at frequencies well below self-resonance (labelled “effective range” on the figure), but beyond, the elements lose connection with physical effects: here because the two-element model fails to account for stored electric energy. [78, Sec. 3-7] explains the same thing in somewhat different though equivalent terms.

We now calculate an expression for this effective (frequency) range.

$$Q_{app} \simeq \Omega Q_{L0} \frac{1 - \Omega^2}{1 + \Omega^2 \left(\frac{Q_{L0}}{Q_0} - 1 \right)} \quad (5.16)$$

$$\simeq \Omega Q_{L0} \left[1 - \Omega^2 \frac{Q_{L0}}{Q_0} \right] \simeq \frac{\omega L}{R_s} \text{ for } \Omega < \Omega_{-1} \quad (5.17)$$

We define the normalized range $(0, \Omega_{-1})$ over which we will accept the last approximation in (5.17) to within a 10% discrepancy. This describes well the effective (frequency) range referred to above. Then $\Omega_{-1} = \sqrt{(0.1Q/Q_L)}$. Suppose $Q_C \approx 0.3Q_L$, a representative ratio for certain inductors on RF-CMOS. This means that $\Omega_{-1} \approx 0.2$.

What of practical value can we extract from measurement of Q_{app} over frequency?

1. The slope $dQ_{app}/d\omega$ close to DC gives L/R_{dc} . Since R_{dc} can be found with a DC ohmmeter, this gives one way to measure inductance L . Of course, ωL may also be read off from $\text{Im}Z(j\omega)$, which is measured to find Q_{app} as in (5.11).
2. In the frequency range $0 < \omega < \omega_{-1}$, the measured Q_{app} will be seen to gradually fall away from the extrapolated straight line $(L/R_{dc})\omega$. This is a measure of the growing contributions of loss due to skin effect and proximity effect. These losses would be evident from measurements of $\text{Re}Z(j\omega)$ when finding Q_{app} .
3. Perhaps most usefully as we show below, $Q_{app} = Q_{tru}$ for $\omega < \omega_{-1}$.
4. If the measurement can extend to frequencies high enough to detect where Q_{app} crosses 0, then we have a means to find the pole- Q of the equivalent circuit Fig. 5.16(b). This is because

$$\text{Im}Z(j\omega) = 0 \text{ when } \omega \simeq \omega_0.$$

While Q_{app} of inductors is measured routinely, it tells the circuit designer that the device-under-test behaves like an ideal inductor—with a resistor in series with it—across frequencies up to 20% or so of self-resonance. This is important to know when the expected use is as a standalone inductor, for example in a matching circuit, a bias circuit, or in power electronics.

But in many cases it is to be used to form a resonator. We have seen that the inductor structure is a resonator to begin with, tuned by its internal capacitances to a self-resonance at ω_0 . With external capacitance, it can be tuned to a lower frequency. To characterize it for this use up to its self-resonance requires a re-definition of Q . This must conform more closely to a physically correct equivalent circuit, and leads us to Q_{tru} .

5.7.3 True Q

Suppose we want to tune an oscillator with the inductor under test to some frequency ω up to ω_0 . To focus on the inductor's own limitations, we will assume that a lossless capacitor may be attached in parallel with the inductor—or, as its dual, in series with it—to satisfy the oscillation criterion, that is $\text{Im}Z(j\omega) = 0$ where Z is the driving point impedance now including the tuning capacitor.

Inserting a lossless C_{tune} across the terminals of an inductor whose equivalent circuit is Fig. 5.16(a) will not change the circuit shape, but merely raise the value of C_3 ; all other elements are unchanged.

With the oscillator tuned to a desired frequency ω , the next matter of concern in an RF use is that oscillator's phase noise. For this we turn to Leeson's classic expression for phase noise [79], which involves Q of the resonator. But what Q ? In arriving at the expression for phase noise, Leeson uses the resonant circuit's rate of change of impedance phase with frequency around resonance. Let us now derive this characteristic for the physically correct equivalent circuit of Fig. 5.16(c).

The circuit consists of a loop of three capacitors, which means two are independent. With the inductor, they will define three pole frequencies for the circuit. Two of the poles will form a

complex conjugate pair in the s -plane, so the third must be real. The driving point impedance is

$$Z(s) = \frac{(R_s + sL)(G + s(C_1 + C_2))}{D(s)} \quad (5.18)$$

$$\begin{aligned} \text{where } D(s) &= G + s[(C_1 + C_2) + R_s G(C_1 + C_3)] \\ &+ s^2 [LG(C_1 + C_3) + R_s(C_1 C_2 + C_2 C_3 + C_3 C_1)] \\ &+ s^3 [L(C_1 C_2 + C_2 C_3 + C_3 C_1)] \end{aligned} \quad (5.19)$$

It will be easier to interpret this expression after factoring its denominator. Assuming that⁴ $R_s G \ll (C_1 + C_2) \div (C_1 + C_3)$, we postulate the factors

$$D(s) \simeq (G + s(C_1 + C_2)) \left(1 + \frac{s}{\omega_0 Q_0} + \frac{s^2}{\omega_0^2} \right) \quad (5.20)$$

where the coefficients ω_0 , Q are to be determined. A pole-zero cancellation in $Z(s)$ is now evident, which leads to a welcome simplification. By equating coefficients of s^3 and s^2 , we get

$$\omega_0^2 = \frac{1}{LC}, \text{ where } C \triangleq (C_1 \parallel C_2) + C_3 \quad (5.21)$$

$$\frac{1}{Q_0} = \frac{G_{eq}}{\omega_0 C} + \frac{R_s}{\omega_0 L} = \frac{1}{Q_{C0}} + \frac{1}{Q_{L0}}, \text{ where } G_{eq} \triangleq G \left(\frac{C_1}{C_1 + C_2} \right)^2 \quad (5.22)$$

$$\text{Alternatively, } \frac{1}{Q_0} = G_{eq} \omega_0 L + \frac{R_s}{\omega_0 L} \quad (5.23)$$

Q_{L0} and Q_{C0} are, to repeat, independent of frequency.

ω_0 is the radial frequency, or undamped natural frequency; Q_0 is their quality factor. In case $C_{tune} = 0$ in Fig. 5.16(a), ω_0 is the self-resonance radial frequency. The approximate equivalent circuit Fig. 5.16(b) may be synthesized from (5.18) and (5.20).

This 2nd-order resonator can tune an oscillator to any frequency $\omega_{osc} \leq \omega_0$. Oscillation requires $\text{Im}Z(j\omega_{osc}) = 0$, which means that after a few steps of algebra,

$$\omega_{osc} = \omega_0 \sqrt{1 - \frac{1}{Q_L Q_0}} \simeq \omega_0 \quad (5.24)$$

⁴This is not always true across the entire range of frequencies to self-resonance, but its results are easy to interpret. We will give a more accurate expression later.

since $Q_{L0} > Q_0 > 3$. This lets us find, from (5.18) and (5.20) but with steps not shown, the sensitivity of phase to frequency at ω_{osc} :

$$\frac{d}{d\omega} \angle Z(j\omega) \simeq -\frac{2Q_0}{\omega_{osc}} \simeq -\frac{2Q_0}{\omega_0} \quad (5.25)$$

But this is exactly the same expression that Leeson employs to derive his expression for phase noise [79]. So we conclude that it is the pole Q in a physically correct equivalent circuit of the inductor that Leeson uses. It stands to reason, then, that $Q_{tru}(\omega_{osc})$ of the inductor should be defined by that pole Q when the inductor is tuned to resonance at ω_{osc} . This definition differs from what is termed “true Q” in [78, Sec. 3-7], which does not consider the contribution of finite Q_C . [80] has explored this definition of quality factor, referring to (5.25) as phase stability. But the equivalent circuit employed does not yield to simple analysis, and the numerical solutions do not bring into clear focus the key physical processes at work. This is what we develop next.

Simple scaling relationships apply when an external capacitor tunes the inductor to the normalized frequency $\Omega \leq 1$. This must mean that after the addition of the external tuning capacitor, C changes into C/Ω^2 whereas G_{eq} is unchanged. Then from (5.22),

$$\frac{1}{Q_{tru}}(\Omega) = \frac{G_{eq}}{\Omega\omega_0\frac{C}{\Omega^2}} + \frac{R_s}{\Omega\omega_0 L} = \frac{\Omega}{Q_{C0}} + \frac{1}{\Omega Q_{L0}} = \frac{1}{Q_C} + \frac{1}{Q_L}. \quad (5.26)$$

Given that $0 < \Omega < 1$, we observe that:

1. If $Q_{C0} < Q_{L0}$, then $Q_{tru} = Q_C \parallel Q_L$ will go through a maximum at the normalized frequency $\Omega = \sqrt{Q_{C0}/Q_{L0}}$. Beyond this frequency the lower Q_C causes Q_{tru} to drop.
2. For $\Omega < \sqrt{Q_{C0}/Q_{L0}}$, $Q_{tru} \simeq \Omega Q_L = Q_{app}(\Omega\omega_0)$. Since it is often the case that $\omega_{-1} < \Omega\omega_0$, measurements of Q_{app} also give Q_{tru} over the “effective range” across which the measured Q_{app} is meaningful.

We also have derived a more accurate expression for Q_{tru} , which takes a different path to an approximate pole-zero cancellation which does not suffer the loss of accuracy that may arise as described in footnote 4. Although it is applied to the same physically-correct equivalent circuit, this simplification leads to a G_{eq} which is frequency-dependent.

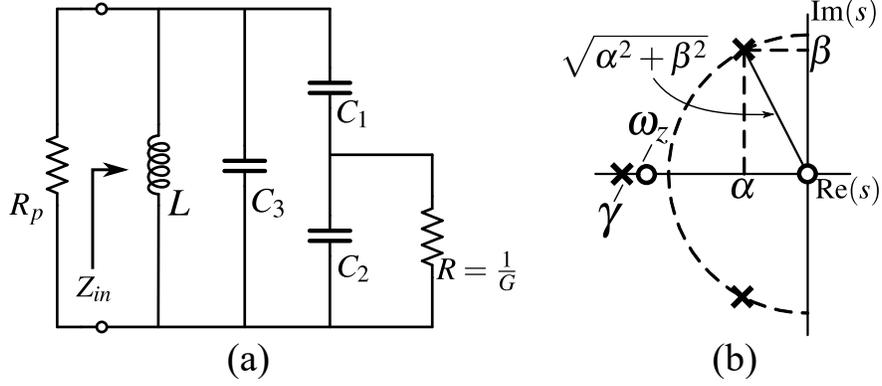


Figure 5.17: (a) Transformed equivalent circuit of on-chip LC resonator from Fig. 5.16(c); (b) Poles and zeros of Z_{in} in (a).

First we transform the L - R sub-circuit in Fig. 5.16(a) to a parallel combination of L and R_p as shown in Fig. 5.17(a), which is only valid when $(\omega L)/R_s > 3$. At the end of the following derivation, we will justify this series-to-parallel transformation. The impedance Z_{in} to the right of R_p in is

$$Z_{in}(s) = \frac{N(s)}{D(s)} = \frac{sL[1 + sR(C_1 + C_2)]}{1 + sR(C_1 + C_2) + s^2L(C_1 + C_3) + s^3LR\sum C_i C_j} \quad (5.27)$$

$$\text{where } \sum C_i C_j \triangleq C_1 C_2 + C_2 C_3 + C_1 C_3 \quad (5.28)$$

By evaluating the coefficients of $D(s)$ in (5.27) in the form of $(s + \gamma) \cdot (s + \alpha - j\beta) \cdot (s + \alpha + j\beta)$, we can get the following equations:

$$\gamma + 2\alpha = \frac{G(C_1 + C_3)}{\sum C_i C_j} \quad (5.29)$$

$$\alpha^2 + \beta^2 + 2\alpha\gamma = \frac{C_1 + C_3}{L\sum C_i C_j} = \frac{1}{LC} \triangleq \omega^2, \quad C \triangleq C_3 + C_1 \parallel C_2 \quad (5.30)$$

$$\gamma(\alpha^2 + \beta^2) = \frac{1}{LR\sum C_i C_j} \quad (5.31)$$

$\omega \leq \omega_0$ (5.21), as $C_{tune} \geq 0$ in Fig. 5.17(a). In addition, we define:

$$\varepsilon = \frac{\omega^2}{2\alpha\gamma}, \quad (5.32)$$

such that from (5.30), we can get:

$$\alpha^2 + \beta^2 = \omega^2 \left(1 - \frac{1}{\varepsilon}\right) \quad (5.33)$$

Next, from (5.31) and (5.30), we can get:

$$\gamma = \frac{GC}{\sum C_i C_j \times \frac{1}{1-\frac{1}{\Omega}}} \quad (5.34)$$

Then, substituting (5.34) to (5.29), we can derive:

$$2\alpha = \frac{n^2 G}{C} \times \frac{1}{1-\frac{1}{\varepsilon}} \left[\frac{\frac{C}{n^2}(C_1+C_3)}{\sum C_i C_j} - \frac{\frac{C^2}{n^2}}{\sum C_i C_j} - \frac{\frac{C}{n^2}(C_1+C_3)\frac{1}{\varepsilon}}{\sum C_i C_j} \right], \text{ where } n \triangleq \frac{C_1}{C_1+C_2} \quad (5.35)$$

With some math, (5.35) can be eventually reduced to:

$$2\alpha = \frac{n^2 G}{C} \frac{1-\frac{\kappa}{n\varepsilon}}{1-\frac{1}{\varepsilon}} \text{ where } \kappa \triangleq 1 + \frac{C_3}{C_1} \quad (5.36)$$

Meanwhile, following (5.34) and $N(s)$ in (5.27), we can find that there is a zero on the negative real axis of the s-plane at:

$$\omega_z = \gamma \times \left(1 - \frac{1}{\varepsilon} \right) \quad (5.37)$$

We now have the following important observations that if $\varepsilon \gg 1$:

1. One real pole and one real zero in (5.27) will approximately cancel each other as shown mathematically in (5.37) and visually in Fig. 5.17(b);
2. The 3rd-order circuit will be reduced to a standard 2nd-order resonator with the resonance frequency is approximately equal to the radial frequency (5.33) of the complex poles.

From (5.34) and (5.36), we can express Ω as:

$$\varepsilon = \underbrace{\frac{\omega RC}{n^2}}_{\lambda_p} \times \underbrace{\omega R(C_1+C_2)}_{\lambda_s} \times \frac{(1-\frac{1}{\varepsilon})^2}{1-\frac{\kappa}{n\varepsilon}} \quad (5.38)$$

which seems rather complicated to be solved in full. However, if we assume $\varepsilon \gg 1$ and neglect the term $(1-1/\varepsilon)^2$, ε can be solved easily as:

$$\varepsilon \approx \lambda_p \times \lambda_s + \frac{\kappa}{n} \quad (5.39)$$

Even if in the worst case when ε is not much greater than 1 ($\varepsilon \sim 5$ for instance), we can successively approximate ε by first evaluating (5.39) and recursively (5.38), until ε converges to a steady final

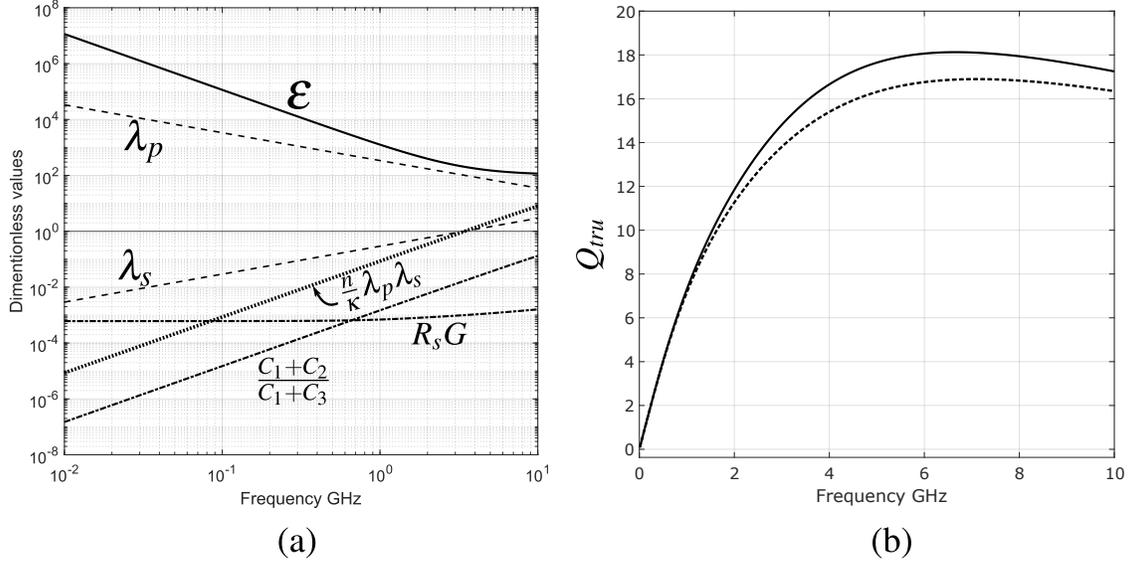


Figure 5.18: (a) Comparison among various dimensionless quantities involved in the derivation of Q_{tru} ; (b) Comparison between the more accurate Q_{tru} (solid line) and the less accurate Q_{tru} (dashed line)

value. But of course, the two aforementioned observations for $\epsilon \gg 1$ will not be rigorous anymore, and the circuit can only be *crudely* reduced to a 2nd-order resonator with large error.

By considering two extremes, we can verify that ϵ is almost always large for parallel resonators formed by on-chip inductors. If a large C_3 (by adding more C_{tune} in parallel with C_s) pulls down the resonance frequency, then in (5.39) $\kappa = (1 + C_3/C_1) \gg 1$; or if ω is large, then $\lambda_p \lambda_s \gg 1$. In either case $\epsilon \gg 1$. [81, Sec. 2.4] has shown that this happens frequently in resonant networks, and that by cancelling the pole and nearby zero, the order of the circuit is reduced with little sacrifice in accuracy.

After pole-zero cancellation, if we fit the remaining terms in $D(s)$ of (5.27) to the standard form

$$1 + \frac{s}{\omega Q_C} + \frac{s^2}{\omega^2} \quad (5.40)$$

we can get:

$$\frac{1}{Q_C} = \frac{G_{eq}}{\omega C} \quad (5.41)$$

for Fig. 5.16(b). According to (5.33) and (5.36), G_{eq} depends on κ , λ_p , and λ_s , and therefore on

frequency:

$$G_{eq} = n^2 G \times \left(1 - \frac{1}{1 + \frac{n}{\kappa} \lambda_p \lambda_s}\right). \quad (5.42)$$

G_{eq} appears in parallel with L and a capacitance $C = C_3 + C_1 \parallel C_2 \approx 1/(\omega^2 L)$ that tunes the inductor to any frequency ω below its self-resonant frequency. (5.42) indicates that when ω is tuned to a lower frequency $G_{eq} \rightarrow 0$ as κ becomes large and $\lambda_p \lambda_s$ becomes small. This means that at lower frequencies where series-to-parallel transformation of the L-R sub-circuit does not hold. The R-C sub-circuit is almost like an ideal capacitor. Then Q_{tru} of this circuit is well-known, as the circuit has become a series resonator.

Hence, we can confidently include the loss in the L-R sub-circuit as a “parallel” loading effect on Q_C in (5.41), we have the complete design-oriented expression for Q_{tru}

$$Q_{tru}(\omega) = \frac{\omega L}{R_s} \parallel \frac{1}{\omega L G_{eq}} = Q_L \parallel Q_C \quad (5.43)$$

When being used in practice (Sec. 5.8), R_s and ωL in (5.43) are changed to real part and imaginary part of the L-R sub-circuit shown in Fig. 5.16(a) to include skin effect and proximity effect.

Fig. 5.18(b) compares Q_{tru} of the reference inductor in Sec. 5.8.1, calculated from the more accurate (5.43) and (5.42) with that from the less accurate (5.26). We can see from Fig. 5.18(a) that beyond 1 GHz, the limitation described in 4 shows up, leading to the discrepancy in Q_{tru} in Fig. 5.18(b). Also, ϵ is always $\gg 1$ as expected.

5.7.4 Measuring Q_{tru} of On-Chip Inductor

Whereas Q_{app} is defined with measurement of impedance in mind, Q_{tru} is defined by poles of a resonator. When a physically correct equivalent circuit with all element values is known, we have shown how to calculate Q_{tru} . But when an equivalent circuit is *not* known, how can Q_{tru} be measured? [82] derives a procedure from the energy definition of Q , which as far as we know is not employed by widely-used lab instruments. In the case studies to follow, we will use it to determine Q_{tru} from the s -parameters simulated at the terminals of the inductor in question.

A low-loss variable capacitor is connected across the inductor. The parallel combination is stimulated with a sinusoidal source at frequency ω_{osc} rad/s. The capacitance is tuned until the

sinusoidal voltage across the resonator's driving point terminals, including the capacitance, is in phase with the current into the terminals. This means that if the driving point admittance $Y(j\omega)$ of a resonant network comprising some connection of inductors and capacitors is $G_{dp} + jB_{dp}$, then $B_{dp}(\omega_{osc}) = 0$. G_{dp} and B_{dp} are effective values, not circuit elements: they will depend on frequency. Following a series of steps given in [82],

$$Q_{tru}(\omega_{osc}) = \frac{\omega_{osc}}{2G_{dp}} \left. \frac{dB_{dp}}{d\omega} \right|_{\omega_{osc}} \quad (5.44)$$

$dB_{dp}/d\omega$ can be measured by a small change in the frequency around ω_{osc} . We can also derive (5.44) heuristically from (5.18) and (5.20) by equating the derivative of phase of the $Y(j\omega)$ at the tuning frequency ω_{osc} with $2Q_{tru}/\omega_{osc}$. When this fundamental definition of Q_{tru} is applied to case study inductors in Sec. 5.8 by using their port characteristics as simulated by Momentum across frequency, it matches very well the equivalent circuit-based definition developed in (5.22) and refined in (5.43). This reassures us that we have defined Q_{tru} in Sec. 5.7.3 with sound reasoning.

At low frequencies when capacitive currents are negligible, $Q_{tru} \approx Q_{app}$.

A network of inductors and capacitors may have multiple resonances (many ω_{osc} 's). For example, a piezoelectric crystal has two resonances; a pair of coupled resonators has three [83]. There is a Q_{tru} associated with each resonance.

The literature does not agree on a single name. In [36], the apparent quality factor is also called inductor's quality factor; [80] calls it the conventional quality factor. The true quality factor is referred to in [36] as the tank's quality factor, and in [84] as the fundamental quality factor.

5.8 Extra Case studies

In the complete research project, we used the *complete equivalent circuit* of Fig. 5.16(a) to study five sets of inductors taken from different sources across many years. Every one of them was fabricated, and then characterized experimentally. Using the methods developed in this paper, we are able to decouple the various sources of loss and thereby understand why a certain spiral geometry is optimal.

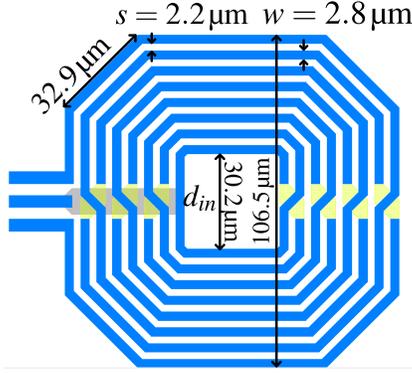


Figure 5.19: Geometry of fabricated reference inductor.

Table 5.1: Parameters for reference 8-turn inductor, and [86]’s inductor.

ρ_m	17 n Ω m
t_m	3.3 μ m
ϵ_{ox}	3.9
t_{ox}	5 μ m
ϵ_{Si}	11.9
t_{Si}	200 μ m
ρ_{Si}	10 Ω cm

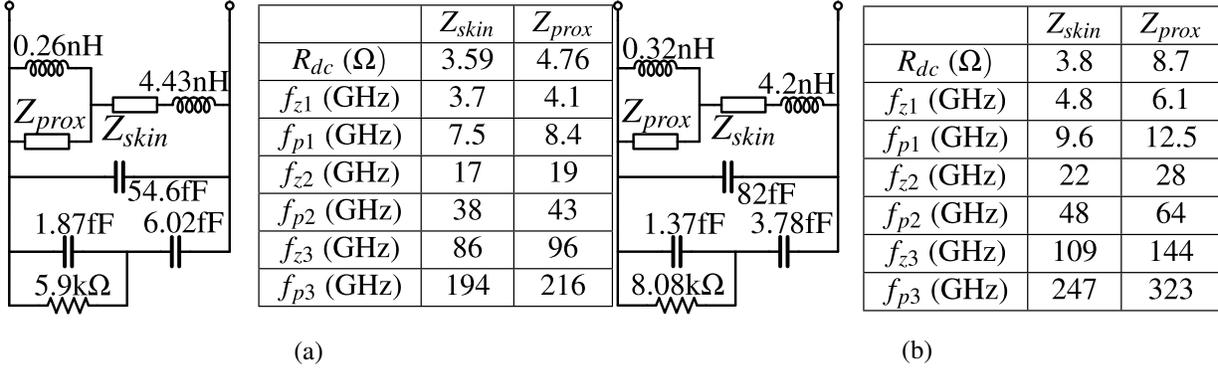


Figure 5.20: (a) Equivalent circuit of the reference 8-turn inductor ($w = 2.8 \mu\text{m}$ and $s = 2.3 \mu\text{m}$); (b) Equivalent circuit of the comparison inductor ($w = 2.2 \mu\text{m}$ and $s = 1 \mu\text{m}$.)

In this thesis, we only include the two extra cases carried out during the author’s Ph.D. study. L_{dc} for the cases below are calculated from [85, Eqn. (2)] with coefficients chosen from [85, Table II] according to the spiral inductor’s shape. For convenience and consistency with Sec. 5.7, we define the self-resonance frequency $f_{sr} \triangleq \omega_0/(2\pi)$.

5.8.1 Inductor Developed at a Semiconductor Company

An 8-turn 4.48 nH inductor for use at 5.5 GHz (Fig. 5.19) was designed at a semiconductor company, and fabricated in the TSMC 65nm process with the process parameters listed in Table 5.1.

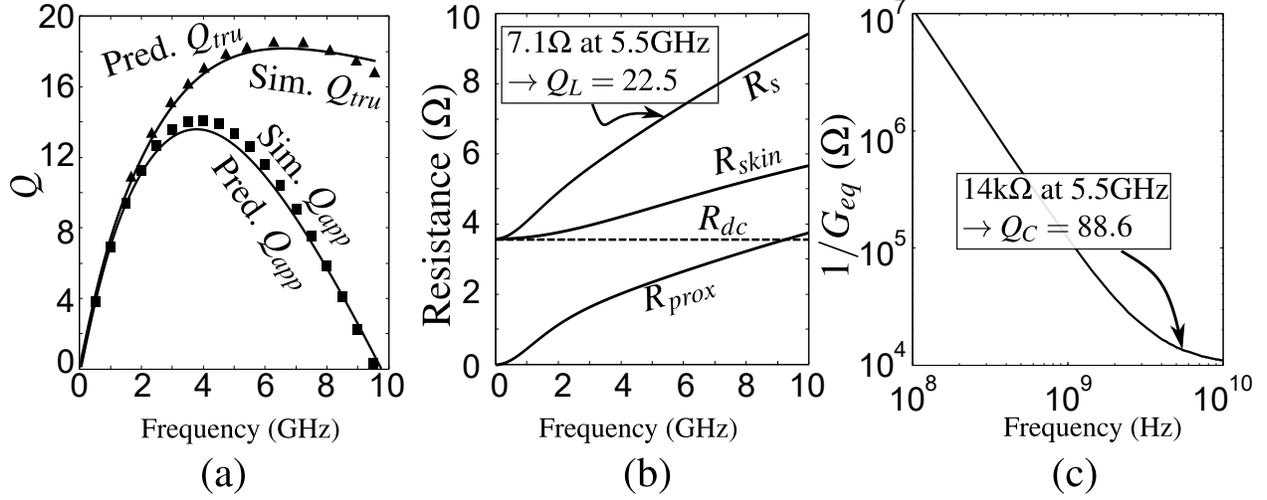


Figure 5.21: Analysis vs. simulation: (a) Comparing Q_{app} (5.11) and Q_{tru} (5.43)(5.44); (b) Contributions to the series loss resistance R_s ; (c) Parallel resistance $1/G_{eq}$ (5.42) (Fig. 5.16(a))

With these data alone, we can construct the equivalent circuit shown in Fig. 5.20(a). Appx. C gives the detailed procedure to calculate parameters. Fig. 5.21(a) compares Q_{app} and Q_{tru} simulated with Momentum and predictions with the equivalent circuit Fig. 5.16(a); the match is remarkably good. To illustrate the insights that the equivalent circuit brings, Fig. 5.21(b) reveals that in this geometry the proximity effect accounts for most of the rise at 5.5 GHz of R_s over R_{dc} . This limits Q_L in (5.43). As to substrate loss, Fig. 5.21(c) shows that $1/G_{eq}$ depends on frequency as discussed at the end of Sec. 5.7.3. Further, Q_{tru} falls at frequencies beyond 8 GHz, as Q_C in (5.43) approaches Q_L . This illustrates the growing prominence of dissipation in the substrate at high frequencies. Q_{app} reaches a null at $f_{sr} = 10$ GHz.

We can go further and use this equivalent circuit to explore the design space of inductor dimensions, by sweeping w and s while holding $N = 8$ for a constant $L_{dc} = 4.48$ nH, and plotting the contours of Q_{app} and Q_{tru} at 5.5 GHz (Fig. 5.22). MATLAB is used to calculate all expressions. The fabricated inductor is labelled “reference”. To keep things simple, the region of $w < s$ is declared unfeasible because it violates the single microstrip approximation of Sec. 5.4. This exploration shows that the geometry of the reference inductor is close to the optimum.

We have designed a “comparison” inductor speculatively (not fabricated) with the same shape and same L_{dc} but with a different metal pitch and space of $w = 2.2$ μm and $s = 1$ μm . This compacts

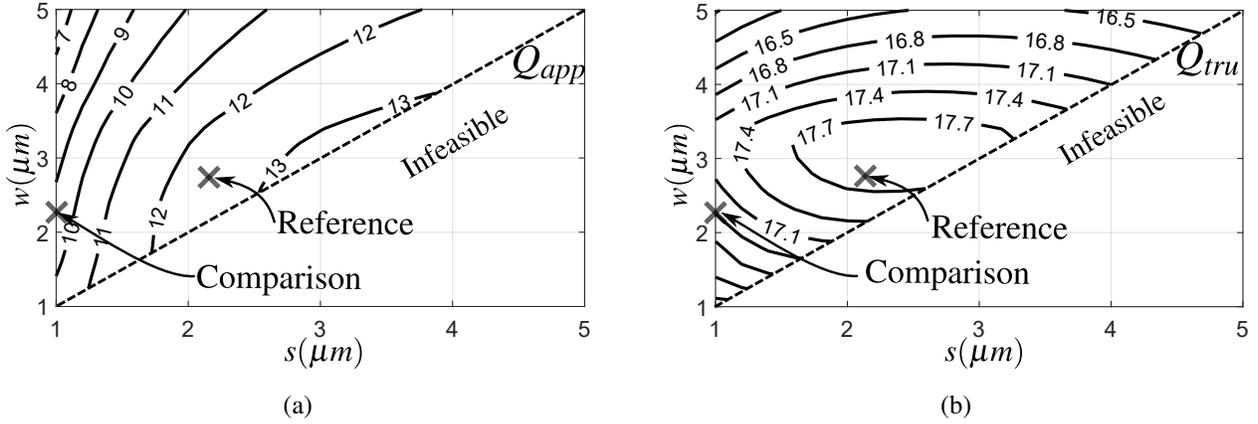


Figure 5.22: (a) Contours of Q_{app} at 5.5 GHz; (b) Contours of Q_{tru} at 5.5 GHz.

its area footprint to 55% of the reference inductor. Calculations following Sec. 5.5 show that the inter-winding capacitance C_s rises 50% due to the closer spacing, from 54.6 fF to 82 fF, thereby lowering f_{sr} . Fig. 5.20(b) gives the equivalent circuit for this comparison inductor. To verify accuracy of the equivalent circuit, the s -parameters of this geometry were obtained from simulations on Momentum. These show that Q_{tru} is about 17 for both inductors. The contours of Q_{tru} calculated from the equivalent circuit are plotted in Fig. 5.22(b). They predict values of 16.8 and 17.7, respectively, for the comparison and reference inductors, while from Fig. 5.22(a) Q_{app} is, respectively 9.5 and 12.5. This is very close to both Q 's derived from simulations on Momentum. The Q_{app} of the comparison inductor is about 20% lower than the reference.

Our equivalent circuit has enabled a rapid exploration of the design space of inductor geometry. By observing how the losses are balanced, we can understand why the reference geometry leads to optimum Q . If Q_{app} is the objective, then the redesigned inductor with nearly half the area will do just as well as the reference geometry.

5.8.2 Square Inductor and Design Space Exploration

Following a similar procedure, we explore various geometries for an 8 nH inductor designed to operate at 2.5 GHz. The inductor originally reported in [40] serves as the reference; Fig. 5.23 shows its geometry. Table 5.2 gives parameters of the BiCMOS process. Fig. 5.24(a) plots the reference

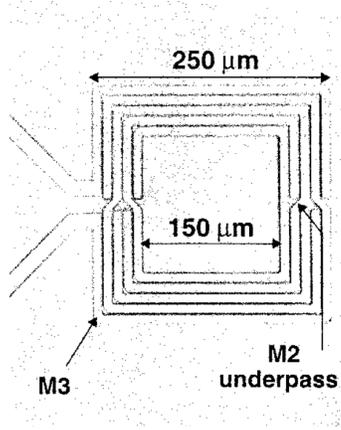


Figure 5.23: Geometry of the reference inductor
A in [40].

Table 5.2: Parameters used in analysis
of [40]’s inductor.

ρ_m	31 nΩm
t_m	2.1 μm
ϵ_{ox}	3.9
t_{ox}	5.6 μm
ϵ_{si}	11.7
t_{si}	200 μm
ρ_{si}	15 Ωcm

Table 5.3: Comparison of important parameters at 2.5 GHz over design space.

	N	w (μm)	s (μm)	L (nH)	R_{skin} (Ω)	R_{prox} (Ω)	R_s (Ω)	R_{DC} (Ω)	R_{eq} (kΩ)	Q_L	Q_C	Q_{tru}	Sim. Q_{tru}	d_{out}^2 (μm ²)	Sim. f_{sr} (GHz)
A	5	8	2.8	8	7.6	1.8	9.4	7.2	8.9	13.4	71.3	11.2	12.2	58597	6.1
B	5	10	3.3	7.9	6.7	1.9	8.6	6.1	7	14.6	56.2	11.6	13.3	71327	5.7
C	5	16	2.8	7.7	5.1	2.3	7.4	4.3	4.5	16.4	37.2	11.4	12.8	103123	4.6
D	4	11.5	3.3	7.9	6.2	1.5	7.7	5.6	5.9	16.2	47.4	12.1	13.8	104904	4.6
E	6	8.3	2.6	7.9	7.4	2.4	9.8	6.9	9	12.7	72.1	10.8	12.3	49292	4.9

Table 5.4: Equivalent circuits’ parameters for 5 inductors as of Fig. 5.16(a)

	Z_{skin}								Z_{prox}											
	L_s	R_{DC}	f_{z1}	f_{p1}	f_{z2}	f_{p2}	f_{z3}	f_{p3}	L_p	R_{DC}	f_{z1}	f_{p1}	f_{z2}	f_{p2}	f_{z3}	f_{p3}	C_1	C_2	1/G	C_s
	nH	Ω	GHz	GHz	GHz	GHz	GHz	GHz	nH	Ω	GHz	GHz	GHz	GHz	GHz	GHz	fF	fF	kΩ	fF
A	7.7	7.1	4.1	7.7	19.1	38.9	87.6	197.3	0.3	3.7	4.1	7.7	19.1	38.9	87.6	197.3	19.0	5.8	2.8	64.4
B	7.6	6.1	3.4	6.1	15.6	31.1	70.1	157.9	0.4	3.2	3.4	6.1	15.6	31.1	70.1	157.9	24.0	6.7	2.5	64.0
C	7.4	4.3	2.2	3.8	10.4	19.4	43.8	98.7	0.6	3.0	2.2	3.8	10.4	19.4	43.8	98.7	37.3	8.7	1.9	88.9
D	7.7	5.6	3.0	5.3	13.8	27.1	60.9	137.3	0.3	2.3	3.0	5.3	13.8	27.1	60.9	137.3	28.6	8.4	2.0	80.7
E	7.6	6.9	4.0	7.4	18.4	37.5	84.4	190.2	0.4	4.8	4.0	7.4	18.4	37.5	84.4	190.2	18.7	5.2	3.2	76.2

inductor’s Q versus frequency, comparing simulation with prediction from our equivalent circuit. For simplicity we have modelled the inductor as circular when actually it is square: this leads to a larger area and a lower substrate resistance, and also to the discrepancy at high frequencies between the predicted and the simulated Q_{tru} . According to Sec. 5.4, the symmetrical reference inductor’s substrate equivalent impedance (αZ_{si} in Fig. 5.16(a)) needs to be scaled down by 4 times from $12Z_{si}$ to $3Z_{si}$ if driven single-endedly. This translates to 4 times C_1 , C_2 , and G in row A of Table 5.4. Then, the predicted single-ended Q_{app} with $3Z_{si}$ matches the published single-ended

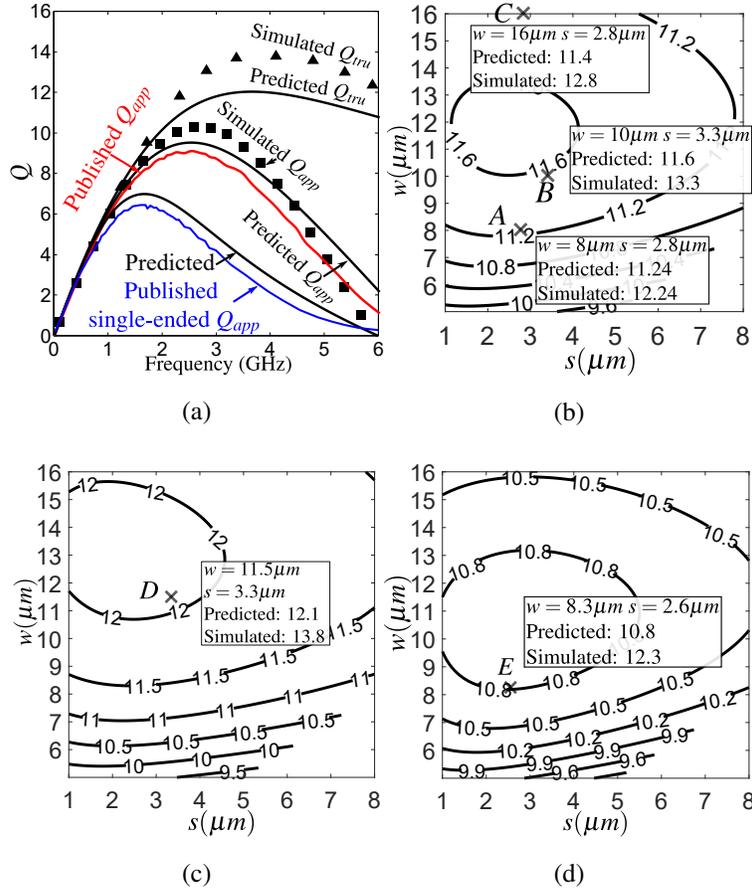


Figure 5.24: (a) Q_{app} (differential and single-ended) and Q_{tru} (differential) of inductor A in [40]: EM simulation vs. analysis. Contours of Q_{tru} for 8 nH inductor with (b) $N = 5$; (c) $N = 4$; (d) $N = 6$.

Q_{app} in [40, Fig. 9]. Our model uses an unambiguous uniform Z_{si} scaling factor of $4\times$, rather than the non-uniform scaling factors in [40, Fig. 10] from parameter fitting. Fig. 5.24(b)-(d) show the contours of Q_{tru} at 2.5 GHz, as w and s are swept for three different spirals with 5, 4 and 6 turns. In Fig. 5.24(b) we see that design A, the reference inductor [40], is close in Q_{tru} to design B which is nearer the optimum; both employ the same number of turns, $N = 5$. We also explore designs C,D,E. EM simulations of these inductor designs with parameters specified in the boxed legends show that D (4 turns) gives the highest Q_{tru} of all—whereas among 5-turn spirals B reaches the highest Q_{tru} . The contours in Fig. 5.24(b-d) are found from our equivalent circuit. They show that all geometries lead to similar quality factors, so the optimum is very broad. When this is so, the design with the most compact footprint will usually be preferred for the valuable benefit of lower

chip area.

With the aid of the equivalent circuits in Table 5.4, we are able to explain in Table 5.3 the trade-offs among the five geometries. The columns labelled R_{skin} through Q_C show the breakdown of losses at 2.5 GHz as predicted by the five equivalent circuits.

- Doubling trace width w from 8 μm in inductor A to 16 μm in inductor C raises Q_L by 22% yet halves Q_C because of the higher substrate loss. The two effects counteract, so no appreciable benefit accrues in Q_{tru} from widening the traces.
- Widening w from 8 μm in A to 10 μm in B lowers the skin effect loss. Spacing traces apart (s) from 2.8 to 3.3 μm maintains proximity effect loss in the now wider traces. Thus inductor B cuts the series loss (R_s) by 9%, which dominates slightly increased the parallel loss. A balance among all losses makes this geometry B a 5-turn optimum for, say, an oscillator operating at 2.5 GHz.
- Inductor D improves upon B by 4% in Q_{tru} , but its much larger area due to less turns lowers its f_{sr} .
- Compared to inductor A, the shorter stretch of metal in inductor E from the extra turn does not lower R_{skin} enough to compensate for the rise in R_{prox} : ultimately Q_L drops by 5% in the equivalent circuit, but Q_{tru} remains the same from Momentum simulation, and almost the same (11.2 vs. 10.8) from the equivalent circuit.

5.8.3 Shielded Inductor

Our equivalent circuit can explain when and how a patterned ground shield (PGS) improves the inductor. We start with an analysis of the shielded inductor reported in [87].

If we increase the pitch of an inductor's traces (larger w and s) while adjusting d_{in} to maintain a constant L_{dc} , Q_L in (5.43) will rise because of lowered R_s . When the silicon substrate is thicker than the inductor diameter, its substrate resistance $1/G_{eq}$ is smaller inversely with its diameter, or, equivalently, with the square root of its area. According to (5.42), (5.43) implies a lower Q_C . This

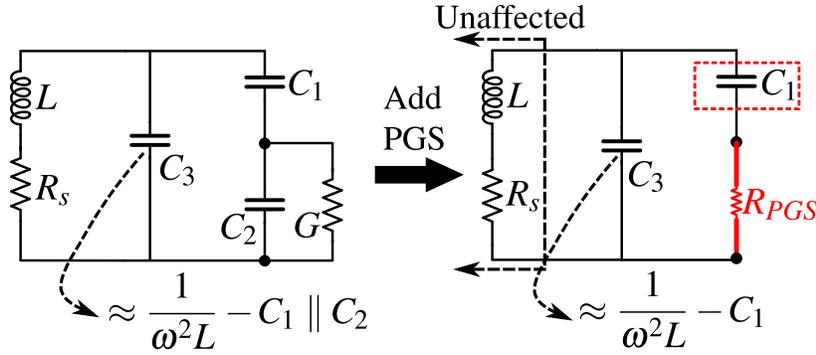


Figure 5.25: PGS principle: for poly PGS, C_1 remains; for metal PGS, C_1 may double. R_{PGS} is usually small and can be neglected for first-cut design.

Table 5.5: Parameters for [87]’s inductor.

ρ_m	28 n Ω m
t_m	2 μ m
ϵ_{ox}	3.9
t_{ox}	5.6 μ m
ϵ_{si}	11.7
t_{si}	200 μ m
ρ_{si}	11 Ω cm

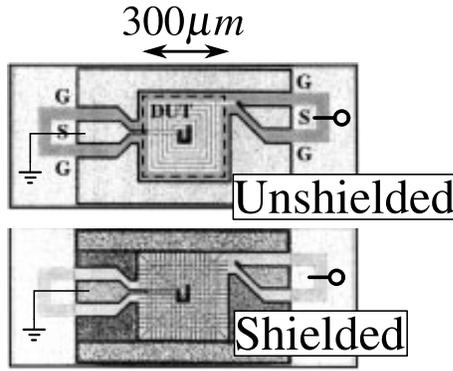
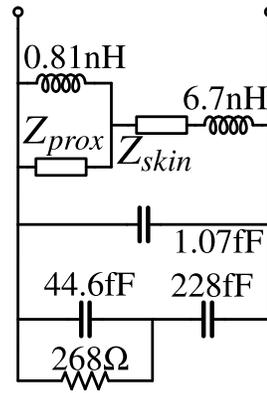


Figure 5.26: Geometry of the shielded and unshielded inductors in [87].



	Z_{skin}	Z_{prox}
R_{dc} (Ω)	3.46	3.3
f_{z1} (GHz)	2.2	2.2
f_{p1} (GHz)	3.8	3.8
f_{z2} (GHz)	10	10
f_{p2} (GHz)	19	19
f_{z3} (GHz)	44	44
f_{p3} (GHz)	98	98

Figure 5.27: Equivalent circuit for [87]’s unshielded inductor.

is more severe for a large unsymmetrical inductor driven from a single end. If $L \approx L_{dc}$ is also large, Q_C may drop to the point that it approaches Q_L , thereby defeating the improvement in Q_L .

Referring to the equivalent circuit in Fig. 5.25, insertion of a patterned ground shield will have three effects. Referring to Fig. 5.25, they are:

1. The vertical electric field lines will terminate on the PGS, which implies that the R - C sub-circuit modeling the silicon substrate is shorted out by a very high conductance G . This requires caution, because if the PGS is on Metal 1, C_1 may rise by as much as $2\times$; whereas if it is on the lowest poly layer, C_1 should remain unchanged.

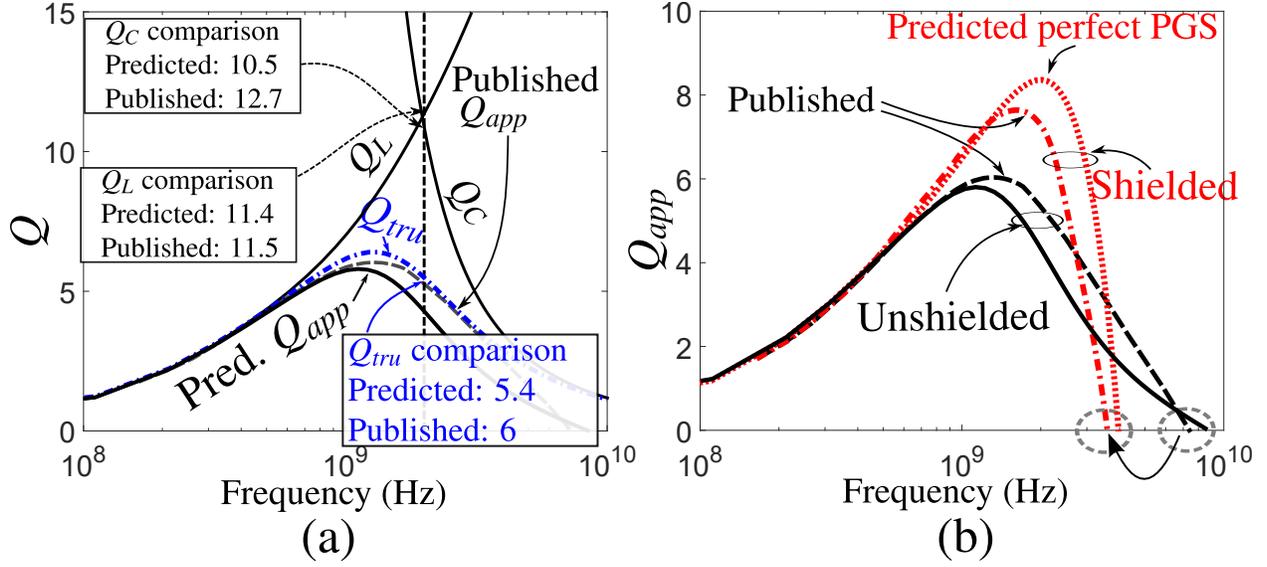


Figure 5.28: (a) Breakdown and comparison of Q_{app} , Q_L , Q_C and Q_{tru} of [87]'s unshielded inductor; (b) comparison between Q_{app} of the shielded and unshielded inductors in [87] and our predictions.

2. Radial slots will block the eddy currents in the shield and maintain the magnetic property of the inductor, which means that the series L - R sub-circuit is unaffected.
3. If $R_{PGS} \rightarrow 0$, then from (5.43) $Q_{tru} \rightarrow Q_L$.

A PGS on metal removes substrate loss at the cost of lower f_{sr} : from (5.21)

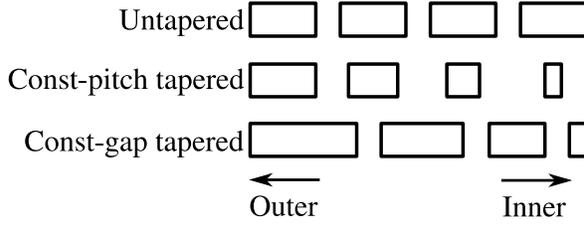
$$\omega_0 = 1/\sqrt{L(C_s + C_1 || C_2)} \approx 1/\sqrt{L(C_s + C_2)} \quad (5.45)$$

without PGS since usually $C_1 \gg C_2$, and it is $\approx 1/\sqrt{L(C_s + C_1)}$ with PGS, where C_s is the interwinding capacitance.

We model the unshielded inductor of [87] with the published process parameters in Table 5.5. By shorting the R - C network in the equivalent circuit of Fig. 5.16(c) with $G \rightarrow \infty$, we predict the improvement brought about by adding the PGS. Fig. 5.26 shows [87]'s inductors used in our comparison; Fig. 5.27 is the equivalent circuit of the unshielded inductor. Since the spiral is unsymmetrical between its terminals and since it is driven from the outer port, we model the substrate as $2Z_{si}$ in the equivalent circuit.

Fig. 5.28(a) shows the predicted quality factors of the unshielded inductor. At 2 GHz, $Q_C \approx Q_L$ in this large inductor with wide traces. The breakout in Fig. 5.28(a) (and common sense) tells us

Table 5.6: Parameters used in analysis of [88]’s inductor.



ρ_m	28 n Ω m
t_m	4 μ m
ϵ_{ox}	3.9
t_{ox}	12 μ m
ϵ_{si}	11.7
t_{si}	200 μ m
ρ_{si}	∞

Figure 5.29: Illustration of two types of tapering.

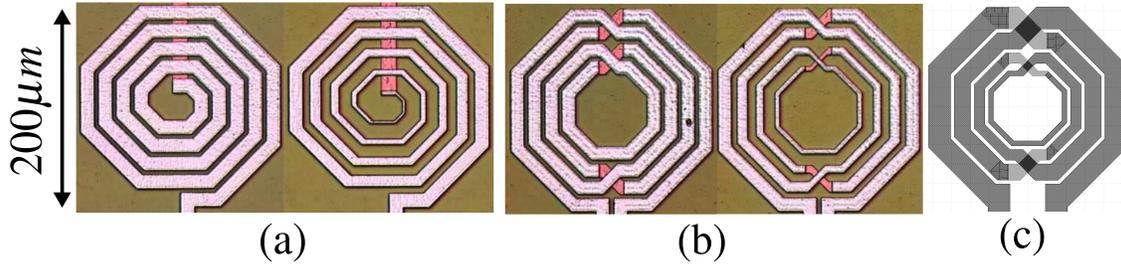


Figure 5.30: Geometry of the untapered and C-P tapered inductors in [88]: (a) 4.5-turn design; (b) 4-turn symmetrical design. (c) Fictitious C-G tapered 4-turn case in Momentum

that ideally a perfect PGS can raise Q_{tru} at 2 GHz from 5.4 to 11.4. The reported Q_{tru} with PGS is 10. It is slightly lower than predicted, we believe, because of the non-zero resistance of the PGS itself (R_{PGS} in Fig. 5.25). [87] gives the measurements of Q_{app} across frequency, so in Fig. 5.28(b) we compare these data with our calculations. At 2 GHz we calculate a Q_{app} with PGS that is slightly higher than the published value but close enough for our purposes. Fig. 5.28(b) also shows that the shield lowers the frequency of self-resonance f_{sr} from 6.8 to 3.6 GHz: this is roughly what we expect from the equivalent circuit, which predicts that it should drop from 8 to 4 GHz.

5.8.4 Tapered Inductor

Whereas a PGS boosts Q_C to raise $Q_{tru}=Q_L||Q_C$, it does not improve Q_L . Inductor tapering is a way to raise Q_L by lowering the spiral’s series loss.

Equation (5.5) tells us that at low frequency the dissipation from the proximity effect rises

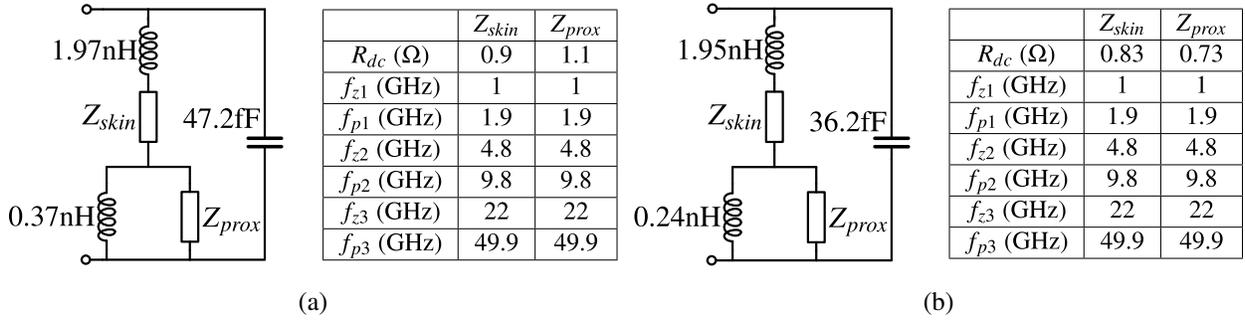


Figure 5.31: Equivalent circuit for (a) 5-turn untapered inductor; (b) 4-turn untapered inductor in [88].

with w^3 . This counteracts the lowered DC resistance by widening traces. What benefits remain will settle earlier in frequency into the regime of gentler rate of rise $\propto \sqrt{f}$. In a multi-turn dense inductor, the loss in the long outer turns arises mainly from skin effect; whereas in the inner turns the proximity effect dominates because of the strong cumulative normal magnetic field. The tapered inductor offers a compromise, with wide outer turns to lower skin effect, and narrow inner turns that lower proximity effect [60]. In this section, we will use the equivalent sub-circuits developed in Sec. 5.3.1 and Sec. 5.3.2 to model these effects, with a separate sub-circuit for each turn. With this we can quantify the benefits that may be achieved.

As shown in Fig. 5.29, tapering can be at constant-pitch (C-P) or with constant-gap (C-G). C-P tapering keeps the centers of the traces at the same position as an untapered reference. By contrast, C-G tapering maintains d_{in} , d_{out} and s but increases turn width outwards from the innermost width w_{in} . C-P tapering maintains L_{dc} , but C-G tapering causes a small drop in L_{dc} as the average diameter of the turns drops, as illustrated by simulation results in Fig. 5.35. Both methods of tapering will lower loss due to proximity effect, but due to smaller w and w/t_m (Fig. 5.2), C-P tapering causes the total skin effect loss to rise in the inner turns. C-G tapering, on the other hand, reapportions the skin effect loss among turns. In most cases, C-G tapering wins over C-P tapering.

Since d_{in} and d_{out} remain almost unchanged before and after tapering, the same C-R-C sub-circuit can model the substrate for both types of tapering, and the reference untapered inductor. But the inter-winding capacitance C_s must be recalculated because the voltage drop on each turn changes.

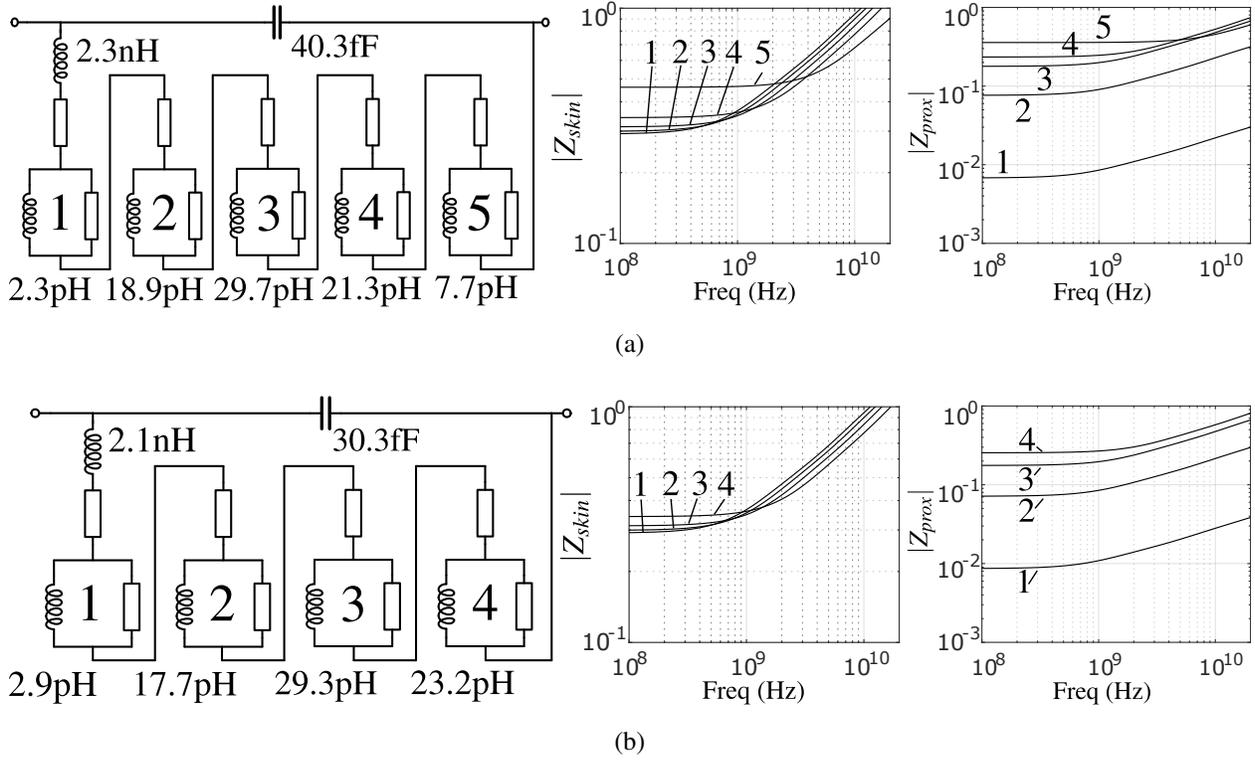


Figure 5.32: Equivalent circuit for (a) 5-turn C-P tapered inductor; (b) 4-turn C-P tapered inductor in [88]. Turn **1** is the outermost.

To demonstrate these tradeoffs, we use our equivalent circuit to analyze the two reference inductors presented in [88] as well as their C-P tapered versions. Fig. 5.30(a,b) shows the layout of the four inductors. For simplicity we round up the 4.5 turns of one spiral to 5. The inductors lie on a 180 nm CMOS HR (high resistivity) substrate whose resistivity is not given, but is usually on the order of 1 k Ω cm; therefore the substrate loss is negligible, that is, $Q_C \gg 1$. Table 5.6 lists process parameters needed for analysis.

Fig. 5.31 shows equivalent circuits of the untapered reference inductors, where the HR substrate behaves as an open circuit and so becomes, in effect, a capacitor in parallel with C_s . Fig. 5.32 shows the equivalent circuits of the C-P tapered inductors: a separate L - R sub-circuit for each turn models losses due to the skin and proximity effect. [88] reports measured Q_{app} of the inductors, which we use to calibrate these equivalent circuits. This enables us to predict Q_{app} of both the untapered and C-P tapered inductors accurately (Fig. 5.34(a)).

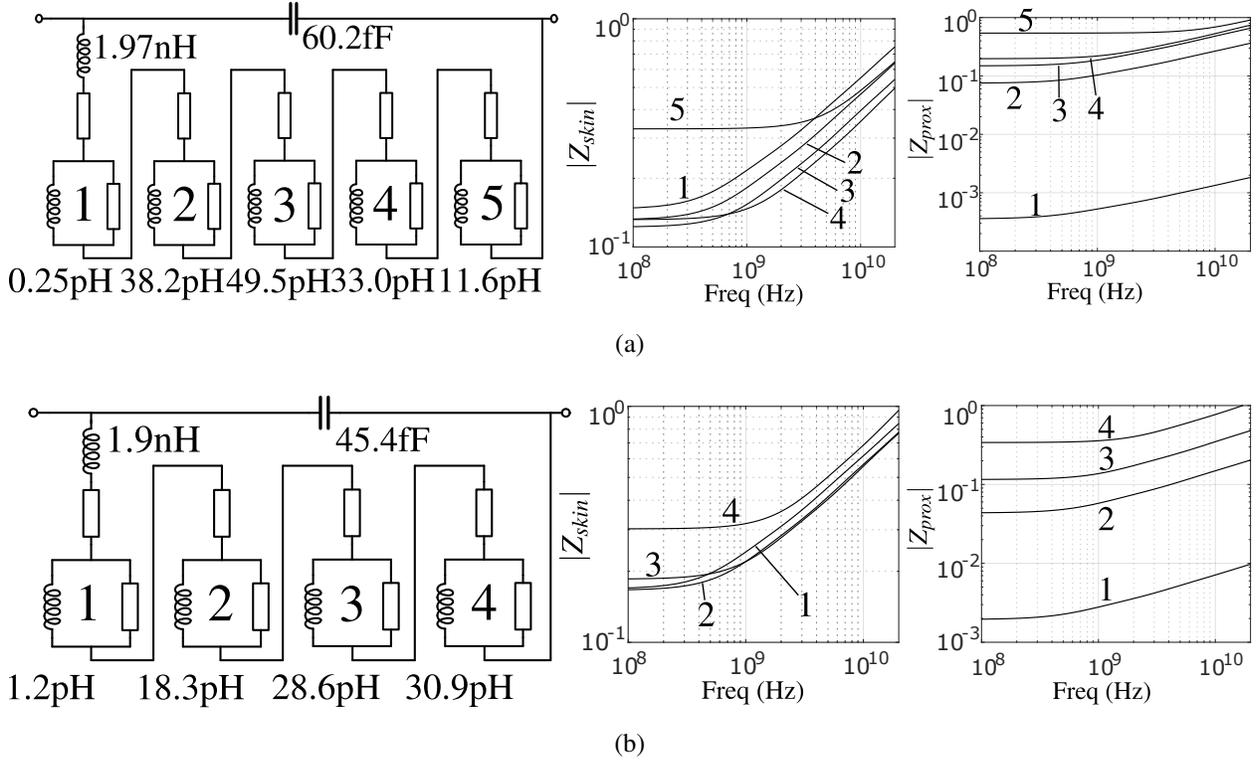


Figure 5.33: Equivalent circuit to analyze C-G tapering on (a) 5-turn inductor; (b) 4-turn inductor, assuming L_{dc} stays at 90% after tapering. Turn 1 is the outermost.

Now we can go one step further and use the equivalent circuits to calculate possible benefits of C-G tapering. Assuming $w_{in}=2\ \mu\text{m}$ for the 5-turn case, $w_{in}=5\ \mu\text{m}$ for the 4-turn case and $d_{out}=200\ \mu\text{m}$ for both, we define the spiral geometry and then derive the equivalent circuits shown in Fig. 5.33. As discussed, a slight drop in L_{dc} limits the expected improvement of Q_{app} . This is hard to predict without using EM simulation (Fig. 5.35). In Fig. 5.34(a), after lowering L_{dc} by 10% as indicated by simulation of the 4-turn case, we foresee significant gains after C-G tapering. The 4-turn case is verified by simulation in Momentum.

Our equivalent circuit captures the changes in series resistances R_{skin} and R_{prox} . Fig. 5.34(c)(d) plot R_{skin} and R_{prox} for each turn. Thus: 1. Since expressions in [65] imply that in a dense multi-turn winding, B_{ext} on each turn rises quasi-linearly from the outermost turn to the innermost turn, we expect from (5.5) that the proximity loss should rise roughly by a quasi-square-law with turn position. This is borne out by the R_{prox} curves of the untapered inductors in Fig. 5.34(d). 2. Fig. 5.34(d) shows that tapering is effective in lowering proximity loss in inner turns. 3. Fig. 5.34(c) shows that

C-P tapering increases R_{skin} of the inner turns, while C-G tapering re-balances R_{skin} of each turn to remain close to the average of R_{skin} across turns of the untapered case.

Fig. 5.34(b) compares contributions from skin and proximity effect on the total resistance of the spirals at 8 GHz. We see that C-P tapering lowers R_{prox} at a price of higher R_{skin} , thus maintaining a near constant loss. Whereas C-G tapering keeps the total R_{skin} constant but lowers the net R_{prox} . We conclude that C-P tapering may improve Q_{app} by some limited amount only at high frequencies; whereas C-G tapering improves Q_{app} at all frequencies.

Tapering can increase Q_L significantly only for multi-turn dense inductors operating at high frequencies (possibly close to f_{sr}) where the proximity effect loss has become equal to or more than the skin effect loss in inner turns. But if that multi-turn dense inductor is built on a $10\ \Omega\text{cm}$ substrate without a PGS, then Q_{tru} is already limited by a low Q_C from substrate loss, so tapering will show little benefit.

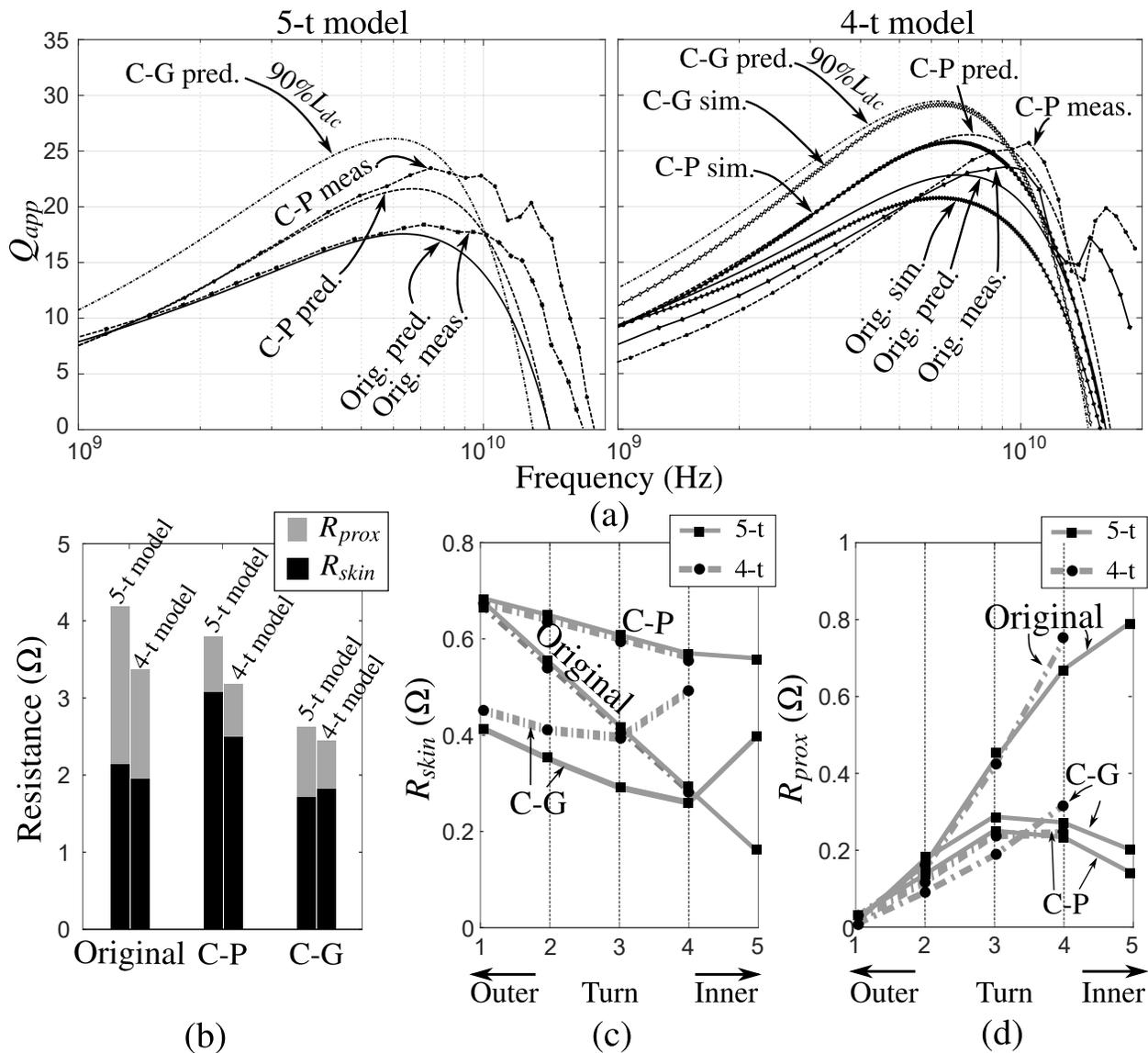


Figure 5.34: Analytical results of [88]'s inductors and our predictions: (a) Comparison of Q_{app} ; (b) Relative contributions to resistance from skin effect and proximity effect at 8 GHz; (c) Resistance breakdown; (d) Proximity effect resistance at 8 GHz. Pred., meas. and sim. stand for prediction, measurement, and simulation. Orig. stands for original untapered.

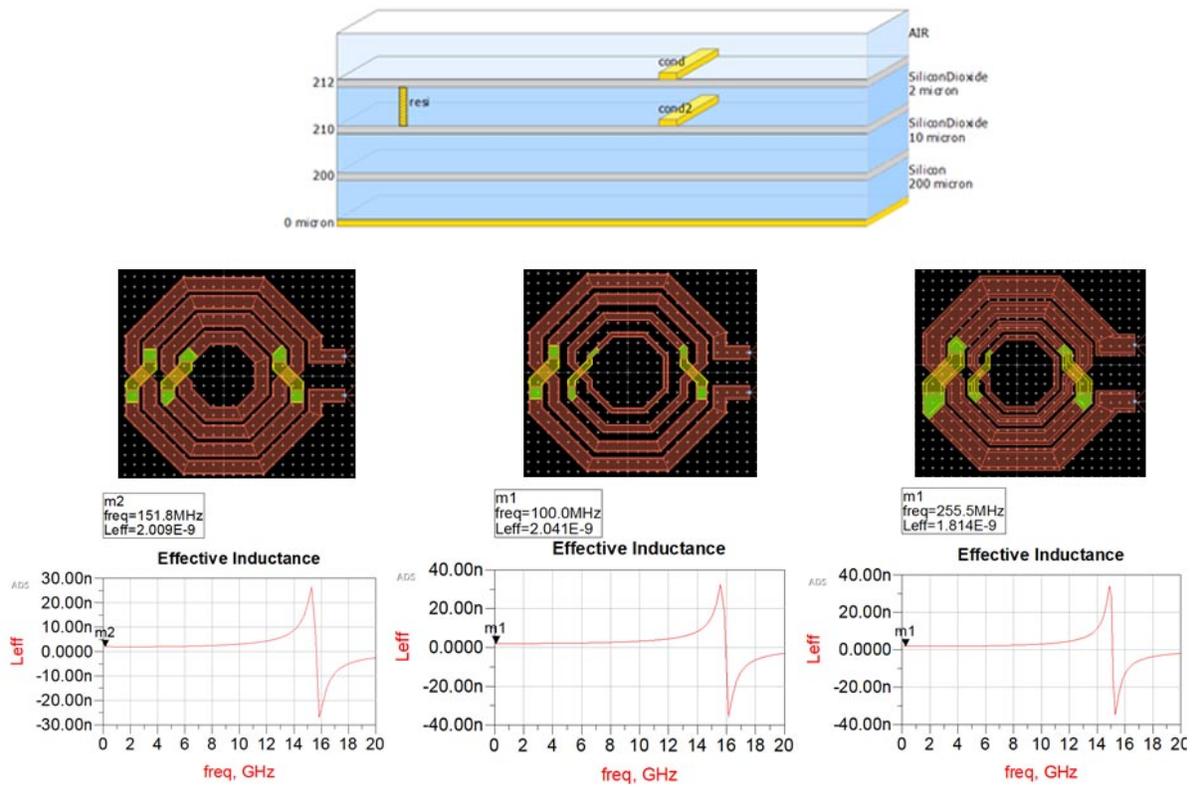


Figure 5.35: Top: Simulation setup. Left: Uniform Inductor. Middle: C-P Tapering. Right: C-G Tapering.

CHAPTER 6

Conclusions

6.1 Conclusion and Future Work of the Envelope-Tracking Supply Modulator Project

Consistent with simulations at the 40nm technology node, trellis-search (TS) shows a small but clear benefit over the hysteresis comparator based-ETSM. As we have shown in our analysis in Sec. 2.5, the overall efficiency of the ETSM still largely depends on the efficiency of the buck converter. So if future technologies enable faster complementary switches, our TS algorithm can be extended to higher bandwidths. Despite the limited improvement on overall efficiency, the most important contribution of our DSP algorithm is a mathematically rigorous proof that the existing hysteresis comparator based-architecture is close to the global optimum. We have successfully found the theoretical upper limit of the 1-inductor architecture of the ETSM, as a valuable guidance for any engineers working on ET in the future.

We have introduced the multibit ETSM architectures in Sec. 1.7.2. Although our initial intention is to avoid more quantization bits which require more off-chip passives, it is worth pointing out that our algorithm can also be generalized to the multibit scenario. Then, the trellis diagram shown in Fig. 2.1(b) will have more states and more interstate candidate branches. The increased number of states can be easily addressed with more hardware, thanks to the parallel structure of the Viterbi Decoder. However, the ACS loop's timing will be more stringent, particularly for the FPGA platform, because the loop has to complete more comparisons and selections in one DSP cycle.

Another potential improvement is found during our lab measurements close to the end of the

project. We discovered that since the resistive drop in the buck converter can be very small for low f_{sw} , so $\langle i_{load}R_s \rangle$ in (4.3) can be as low as 1 LSB. This will render the calibration difficult and inaccurate. Since our TS algorithm naturally needs oversampling, we could apply $\Sigma\Delta$ dithering on V_H and V_L in (4.3) with respect to a finer reference. The actual quantization bits will be 1 or 2 LSB(s), but the oversampling will enable effective resolution below 1 LSB. Then we can achieve more accurate representations on V_H and V_L without overly increasing the number of bits for the states in Fig. 2.1(b). The potential modification for the pre-processor is shown in Fig. 6.1.

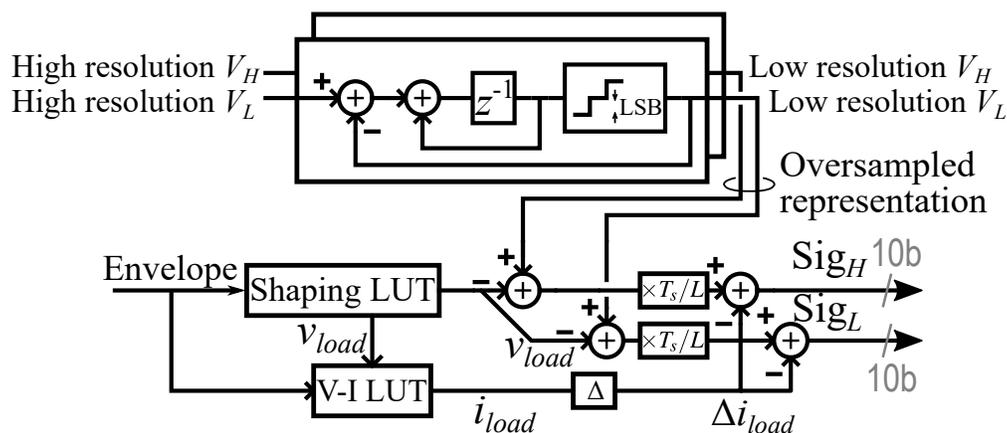


Figure 6.1: Pre-processor with dithering for future work

6.2 Conclusion and Future Work of the On-Chip Inductor Modeling Project

We have developed an approximate equivalent circuit for on-chip inductors fabricated on modern IC processes, that employ substrates of moderate-to-high resistivity. The elements in the sub-circuits modeling the skin effect, the proximity effect, and the substrate loss are calculated in closed-form expressions. Our analysis on the skin effect unifies past works. It reveals the true frequency characteristic of the proximity effect. We also show a simplified way to model the substrate: the approximate microstrip method. Using the equivalent circuit, fast exploration of the inductor geometry first-cut optimization becomes possible without EM simulation.

Definitions clarify the difference between the apparent and true quality factor, and attribute the shortcomings of the former to a non-physical equivalent circuit. Expressions for the quality

factor enable breakouts of loss contributions from various sources. All this culminates in the case studies, where with our equivalent circuits we can explain the trade-offs that lead to optimum spiral geometries, the effects of patterned ground shields, and tapering.

Here is a list of suggestions to future researchers based on our lessons learnt:

1. We do not expect equivalent circuits to replace today's fast and accurate inductor simulators. But simple estimation of losses helps untangle their relative importance approximately. In particular, the metal losses read quickly from the universal curves in Fig. 5.2 and Fig. 5.8 and the substrate loss estimated from the 3rd paragraph of Sec. 5.4 will give a good back-of-the-envelope estimation on Q_{app} or Q_{tru} of an inductor.
2. Q_{app} or Q_{tru} should be chosen properly as discussed in Sec. 5.7.2 and Sec. 5.7.3. When Q_{tru} is the objective of optimization, Sec. 5.7.4 gives a handy way of simulating in Spectre with (5.44).
3. The inductor's design space typically has broad optimum, so a rough optimization is sufficient for most of the projects. Particularly for LC oscillators, where absolute phase noise depends only on R_s of the inductor and thus small single-turn inductors are always used, the design space becomes very limited.
4. The conclusion drawn in Sec. 5.6 may be further supported if we could dissect the substrate into ring filaments and use a distributed equivalent circuit similar to Fig. 5.7 to solve the eddy current density in the substrate. The difficulties are how to model the mutual inductance between rings and how to calculate the flux enclosed by each air-core ring. They may need some heuristic approximations. But we believe that solving hard EM problems for lumped devices with approximate equivalent circuits can bridge the gap between EM engineering and circuit engineering.

APPENDIX A

Important Verilog Codes

A.1 Verilog Codes for BMU and TMU

```
1 module BMU_ACS_v3 #(
2     parameter SIG_WIDTH = 10,
3     parameter TRC_WIDTH = 16,
4     parameter NUM_STATE = 16
5 ) (
6     input wire signed [SIG_WIDTH-1:0] sig_in,
7     input wire signed [SIG_WIDTH-1:0] penalty,
8     input wire signed [TRC_WIDTH-1:0] threshold,
9     input wire signed [TRC_WIDTH-1:0] step,
10    input wire signed [SIG_WIDTH-1:0] v_high,
11    input wire signed [SIG_WIDTH-1:0] v_low,
12    output wire [0:0] error_direction_out,
13    output wire [0:NUM_STATE-1] branch_selection_out,
14    input wire rst, clk
15 );
16
17 integer k=0; // for loop index // NOT SYNTHESIZED
18
19 reg signed [TRC_WIDTH-1:0] error_reg [0:NUM_STATE-1];
20 reg signed [TRC_WIDTH-1:0] error_reg_abs_no_penalty [0:NUM_STATE-1];
21 reg signed [TRC_WIDTH-1:0] error_reg_abs_plus_penalty [0:NUM_STATE-1];
22 reg signed [0:0] error_direction_reg;
23 reg signed [0:0] error_direction_reg_old;
24 reg signed [0:0] error_direction_reg_old_old;
25
26 reg [0:0] branch_select_reg [0:NUM_STATE-1];
27
28 reg signed [TRC_WIDTH-1:0] trace_metric_reg [0:NUM_STATE-1];
29 reg signed [TRC_WIDTH-1:0] trace_metric_reg_new_0 [0:NUM_STATE-1]; // infer as wire
30 reg signed [TRC_WIDTH-1:0] trace_metric_reg_new_1 [0:NUM_STATE-1]; // infer as wire
31
```

```

32 // Added for fast search window tracking:
33 wire signed [TRC_WIDTH-1:0] error_reg_abs_first;
34 wire signed [TRC_WIDTH-1:0] error_reg_abs_last;
35
36 assign error_reg_abs_first = error_reg[0][TRC_WIDTH-1] ? ~error_reg[0] : error_reg[0];
37 assign error_reg_abs_last = error_reg[NUM_STATE-1][TRC_WIDTH-1] ? ~error_reg[NUM_STATE-1
    ] : error_reg[NUM_STATE-1];
38
39 // wire going to TBU shift reg memory
40 // NOTE: this aligns with branch_select_reg_Q
41 // NOTE: error_direction_reg_old aligns with branch_select_reg_D
42 assign error_direction_out = error_direction_reg_old_old;
43 genvar i;
44 generate
45 for (i = 0; i < NUM_STATE ; i = i + 1) begin
46     assign branch_selection_out[i] = branch_select_reg[i][0];
47 end
48 endgenerate
49
50 // combinational: Add
51 // trace_metric has more bits, so when add, need to pad MSB with zeros (abs value is
    always 0)
52 always @(*) begin
53     if (error_direction_reg_old) begin // comes from UP state transition
54         for (k = 0; k < NUM_STATE-1; k = k+1) begin
55             trace_metric_reg_new_0[k] <= trace_metric_reg[k+1] + (branch_select_reg[k+1]
                ? error_reg_abs_plus_penalty[k] : error_reg_abs_no_penalty[k] );
56         end
57         trace_metric_reg_new_0[NUM_STATE-1] <= {1'b0, {(TRC_WIDTH-1){1'b1}}};
58         for (k = 0; k < NUM_STATE; k = k+1) begin
59             trace_metric_reg_new_1[k] <= trace_metric_reg[k] + (branch_select_reg[k] ?
                error_reg_abs_no_penalty[k] : error_reg_abs_plus_penalty[k] );
60         end
61     end else begin // comes from FLAT state transition
62         for (k = 0; k < NUM_STATE; k = k+1) begin
63             trace_metric_reg_new_0[k] <= trace_metric_reg[k] + (branch_select_reg[k] ?
                error_reg_abs_plus_penalty[k] : error_reg_abs_no_penalty[k] );
64         end
65         for (k = 1; k < NUM_STATE; k = k+1) begin
66             trace_metric_reg_new_1[k] <= trace_metric_reg[k-1] + (branch_select_reg[k-1]
                ? error_reg_abs_no_penalty[k] : error_reg_abs_plus_penalty[k] );
67         end
68         trace_metric_reg_new_1[0] <= {1'b0, {(TRC_WIDTH-1){1'b1}}};

```

```

69     end // if-else
70 end
71
72 // sequential: Compare and Select
73 reg signed [TRC_WIDTH-1:0] pipe_line_out_reg [0:clogb2(NUM_STATE-1)] [0:NUM_STATE-1];
74 wire signed [TRC_WIDTH-1:0] pipe_line_out_clip;
75
76 assign pipe_line_out_clip = pipe_line_out_reg[0][0] < threshold ? pipe_line_out_reg[0][0
    ] : step;
77
78 always @(posedge clk) begin // Register: branch_select_reg; trace_metric_reg
79     if (rst) begin
80         for (k = 0; k < NUM_STATE; k = k+1) begin
81             trace_metric_reg[k] <= 0;
82             if (k < (NUM_STATE>>1))
83                 branch_select_reg[k] <= 1'b0;
84             else
85                 branch_select_reg[k] <= 1'b1;
86         end
87     end else begin
88         if (error_direction_reg_old) begin // comes from UP state transition
89             for (k = 0; k < NUM_STATE-1; k = k+1) begin
90                 trace_metric_reg[k] <= (trace_metric_reg_new_0[k] <
                    trace_metric_reg_new_1[k] ? trace_metric_reg_new_0[k] :
                    trace_metric_reg_new_1[k]) - pipe_line_out_clip;
91                 branch_select_reg[k] <= trace_metric_reg_new_0[k] >
                    trace_metric_reg_new_1[k] ? 1'b1 : 1'b0;
92             end
93             trace_metric_reg[NUM_STATE-1] <= trace_metric_reg_new_1[NUM_STATE-1] -
                pipe_line_out_clip;
94             branch_select_reg[NUM_STATE-1] <= 1'b1;
95         end else begin // comes from FLAT state transition
96             for (k = 1; k < NUM_STATE; k = k+1) begin
97                 trace_metric_reg[k] <= (trace_metric_reg_new_0[k] <
                    trace_metric_reg_new_1[k] ? trace_metric_reg_new_0[k] :
                    trace_metric_reg_new_1[k]) - pipe_line_out_clip;
98                 branch_select_reg[k] <= trace_metric_reg_new_0[k] >
                    trace_metric_reg_new_1[k] ? 1'b1 : 1'b0;
99             end
100             trace_metric_reg[0] <= trace_metric_reg_new_0[0] - pipe_line_out_clip;
101             branch_select_reg[0] <= 1'b0;
102         end // if-else
103     end

```

```

104     end
105
106     // sequential: posedge clk
107     // error_reg;
108     // error_direction_reg, error_direction_reg_old, error_direction_reg_old_old
109     // error_reg_abs_no_penalty, error_reg_abs_plus_penalty
110     always @(posedge clk) begin
111         if (rst) begin
112             for (k = 0; k < NUM_STATE; k = k+1) begin
113                 error_reg[k] <= (k - (NUM_STATE>>1)) * (v_high - v_low);
114                 error_direction_reg <= 0;
115                 error_direction_reg_old <= 0;
116                 error_direction_reg_old_old <= 0;
117             end
118             for (k = NUM_STATE>>1; k < NUM_STATE; k = k+1) begin
119                 error_reg_abs_no_penalty[k] <= (k - (NUM_STATE>>1)) * (v_high - v_low);
120                 error_reg_abs_plus_penalty[k] <= (k - (NUM_STATE>>1)) * (v_high - v_low) +
121                     penalty;
122             end
123             for (k = 0; k < NUM_STATE>>1; k = k+1) begin
124                 error_reg_abs_no_penalty[k] <= ((NUM_STATE>>1) - k) * (v_high - v_low);
125                 error_reg_abs_plus_penalty[k] <= ((NUM_STATE>>1) - k) * (v_high - v_low) +
126                     penalty;
127             end
128             end else begin
129                 for (k = 0; k < NUM_STATE; k = k+1) begin
130                     error_reg[k] <= error_reg_abs_last > error_reg_abs_first ? error_reg[k] -
131                         sig_in + v_low : error_reg[k] - sig_in + v_high;
132                     error_direction_reg <= ~(error_reg_abs_last > error_reg_abs_first);
133                     error_direction_reg_old <= error_direction_reg;
134                     error_direction_reg_old_old <= error_direction_reg_old;
135
136                     error_reg_abs_no_penalty[k] <= (error_reg[k][TRC_WIDTH-1]==0 ? error_reg[k]
137                         : -error_reg[k]);
138                     error_reg_abs_plus_penalty[k] <= (error_reg[k][TRC_WIDTH-1]==0 ? error_reg[k]
139                         : -error_reg[k]) + penalty;
140                 end
141             end // if-else
142         end // always@
143
144     // trace_matric_reg minimization pipeline
145

```

```

142     integer row,col;
143
144     // Minimization pipeline
145     always @(posedge clk) begin
146         if (rst) begin
147             for (row = 0; row < clogb2(NUM_STATE-1); row = row + 1) begin
148                 for (col = 0; col < (1<<row); col = col + 1) begin
149                     pipe_line_out_reg[row][col] <= 0;
150                 end
151             end
152         end else begin
153             for (row = 0; row < clogb2(NUM_STATE-1); row = row + 1) begin
154                 for (col = 0; col < (1<<row); col = col + 1) begin
155                     pipe_line_out_reg[row][col] <= pipe_line_out_reg[row+1][col<<1] <
156                         pipe_line_out_reg[row+1][(col<<1)+1] ? pipe_line_out_reg[row+1][col<
157                             <1] : pipe_line_out_reg[row+1][(col<<1)+1];
158                 end
159             end
160         end
161     end
162
163     // Pipeline input (count = NUM_STATE) wiring assignment
164     always @(*) begin
165         for (col = 0; col < NUM_STATE; col = col + 1) begin
166             pipe_line_out_reg[clogb2(NUM_STATE-1)][col] <= trace_metric_reg[col];
167         end
168     end
169
170
171     function integer clogb2;
172     input integer depth;
173     for (clogb2=0; depth>0; clogb2=clogb2+1)
174         depth = depth >> 1;
175     endfunction
176
177 endmodule

```

A.2 Verilog Codes for TBU

```

1 module TBU_shift_reg_simple #(
2     parameter NUM_STATE = 16,
3     parameter SEARCH_DEPTH = 4
4 ) (
5     input wire clk,
6     input wire [0:NUM_STATE-1] branch_selection_in,
7     input wire error_direction_in,
8     output wire vsw_out
9 );
10
11 reg [0:0] error_direction_memory_peek [0:SEARCH_DEPTH-1];
12 reg [0:0] error_direction_memory_gap [0:SEARCH_DEPTH-1];
13
14 reg [0:NUM_STATE-1] branch_memory_peek [0:SEARCH_DEPTH-1];
15 reg [0:NUM_STATE-1] branch_memory_gap [0:SEARCH_DEPTH-1];
16 reg [clogb2(NUM_STATE-1)-1:0] MUX_index_reg [0:SEARCH_DEPTH-1];
17
18 integer k=0;
19
20 assign vsw_out = branch_memory_peek[SEARCH_DEPTH-1][MUX_index_reg[SEARCH_DEPTH-1]];
21
22 always @(posedge clk) begin
23     // absorb results from BMU_ACS
24     MUX_index_reg[0] <= {clogb2(NUM_STATE-1){1'b0}};
25     error_direction_memory_peek[0] <= error_direction_in;
26     branch_memory_peek[0] <= branch_selection_in;
27
28     for (k = 1; k < SEARCH_DEPTH; k = k+1) begin
29         branch_memory_peek[k] <= branch_memory_gap[k-1];
30         error_direction_memory_peek[k] <= error_direction_memory_gap[k-1];
31     end
32
33     for (k = 0; k < SEARCH_DEPTH; k = k+1) begin
34         branch_memory_gap[k] <= branch_memory_peek[k];
35         error_direction_memory_gap[k] <= error_direction_memory_peek[k];
36     end
37
38     for (k = 1; k < SEARCH_DEPTH; k = k+1) begin
39         if (error_direction_memory_peek[k-1]) begin
40             if (branch_memory_peek[k-1][MUX_index_reg[k-1]])
41                 MUX_index_reg[k] <= MUX_index_reg[k-1];
42             else
43                 MUX_index_reg[k] <= MUX_index_reg[k-1] + 1;

```

```

44     end else begin
45         if (branch_memory_peek[k-1][MUX_index_reg[k-1]])
46             MUX_index_reg[k] <= MUX_index_reg[k-1] - 1;
47         else
48             MUX_index_reg[k] <= MUX_index_reg[k-1];
49     end
50 end
51
52 end // always@
53
54
55 // LOG base2 function def
56 function integer clogb2;
57     input integer depth;
58     for (clogb2=0; depth>0; clogb2=clogb2+1)
59         depth = depth >> 1;
60 endfunction
61
62 endmodule

```

A.3 Verilog Codes for Delay Lines between TMU and TBU

```

1  module DL #(
2      parameter DATA_WIDTH = 10,
3      parameter DELAY = 2
4  ) (
5      input wire clk,
6      input wire [DATA_WIDTH-1:0] da,
7      output wire [DATA_WIDTH-1:0] qa
8  );
9
10 reg [DATA_WIDTH-1:0] delay [0:DELAY-1];
11 integer i = 0;
12
13 always @(posedge clk) begin
14     delay[0] <= da;
15     for (i=0; i<(DELAY-1); i=i+1) begin
16         delay[i+1]<=delay[i];
17     end
18 end
19

```

```

20 assign qa=delay [DELAY-1];
21
22 endmodule

```

A.4 Verilog Codes for VA Wrapper

```

1
2 module Viterbi_TOP #(
3     parameter INPUT_SIG_WIDTH = 10,
4     parameter TRC_WIDTH = 16,
5     parameter NUM_STATE = 64,
6     parameter SEARCH_DEPTH = 256
7 ) (
8     input wire [INPUT_SIG_WIDTH-1:0] input_sig,
9     input wire [INPUT_SIG_WIDTH-1:0] v_high,
10    input wire [INPUT_SIG_WIDTH-1:0] v_low,
11    input wire [TRC_WIDTH-1:0] threshold,
12    input wire [TRC_WIDTH-1:0] step,
13    input wire [INPUT_SIG_WIDTH-1:0] penalty,
14    input wire dsp_clk,
15    input wire pp_clk,
16    input wire rst_single_2,
17    output wire [0:0] vsw,
18    output wire [15:0] monitor
19 );
20
21 wire [NUM_STATE-1:0] branch_selection_BMU_ACS_2_TBU;
22 wire [0:0] error_direction_BMU_ACS_2_TBU;
23
24 wire [NUM_STATE-1:0] branch_selection_BMU_ACS;
25 wire [0:0] error_direction_BMU_ACS;
26
27 wire [NUM_STATE-1:0] branch_selection_TBU;
28 wire [0:0] error_direction_TBU;
29
30 assign monitor = branch_selection_BMU_ACS;
31
32 BMU_ACS_v3 #(
33     .SIG_WIDTH(INPUT_SIG_WIDTH),
34     .TRC_WIDTH(TRC_WIDTH),
35     .NUM_STATE(NUM_STATE)

```

```

36 ) BMU_ACS_v3_DUT (
37     .sig_in(input_sig),
38     .threshold(threshold),
39     .step(step),
40     .penalty(penalty),
41     .v_high(v_high),
42     .v_low(v_low),
43     .branch_selection_out(branch_selection_BMU_ACS),
44     .error_direction_out(error_direction_BMU_ACS),
45     .rst(rst_single_2),.clk(dsp_clk)
46 );
47
48 TBU_shift_reg_simple #(
49     .NUM_STATE(NUM_STATE),
50     .SEARCH_DEPTH(SEARCH_DEPTH)
51 ) TBU_shift_reg_DUT (
52     .clk(pp_clk),
53     .branch_selection_in(branch_selection_TBU),
54     .error_direction_in(error_direction_TBU),
55     .vsw_out(vsw)
56 );
57
58 DL #(
59     .DATA_WIDTH(NUM_STATE),
60     .DELAY(2)
61 ) DL_dsp_branch_selection (
62     .clk(dsp_clk),
63     .da(branch_selection_BMU_ACS),
64     .qa(branch_selection_BMU_ACS_2_TBU)
65 );
66
67 DL #(
68     .DATA_WIDTH(1),
69     .DELAY(2)
70 ) DL_dsp_error_direction (
71     .clk(dsp_clk),
72     .da(error_direction_BMU_ACS),
73     .qa(error_direction_BMU_ACS_2_TBU)
74 );
75
76 DL #(
77     .DATA_WIDTH(NUM_STATE),
78     .DELAY(2)

```

```
79 ) DL_pp_branch_selection (
80     .clk(pp_clk),
81     .da(branch_selection_BMU_ACS_2_TBU),
82     .qa(branch_selection_TBU)
83 );
84
85 DL #(
86     .DATA_WIDTH(1),
87     .DELAY(2)
88 ) DL_pp_error_direction (
89     .clk(pp_clk),
90     .da(error_direction_BMU_ACS_2_TBU),
91     .qa(error_direction_TBU)
92 );
93
94 endmodule
```

APPENDIX B

Derivation from Proximity Effect's Universal Curves to the Equivalent Circuits

$$\hat{P}_{prox} = \frac{P_{prox}}{|B_{ext}|^2 l_{tt}} \quad (\text{B.1})$$

According to the asymptote in Fig. 5.8(b):

$$\frac{\hat{P}_{prox}}{\frac{\rho_m w}{t_m}} = \frac{0.166}{\mu_0^2} \times x^4, \text{ where } x = \frac{\sqrt{\omega t_m}}{\delta} \quad (\text{B.2})$$

$$= \frac{0.166}{\mu_0^2} \times \frac{w^2 t_m^2}{\frac{4\rho_m^2}{\omega^2 \mu_0^2}} \quad (\text{B.3})$$

$$\Rightarrow \hat{P}_{prox} = \frac{\rho_m w}{t_m} \frac{0.166}{4} \times \omega^2 \times w^2 \times t_m^2 \times \frac{1}{\rho_m^2} \quad (\text{B.4})$$

$$= \frac{0.166}{4} \times \omega^2 \frac{w^3 t_m}{\rho_m} \quad (\text{B.5})$$

$$\therefore P_{prox} = l_{tt} \times |B_{ext}|^2 \times \frac{0.166}{4} \times \omega^2 \frac{w^3 t_m}{\rho_m} = \frac{1}{2} |I_{even}|^2 \frac{\omega^2 L_p^2}{R} \quad (\text{B.6})$$

This should give L_p^2/R , but we found (5.6) works better with doubling of power. As a reminder, the dimension of B is H A m^{-1} .

APPENDIX C

Calculation of Sec. 5.8.1's Reference Equivalent Circuit

We present below the details of finding parameter values for the complete equivalent circuit of Fig. 5.16(a). This 4.48 nH inductor of symmetrical geometry is driven differentially (Sec. 5.8.1). Geometry parameters: $s = 2.2 \mu\text{m}$, $d_{in} = 30.2 \mu\text{m}$, $N = 8$, $w = 2.8 \mu\text{m}$. Process parameters: $t_m = 3.3 \mu\text{m}$, $\rho_m = 17 \text{ n}\Omega\text{cm}$, $t_{ox} = 5 \mu\text{m}$, $\epsilon_{ox} = 3.9$, $t_{sub} = 200 \mu\text{m}$, $\rho_{sub} = 10 \Omega\text{cm}$, and $\epsilon_{sub} = 11.7$.

Using [85, Eqn. (2)], DC inductance is calculated from the spiral's geometry. The shape of the spiral is between square and octagon but not hexagon, so we choose the averaged coefficients listed in [85, Table II] for square and octagon: $c_1 = 1.17$, $c_2 = 2.18$, $c_3 = 0.09$, $c_4 = 0.16$. The calculated L_{dc} is 4.69 nH, which is close to the measured inductance of 4.48 nH. This averaging on geometry works well. Accurate L_{dc} is not very important for our purposes: a rough estimation that is consistent with commonsense is good enough. For other regular shapes, normal coefficients in [85, Table II] can be adopted similarly for use with [85, Eqn. (2)].

The length of each turn is measured by the ruler in Momentum to be

$$\mathbf{l} = 2 \times [180 \ 163.37 \ 146.74 \ 130.11 \ 113.49 \ 96.86 \ 80.23 \ 63.6] \mu\text{m}.$$

They sum up across the entire inductor to $l_{tt} = 1949 \mu\text{m}$ and an R_{dc} of 3.6Ω . For regular shapes, \mathbf{l} can be derived from geometrical parameters directly.

First we calculate the physical oxide capacitance between metals, which will be used to estimate C_s and C_{ox} . Using [72, Eqn. (9~18), Tab. II] and the relevant geometric and process parameters, we calculate the coupling capacitance per unit length between two turns $C_{couple} = 8.428 \times 10^{-11} \text{ F/m}$. The bottom plate capacitance over oxide per unit length of the sandwiched turns is $C_{bot,in} = 3.293 \times 10^{-11} \text{ F/m}$, and for the edge turns is $C_{bot,out} = 7.025 \times 10^{-11} \text{ F/m}$. Using the method of [71], the voltage of each half section is $\mathbf{v} = V_0 \times [+0.93 \ -0.81 \ +0.67 \ -0.55 \ +$

$0.42 - 0.29 + 0.16 - 0.03]^T$. The alternating signs in \mathbf{v} are a result of windings that are symmetric around the spiral's center. From \mathbf{v} , \mathbf{l} and C_{couple} , we can derive the total electrical energy stored in the inter-winding capacitance across each pair of adjacent turns, concluding with an effective lumped $C_s = 54.55$ fF across the differential terminals that stores the same amount of energy for a total voltage drop of $2V_0$.

To calculate the total common-mode oxide capacitance, the inductor geometry is assumed circular with unchanged diameter and pitch. Then, the length of each turn is πd_i , where d_i is the diameter of each turn. The oxide capacitance contributed by all inner (sandwiched) turns is 42.21 fF and by the edge turns it is 30.02 fF. Then from Sec. 5.4, the total oxide capacitance for the R - C sub network is $(42.21+30.02)/12=6.02$ fF.

Next, following Sec. 5.4 we approximate the entire inductor as a ring of microstrip line with width $37.8 \mu\text{m}$ and length $213.63 \mu\text{m}$. The substrate capacitance and conductance per unit length are calculated using the expressions in [89], listed as [70, Eqn. (14~17)]. They are, respectively, 1.051×10^{-10} F/m and 9.53 S/m. Thus the total substrate capacitance and resistance for the R - C sub network are, respectively, 1.87 fF and 5.90 k Ω (Sec. 5.4).

Although it is non-physical to assume that the inductor is circular with the same diameter as the actual structure, we found that this approximation gives the best fit for all cases, so we suggest it as a rule of thumb.

For the skin effect network, we first calculate the reference frequency $f_0 = 1.797$ GHz and, using the expressions in Fig. 5.3, an adjusting factor $a = 0.9963$. Then, the relations marked on Fig. 5.3 lead to the poles and zeros that comprise (5.2). They are, from frequency of the first zero to last pole, [3.693 7.535 17.03 38.23 86.11 194.0] GHz. Thus Z_{skin} is fully known from R_{dc} and (5.2). A RL network may be synthesized approximating the impedance Z_{skin} following the steps in [56].

For the proximity effect network, we need to calculate $|B_i/I|^2 \cdot l_i$ for each turn. When using [65, Eqn. (11)], we temporarily assume the inductor is circular with the same diameter as the actual structure. Then B_i/I of each turn is calculated by averaging the B-field at outer edge and the inner edge of each metal trace. Those B-fields for each turn are calculated by sum-

ming vertical B -fields created by all turns, including the turn in question. This fully interconnected calculation can be implemented by an $N \times N$ loop in MATLAB. However, we use the actual length l_i for turn i . In other words, the only simple way to calculate B_i/I is to assume that the inductor is circular, but to be consistent with the calculation of skin effect loss we use l_i taken from the actual geometry. For this case, from outermost turn to innermost turn, $\mathbf{B}_i/I = [-0.043 \ 0.014 \ 0.049 \ 0.0784 \ 0.106 \ 0.136 \ 0.170 \ 0.224] \text{ H/m}^2$. Then the sum in (5.7) is $2.024 \times 10^{-5} \text{ H}^2/\text{m}^3$, and from (5.5), $L^2/R = 1.432 \times 10^{-20} \text{ H}^2/\Omega$. Since $w < t_m$ for this case, expressions in Fig. 5.10 provide $F_1 = 1.116$, $R/L = 1.824 \times 10^{10} \text{ s}^{-1}$, and an adjusting factor $a = 1$ for proximity effect. From L^2/R and R/L , we get $L \Rightarrow L_p = 262 \text{ pH}$ and $R \Rightarrow Z_{prox}|_{DC} = 4.76 \Omega$. Using $a = 1$, $f_0 = F_1 \cdot 1.797 = 2 \text{ GHz}$ and the other relations for skin effect given in Fig. 5.3, we can calculate the poles and zeros for Z_{prox} . They are, from frequency of first zero to last pole, $[4.106 \ 8.407 \ 18.94 \ 42.65 \ 96.08 \ 216.4] \text{ GHz}$. Thus Z_{prox} is fully known from $Z_{prox}|_{DC} = 4.76 \Omega$ and (5.2). Last, using the L_{dc} calculated at the beginning of this section, $L_s = 4.69 - 0.262 = 4.429 \text{ nH}$. The effective inductance of Z_{skin} is ignored, because it models L_{int} at DC, but as discussed in Sec. 5.3.1 this is much smaller than L_{dc} .

With values of all the components in the equivalent circuit now at hand, we can derive Q_{app} at any frequency as $\text{Im}(Z)/\text{Re}(Z)$, and Q_{tru} from (5.43).

REFERENCES

- [1] W. H. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proceedings of the Institute of Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, 1936.
- [2] H. Chireix, "High Power Outphasing Modulation," *Proceedings of the Institute of Radio Engineers*, vol. 23, no. 11, pp. 1370–1392, 1935.
- [3] S. Moloudi and A. A. Abidi, "The Outphasing RF Power Amplifier: A Comprehensive Analysis and a Class-B CMOS Realization," *IEEE Trans. Syst. Sci. Cybern.*, vol. 48, no. 6, pp. 1357–1369, 2013.
- [4] L. R. Kahn, "Single-Sideband Transmission by Envelope Elimination and Restoration," *Proc. of IRE*, vol. 40, no. 7, pp. 803–806, 1952.
- [5] F. Ellinger, *Radio Frequency Integrated Circuits and Technologies*. Springer, 2008.
- [6] J. Paek, D. Kim, J. Bang, J. Baek, J. Choi, T. Nomiyama, J. Han, Y. Choo, Y. Youn, E. Park, S. Lee, I. Kim, J. Lee, T. B. Cho, and I. Kang, "An 88%-Efficiency Supply Modulator Achieving 1.08 μ s/V Fast Transition and 100MHz Envelope-Tracking Bandwidth for 5G New Radio RF Power Amplifier," in *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2019, pp. 238–240.
- [7] T. Kwak, M. Lee, and G. Cho, "A 2 W CMOS Hybrid Switching Amplitude Modulator for EDGE Polar Transmitters," *IEEE Trans. Syst. Sci. Cybern.*, vol. 42, no. 12, pp. 2666–2676, 2007.
- [8] P. Y. Wu and P. K. T. Mok, "A Two-Phase Switching Hybrid Supply Modulator for RF Power Amplifiers With 9% Efficiency Improvement," *IEEE Trans. Syst. Sci. Cybern.*, vol. 45, no. 12, pp. 2543–2556, 2010.
- [9] R. Shrestha, R. van der Zee, A. de Graauw, and B. Nauta, "A Wideband Supply Modulator for 20 MHz RF Bandwidth Polar PAs in 65 nm CMOS," *IEEE Trans. Syst. Sci. Cybern.*, vol. 44, no. 4, pp. 1272–1280, 2009.
- [10] M. Hassan, L. E. Larson, V. W. Leung, and P. M. Asbeck, "A Combined Series-Parallel Hybrid Envelope Amplifier for Envelope Tracking Mobile Terminal RF Power Amplifier Applications," *IEEE Trans. Syst. Sci. Cybern.*, vol. 47, no. 5, pp. 1185–1198, 2012.
- [11] H. He, T. Ge, and J. Chang, "A review on supply modulators for Envelope-Tracking Power Amplifiers," in *2016 International Symposium on Integrated Circuits (ISIC)*, 2016, pp. 1–4.
- [12] B. Park, D. Kim, S. Kim, Y. Cho, J. Kim, D. Kang, S. Jin, K. Moon, and B. Kim, "High-Performance CMOS Power Amplifier With Improved Envelope Tracking Supply Modulator," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 3, pp. 798–809, 2016.

- [13] C. Ho, S. Lin, C. Meng, H. Hong, S. Yan, T. Kuo, C. Peng, C. Hsiao, H. Chen, D. Sung, and C. Kuan, "An 87.1% Efficiency RF-PA Envelope-Tracking Modulator for 80MHz LTE-Advanced Transmitter and 31dBm PA Output Power for HPUE in 0.153 μ m CMOS," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, 2018, pp. 432–434.
- [14] M. Hassan, P. M. Asbeck, and L. E. Larson, "A CMOS dual-switching power-supply modulator with 8% efficiency improvement for 20MHz LTE Envelope Tracking RF power amplifiers," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 366–367.
- [15] P. Mahmoudidaryan, D. Mandal, B. Bakaloglu, and S. Kiaei, "A 91%-Efficiency Envelope-Tracking Modulator Using Hysteresis-Controlled Three-Level Switching Regulator and Slew-Rate-Enhanced Linear Amplifier for LTE-80MHz Applications," in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2019, pp. 428–430.
- [16] J. Choi and G. Cho, "A Hybrid Power Amplifier Using 3-Phase 3-Level Class-D with 200nH Inductors and Current Balancing Technique," in *2017 Symposium on VLSI Circuits*, 2017, pp. C138–C139.
- [17] H. Gwon, J. Bang, K. Yoon, S. Park, S. Park, M. Jung, S. Lee, M. Kim, S. Hong, and G. Cho, "2-Phase 3-Level ETSM With Mismatch-Free Duty Cycles Achieving 88.6% Peak Efficiency for a 20-MHz LTE RF Power Amplifier," *IEEE Trans. Power Electron.*, vol. 33, no. 4, pp. 2815–2819, 2018.
- [18] C. Hsia, A. Zhu, J. J. Yan, P. Draxler, D. F. Kimball, S. Lanfranco, and P. M. Asbeck, "Digitally Assisted Dual-Switch High-Efficiency Envelope Amplifier for Envelope-Tracking Base-Station Power Amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 11, pp. 2943–2952, 2011.
- [19] R. Wood, "Magnetic Megabits," *IEEE Spectr.*, vol. 27, no. 5, pp. 32–33, 1990.
- [20] G. D. Forney, "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [21] A. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, 1967.
- [22] C. Rader, "Memory Management in a Viterbi Decoder," *IEEE Trans. Commun.*, vol. 29, no. 9, pp. 1399–1401, 1981.
- [23] P. J. Black and T. H. Meng, "A 140-Mb/s, 32-State, Radix-4 Viterbi Decoder," *IEEE Trans. Syst. Sci. Cybern.*, vol. 27, no. 12, pp. 1877–1885, 1992.
- [24] G. Fettweis and H. Meyr, "High-Speed Parallel Viterbi Decoding: Algorithm and VLSI-Architecture," *IEEE Commun. Mag.*, vol. 29, no. 5, pp. 46–55, 1991.
- [25] E. Janssen and A. van Roermund, *Look-Ahead Based Sigma-Delta Modulation*, ser. Analog Circuits and Signal Processing. Springer Netherlands, 2011.

- [26] M. Bathily, B. Allard, F. Hasbani, V. Pinon, and J. Verdier, "Design Flow for High Switching Frequency and Large-Bandwidth Analog DC/DC Step-Down Converters for a Polar Transmitter," *IEEE Trans. Power Electron.*, vol. 27, no. 2, pp. 838–847, 2012.
- [27] W. Sansen, *Analog Design Essentials*. Springer US, 2007.
- [28] H. He, Y. Kang, T. Ge, L. Guo, and J. S. Chang, "A 2.5-W 40-MHz-Bandwidth Hybrid Supply Modulator With 91% Peak Efficiency, 3-V Output Swing, and 4-mV Output Ripple at 3.6-V Supply," *IEEE Trans. Power Electron.*, vol. 34, no. 1, pp. 712–723, 2019.
- [29] D. Chowdhury, S. R. Mundlapudi, and A. Afsahi, "A fully integrated reconfigurable wide-band envelope-tracking SoC for high-bandwidth WLAN applications in a 28nm CMOS technology," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 34–35.
- [30] W. Tsai, C. Liou, Z. Peng, and S. Mao, "Wide-Bandwidth and High-Linearity Envelope-Tracking Front-End Module for LTE-A Carrier Aggregation Applications," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4657–4668, 2017.
- [31] M. Tan and W. Ki, "A 100 MHz Hybrid Supply Modulator With Ripple-Current-Based PWM Control," *IEEE J. Solid-State Circuits*, vol. 52, no. 2, pp. 569–578, 2017.
- [32] C. Kim, C. Chae, Y. Yuk, C. M. Thomas, Y. Kim, J. Kwon, S. Ha, G. Cauwenberghs, and G. Cho, "A 500-MHz Bandwidth 7.5-mVpp Ripple Power-Amplifier Supply Modulator for RF Polar Transmitters," *IEEE J. Solid-State Circuits*, vol. 53, no. 6, pp. 1653–1665, 2018.
- [33] A. M. Niknejad and R. G. Meyer, "Analysis, Design, and Optimization of Spiral Inductors and Transformers for Si RF ICs," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1470–1481, Oct 1998.
- [34] ———, "Analysis of Eddy-Current Losses over Conductive Substrates with Applications to Monolithic Inductors and Transformers," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 1, pp. 166–176, Jan 2001.
- [35] B. Rejaei, "Mixed-Potential Volume Integral-Equation Approach for Circular Spiral Inductors," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 8, pp. 1820–1829, Aug 2004.
- [36] C. P. Yue and S. S. Wong, "Physical Modeling of Spiral Inductors on Silicon," *IEEE Trans. Electron Devices*, vol. 47, no. 3, pp. 560–568, Mar 2000.
- [37] A. Scuderi, T. Biondi, E. Ragonese, and G. Palmisano, "A Lumped Scalable Model for Silicon Integrated Spiral Inductors," *IEEE Trans. Circuits Syst. I*, vol. 51, no. 6, pp. 1203–1209, June 2004.
- [38] K. Tong and C. Tsui, "A Physical Analytical Model of Multilayer On-Chip Inductors," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1143–1149, April 2005.

- [39] J. Sieiro, J. López-Villegas, J. Cabanillas, J. Osorio, and J. Samitier, “A Physical Frequency-Dependent Compact Model for RF Integrated Inductors,” *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 1, pp. 384–392, Jan 2002.
- [40] M. Danesh and J. R. Long, “Differentially Driven Symmetric Microstrip Inductors,” *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 1, pp. 332–341, Jan 2002.
- [41] Y. Cao, R. Groves, X. Huang, N. Zamdmer, J.-O. Plouchart, R. Wachnik, T.-J. King, and C. Hu, “Frequency-Independent Equivalent-Circuit Model for On-Chip Spiral Inductors,” *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 419–426, Mar 2003.
- [42] A. C. Watson, D. Melendy, P. Francis, K. Hwang, and A. Weisshaar, “A Comprehensive Compact-Modeling Methodology for Spiral Inductors in Silicon-Based RFICs,” *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 3, pp. 849–857, March 2004.
- [43] C. Wang, H. Liao, C. Li, R. Huang, W. Wong, X. Zhang, and Y. Wang, “A Wideband Predictive Double- π Equivalent-Circuit Model for On-Chip Spiral Inductors,” *IEEE Trans. Electron Devices*, vol. 56, no. 4, pp. 609–619, April 2009.
- [44] J. Gil and H. Shin, “Simple Wide-Band On-Chip Inductor Model for Silicon-Based RF ICs,” in *Simulation of Semiconductor Processes and Devices, 2003. SISPAD 2003. International Conference on*, Sept 2003, pp. 35–38.
- [45] X. Huo, P. C. H. Chan, K. J. Chen, and H. C. Luong, “A Physical Model for On-Chip Spiral Inductors With Accurate Substrate Modeling,” *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp. 2942–2949, Dec 2006.
- [46] F. Huang, J. Lu, N. Jiang, X. Zhang, W. Wu, and Y. Wang, “Frequency-independent asymmetric double- π equivalent circuit for on-chip spiral inductors: Physics-based modeling and parameter extraction,” *IEEE J. Solid-State Circuits*, vol. 41, no. 10, pp. 2272–2283, Oct 2006.
- [47] W. Leng, “Design-Oriented Modeling and Optimization of On-Chip Inductors,” Master’s thesis, UCLA, <https://escholarship.org/uc/item/2nh4g9f4>, 6 2015.
- [48] P. Silvester, “Modal Network Theory of Skin Effect in Flat Conductors,” *Proc. IEEE*, vol. 54, no. 9, pp. 1147–1151, Sept 1966.
- [49] G. Antonini, A. Orlandi, and C. R. Paul, “Internal Impedance of Conductors of Rectangular Cross Section,” *IEEE Trans. Microw. Theory Techn.*, vol. 47, no. 7, pp. 979–985, Jul 1999.
- [50] E. B. Rosa, “The self and mutual inductances of linear conductors,” *Bulletin of the Bureau of Standards*, vol. 4, no. 2, pp. 301–344, 1908.
- [51] H. B. Dwight, “Geometric Mean Distances for Rectangular Conductors,” *Trans. American Inst. of Electr. Engineers*, vol. 65, no. 8, pp. 536–538, Aug 1946.
- [52] W. H. Hayt, Jr. and J. A. Buck, *Engineering Electromagnetics*, 6th ed. New York, NY: McGraw-Hill, 2001.

- [53] C. Holloway and E. F. Kuester, “DC Internal Inductance for a Conductor of Rectangular Cross Section,” *IEEE Trans. Electromagn. Compat.*, vol. 51, no. 2, pp. 338–344, May 2009.
- [54] F. E. Terman, *Radio Engineer’s Handbook*. New York: McGraw-Hill, 1943.
- [55] H. Statz, E. A. Guillemin, and R. A. Pucel, “Design Considerations of Junction Transistors at Higher Frequencies: Based upon an Accurate Equivalent Circuit,” *Proc. of the IRE*, vol. 42, no. 11, pp. 1620–1628, Nov 1954.
- [56] M. E. Van Valkenburg, *Introduction to Modern Network Synthesis*. New York, NY: Wiley, 1960.
- [57] S. Kim and D. P. Neikirk, “Compact Equivalent Circuit Model for the Skin Effect,” in *IEEE Microwave Symposium*, vol. 3, San Francisco, CA, June 1996, pp. 1815–1818 vol.3.
- [58] H. Wheeler, “Formulas for the Skin Effect,” *Proc. IEEE*, vol. 30, no. 9, pp. 412–424, Sept 1942.
- [59] S. Mei and Y. I. Ismail, “Modeling Skin and Proximity Effects with Reduced Realizable RL Circuits,” *IEEE Trans. VLSI Syst.*, vol. 12, no. 4, pp. 437–447, April 2004.
- [60] J. M. López-Villegas, J. Samitier, C. Cane, P. Losantos, and J. Bausells, “Improvement of the Quality Factor of RF Integrated Inductors by Layout Optimization,” *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 1, pp. 76–83, Jan 2000.
- [61] W. Kuhn and N. Ibrahim, “Analysis of Current Crowding Effects in Multiturn Spiral Inductors,” *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 1, pp. 31–38, Jan 2001.
- [62] P. L. Dowell, “Effects of eddy currents in transformer windings,” *Proc. of IEE*, vol. 113, no. 8, pp. 1387–1394, August 1966.
- [63] J. Ferreira, “Analytical Computation of AC resistance of Round and Rectangular Litz Wire Windings,” *IEE Proc. B: Electric Power Applications*, vol. 139, no. 1, pp. 21–25, Jan 1992.
- [64] R. W. Erickson and D. Maksimović, *Fundamentals of Power Electronics*, 2nd ed. Norwell, MA: Kluwer, 2001.
- [65] D. Montgomery and J. Terrell, *Some Useful Information for the Design of Air-Core Solenoids: Part I: Relationships between Magnetic Field, Power, Ampere-Turns and Current Density. Part II: Homogenous Magnetic Fields*, ser. AFOSR-1525. MIT National Magnet Laboratory, 1961.
- [66] MIT Dept. of EE Staff, *Electric Circuits*. New York, NY: Wiley, 1943.
- [67] M. E. Van Valkenburg, *Network Analysis*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1974.
- [68] I. Lope, C. Carretero, J. Acero, R. Alonso, and J. M. Burdío, “AC Power Losses Model for Planar Windings With Rectangular Cross-Sectional Conductors,” *IEEE Trans. Power Electron.*, vol. 29, no. 1, pp. 23–28, Jan 2014.

- [69] B. Ufnalski, “Foster and Cauer Equivalent Networks,” <https://www.mathworks.com/matlabcentral/fileexchange/48042-foster-and-cauer-equivalent-networks>, Aug. 2020.
- [70] Y. Eo and W. R. Eisenstadt, “High-Speed VLSI Interconnect Modeling Based on S-Parameter Measurements,” *IEEE Trans. Compon., Hybrids, Manuf. Technol.*, vol. 16, no. 5, pp. 555–562, Aug 1993.
- [71] C.-H. Wu, C.-C. Tang, and S.-I. Liu, “Analysis of On-Chip Spiral Inductors using the Distributed Capacitance Model,” *IEEE J. Solid-State Circuits*, vol. 38, no. 6, pp. 1040–1044, June 2003.
- [72] W. Zhao, X. Li, S. Gu, S. Kang, M. Nowak, and Y. Cao, “Field-Based Capacitance Modeling for Sub-65-nm On-Chip Interconnect,” *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1862–1872, Sept 2009.
- [73] J. Lee, A. Kral, A. A. Abidi, and N. G. Alexopoulos, “Design of Spiral Inductors on Silicon Substrates with a Fast Simulator,” in *Proc. of European Solid-State Circuits Conf.*, The Hague, The Netherlands, Sept 1998, pp. 328–331.
- [74] J. N. Burghartz and B. Rejaei, “On the Design of RF Spiral Inductors on Silicon,” *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 718–729, March 2003.
- [75] Y. Liu, S. A. Sebo, R. Caldecott, D. G. Kastan, and S. E. Wright, “Modeling of Converter Transformers using Frequency Domain Terminal Impedance Measurements,” *IEEE Trans. on Power Delivery*, vol. 8, no. 1, pp. 66–72, 1993.
- [76] R. M. Fano, L. J. Chu, and R. B. Adler, *Electromagnetic Fields, Energy, and Forces*. New York, NY: Wiley, 1960.
- [77] “Keysight Impedance Measurement Handbook,” Application Note 5950-3000, Keysight Technologies, Santa Rosa, CA, Nov. 2016.
- [78] F. E. Terman and J. M. Pettit, *Electronic Measurements*, 2nd ed. New York: McGraw-Hill, 1952.
- [79] D. B. Leeson, “A simple model of feedback oscillator noise spectrum,” *Proc. IEEE*, vol. 54, no. 2, pp. 329–330, 1966.
- [80] K. K. O, “Estimation Methods for Quality Factors of Inductors Fabricated in Silicon Integrated Circuit Process Technologies,” *IEEE J. Solid-State Circuits*, vol. 33, no. 8, pp. 1249–1252, Aug 1998.
- [81] K. K. Clarke and D. T. Hess, *Communication Circuits: Analysis and Design*. Reading, MA: Addison-Wesley, 1971.
- [82] R. Beringer, “Resonant Cavities as Microwave Circuit Elements,” in *Principles of Microwave Circuits*, ser. MIT Rad Lab, C. Montgomery, R. Dicke, and E. Purcell, Eds. New York, NY: McGraw-Hill, 1948, vol. 8, pp. 209–237.

- [83] J. Pan, A. A. Abidi, W. Jiang, and D. Marković, “Simultaneous Transmission of up to 94 mW Self-Regulated Wireless Power and up to 5 Mb/s Reverse Data over a Single Pair of Coils,” *IEEE Trans. Syst. Sci. Cybern.*, vol. 54, no. 4, pp. 1003–1016, April 2019.
- [84] A. Zolfaghari, A. Chan, and B. Razavi, “Stacked Inductors and Transformers in CMOS Technology,” *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 620–628, Apr 2001.
- [85] S. S. Mohan, M. del Mar Hershenson, S. P. Boyd, and T. H. Lee, “Simple Accurate Expressions for Planar Spiral Inductances,” *IEEE J. Solid-State Circuits*, vol. 34, no. 10, pp. 1419–1424, Oct 1999.
- [86] M. Tohidian, S. Mehr, and R. B. Staszewski, “Dual-Core High-Swing Class-C Oscillator with Ultra-Low Phase Noise,” in *IEEE Radio Frequency Integrated Circuits Symp.*, Seattle, WA, June 2013, pp. 243–246.
- [87] C. P. Yue and S. S. Wong, “On-Chip Spiral Inductors with Patterned Ground Shields for Si-Based RF ICs,” *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 743–752, May 1998.
- [88] V. N. R. Vanukuru, “High-Q Inductors Utilizing Thick Metals and Dense-Tapered Spirals,” *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 3095–3099, Sept 2015.
- [89] H. Hasegawa, M. Furukawa, and H. Yanai, “Properties of Microstrip Line on Si-SiO₂ System,” *IEEE Trans. Microw. Theory Techn.*, vol. 19, no. 11, pp. 869–881, Nov 1971.