High Linearity Full Duplex System Implemented with Novel Impedance Matching Network

Fucheng Yin

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Electrical Engineering

University of Washington

2022

Committee:

Jacques Christophe Rudell

Yasuo Kuga

Program Authorized to Offer Degree:

Department of Electrical Engineering

University of Washington

Abstract

High Linearity Full Duplex System Implemented with Novel Impedance Matching Network

Fucheng Yin

Chair of the Supervisory Committee:
Professor Jacques C. Rudell
Department of Electrical Engineering

As the era of 5G internet begins to rise, the available spectrum is getting more congested and expensive. The full-duplex technology has evolved as one solution to help improve spectral congestion for mobile operators and infrastructure vendors, due to its capability to double the spectral efficiency. In this thesis, all discussions surround the design of the front-end system of full-duplex radios aiming at extremely high IIP3 performance. With this premise, we study the structure of an electrical balanced duplexer via systemically analyzing its loss mechanism, hereby improving the insertion loss. In addition, a novel topology of the impedance matching network that promises a broad tuning range while occupying a reasonable die area has been explored. We tape-outed this work in the TSMC general mixed-signal 45nm technology node, and its measurement result is presented at the end of this thesis.

TABLE OF CONTENTS

LIST OF FIGUIRES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have been incredibly fortunate enough to receive guidance from my supervisor, Professor Rudell, who is always patient and supportive. I want to express my deepest gratitude here. I also want to express my thanks to my colleagues, my family, and whoever is willing to give me a helping hand during my life of study and research.

1 Introduction

1.1 Full-Duplex (FD) Wireless Communication in 5G

With the proliferation of Internet usage, the explosion of wireless mobile devices and services, 5th generation (5G) communication system provides an immense impetus to hastily develop and deploy for half of the United States by early 2021. As 5G features with high-data rates and low latency communications for extreme coverage that accommodate users for ubiquitous and instantaneous Internet access, a significantly higher system capacity of $1000X$ is obliged to be part of the system requirements. In order to manage this $1000X$ traffic surge, the wireless spectrum is taking the burden of the physical limitations, unlikely seen in previous generations of wireless systems. In regard to addressing many engineering issues, researchers have been seeking ways to increase spectral efficiency, one of which is the use full-duplex (FD) communication systems. The concept of FD communication can be tracked back to the 1800s century enables the simultaneous transmission and reception for signals so as to promisingly double the capacity of the spectrum[25]. While FD radios aim to salvage the scarcity of the spectrum space and reliably shines some light on the gateway of the technology innovations for 5G, it does face a number of technological difficulties which await to be resolved. One of the cardinal challenges to realizing FD radios is self-interference (SI) which refers to the significant leakage signals from the transmitting chain to the receiving chain, desensitizing the desired received signal, in some cases even saturating the receiver rendering it useless. To picture this issue more vividly, we can calculate and compare the Shannon capacities of an FD system with a poor SI performance and that of a standard half-duplex system. To a typical Wi-Fi protocol, the average power of the transmitting signal is north of $20(dBm)$. Meanwhile, the thermal noise level

of a signal with $20(MHz)$ bandwidth is approximately $-101(dBm)$. In order to detect a fine signal from the antenna, the SI cancelation ought to be larger than $20-(-91)=111(dB)$, considering the detectable signal to be higher than the noise floor by $10(dBm)$. Namely, as long as the SI cancelation is less than $111(dB)$, the received signal will become overwhelmed by the sum of thermal noise and the TX leakage. For the sake of argument, we could assume the cumulative SI cancelation is precisely $111(dB)$, or the $SNR=0(dB)$, which will map into a Shannon capacity of $\frac{log_{10}(1+SNR\times 2)}{log_{10}(2)}=1.58(\frac{bits}{s\cdot Hz})$. In contrast, a half-duplex system with a $SNR=5(dB)$ is entitled to a Shannon capacity of $\frac{log_{10}(1+SNR)}{log_{10}(2)}=2.06(\frac{bits}{s\cdot Hz})$. With that being laid down, we could conclude that the SI performance is crucial to the FD system [26].

Additionally, the fact that the ownership of spectrum license is a multi-billion business makes the FD communication method, which potentially doubles the spectral utilization efficiency, an appealing research topic.

1.2 Self-Interference (SI) Cancellation

From the last section, the importance of the SI cancelation performance was described. There has been a great amount of literature that tries to improve the depth of cancelation. Prior SI cancelation strategies can be categorized into three distinct areas: antenna cancelation, analog cancelation, and digital cancelation.

The basic concept for antenna cancelation is that two or more antennas are used to transmit the TX signal. A set of antennas (TX) transmit the signal, and their leakages (SI) add up out-of-phase on the receiving antenna (RX), thus reducing the coupling of the self-interference to the receiver channel, while the desired received signal capture by another set of antenna (ANT) adds in phase on the RX, thus preserving the desired received signal information [3], [9], [19]. The

main drawbacks of multiple antenna cancellation technique are 1) the relatively large size of the design due to the multi-antenna and 2) the sensitivity of the cancellation precision subject to the environmental variance.

The digital SI cancellation mainly focuses on the base-band frequency in the digital domain performed by signal processing techniques and ADC circuitry. In a nutshell, the SI signal is modeled as a function of the transmitted and received signal, later used to calculate the residual SI signal and substrate from the raw received signal. The limitations of the digital cancellation are that it can only remove the SI within the dynamic range of the ADC and cannot alleviate the noise of the transmitter, because the algorithm of the digital model is unable to predict the amplitude of the noise. unpredictability [11], [12].

In contrast to the digital SI cancellation, noise cancellation is feasible for the analog SI cancellation approach. The condensed notion of the analog cancellation is to copy the interfering signals from the critical nodes on the transmitting chain, cast them to the corresponding nodes on the receiving chain via some tunable or auxiliary analog circuit blocks. As a result, the system can cancel the noise or SI around the carrier frequency with a wide bandwidth. A promising cancellation depth of $+70(dB)$ has been proven in [1]. Apart from the SI cancellation, the $IP3$ performance is another extreme aspect of a transceiver system, whose importance and influence will be elaborated on in the following chapter. Therefore, this thesis is laser-focused on exploring and designing a high IP3 FD system that is implemented by the analog cancellation technique.

1.3 Organization of Thesis

This paper will start by discussing, in chapter 2, the design of the electrical-balance duplexer or EBD, which is a principal component being the first barrier or filter to the SI signal. Other than

the cancellation performance, the EBD also needs to concern the insertion loss. Hence, this chapter will discuss what are the theoretical the minimum insertion loss of the EBD and how to improve it in practice. Chapter 3 discusses the design of an impedance matching network that serves as the passive analog SI cancellation. The traditional design of a matching network is implemented in the LC ladder topology, which requires multiple inductors which leads to solutions that occupy a large chip area. Therefore, in this work, we try to seek LC ladder topologizes and solutions which are more area-friendly. In addition, this chapter also contemplates how to boost up the linearity of the matching network, which is a fundamentally a critical issue in FD radios system.

In addition, in the last section of each chapter, a summary table regarding the simulation result of the topic of the chapter will be described. And finally, chapter 4 will thoroughly review the measurement results of the tape-out for the design and draw a conclusion on how to improve the EBD system in the future.

2 The Design of New EBD

2.1 Minimum IL from EBD

In the last decade, electrical-balance duplexer (EBD) have been investigated as a potential solution for full-duplex communication on monolithic ICs. It is a promising alternative solution to the surface-acoustic wave (SAW) filters, which are bulky, expensive, hard to implement on-chip, and losing the frequency selectivity beyond 2.5 (GHz). EBD debuts in a tunable RF front-end concept that seeks to address several crucial obstacles of 4G and 5G mobile systems. The below Figure 2.1 is to assist in understanding the functionality or role of EBD in a full-duplex system.

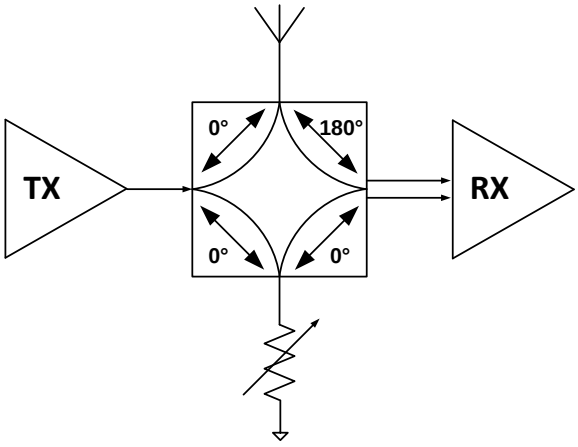

Figure 2.1 Typical FD System Implemented with EBD

An EBD is essentially a six-terminal hybrid transformer, with one terminal, which is omitted in the figure, tied to the ground, and other rest are connected to the 1) output of the transmitter, 2) & 3) the differential inputs of the receiver, 4) the antenna, and 5) an impedance balancing network (Z_{bal} network), respectively. In the receiving mode, the signal caught by the antenna from the free space is delivered from the EBD to the receiving chain. As you could observe, a partition of the receiving energy will be dissipated on both the TX chain and the

impedance matching network. On the other hand, in the transmitting mode, the power output generated from the power amplifier will split into two portions, one of which is emitted by the antenna, while the other is unwantedly exhausted on the Z_{bal} network. Due to the fact that the terminals are on the same silicon substrate, a fraction of the transmitting power will leak its way through to the receiving side and saturate the LNA. The objective of the EBD is to balance the impedance of the antenna with the Z_{bal} so as to cancel or minimize the undesired leakage or so-called self-interference from the TX side. In contrast to SAW filter, an EBD can be implemented on-chip and is able to propagate signals from the transmitter to antenna and from antenna to receiver over a broad bandwidth.

As our preceding discussion dictates that there are insertion losses (IL) on thetransmitting path and receiving path. What is most meaningful for us is to understand methods to minimize the insertion loss and the feasibility of implementing practical solutions. We will start by discussing methods to minimize the IL from TX to ANT.

2.1.1 TX IL

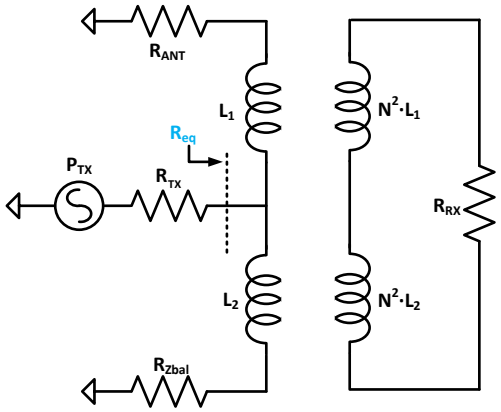

Figure 2.2 Schematic for Calculating TX IL

Referring to Figure 2.2, if we now assume the power of TX won't couple over to the RX side, yet entirely absorbed by the antenna and the Z_{bal} network, the IL of TX can be equal to 3dB under the circumstance of the impedance of the antenna being identical to that of Z_{bal} network. However, could we do better? As we know, once the impedances defined as R_{TX} and R_{eq} are equal (ignoring the imaginary part for now), the maximum power delivery condition is met. On the premise of maximum power transfer, if R_{ANT} is smaller than R_{Zbal}, the current through the antenna, $I_{ANT,}$ will become larger. Hence, the power delivered to the antenna, $P_{ANT} = I_{ANT}^2 \cdot R_{ANT}$, is bound to be larger. In other words, if R_{Zbal} keeps increasing, the IL of TX will also reduce. A new problem, however, emerges from this observation, specifically how to prevent the energy of TX from coupling to the RX side without R_{ANT} being equal to R_{Zbal}?

To address this question, R_{ANT} is assumed to be equal to $(1-\varepsilon)\cdot R$, while R_{Zbal} equal to $(1+\varepsilon)\cdot R$, where the ε represents the error between the impedances of two ports, with Figure 2.3 illustrated below.

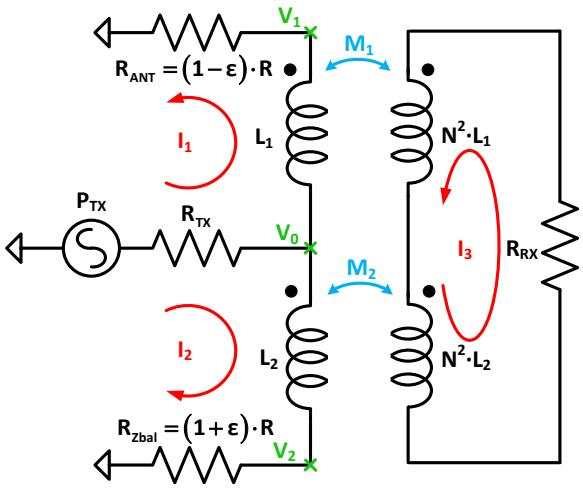

Figure 2.3 Circuit Analysis for TX IL

According to KVL, the voltage loop of the secondary winding can be established with:

$$(I_3 \cdot N^2 L_1 \cdot s + M_1 \cdot I_1 \cdot s) + (I_3 \cdot N^2 L_2 \cdot s - M_2 \cdot I_2 \cdot s) + I_3 \cdot R_{RX} = 0 \qquad \textit{Eq. 2.1}$$

, where: $M_1 = k \cdot \sqrt{L_1 \cdot N^2 L_1} = k \cdot L_1; M_2 = k \cdot L_2$

As our assumption is no TX leakage on the secondary side, I_3 is obviously equal to 0.

Therefore,

$$L_1 \cdot I_1 \cdot s = L_2 \cdot I_2 \cdot s \qquad \textit{Eq. 2.2}$$

In other words, the voltage drops on either side of the V_0 is identical:

$$V_1 = V_2 \qquad \textit{Eq. 2.3}$$

Besides,

$$I_1 = V_1/R_{ANT} = V_1/(1-\varepsilon) \cdot R$$
$$I_2 = V_2/R_{ANT} = V_2/(1+\varepsilon) \cdot R \qquad \textit{Eq. 2.4 a) \& b)}$$

Now plug Eq. 2.4 a) & b) into Eq. 2.2:

$$L_1 \cdot V_1/(1-\varepsilon) \cdot R = L_2 \cdot V_2/(1+\varepsilon) \cdot R \qquad \textit{Eq. 2.5}$$

The relationship between L_1 and L_2 can be eventually derived as:

$$\frac{L_1}{L_2} = \frac{(1-\varepsilon)}{(1+\varepsilon)} \qquad \textit{Eq. 2.6}$$

As long as Eq. 2.6 holds, there will not be any TX power coupling over to the RX side. Under such a circumstance, the maximum power delivered from TX to ANT is $P_{TX} \cdot \frac{1+\varepsilon}{2}$. According to this finding, $TX\ IL$ can be reduced by increasing the error term, ε.

2.1.2 RX IL

The next question waiting to resolve is to determine the value of R_{TX} by the maximum power transfer theorem. If the impacts of L_1 and L_2 are ignored for now, R_{TX} is simply equal to the parallel of R_{ANT} and R_{Zbal}.

$$R_{TX} = \frac{(1-\varepsilon^2)}{2} \cdot R \qquad \textit{Eq. 2.7}$$

The last element in the EBD system needed to be determined is R_{RX}.

To identify the value of R_{RX}, we could observe the system again from the perspective of energy conservation.

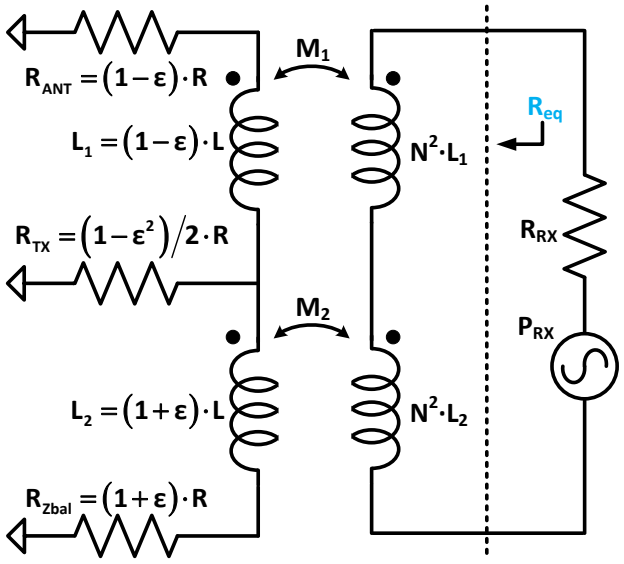

Figure 2.4 Schematic for Calculating RX IL

In Figure 2.4, we re-examine the EBD system with P_{RX} as the input signal.

As the discussion above describes, if the power is injected from the TX port, no energy shall leak to the RX side with the proper ratio of L_1 to L_2. Since this is a reciprocal system, a conclusion can be drawn that if the signal is imported from the RX side, no power can be delivered across R_{TX}

(TX port). In other words, all signals will land on R_{ANT} and R_{Zbal} and be rearranged according to their impedance ratio. Namely, the more energy that can be propagated to the primary side of the EBD, the more energy will be delivered to R_{ANT}. Here again, fulfilling the maximum power transferring condition is the gateway to the lowest $RX\ IL$. Since the total resistance on the primary side can be treated as a series resistor of R_{ANT} and R_{Zbal}, R_{eq} can be calculated by utilizing the reflecting impedance of the transformer.

$$R_{eq} = (\ R_{Zbal} + \ R_{ANT}) \cdot N^4 = 2 \cdot R \cdot N^4 \qquad \textit{Eq. 2.8}$$

With the matching between $R_{RX} = R_{eq}$, the maximum power propagation from ANT to RX is $P_{ANT} \cdot \frac{1-\varepsilon}{2}$ decreasing as the error term, ε, largens, which is an opposite result from $TX\ IL$. Based on this observation, we realize there is a trade-off between $TX\ IL$ and $RX\ IL$.

Thus far, all the parameters in a basic EBD system under ideal conditions have been determined. The complete schematic is illustrated below:

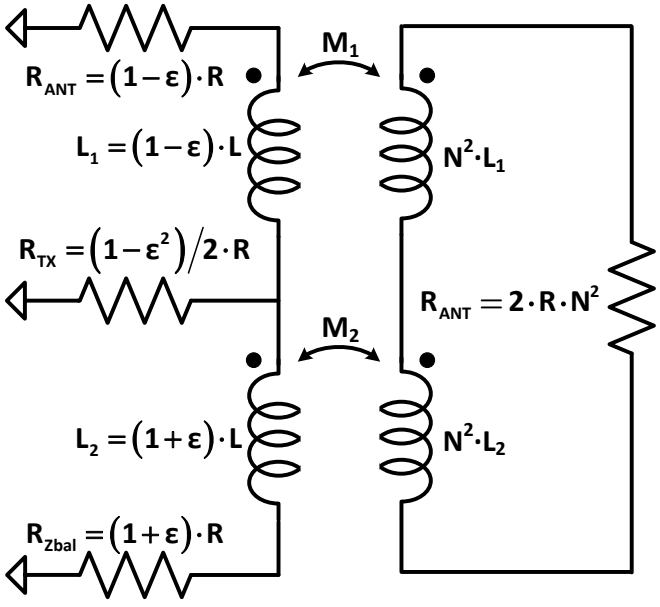

Figure 2.5 Complete Design for EBD Loading Arrangement

2.1.3 Re-examine IL under Practical Condition

The discussion above was deducted by assuming an ideal EBD that doesn't include the non-ideal effects, such as the inductance of the EBD itself, the coupling capacitances between primary and secondary windings, and the quality factor of the EBD. In this section, these non-idealities will gradually be taken into account.

2.1.3.1 Inductance of EBD

In the last section, the self-inductance of the EBD is ignored by which the $TX\ IL$ can theoretically be as low as 3dB. In practice, it is quite straightforward that a smaller inductance from the EBD will have less of an impact on the $TX\ IL$ as, in an extreme case, a zero self-inductance could lead to the ideal value of 3dB $TX\ IL$. However, in contrast to $TX\ IL$, a large inductance of EBD is in favor of the $RX\ IL$, and the reason is the following. To simplify the question, let us just examine the reflecting impedance of a 3-terminals transformer shown in Figure 2.6.

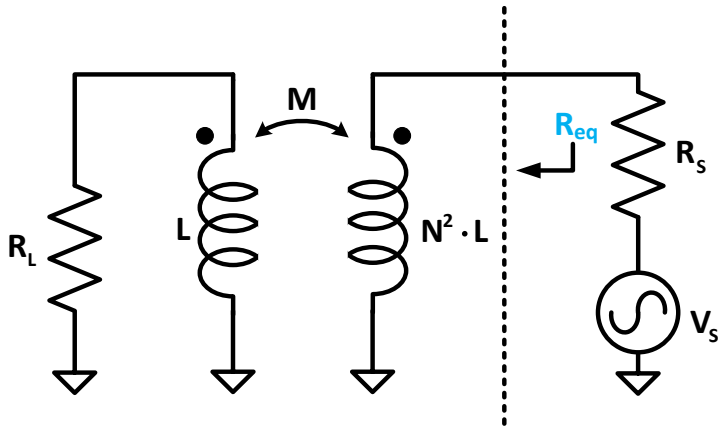

Figure 2.6 Considering the Value of Inductance

It is well-known that the reflecting resistance of an ideal transformer, $R_{eq} = N^2 \cdot R_L$, where the definition of an ideal transformer implies the inductances of the windings are

significantly larger than R_L and R_S. The upcoming question is to find an estimate of the transformer's inductance that could result in a reflecting impedance close to the ideal value. To investigate this problem quantitatively, we are going to solve the mathematical equation of R_{eq} with the circuit illustrated in Figure 2.7, plug into some reasonable numbers, and analyze the results.

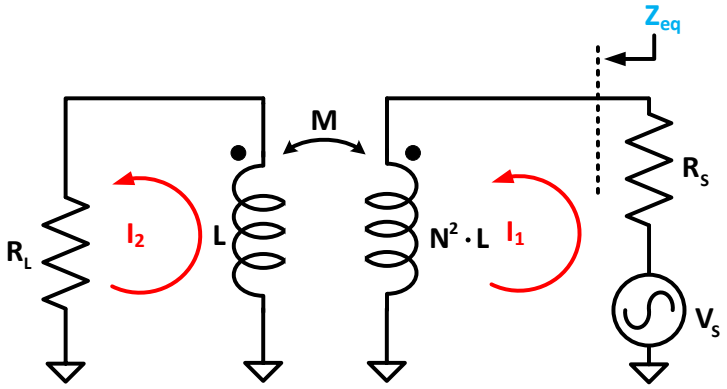

Figure 2.7 Circuit Analysis after Considering Self Inductance

According to KVL,

$$V_S = I_1 \cdot R_S + I_1 \cdot N^2 L \cdot s - M \cdot I_2 \cdot s \qquad \textit{Eq. 2.9}$$

$$0 = I_2 \cdot L \cdot s - M \cdot I_1 \cdot s + I_2 \cdot R_L \qquad \textit{Eq. 2.10}$$

Where, assuming $K = 1$ for simplification, $M = K \cdot \sqrt{L \cdot N^2 L} = N \cdot L$.

Therefore,

$$Z_{eq} = \frac{R_L \cdot (R_S + L_1 \cdot s) + s \cdot [L_2 \cdot (R_S + L_1 \cdot s) - M^2 \cdot s]}{R_L + L_2 \cdot s} - R_S \qquad \textit{Eq. 2.11}$$

Furthermore,

$$Real(Z_{eq}) = \frac{\omega^2 \cdot R_L \cdot M^2}{R_L^2 + \omega^2 \cdot L_2^2} \qquad \textit{Eq. 2.12}$$

$$Imag(Z_{eq}) = j \cdot \frac{\omega \cdot [R_L^2 \cdot R_S + \omega^2 \cdot (M^2 - L_1 \cdot L_2)]}{R_L^2 + \omega^2 \cdot L_2^2} \qquad \textit{Eq. 2.13}$$

Here, we assume: $R_L = R_S = 50(\Omega), \omega = 2\pi \cdot 2.4G\left(\frac{rad}{s}\right), L_1 = 2 \cdot L_2, M = \sqrt{2} \cdot L_2$

Given the above values, we could plot the ideal reflecting impedance, $N^2 \cdot R_L$ and the actual impedance implemented with an on-chip spiral transformer on the same plot in a semi-logarithmic scale and examine the difference between the two. As Figure 2.8 shows, it is evident that the real part of the impedance seen by the EBD (blue curve) is gradually ascending to the ideal value (100Ω) as the inductance becomes insufferably bulk.

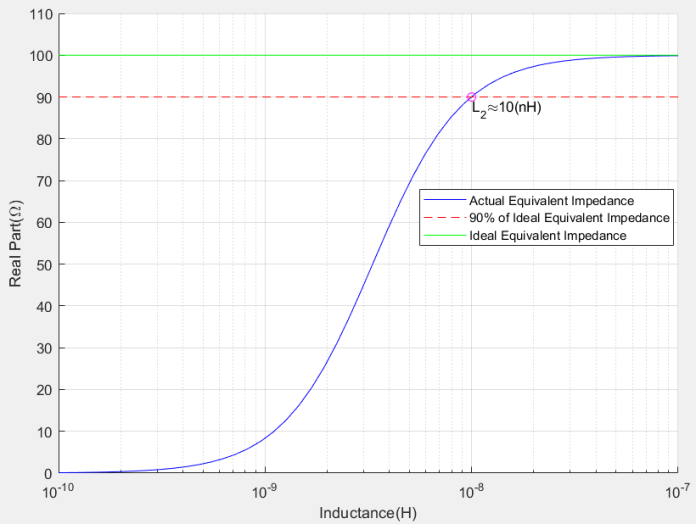

Figure 2.8 Curve of Reflected Resistance with respect to Inductance

Even though we merely try to approach 90% of the ideal value, an approximately $10(nH)$ L_2 is needed, namely, $20(nH)$ for L_1. In addition, to an EBD, the primary winding is twice of the L_1, and secondary winding is twice of the L_2. The main drawbacks of these bulky inductances are

twofold, one of which is the undeniably enormous die area. Practically speaking, the most apparent issue of implementing such a large EBD on a monolithic chip is the area occupied by such a passive device and the associated cost penalty when fabricating the chip for commercial applications. In addition, there is another slightly subtle issue. The parasitic capacitances between a large transformer and the substrate will result in a low self-resonate frequency (SRF). In that case, the Q-factor and the inductance become inherently unstable and are hard to anticipate in applications. These drawbacks imply that to fulfill the impedance matching condition at the $EBD - RX$ interface, a compensation technique is desired. The typical approach to resolve this issue uses a parallel capacitance to resonate out the inductance at the interface. The rationale behind this approach is that, with the parallel capacitor resonating the imaginary part, the voltage across the EBD is equal to $Q^2 \cdot V$, a much larger value than V_{R_S} such that the effect of R_S is negligible.

2.1.3.2 Coupling Parasitic Capacitors

An EBD or on-chip spiral transformer is a physical component that consists of two tightly interlocked inductors. Compared to a discrete transformer that uses a solid metal core with nearly infinite magnetic permeability, the on-chip spiral transformer utilizes its very silicon substrate, which is lossy and embraces a relatively low permeability, as the shared magnetic flux tunnel for the primary and secondary windings. The mutual coupling coefficient, K, of the transformer is defined by how strong the magnetic flux can be coupled from one inductor to another. The lower the K is, the more power leakage is introduced by the transformer. Intuitively speaking, in order to minimize the TX and $ANT\ IL$, we should try to preserve the maximum energy throughout the propagation or remain a high K. However, the pitfall of implementing an EBD with a high coupling coefficient, K, is the significant distributed parasitic capacitances embedding between

the two windings. Figure 2.9a illustrates the distributed parasitic capacitors that are larger as K increases. For further discussion, we use three equivalent capacitors to manifest the overall impact of the distributed capacitors, as shown in Figure 2.9b.

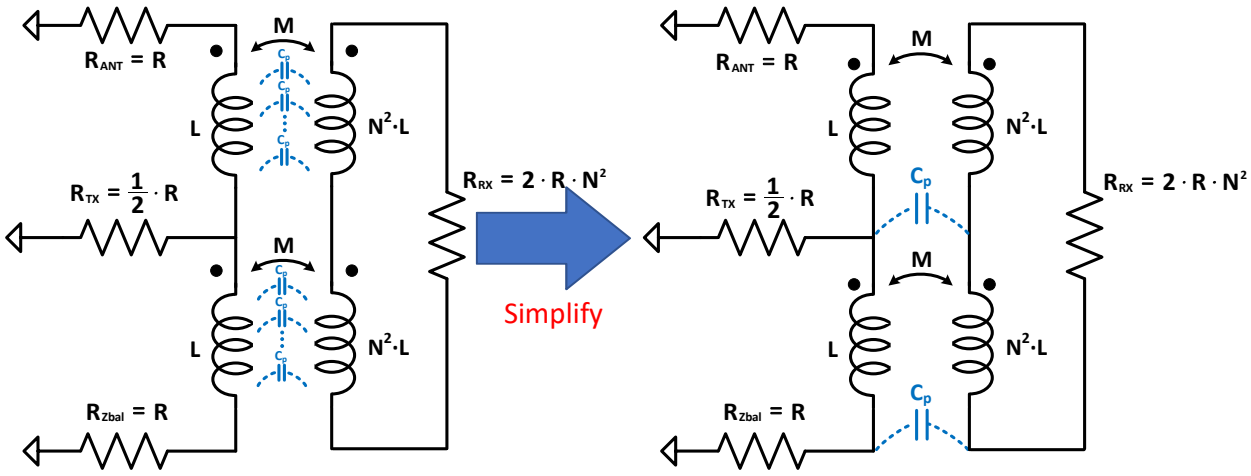

Figure 2.9a(Left) EBD with Distributed Caps, b(Right) EBD with a Single Equivalent Cap

These additional parasitic capacitors force us to re-consider the EBD matching condition. Yet, the values of these parasitic capacitors are merely an approximation on the order of fFs which has limited impact on the entire matching network.

2.1.3.3 Quality Factor

The quality factor (Q-factor) can be titled as the most important property of a transformer or an EBD, as the Q-factor is the dominant influence in determining the amount of the insertion loss. In other words, the Q-factor is a measure of energy loss or dissipation. A higher number of Q results in a lower loss. Therefore, one of our main targets for designing monolithic transformers is to improve or maximize the Q-factor. To do so, we will first examine the stem of the energy loss in a coil of a transformer. The formula of the Q-factor for an on-chip inductor is described as:

$$Q = \frac{\omega \cdot L}{R} \times substrate\ loss \times self - resonance\ factor \qquad \textit{Eq. 2.14}$$

Where the L and R represent the inductance and resistance of the coil, respectively. For a discrete inductor, the first term is sufficient to define the Q-factor. However, due to the introduction of the silicon substrate, the on-chip inductor will include two more terms. The term, $substrate\ loss$ means the remarkable loss implanted by the lossy substrate. The diagrams below elaborate the equivalent circuit of an on-chip symmetric center-tapped inductor, which is also used for the simulations by Cadence. Figure 2.10a is a compact circuit diagram that includes the mutual inductors and resistors between every two ports to emulate the impacts of eddy current losses and skin effect. In Cadence, the converted model, as illustrated in Figure 2.10b, is utilized for the simulations.

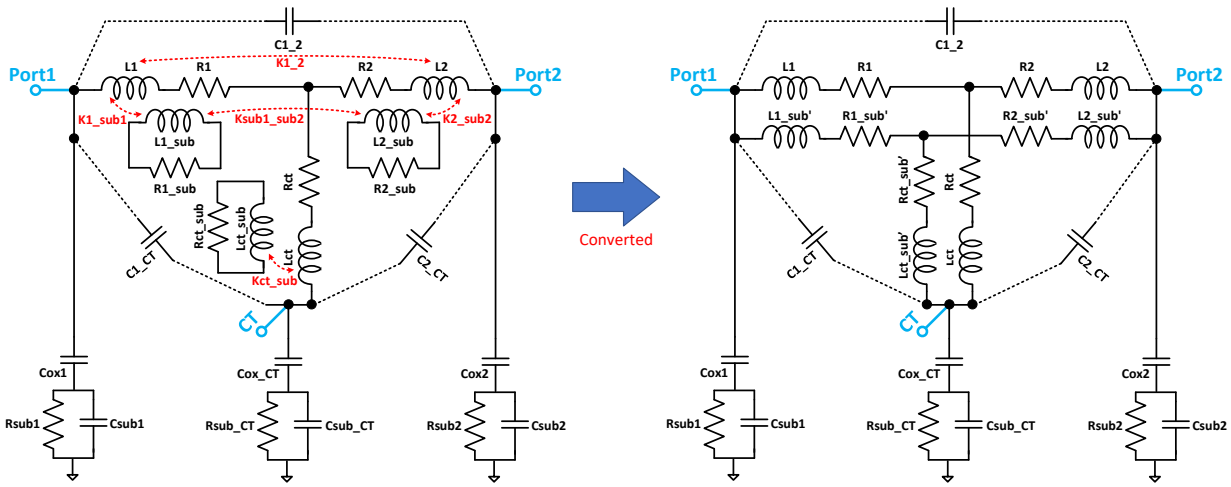

Figure 2.10a(Left) Model of EBD, b(Right) Simplified Model of EBD

In this analytical model, the C_{1_2}, $C_{1_{CT}}$, and $C_{2_{CT}}$ represent the coupling capacitances between each pair of ports, which are usually negligible as the distance from one port to another is remarkably large in general. R_1 and R_2 are the ohm resistor stemmed from the metal tracks. They obey a relatively simple relationship between the width and the length of the inductor yet

are the dominant factor in determining the Q-factor. In order to minimize these resistors or maximize the Q-factor, modern technology often uses the ultra-thick layer as the primary routing layer for the transformer. As a comparison, the choice for the connecting layer in a transformer is controversial in which layer to use, and in practice, the PDK provided by the chip manufacturer indeed includes some choices of inductors for different connecting layers. The thickness of the aluminum layer (AP layer) is almost twice that of the M_Z layer (the layer below the ultra-thick layer), which implies, with the same width, the conducting area in the AP layer for the current is almost double. Nevertheless, the conductivity of the M_Z is substantially higher than that of the AP layer. Thus, with the exact dimensions, the Mz layer eventually ends up with a lower resistance than the AP layer. Hence, M_Z layer usually is preferable in wiring the inductor or transformer. The inductor or transformer can use the AP layer for the connecting layer, however, does have one advantage, whose reason is following. As the position of AP layer are higher than the M_Z layer, its associated parasitic substrate capacitance will end up considerably smaller. Therefore, some experienced engineers take advantage of this property to boost up the Q-factor. The practice is to add an AP layer right on top of the entire transformer so as to achieve the lowest resistance. As only modest parasitic capacitance embedding on the AP layer, we can obtain a valuable elevation on the Q-factor by trading with a decent amount of degradation on the SRF. Last but not least, the set of components that consists of the oxide capacitance, C_{ox}, the substrate capacitance, C_{sub}, and the eddy resistance, C_{sub}, plays a pivotal role in both Q-factor and SRF. When the operating frequency is relatively low, the C_{ox} will block most of the power from leaking into the substrate. However, since the reactance within this frequency range is still modest, the quality factor will not be maximized. The Q-factor will exhibit a maximum value at a high frequency where the reactance of the inductor increases to a fairly large amount while the injected

power essentially passes through the path of $L_1(L_2)$. Owing to this nature of the loss mechanism, the inductors (transformers) in mmWave and microwave designs typically have a much larger Q-factor than in RF circuits.

2.1.3.4 *Calculation of IL under Practical Conditions*

After examining the practical factors that affect the IL, we could take them into account and solve for the IL once more. Before we dive into deriving the equation of $TX\ IL$, we must make two adjustments to EBD system. One is to use an equivalent resistance, R_Q, that represents the loss mechanism of all sorts. The other is, when the antenna inputs the signal, to change the RX output to a common mode for the consideration of common-mode loss, where the input impedance of the LNA is not fully differential. Therefore, we will use the Figure 2.11 below to re-examine the $TX\ IL$ and $RX\ IL$.

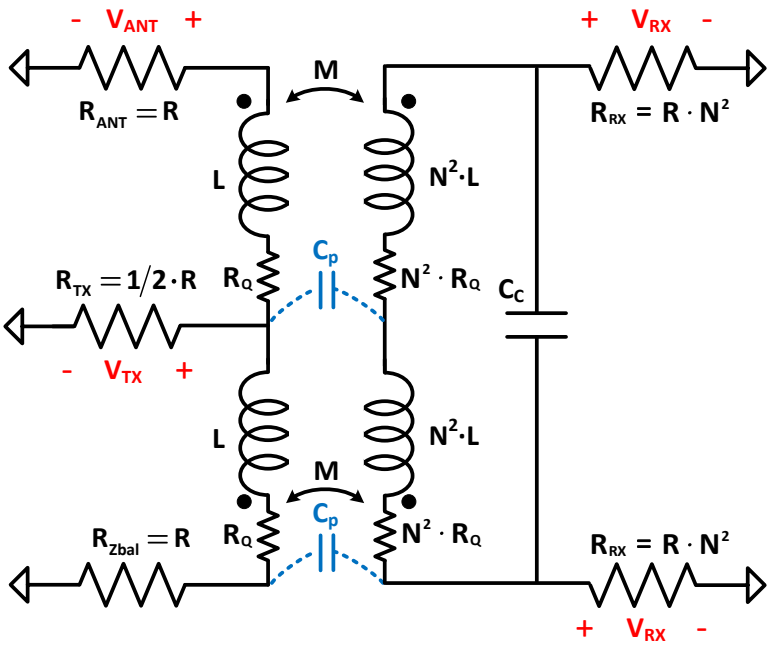

Figure 2.11 Re-examining TX & RX IL

Since now, with all the non-ideal elements considered in the analysis, it is complicated for us to calculate the IL by means of energy conservation. Here, we adopt the approach of a 2-port network as an alternative to solve the IL.

To a 2-port network, the typical way to solve for the IL or insertion loss, Sij is utilizing the matching theory. Suppose we may take a look at the diagram below. In that case, an insertion loss is calculated by placing a matched load on the terminal and solved for the incident voltage and reflecting voltage ratio.

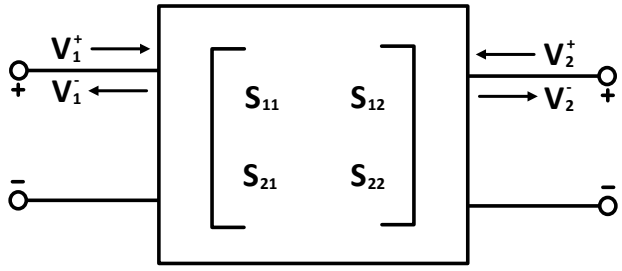

Figure 2.12 Two-port S Network Diagram

$$\begin{cases} V_1^- = S_{11} \cdot V_1^+ + S_{12} \cdot V_2^+ \\ V_2^- = S_{21} \cdot V_1^+ + S_{22} \cdot V_2^+ \end{cases}$$

Error! No sequence specified. ***Eq. 2.15 a & b***

Namely, the S parameter matrix can be calculated as:

$$S_{11} = \left.\frac{V_1^-}{V_1^+}\right|_{V_2^+=0} \quad \textbf{Eq. 2.16a} \qquad S_{12} = \left.\frac{V_1^-}{V_2^+}\right|_{V_1^+=0} \quad \textbf{Eq. 2.16b}$$

$$S_{21} = \left.\frac{V_2^-}{V_1^+}\right|_{V_2^+=0} \quad \textbf{Eq. 2.16c} \qquad S_{22} = \left.\frac{V_2^-}{V_2^+}\right|_{V_1^+=0} \quad \textbf{Eq. 2.16d}$$

By setting the terminals with a matched load, V_1^+ or V_2^+ can be vanished on the right-hand side so that all the scattering parameters can be found. However, this approach will necessitate a tremendous amount of work as we have to first find out the input impedance seen at the terminal,

then solving for the reflecting coefficient, and then calculate the voltage potential on the other terminal in the case of transmission loss. To reduce the unnecessary workload, a better approach will be to solve for the Y parameters first and convert the results to S parameters later. A general Y parameter 2-port network is illustrated below.

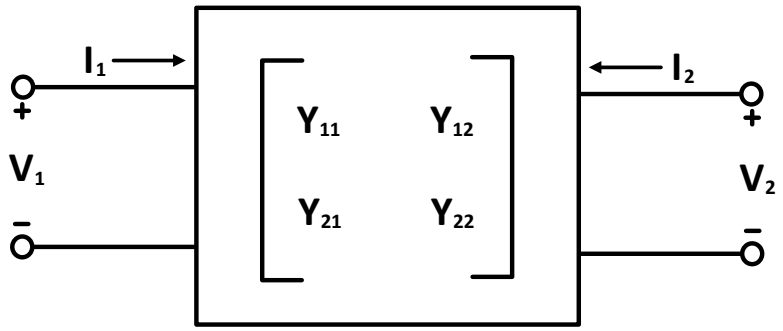

Figure 2.13 Two-port S Network Diagram

$$\begin{cases} I_1 = Y_{11} \cdot V_1 + Y_{12} \cdot V_2 \\ I_2 = Y_{21} \cdot V_1 + Y_{22} \cdot V_2 \end{cases} \qquad \textit{Eq. 2.17 a \& b}$$

The beauty of the utilization of Y parameter is that it does not require us to match the input impedance seen by the terminal, while the Y matrix can be solved by simply shorting one side of the terminal at a time. More specifically, Y matrix is equal to:

$$Y_{11} = \frac{I_1}{V_1}\bigg|_{V_2=0} \qquad \textit{Eq. 2.18a} \qquad Y_{12} = \frac{I_1}{V_2}\bigg|_{V_1=0} \qquad \textit{Eq. 2.18b}$$

$$Y_{21} = \frac{I_2}{V_1}\bigg|_{V_2=0} \qquad \textit{Eq. 2.18c} \qquad Y_{22} = \frac{I_2}{V_2}\bigg|_{V_1=0} \qquad \textit{Eq. 2.18d}$$

Consequently, the S parameters can be found as:

$$S_{11} = \frac{(1 - Y_{11}Z_{S1}^*)(1 + Y_{11}Z_{S2}) + Y_{12}Y_{21}Z_{S1}^*Z_{S2}}{(1 + Y_{11}Z_{S1})(1 + Y_{22}Z_{S2}) - Y_{12}Y_{21}Z_{S1}Z_{S2}} \qquad \textit{Eq. 2.19a}$$

$$S_{21} = \frac{-2Y_{21}\sqrt{R_{S1} \cdot R_{S2}}}{(1 + Y_{11}Z_{S1})(1 + Y_{22}Z_{S2}) - Y_{12}Y_{21}Z_{S1}Z_{S2}} \qquad \textit{Eq. 2.19b}$$

$$S_{12} = \frac{-2Y_{12}\sqrt{R_{S1} \cdot R_{S2}}}{(1 + Y_{11}Z_{S1})(1 + Y_{22}Z_{S2}) - Y_{12}Y_{21}Z_{S1}Z_{S2}}$$

Eq. 2.19c

$$S_{22} = \frac{(1 + Y_{11}Z_{S1})(1 - Y_{22}Z_{S2}^*) + Y_{12}Y_{21}Z_{S1}Z_{S2}^*}{(1 + Y_{11}Z_{S1})(1 + Y_{22}Z_{S2}) - Y_{12}Y_{21}Z_{S1}Z_{S2}}$$

Eq. 2.19d

After developing the approach of finding the S parameters, we could now turn to look at the 2-port network blocks for $S_{TX-to-ANT}$ and $S_{ANT-to-RX}$.

As we re-visit Figure 2.11, we observe there is a strong symmetry regarding this circuit when V_{TX} as the input signal. If we might take a closer look at this circuit, we could soon realize the V_{RX1} and V_{RX2} of the LNA are identical. Therefore, the LNA on the secondary side of the EBD will not consume any energy from the TX signal. With this being said, the LNA impedance and the compensation capacitor, C_C, can be dropped off from the analysis. After that, by flipping over the circuit vertically, the $TX\ IL$ could be reduced to the circuit shown Figure 2.14.

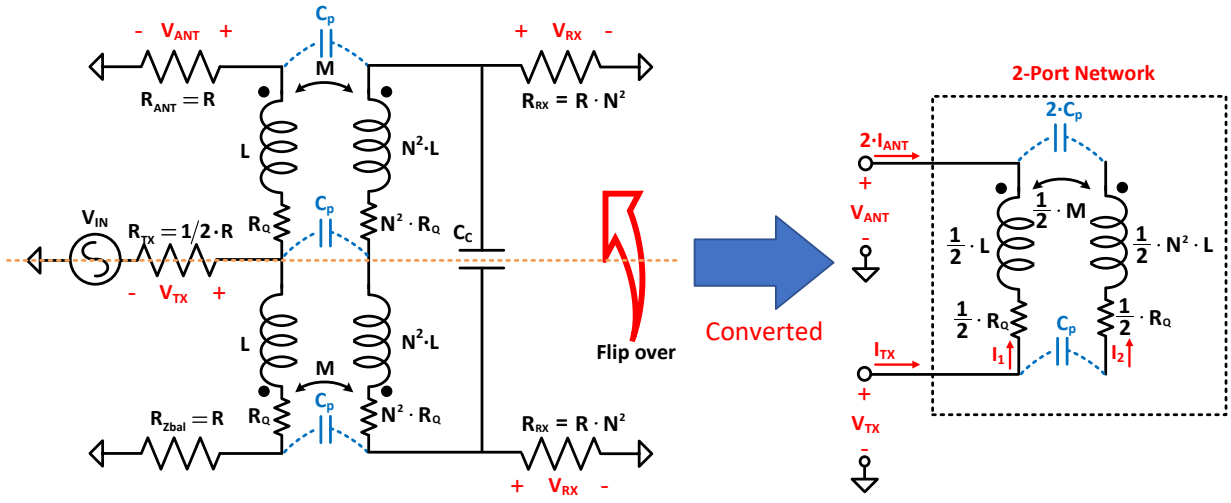

Figure 2.14 Circuit Transformation for TX IL

Two cardinal observations should be mentioned here. One of them is that the incident current from the antenna becomes twice as big as it should be. This discrepancy stems from the fact that now the antenna's impedance becomes one-half after it parallels with the impedance of the Z balanced network. As a result, the insertion loss, $S_{TX-to-ANT}$, will become less since the

antenna terminal occupies the energy of the $Zbal$ network as well. We could make use of this from the standpoint of energy conservation to correct the discrepancy in the solution. Since the powers or energies delivered on the antenna and the $Zbal$ network are identical, we could time a factor of ½ in voltage or $-3dB$ in power to the calculated $S_{TX-to-ANT}$ to resume the correct value. The other crucial observation or coincidence is that the impedance of the TX port is equal to that of the antenna now. Thus, combining the fact that it is a reciprocal system, we can acknowledge that $Y_{11} = Y_{12} = Y_{21} = Y_{22} = Y_{ij}$. Finally, the Y_{ij} can be calculated using the following equations developed from the KVL:

$$-V_{TX} + \frac{1}{2} \cdot M \cdot s \cdot I_2 + \frac{1}{2}\left(L \cdot s + R_q\right) \cdot I_1 = 0 \qquad \textit{Eq. 2.20}$$

$$-V_{TX} + \frac{1}{2} \cdot M \cdot s \cdot I_1 + \frac{I_2}{s \cdot C_p} + \frac{I_2}{s \cdot 2C_p} + \frac{1}{2}N^2\left(L \cdot s + R_q\right) \cdot I_2 = 0 \qquad \textit{Eq. 2.21}$$

$$I_{TX} = I_1 + I_2 \qquad \textit{Eq. 2.22}$$

Therefore,

$$Y_{ij} = \frac{I_{TX}}{V_{TX}} = \frac{6 + C_p s\left[(2 + 2N^2)R_q + L(2 - 4kN + 2N^2)s\right]}{C_p s N^2 R_q^2 + R_q\left(3 + 2C_p L N^2 s\right) + Ls\left[3 + C_p(1 - k^2)LN^2 s\right]} \qquad \textit{Eq. 2.23}$$

Furthermore,

$$S_{TX-to-ANT} = \frac{1}{\sqrt{2}} \frac{100\left\{-3 - C_p s\left[\left(1+N^2\right)R_q + Ls\left(1 - 2kN + N^2\right)\right]\right\}}{300 + C_p s N^2 R_q^2 + Ls\left\{3 + C_p s\left[1 - 2kN + N^2\right] - LsN^2\left(k^2 - 1\right)\right\} + R_q\left\{3 + 2C_p s\left[50 + N^2(50 + Ls)\right]\right\}}$$

Eq. 2.24

Next, we will calculate the insertion loss from the antenna to the receiver side, $S_{ANT-to-RX}$, with the aid of Figure 2.15.

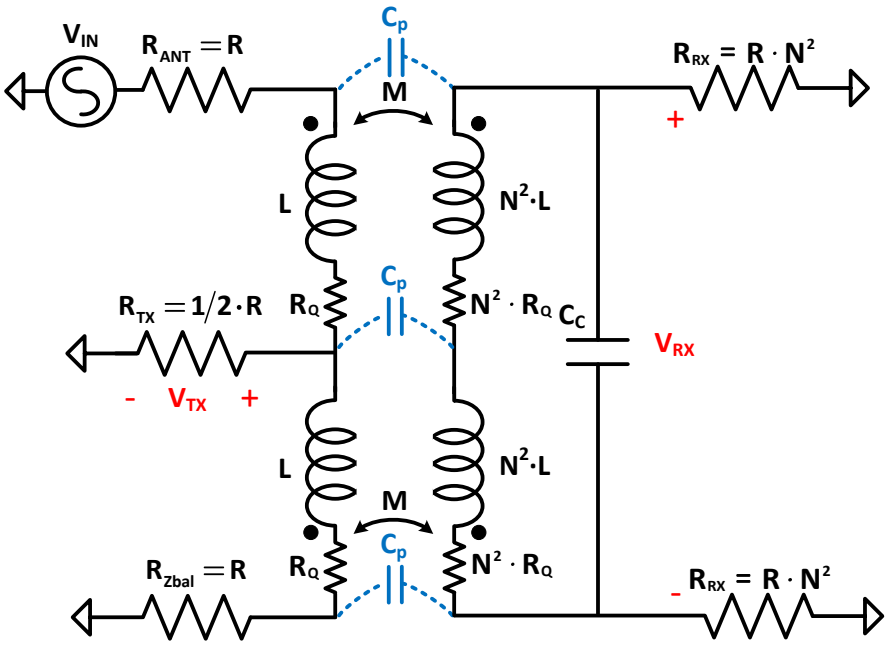

Figure 2.15 Schematic for Calculating RX IL

When first looking at this circuit, the analysis does not seem to be trivial. Due to the position of the input signal, the EBD system loses its symmetry. We could, however, add another opposite input signal to recover the symmetry and perform a much simpler analysis. This idea is demonstrated in Figure 2.16 below.

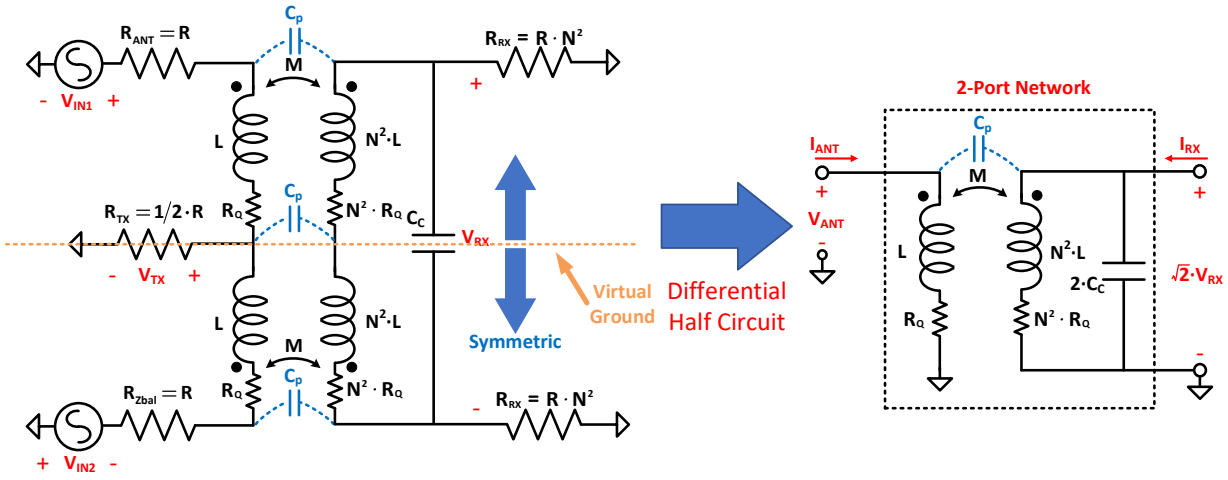

Figure 2.16 Circuit Transformation for RX IL

With the additional signal source, the delivered power on the RX side is doubled, and the peak voltage amplitude is $\sqrt{2}$ times of the original signal. From the fully differential standpoint, this circuit becomes perfectly symmetric, and the central line appears as a virtual ground. Thus, we could employ the technique of differential circuit analysis and transform the current circuit to the one on its right. Stated differently, the peak voltage on the RX side has become $\sqrt{2}$ times larger. Therefore, at the end of the calculation, the $RX\ IL$, $S_{ANT-to-RX}$, needs to be subtracted $3dB$ to regain the correct number.

The Y parameters technique is used here as well. First, by shorting the RX port to the ground, the 2-port network is reduced to Figure 2.17, where $2 \cdot C_C$ is vanished since the voltage difference between its two terminals is identical. Moreover, C_p can be treated as an additional component paralleled with the 2-port network and excluded from the circuit analysis and added back afterward.

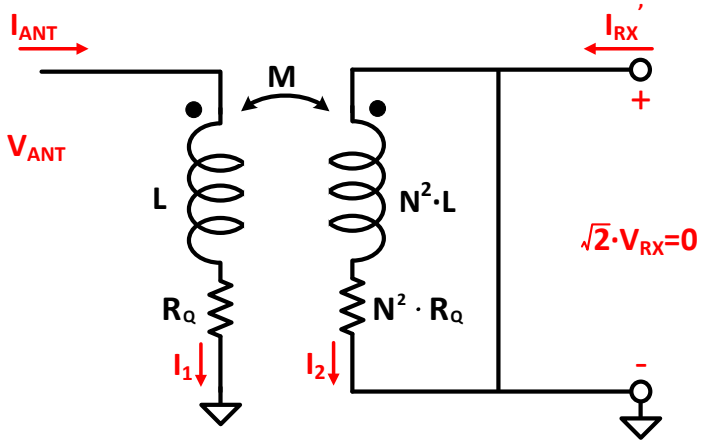

Figure 2.17 Simplified Circuit for RX IL

According to KVL and VCL, we could write down:

$$-V_{ANT} + M \cdot s \cdot I_2 + \left(L \cdot s + R_q\right) \cdot I_1 = 0 \qquad \textit{Eq. 2.25}$$

$$M \cdot s \cdot I_1 + N^2(L \cdot s + R_q) \cdot I_2 = 0 \qquad \textit{Eq. 2.26}$$

If we assume ANT port is port 1, RX port is port 2,

$$Y_{11} = sC_p + \frac{I_1}{V_{ANT}} = sC_p + \frac{Ls + R_q}{[R_q + (1-k)Ls][R_q + (1+k)Ls]} \qquad \textit{Eq. 2.27a}$$

$$Y_{21} = sC_p - \frac{I_2}{V_{ANT}} = sC_p + \frac{kLs}{N[R_q + (1-k)Ls][R_q + (1+k)Ls]} \qquad \textit{Eq. 2.27b}$$

Similarly, by shorting the ANT port to the ground, we could obtain the rest of the Y parameters:

$$Y_{22} = s(2C_C + C_p) + \frac{Ls + R_q}{N^2[R_q + (1-k)Ls][R_q + (1+k)Ls]} \qquad \textit{Eq. 2.27c}$$

$$Y_{12} = Y_{21} = sC_p + \frac{kLs}{N[R_q + (1-k)Ls][R_q + (1+k)Ls]} \qquad \textit{Eq. 2.27d}$$

As a result,

$$S_{ANT-to-RX} = \frac{1}{\sqrt{2}} \frac{-100Ns\left(C_p + \frac{kL}{NAB}\right)}{-2500\left(C_pNs + \frac{kLs}{AB}\right)^2 + \left(1 + 50C_ps + \frac{25}{A} + \frac{25}{B}\right)\left[1 + 50(2C_C + C_p)N^2s + \frac{50(R_q + Ls)}{AB}\right]}$$

Eq. 2.28

Where, $A = R_q + (1-k)Ls$ and $B = R_q + (1+k)Ls$

So far, we have derived the equations for $S_{TX-to-ANT}$ and $S_{ANT-to-RX}$ by taking the non-ideal factors into account.

2.2 Post-Layout Simulation and Specification

The EBD is first designed in Cadence, including the guard ring, and then transformed to $HFSS$ to obtain the SP file, which is then imported back to Cadence for further simulation. Figure 2.18 below is the entire layout of the EBD.

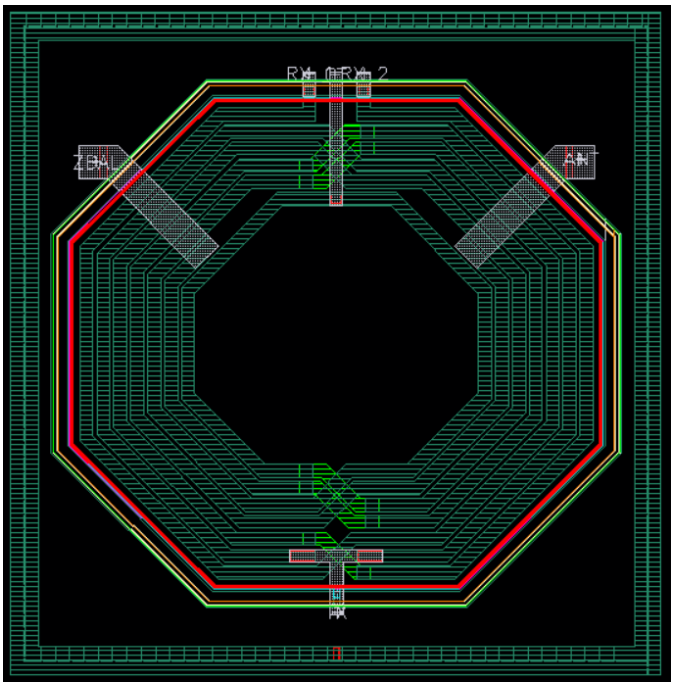

Figure 2.18 Layout of New Designed EBD

The total size of the EBD is roughly $500(um)$ by $500(um)$, or $0.25(mm^2)$ using a TSMC 45nm CMOS process. The primary routing of the EBD is constructed with the highest metal layer M10 to achieve the highest quality factor. Out of consideration for the self-interference, ports $Zbal$ and ANT are placed far apart from the TX port. The inductances of the EBD are shown below, where at $2.4(GHz)$, the primary winding is $3.5(nH)$, while the secondary $5.8(nH)$.

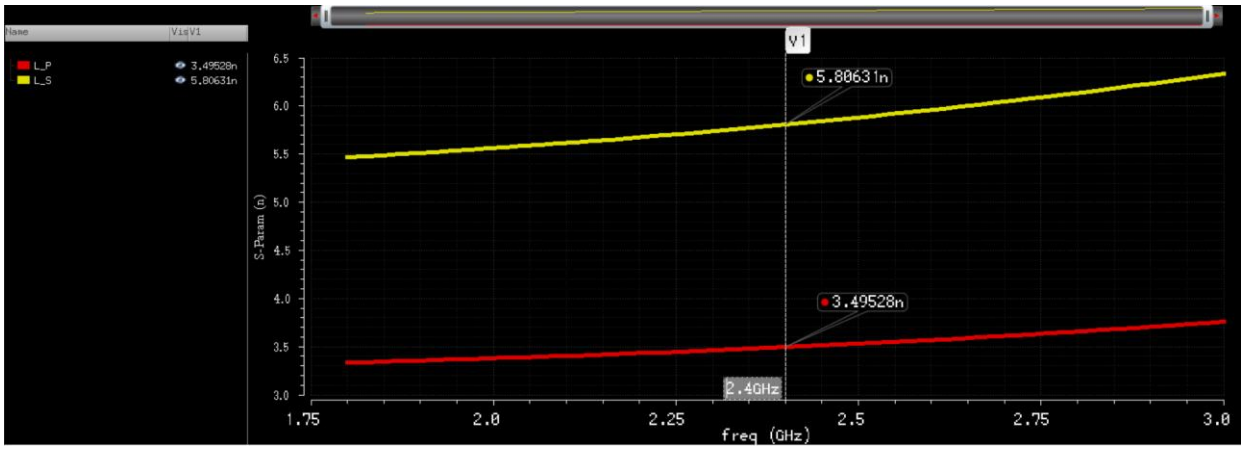

Figure 2.19 Inductances of the New Designed EBD

The simulation result of the quality factor is provided in Figure 2.20. The primary side has a Q-factor of 10.3, while the secondary side 13.6.

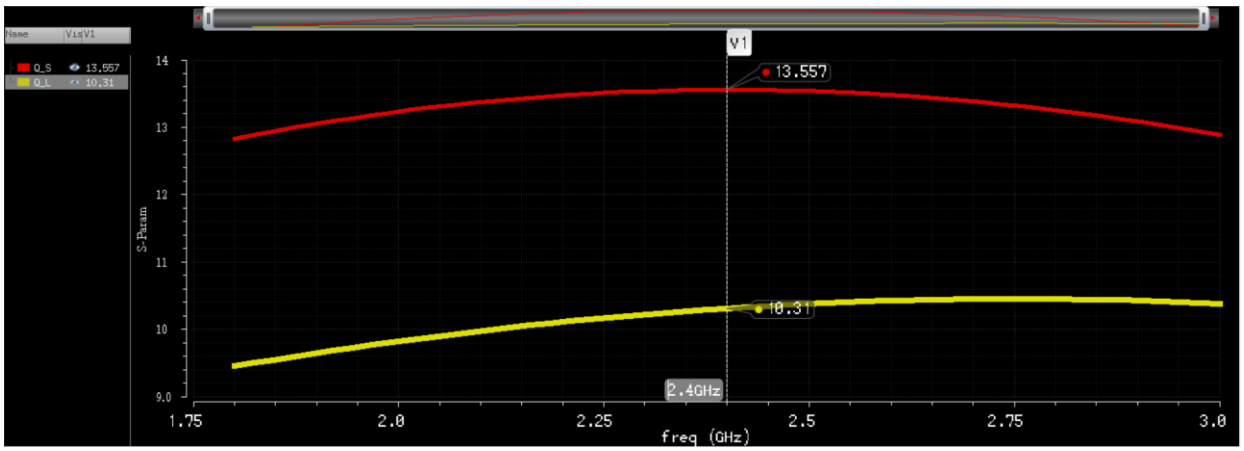

Figure 2.20 Q-factor of the New Designed EBD

The coupling coefficient that has around 0.82 is shown in Figure 2.21.

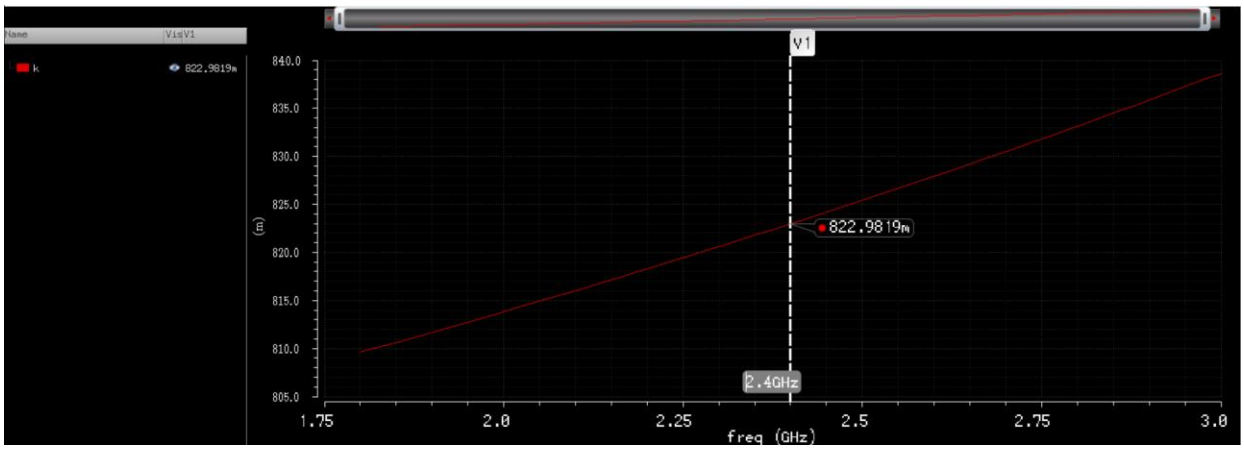

Figure 2.21 Coupling Coefficient of the New Designed EBD

3 Design of a High Linear Z_{bal} Network

In this chapter, we are going to contemplate the design of $Zbal$ network. First, we need to determine how high an IP3 value is required for a full-duplex system based on the WiFi-6 (802.11ax) standard. Second, we will start discussing the source of the nonlinearity in a full-duplex system. As a start, the passive device will be scrutinized in terms of voltage dependence. And then, we will fathom the issue of the discontinuity in the BSIM4 model followed by a schematic solution to overcome the obstacle where the harmonic balanced fails to simulate the $IP3$ number of the RF switch. Next, after we achieved simulating the $IP3$ of an RF switch, we shall officially start the design of the Z_{bal} network. The design process consists of the high-linearity RF switch and the impedance network topology. Lastly, the post-layout specifications and the simulation results of the tuning range will be presented.

3.1 The Necessity of High Linearity in FD Communication

In other communication systems, the $TX - RX$ feedthrough as crucial compared to a full-duplex radio. The reason that sets the FD system apart from the others is essentially the physical distance between the transmitting chain and the receiving chain. The leakage power from the TX side, SI, can contribute significant distortion in the receive signal path as the SI passes through any non-linearity in the signal path. Significant distortion is potentially created via the coupling from the shared substrate, which desensitizes the RX input. In order to better visualize the impact of the SI, we could use some data based on the Wi-Fi 802.11ax specifications and the transceiver scheme shown in Figure 3.1 to calculate the targeted $IP3$ number.

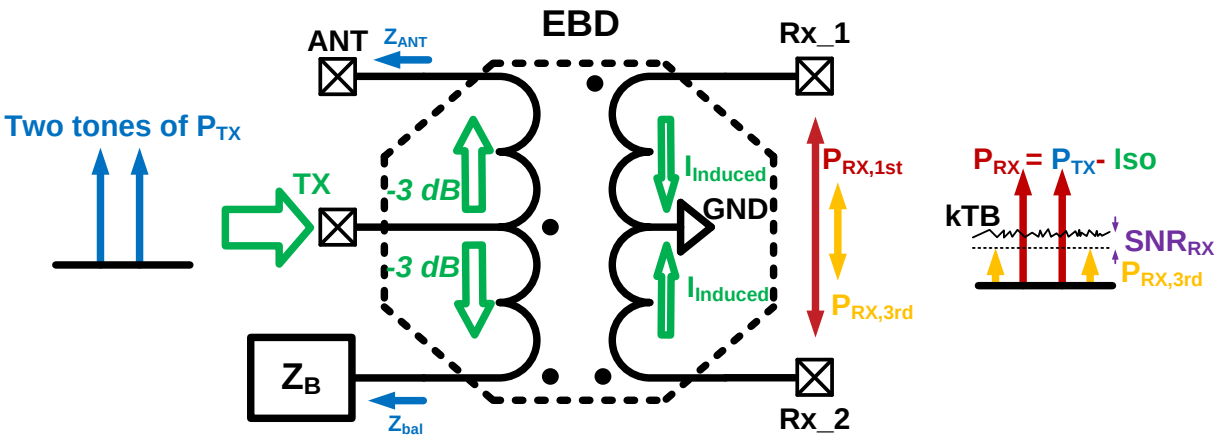

Figure 3.1. Consideration for FD EBD System

Specifications	Values
TX power (P_{TX})	20 dBm
TX-to-RX isolation (ISO)	40 dB
RX Sensitivity (P_{RX})	-80 dBm
RX SNR (SNR_{RX})	10 dB
Noise Floor (kTB) (assuming BW = 160MHz)	-92 dBm

Table 1. Specifications of FD EBD System

Given these specifications, the desired $IIP3$ referred to the TX output can be reckoned as:

$$\boldsymbol{IIP_3 = P_{TX} + \frac{[P_{RX} - (kTB - SNR_{RX})]}{2}} \qquad \textit{Eq. 3.1a}$$

$$\boldsymbol{= P_{TX} + \frac{[(P_{TX} - ISO) - (kTB - SNR_{RX})]}{2}} \qquad \textit{Eq. 3.1b}$$

$$\boldsymbol{= 61(dBm)}$$

In summary, under the circumstance where the $IIP3$ of the $Zbal$ is higher than $61(dBm)$, the resulted distortion from the TX path shall be lower than the noise floor by $10(dB)$.

3.2 Distortion from Passive Devices

It is well-known that active devices are inherently nonlinear, while passive components are very linear. The motivates researchers to explore circuit topologies for FD front-end systems which are realized using passive components as much as possible.

3.2.1 Junction Capacitors

We will start the discussion regarding the nonlinearity effect in passive devices with junction capacitors. As we know, the junction capacitors are strongly nonlinear and are prevalent in the $P-N$ junction or the depletion region.

In general, the formula for the junction capacitance is described as:

$$C_j = \frac{C_{jo} \cdot AS}{\left(1 - \frac{v_B}{\phi}\right)^{MJ}} + \frac{C_{jsw} \cdot PS}{\left(1 - \frac{v_B}{\phi}\right)^{MJSW}} \qquad \textit{Eq. 3.2}$$

There are essentially two takeaways in the above formula. First, the nonlinearity of the junction capacitor sources from the exponential term in the denominator. Second, the capacitance consists of the contacting area and perimeter. Hence, in order to minimize the junction capacitance, we could increase the distance between the P-type and N-type materials and reduce the depletion area and perimeter.

3.2.2 Resistors

Resistors have been treated as extremely linear devices for most applications. In reality, the biased voltage applied on the terminals of the on-chip resistor will significantly impact the resistance, and suggests that this class of resistors have a significant non-linearity. To quantify the resistor's nonlinearity and voltage dependence, we have developed a voltage coefficient on different types of resistors to determine their $IP3$ performance.

There are mainly four different types of resistors in TSMC 40nm mixed-signal PDK including the metal film resistor, diffusion resistor, N-well resistor, and polysilicon resistor. Each type has its own mechanism characterizing the voltage dependence. In nature, the metal resistor should have an unprecedented number of $IP3$ since it is basically a piece of a metal strip. The only element contributing a nonlinearity in a metal resistor is the associated capacitances between the metal and the substrate, which can be minimized by using the upper-level metal. In short, the metal film resistors have an exceptional $IP3$ performance. Nevertheless, the applications of metal film resistors are infinitesimally small due to the low inherent resistance. A straightforward example is that with a minimum width of $70(nm)$, a length of $10(\mu m)$ metal one resistor has a resistance of $32(\Omega)$. Not only is the length fairly long, but also it has an abnormal length-width ratio (≈ 150) which is a strongly function of the fabrication process. In short, we seldomly use metal film resistors.

Other types of resistors feature relatively high sheet resistance, given the same dimensions and are commonly employed in various circuitry. Nonetheless, they have their disadvantage as well. Nothing is ever free or without consequence, and the penalty here is voltage dependence or nonlinearity. As simple as it can be, the very source of the nonlinearity in an on-chip resistor is

the parasitic junction capacitors. Next, we will verify this argument via analyzing the physical structures of some of the resistors.

3.2.2.1 N-Diffusion Resistor

First, we will start off our discussion with the N-Diffusion resistor, whose main resistance is determined by the N^+ body. In the TSMC 40nm node, there is usually an option for the silicide layer. The silicon layer is a metal contact formed during the annealing or the etching process to lower the connection resistance. But unfortunately, the silicon layer will become a source of nonlinearity owing to the following reason. The component of silicon is, in fact, categorized as a type of metallic component. With that being said, the silicide layer (metal) will form a $Schottky$ diode with the N^+ body (semiconductor), which in our case is the N-well. If we could turn our attention to Figure 3.2 and Figure 3.3, the side views of the physical structures for the N-Well resistors with and without silicide layer are presented, from which we could easily observe the impact of a nonlinearity associated with the silicide.

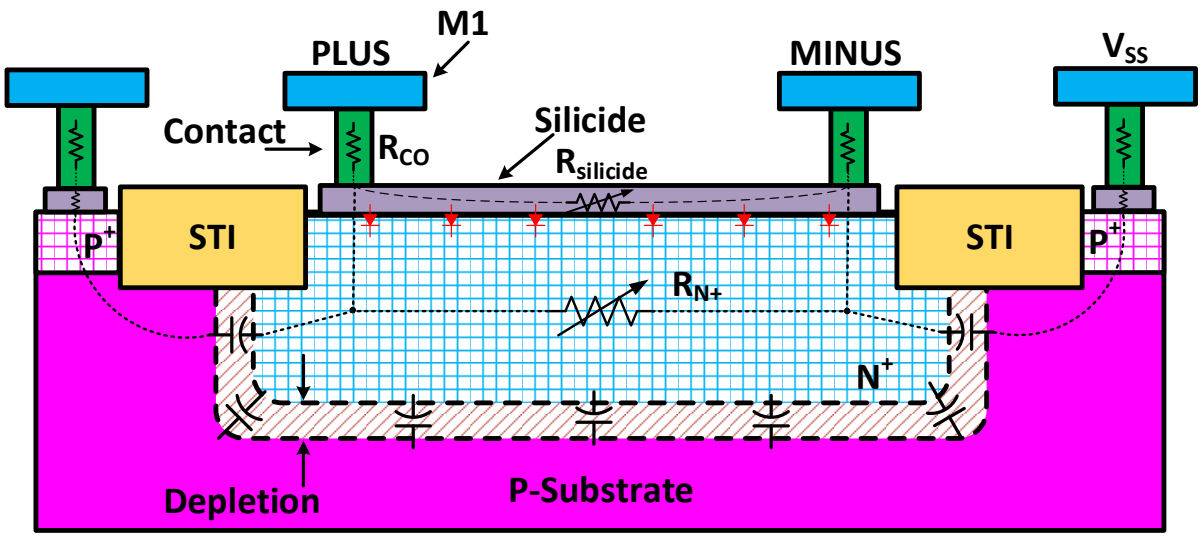

Figure 3.2 Side View of N-Diffusion Resistor with Silicide

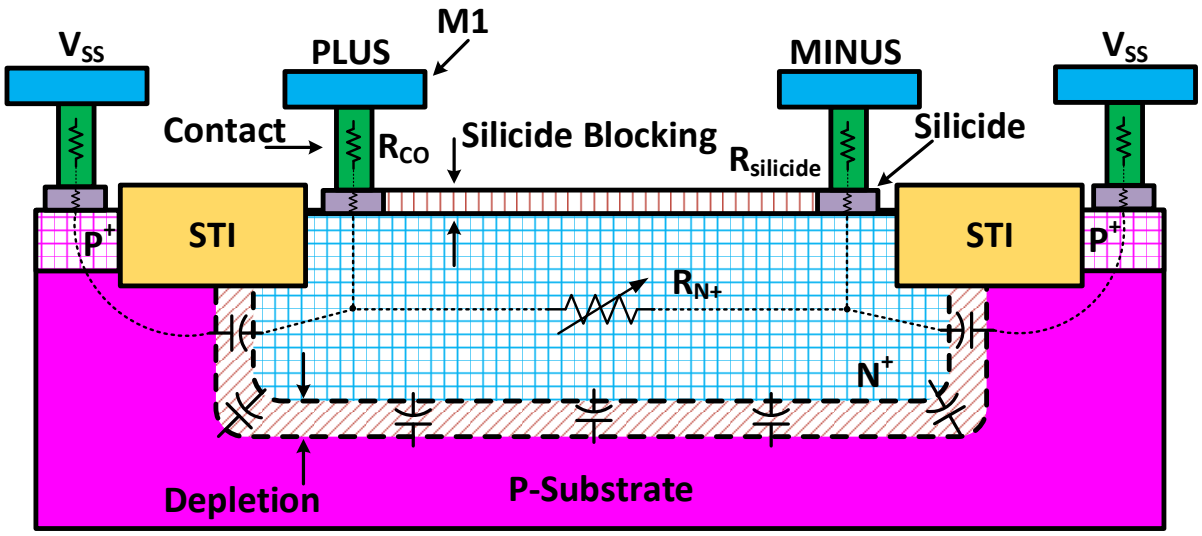

Figure 3.3 Side View of N-Diffusion Resistor without Silicide

In Figure 3.2 and Figure 3.3, we choose to use a voltage-dependent lump resistor to represent the distributed $Schottky$ diodes. On the one hand, the presence of the silicide-silicon interface will introduce first-order nonlinearity. On the other hand, the formation of PN junction from the N-well and the P-substrate will adversely add another degree of nonlinearity. As the major resistance of the N-diffusion resistors with silicide is comprised by the paralleled $R_{silicide}$ and R_{N+}, this type of on-chip resistor is not an attractive choice to achieve an acceptable linearity performance.

In fact, we could draw another conclusion from the above discussion that if a resistor without the silicide layer is replaced by a silicon blocking layer (insulator), the linearity will be improved in most cases.

3.2.2.2 *N-Well Resistor*

Secondly, let us look at the N-Well resistor, whose sheet resistance is usually the lowest other than the metal film resistor. The physical structure of the N-Well resistor without silicide is illustrated below (Figure 3.4).

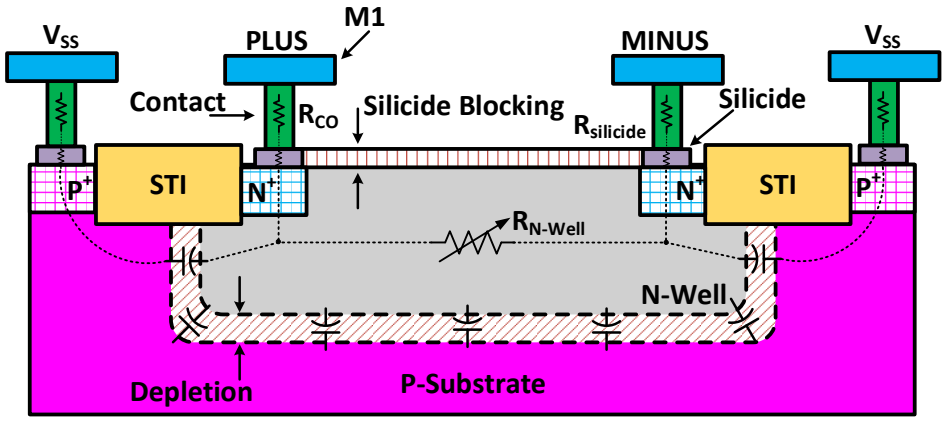

Figure 3.4 Side View of N-Well Resistor without Silicide

The main difference between the N-Well resistor and the N-Diffusion resistor is that the N-well-based device becomes the major conductive material. Because of the N-Well being less conductive than the N-Diffusion or N+ material, this type of resistor has a higher sheet resistance than the diffusion resistor. Owing to the low conductivity of the N-Well body, given the same amount of current, a higher percentage of current will flow through the PN junctions compared to the diffusion resistor. Consequently, more nonlinear distortion products are generated when current passes through the depletion layer. Therefore, the N-Well resistor has been proven infamously nonlinear.

3.2.2.3 *Polysilicon Resistor*

The polysilicon resistor finds ubiquitous use in IC designs due to its large sheet resistance and low-temperature coefficient. Typically, people do not cast attention on the linearity

of a polysilicon resistor as they assume it will be highly linear. However, while applying a voltage across the polysilicon resistor, the value of the resistance will vary, which is defined as the term of voltage coefficient. The resistance of a polysilicon resistor will have a positive voltage coefficient with a relatively high doping concentration while manifests a negative coefficient when the doping concentration is low. This pattern of voltage coefficients is related to the grain boundaries that contain the grain of the polysilicon [17]. When the polysilicon is doped lightly, most of the carriers will be trapped by the grain boundaries due to the grain boundary defect such that the lateral motion of the carriers will be hindered. Yet, when a voltage is applied across the resistor, the carriers can easily travel through the grain boundaries and obtain kinetic energy from the process. As a result, polysilicon with a lower amount of impurity-doped has a negative voltage coefficient on the resistance. In contrast, if the polysilicon is doped with a high impurity, the grain boundaries can capture more carriers and be entirely saturated. In this case, the leftover free carriers will collide with the confined carriers. The stronger the applied voltage is, the stronger this effect will be. Thus, the aftermath of this process is a negative voltage coefficient.

By examining the observations and discussion above, we could conclude that under no circumstance are the polysilicon resistors linear. Figure 3.5 below is the side view of a polysilicon resistor.

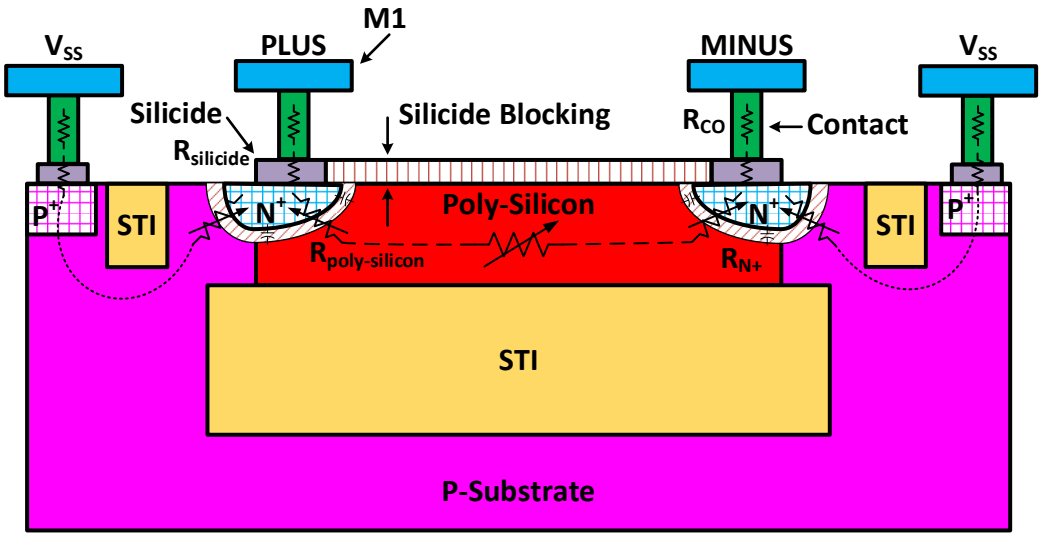

Figure 3.5 Side View of Polysilicon Resistor

From this figure, we could now locate that the majority of the carriers that interact with grain boundaries in the polysilicon come from the contact material. The nonlinearity of a polysilicon resistor is governed by the paralleled junction resistance and the voltage-dependent polysilicon resistance.

3.2.2.4 Summary Table of Resistors Linearity

With respect to the discussion at the end of the last section, we claimed that the nonlinearity in all types of on-chip resistors is somehow relevant to the junction capacitances. After contemplating the source of nonlinearity, we can validate this argument and propose a strategy to improve the nonlinearities for all resistors, which is to increase the sizes of the resistor.

The junction capacitance is determined by the total area of the contact interface between the two materials, one of which is always the main conductive material of the resistor. Therefore, as the resistor's width and length are increased, the percentage of the contact area that generates the nonlinear effect is decreased.

Here, we will provide a summary table of the voltage coefficients for different types of resistors with two different sizes from the PDK of TSMC 45nm.

V_R is swept from 0(V) to 0.9(V)		
Type of Resistor	Variety **(%/V)**	
	Original Size ($2\mu m\ X\ 10\mu m$)	Increased Size ($5\mu m\ X\ 25\mu m$)
N^+ Diffusion	6.443	1.956
P^+ Diffusion	6.627	1.513
N Well under OD	3.447	7.06
N Well under STI	0.9414	6.489
N polysilicon	0.9981	0.1691
P polysilicon	0.1399	0.0193

Table 2 Voltage Coefficients of On-Chip Resistors

3.3 Topologies for Broad Tuning Range

In this work, we utilize an impedance matching network (Z_{Bal} network) to match the antenna's impedance, to cancel out the large TX signal and prevent it from leaking into the RX side. In this section, we will discuss a new approach of the Z_{Bal} network to dramatically reduce the area while achieving an even wider tuning range.

3.3.1 Traditional Approach

The design of Z_{Bal} network traditionally adopts the LC ladder network [13, 14, 15], which use a $50(\Omega)$ resistor as the origin and stacks several tunable LC units from there. Figure 3.16 illustrates this idea, where the tunable inductor consists of a fixed value inductor paralleled with a tunable capacitor.

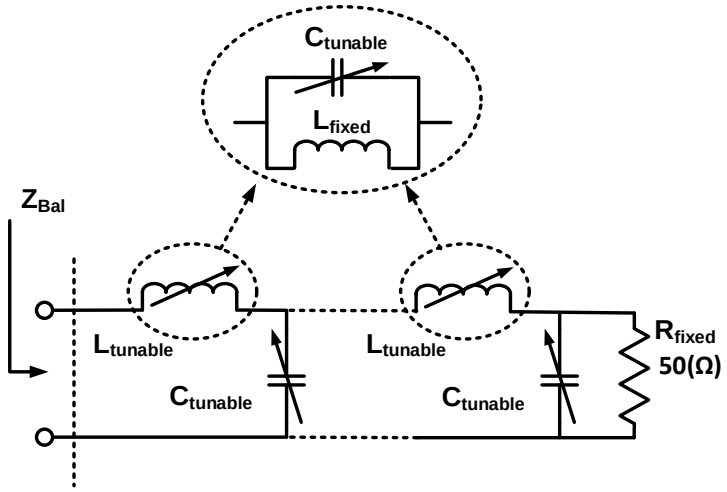

Figure 3.6 LC Ladder Matching Network

This approach has the advantage of a smooth group delay in terms of the frequency response and low sensitivity to non-exact component values. However, the synthesis procedure of this matching network is fairly complicated since the variation of the capacitance is hard to map into the Smith Chart due to the parasitic components. However, formal methods to obtain component values have been established dating back to the early twentieth century [20-23]. In other words, the tuning range is difficult to predict without a series of meticulous analyses and simulations. In some cases, adding more tunable LC units will not increase but reduce the tuning range. In addition, the required area for the tunable LC ladder is vast due to the utilization of the on-chip inductors.

3.3.2 Novel Admittance Approach

In this work, we adopt a new approach to realize the Z_{Bal} network that overcomes the general drawbacks in the LC ladder network. The very reason that an LC network occupies such a large die area for a particular tuning range, and is hard to synthesize, relates to the lack of direct control on the real part impedance. In a LC network, the change of a real part is realized by

sweeping along the imaginary part in a zigzag course on the Smith Chart. Namely, if we could discover a way to change the real part directly, the issues with the traditional matching networks might potentially be resolved.

Here, we propose a novel approach to realize a Z_{Bal} network from the perspective of admittance instead of impedance. The novelty of the admittance standpoint is the independence of the paralleled components. If we treat each tunable unit in the Z_{Bal} network as a two-port network. Then, by simply considering a case for two two-port networks, each of them has an impedance matrix described by:

$$Z1 = \begin{bmatrix} Z1_{11} & Z1_{12} \\ Z1_{21} & Z1_{22} \end{bmatrix} \text{ and } Z2 = \begin{bmatrix} Z2_{11} & Z2_{12} \\ Z2_{21} & Z2_{22} \end{bmatrix} \qquad \textit{Eq. 3.3}$$

When we place these two networks in series, the sum of the impedance matrix is equal to:

$$Z1 \cdot Z2 = \begin{bmatrix} Z1_{11} \cdot Z2_{11} + Z1_{12} \cdot Z2_{21} & Z1_{11} \cdot Z2_{12} + Z1_{12} \cdot Z2_{22} \\ Z1_{21} \cdot Z2_{11} + Z1_{22} \cdot Z2_{21} & Z1_{21} \cdot Z2_{12} + Z1_{22} \cdot Z2_{22} \end{bmatrix} \qquad \textit{Eq. 3.4}$$

As simple as it appears to be, the sum of the two-port networks cannot be easily predicted, which is why the design and the analysis of a LC network are complex. In contrast to an LC or impedance network, the sum of paralleled admittance networks can perform a simple summation:

$$Y1 + Y2 = \begin{bmatrix} Y1_{11} & Y1_{12} \\ Y1_{21} & Y1_{22} \end{bmatrix} + \begin{bmatrix} Y2_{11} & Y2_{12} \\ Y2_{21} & Y2_{22} \end{bmatrix} = \begin{bmatrix} Y1_{11} + Y2_{11} & Y1_{12} + Y2_{12} \\ Y1_{21} + Y2_{21} & Y1_{22} + Y2_{22} \end{bmatrix} \qquad \textit{Eq. 3.5}$$

The advantage of an admittance network is that they are independent of each other. Therefore, no matter how many tunable units are added to the design, the total admittance is well controlled and able to cover an extensive range with a minimal number of devices.

The next challenge to address when synthesizing the matching network is the design of the tunable units. In order to cover the entire desired tuning range, we essentially need three

components: a tunable capacitor, a tunable resistor, and a fixed value inductor. In terms of resolution, 6 bits of tuning capability is sufficient for our design. The complete schematic of the admittance Z_{Bal} network is illustrated below:

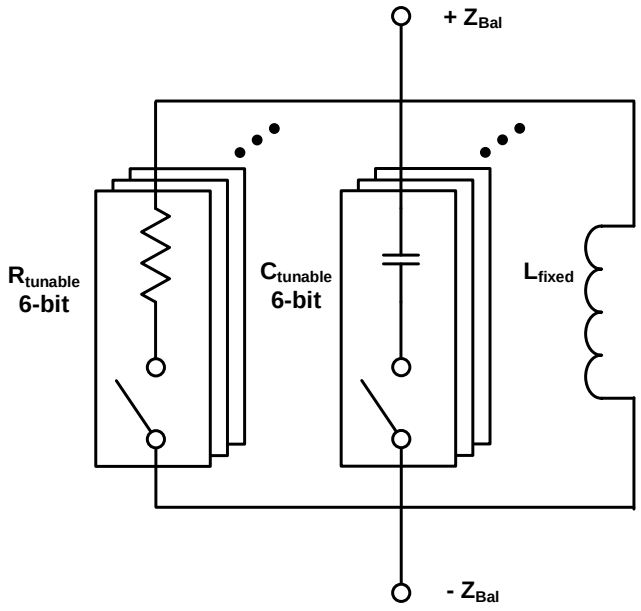

Figure 3.7 Admittance Approach of Matching Network

Both the tunable units of resistor and capacitor are made of a passive components in series with an RF switch. Ideally, when the RF switch has no feedthrough via the parasitic capacitance, zero on-resistance, and infinite off-resistance, the equivalent admittance of the network is just some resistors and capacitors in parallel with an inductor. Figure 3.8 illustrates the transition path on a Smith Chart from the original impedance Z_o, to the targeted impedance Z', with the two matching networks, the traditional impedance or LC approach (①) and the novel admittance approach (②).

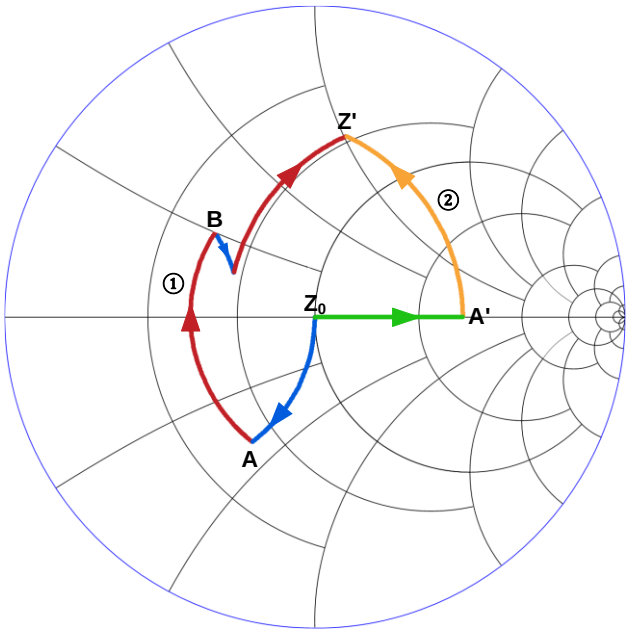

Figure 3.8 Impedance Track on Smith Chart

In path ①, it requires four steps (*increasing paralleled capacitance -> increasing serial inductance -> increasing paralleled capacitance -> increasing serial inductance*) changing the impedance from the origin to the targeted location. In contrast to the LC network, the admittance approach, path ②, merely takes two steps to reach the desired location (*decreasing paralleled resistor -> decreasing paralleled inductance*). From the above observation, it is unquestionable that the adaptive impedances can be clearly mapped on the Smith Chart with the admittance topology.

3.4 The Core of the Z_{bal} Network – High linearity Switches

In section 3.1, we have discussed the need for high linearity in full-duplex systems based on the $TX\ SI$, EBD cancellation, and noise floor specifications. Overall, the nonlinearity in the designed EBD system mainly stems from the MOS RF switch, since that is the only circuit block containing the active devices. In other words, in order to fulfill the extremely high $IP3$

requirement, we would to focus the design on realizing cutting-edge RF switches with excellent linearity.

3.4.1 Realization of High Linearity

The topology of the RF switch we use in this design is shown in Figure 3.9below.

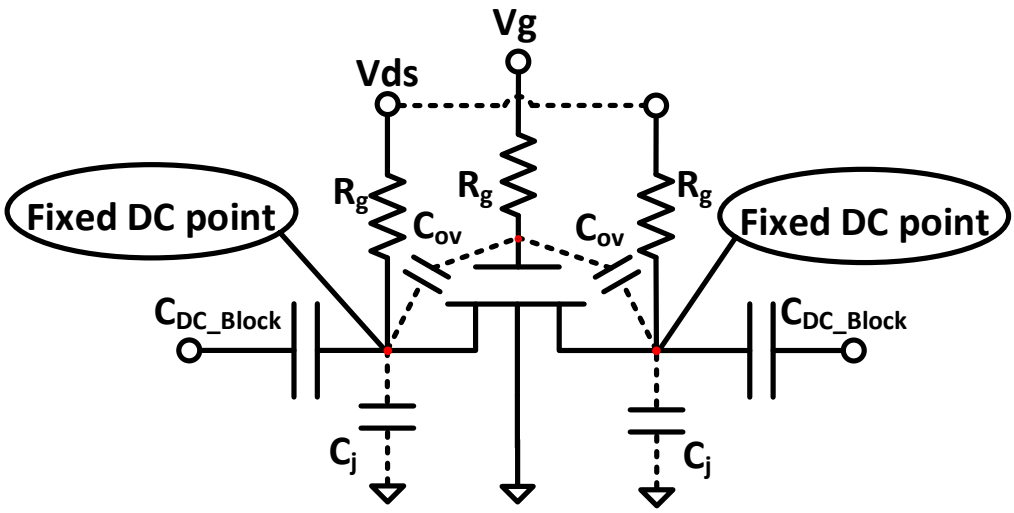

Figure 3.9. RF Switch Unit

The C_{ov} and C_j represent the gate-drain/source overlay capacitors and the bulk-drain/source junction capacitors, respectively. Two large values of DC blocking capacitors are placed on each side of the transistor to eliminate the intervene from the DC voltage. Therefore, the DC operating points of the three terminals of the transistor can be set to the desired value without interfering with the peripheral circuits. During the operation, a DC voltage of V_g will be set to V_{DD} (logic "1") and the V_{ds} set to 0 (logic "0") to turn on the switch, and vice versa to turn off the switch. These control voltages are applied at the gate and drain/source terminals through the large value resistors, R_g, whose value should be designed such that R_g is large enough to float the terminals within the frequency of interest. Normally, in order to maintain a decent switching time, R_g cannot be so large that the $R_g \cdot C_g$ (the sum capacitance at the gate) constant is

compatible with the operating time period. In our design, however, the switching time is not a major consideration, so that R_g can be set to a substantial value as long as it has a reasonable area.

With this setup for the RF switch, it can result in an outstanding $IP3$ performance. Assuming the nonlinearity mainly stems from the junction capacitances and the on-resistance, the smaller the voltage swing applied on these components, the better the $IP3$ number will be. Using Figure 3.10 below, we see that adding a passive component with a high impedance prior to the switch, has the effect of significantly reducing the voltage drop across the switch.

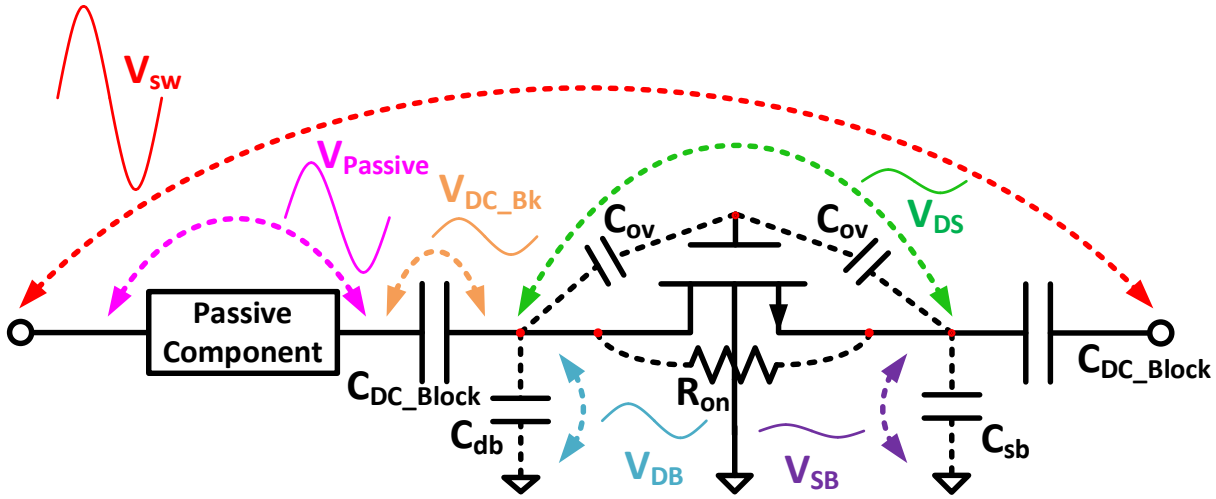

Figure 3.10. Conceptual schematic illustrating the voltage swing on an RF Switch

By observing Figure 3.10, we notice:

$$V_{SW} = V_{Passive} + V_{DC_{Bk}} + V_{DB} \qquad \textit{Eq. 3.6}$$

$$V_{DB} = V_{DS} + V_{SB} \qquad \textit{Eq. 3.7}$$

As long as we keep $V_{Passive}$ as the dominant proportion of the total voltage drop, V_{SW}, V_{DB} will be minimized, and the nonlinearity originated from the C_{db} will bear less weight on the overall switch linearity. Similarly, we soon realize the voltage across the on-resistance, R_{on}, is just a fraction of the V_{DB}, which implies the nonlinearity from the resistor can also be alleviated.

In our particular design, the series passive component should be a polysilicon resistor which inherently has a favorable $IP3$ performance based on the preceding discussion.

3.4.2 Breakdown Voltage

After examining the prototype of the RF switch, we shall now consider the feasibility of the implementation. Design-wise, there are two more practical concerns regarding the performances of the switches other than the linearity, one of which is the breakdown voltage of the transistor. In the latest TSMC 45nm PDK, the customary voltage for the core device is $0.9(V)$, and that of IO devices is $1.8(V)$, whereas the PA transmitting power in the Wi-Fi 802.ax standard is as high as $20(dBm)$, or $3.16(V)$ in voltage assuming a 50Ω environment. Therefore, if we merely use one transistor as the core element of the RF switch, neither the core nor IO devices could sustain the stress of such high voltages. The simple, yet effective, strategy for overcoming this large signal swing is to stack two transistors in series, as shown in Figure 3.11.

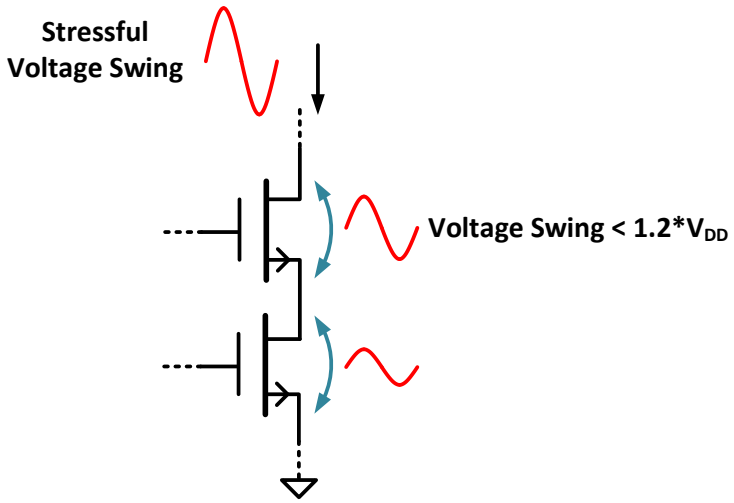

Figure 3.11. Stacked MOSFET Unit

Due to the virtues of the two-port network, the voltage arrangements of the two transistors are not even. Apparently, the transistor on the top is bound to suffer a higher voltage swing, which

we have to guarantee to be 1.2 times less than the supply voltage, V_{DD}, according to the TSMC 45nm user manual. Since we only stacked two transistors, an IO device that can tolerate $1.8(V) \times 1.2 = 2.16(V)$ is a better choice in our work.

3.4.3 Transistor Sizing

Another practical issue related to the design of an RF switch deals with the amount of feedthrough and the isolation. The RF switch is serves as a means of smooth transition between ON and OFF states. The switch is supposed to have a negligible resistance when it is turned on while it should behave as an open circuit when it is turned off. When the aspect ratio of a MOSFET is large, the on resistance is bound to be small, yet with a large parasitic capacitor through that signal leaks. On the other hand, a transistor with a small width over length ratio implying a large on-resistance has the properties of a lower parasitic capacitance that would perform better in terms of isolation when it is turned off.

This is a typical trade-off problem in designing the RF switches. With regard to this work, the approach we use to determine the switch sizes is as follows. As our final goal is to design an impedance matching network, its tuning range is of utmost importance to the EBD system. If we assume the targeted VSWR is 2:1, the tunable real part shall be mapped to $25(\Omega)$ to $100(\Omega)$ with respect to $50(\Omega)$. The south side of the impedance ($25(\Omega)$) is dominated by the serial resistors in the tunable resistance unit under the circumstance where the switches are on. Therefore, by combining the fact of 6-bits of resolution, the on-resistance of switches can be calculated as:

$$30(\Omega) = \frac{R_{on} + R_{serial}}{64} \Rightarrow R_{on} = 30(\Omega) \times 64 - R_{serial} \qquad \textit{Eq. 3.8}$$

In the equation above, we use resistance of $30(\Omega)$ instead of $25(\Omega)$. The reason for setting this $5(\Omega)$ margin is that the tunable capacitance unit, in fact, has feedthrough to some extent.

After we determine what on-resistance we should use, the next question is how to minimize the parasitic capacitance. As we have discussed before, the junction capacitors, C_{db} and C_{sb}, are another source of nonlinearity. Presumably, if these capacitors could become so small that they can be treated as an open circuit, the linearity will improve significantly. Here, we utilize two approaches to minimize these junction capacitances.

In our work, the RF switches are implemented with the six terminal transistors, or so-called triple-well devices, whose cross-sectional view and equivalent schematic (*the two RF floating resistors are not shown in the cross-sectional view) are shown (Figure 3.12) below:

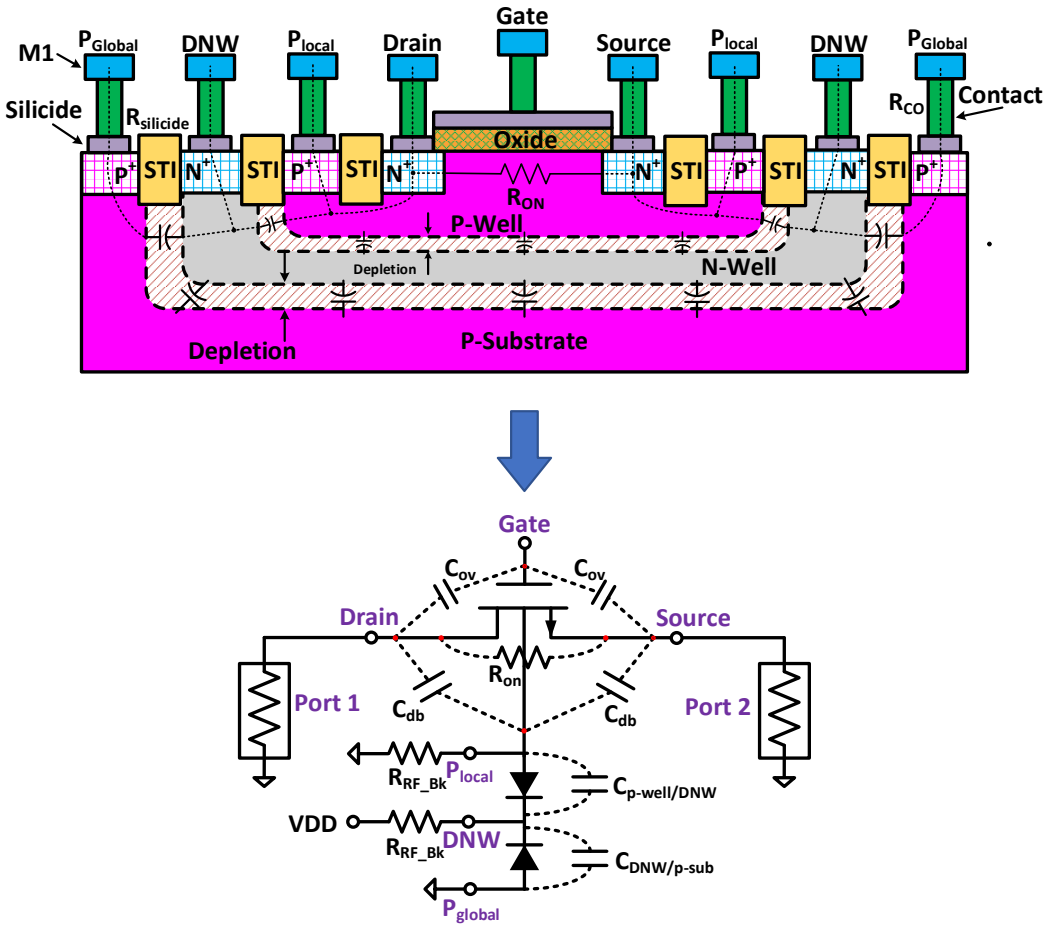

Figure 3.12. Bootstrapping a triple-well device

The terminals P_{local} and DNW are biased through the RF blocking resistors, $R_{RF_{Bk}}$ in order to set to the proper DC voltage levels, ground, and VDD, respectively. Under this setup, the parasitic diodes are forced to operate in the reverse biased region, and the buried deep N-well segregates the local body of the NMOS from the global substrate. Therefore, the nonlinear effect from the junction capacitor, C_{db}, is highly suppressed.

Another strategy we adopt to improve the linearity is customizing our own triple-well device instead of directly using the existing device provided by the PDK. By realizing a custom device, the PN junction areas can be reduced, thus minimizing the P-well/DNW and DNW/P-sub area and minimize the impact of the junction capacitors. Figure 3.13 below is the $IP3$ simulation result when the switches are turned off for the normal 4-terminal NMOS, 6-terminal NMOS, and the custom 6-terminal NMOS.

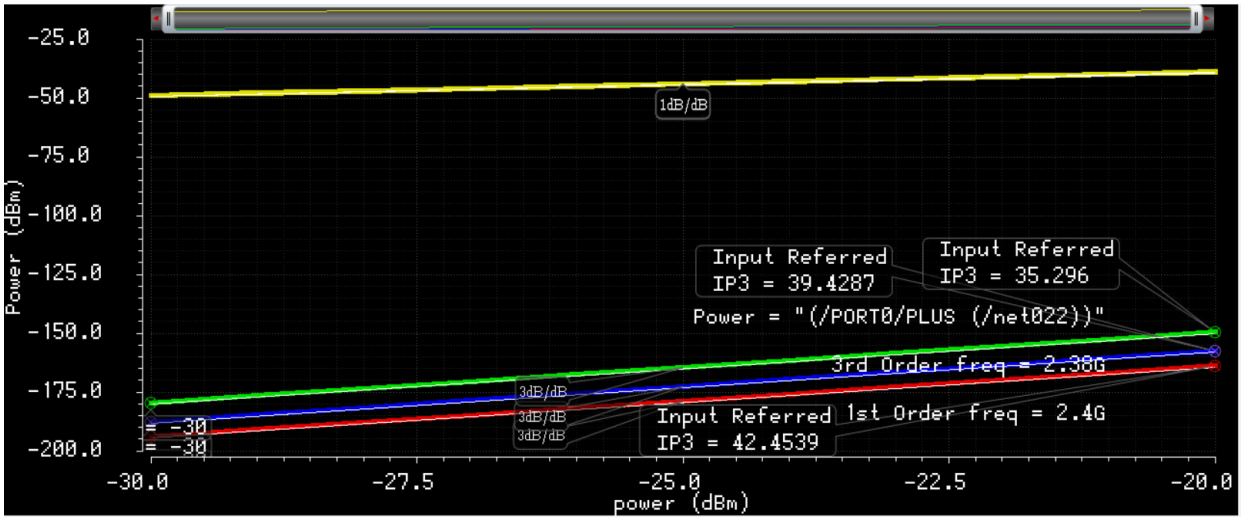

Figure 3.13. IP3 Comparison of the switches in the off-state.

In this graph, we observe that from top to bottom, the $IP3$ numbers for these three devices are $35.3(dBm)$, $39.4(dBm)$, $42.5(dBm)$, respectively. As we anticipated, the custom 6-terminal NMOS provides the best linearity. Although one might claim a $3(dBm)$ improvement

is not that impressive, yet this improvement is measured for merely one RF switch, and there is a total of 128 in this design.

3.4.4 Inherent Issue of BSIM Model for Linearity

Now that we have established a ground rule for determining the size of the switches, we would like to simulate the actual $IP3$ numbers for the designed RF switch. Nevertheless, one serious issue arises under the condition when the switch is turned on. Below, in Figure 3.14, we observe the $IP3$ simulation results when the switch is turned using the testbench of Figure 3.9.

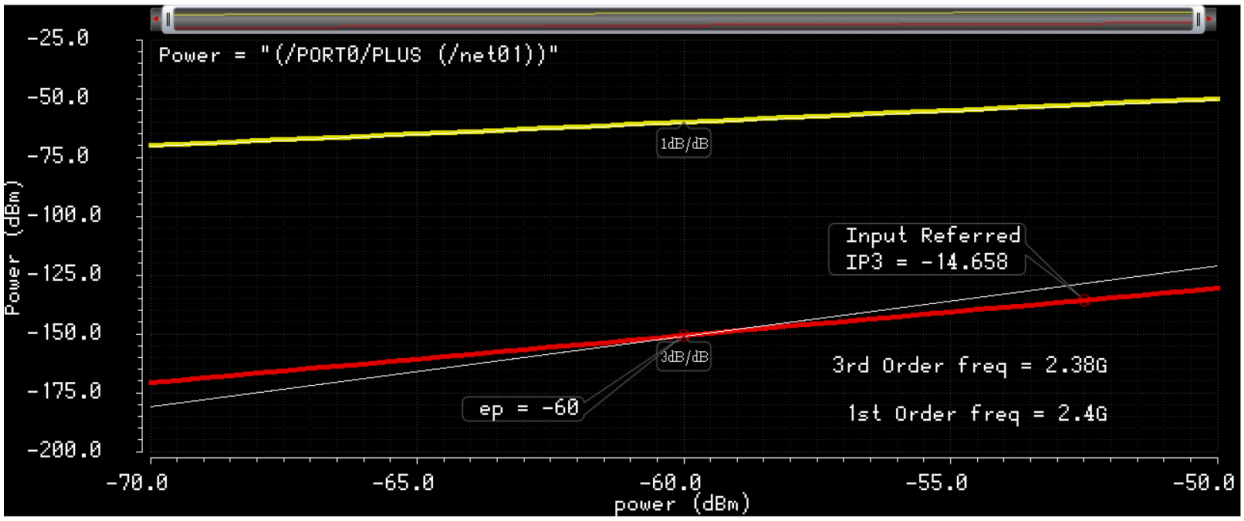

Figure 3.14. IP3 Simulation of the switch in the on state.

In this simulation, the 3^{rd} harmonic frequency element does not increase by a slope of $3(dB)$ but $2(dB)$. In conclusion, this issue is caused by a singularity in the transistor models around $V_{DS} = 0$, which is a historical problem with the BSIM4 model[18]. In the BSIM4 model, the drain and source terminals are not precisely symmetric, such as the drain and source internal resistance, but directional. The equation that governs the parameters relevant to the drain and source voltage is piecewise to be mirroring to each other, meaning $f(v_{DS}) = f(v_{SD})$, an even function.

Now, let us examine this statement from a mathematical standpoint. According to the BSIM4 manual, the equation of the drain current, I_D, under the circumstance of $V_{DS} = 0$ is:

$$I_D = \mu_{eff} \cdot C_{ox} \cdot \frac{W}{L} \cdot \frac{1}{1 + \frac{|v_{DS}|}{E_{sat} \cdot L}} \cdot \left(v_{GS} - V_{th} - A_{bulk} \cdot \frac{|v_{DS}|}{2} \right) \cdot |v_{DS}| \qquad \textit{Eq. 3.9}$$

In this discussion, we will utilize the particular test bench shown in Figure 3.15 measuring the drain current, where 1) the DC voltage, $V_{DS} = 0$, due the presence of the large DC blocking capacitors and the biasing resistors, R_g, 2) the AC swing of $v_{gs} = \left|\frac{v_{ds}}{2}\right|$, $K = \mu_{eff} \cdot C_{ox} \cdot \frac{W}{L}$,and 3) $v_{GS} = V_G + v_{gs}$.

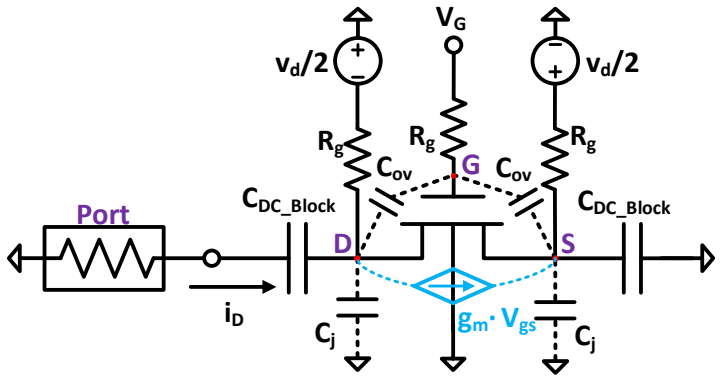

Figure 3.15. Testbench to study the drain current discontinuity.

If we calculate the derivative of Eq. 3.9,

when v_{DS} tends to 0 from the left side:

$$\frac{dI_D}{dv_{DS}} = \frac{K \cdot E_{sat} L \left[-\frac{1}{2}(1 - A_{bulk}) v_{DS}^2 + E_{sat} L (1 - A) v_{DS} + (V_G - V_{th}) \right]}{(E_{sat} L - v_{DS})^2} \qquad \textit{Eq. 3.10}$$

when v_{DS} tends to 0 from the right side:

$$\frac{dI_D}{dv_{DS}} = \frac{K \cdot E_{sat}L\left[\frac{1}{2}(1 - A_{bulk})v_{DS}^2 + E_{sat}L(1 - A)v_{DS} - (V_G - V_{th})\right]}{(E_{sat}L + v_{DS})^2} \quad \textit{Eq. 3.11}$$

Therefore, by plugging $v_{DS} = 0$ into the Eq. 3.10 and Eq. 3.11:

When v_{DS} tends to 0 from the left side:

$$\frac{dI_D}{dv_{DS}} = -K \cdot (V_G - V_{th}) \quad \textit{Eq. 3.12}$$

when v_{DS} tends to 0 from the right side:

$$\frac{dI_D}{dv_{DS}} = +K \cdot (V_G - V_{th}) \quad \textit{Eq. 3.13}$$

By observing Eq. 3.12 and Eq. 3.13, we notice that when $v_{DS} \to -0$, the derivative is a negative number, while when $v_{DS} \to +0$ or $v_{SD} \to -0$, that will appear to be positive. Confirmatively, the slope of Figure 3.16 below verifies this statement, where the discontinuity at $v_{DS} = 0$ not also appears in the derivative of drain current, $\frac{dI_D}{dv_{DS}}$, but also the effected unified mobility, μ_{eff}, and the threshold voltage, V_{th}.

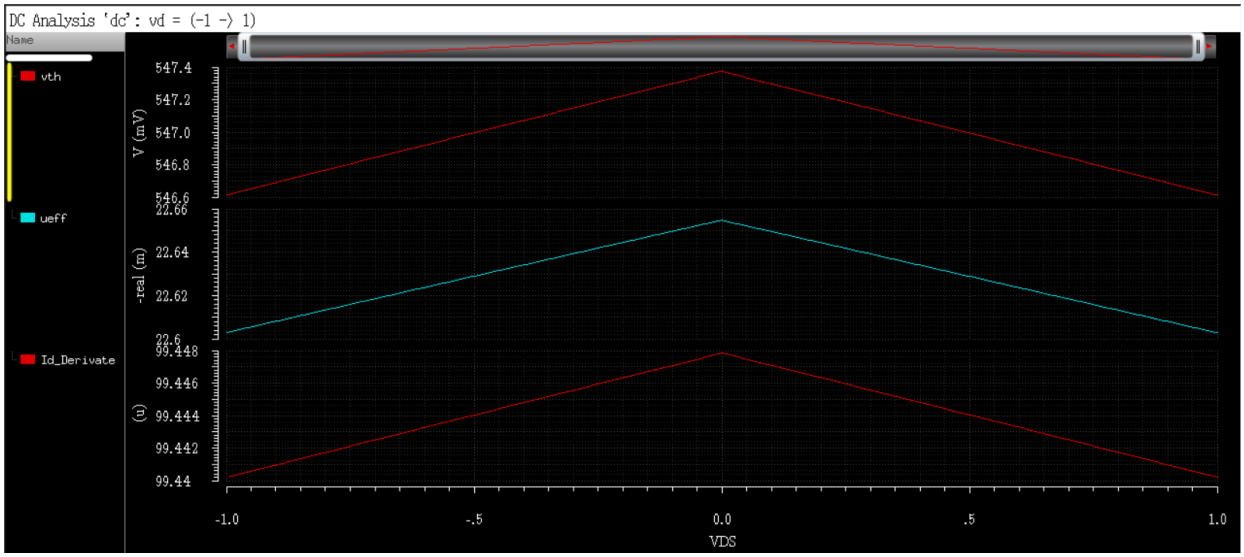

Figure 3.16 DC Simulation for Symmetric Testing

How would this discontinuity invoke a peculiar phenomenon on the $3rd$ harmonic extrapolation line in Figure 3.14? To answer this question, we would have to understand the algorithms used in the harmonic-balance simulator. The harmonic balance method bypasses the tedious and time-consuming calculation for the phase of the initial transient response rather than solving the steady-state response of the nonlinear circuit. Therefore, the harmonic balance method will not perform the transient simulation, but instead, calculate the frequency components by means of Fourier transform after a DC simulation estimates the iterative solution. We will illustrate this method by taking the RF switch as an example. In the upper circuit of Figure 3.17, where the DC biasing circuit elements are omitted, a linear current, $\overline{I_{Rs}}$, is going through the source resistor, R_S. Since there are two nonlinear elements, the dependent current source of the MOSFET and the junction capacitor, the spectral elements of the $\overline{I_{Rs}}$ contains a number of harmonic frequencies other than the original tone from the voltage source.

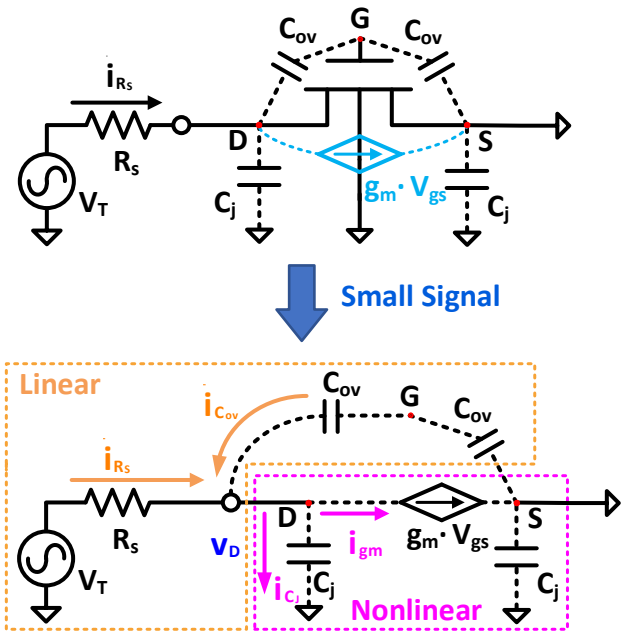

Figure 3.17: Model of HB Analysis for an RF Switch

If we partition the small signal of the circuit into the linear and nonlinear subcircuits shown in Figure 3.17, presumably, we could apply the KCL and KVL equations in the frequency domain. The core formula in this circuit analysis will be:

$$\overline{I_{Rs}} + \overline{I_{Cov}} = \widetilde{I_{gm}} + \widetilde{I_{CJ}} \qquad \textit{Eq. 3.14}$$

By noticing all the currents in Eq. 3.14, in fact, contain infinite tones, this is the reason why we perform a harmonic balance simulation in Cadence, we have to define the maximum tone for the frequency, which will affect both the simulation time and accuracy. In addition, if we consider the KCL for each tone, we could write:

$$E(f) = \overline{I_{Rs}}|_{f=f_K} + \overline{I_{Cov}}|_{f=f_K} - \widetilde{I_{gm}}|_{f=f_K} + \widetilde{I_{CJ}}|_{f=f_K} \qquad \textit{Eq. 3.15}$$

The equation $E(f)$ is an error equation with respect to every tone, where $k = 0$ (DC offset), $1, 2, \cdots K$. In order to solve for the amplitude and phase for each current element, the Newton-Raphson iterative technique can be employed here:

$$\begin{bmatrix} v_D|_{f=f_0} \\ \vdots \\ v_D|_{f=f_K} \end{bmatrix}_{i+1} = \begin{bmatrix} v_D|_{f=f_0} \\ \vdots \\ v_D|_{f=f_K} \end{bmatrix}_i - J\left(\begin{bmatrix} v_D(f)|_{f=f_0} \\ \vdots \\ v_D(f)|_{f=f_K} \end{bmatrix}\right)_i^{-1} \begin{bmatrix} E(f)|_{f=f_0} \\ \vdots \\ E(f)|_{f=f_K} \end{bmatrix}_i \qquad \textit{Eq. 3.16}$$

Where the operator J represents the $Jacobian$ of the error function:

$$J\left(\left[v_D|_{f=f_0}, \cdots, v_D|_{f=f_K}\right]_i\right)_i^T = \begin{bmatrix} \left(\frac{\partial E(f)|_{f=f_0}}{\partial v_D(f)|_{f=f_0}}\right)^T & \cdots & \left(\frac{\partial E(f)|_{f=f_0}}{\partial v_D(f)|_{f=f_K}}\right)^T \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial E(f)|_{f=f_K}}{\partial v_D(f)|_{f=f_0}}\right)^T & \cdots & \left(\frac{\partial E(f)|_{f=f_K}}{\partial v_D(f)|_{f=f_K}}\right)^T \end{bmatrix}_i \qquad \textit{Eq. 3.17}$$

Unfortunately, as we have discussed, the derivative of the I_{gm} doesn't exist at $v_D = 0$. Hence, Cadence fails to calculate the $Jacobian$ factor, and the extrapolation of the $3rd$ harmonic becomes meaningless. In fact, any harmonic frequencies are improperly plotted using the Cadence Harmonic Balance method. In summary, this observation describes the discontinuity of the BSIM4 model.

3.4.5 Modifying BSIM4 Model by Schematic Approach and Gummel Symmetry Test

Accurate measurement for the $IP3$ number is enormously essential to our design. Through the meticulous analysis above, we discern this discontinuity in the derivative of the drain current at $V_{DS} = 0$ arises from a flaw that the drain equation uses for absolute value of V_{DS}. Hence, if we could manipulate the equation and create a smooth transition zone for the derivative of the drain current, then the 3^{rd} harmonic frequency element will be resumed to climb at a $3dB$ pace. In order to escape from the kink of the derivative, we could apply a small V_{DS} on the transistor, which should be small enough to not disturb the DC biasing condition, yet large enough to elude the adjacent area of the singularity. Figure 3.18 below illustrates the circuit implementation in Cadence for this idea [18].

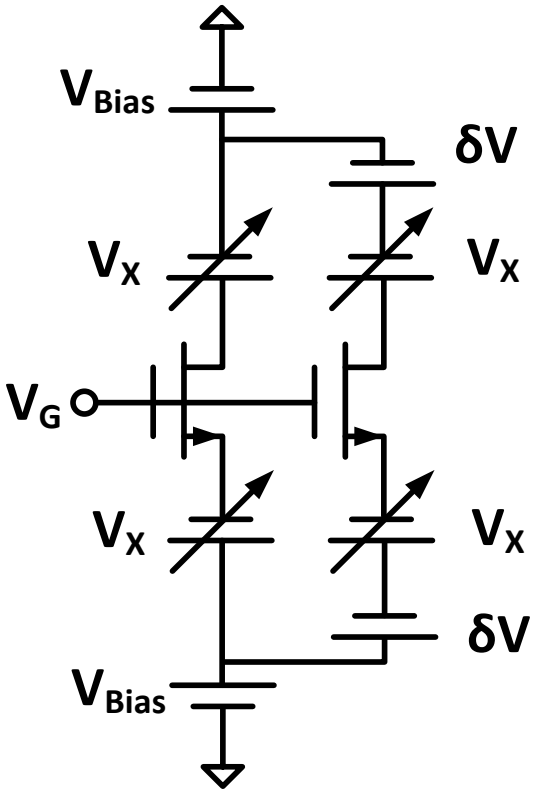

Figure 3.18 Schematic for Gummel Test

3.4.6 An Analytical Method to Calculate Linearity in RF Switches

Now that we have developed a promising model to accurately simulate the $IP3$ number for the RF switches, we should derive the mathematical formula to better guide us to analyze the nonlinearity of the RF switch. Shown in Figure 3.19 are all the contributors to the distortion that arise from the RF switch. The circuit is analyzed under a small AC signal condition so that some of the components, such as the large RF block resistors, the DC blocking capacitor, and the triple-well substrate, are omitted from the diagram. The very foundation that supports the application of small-signal analysis is the small voltage swing on the switches. Namely, we could presume the DC operating point is constant. Hereby, the on-resistance is translated into the transconductance, g_m, under the circuit condition of $|V_{ds}| \approx 0$, which is employed in the below calculation.

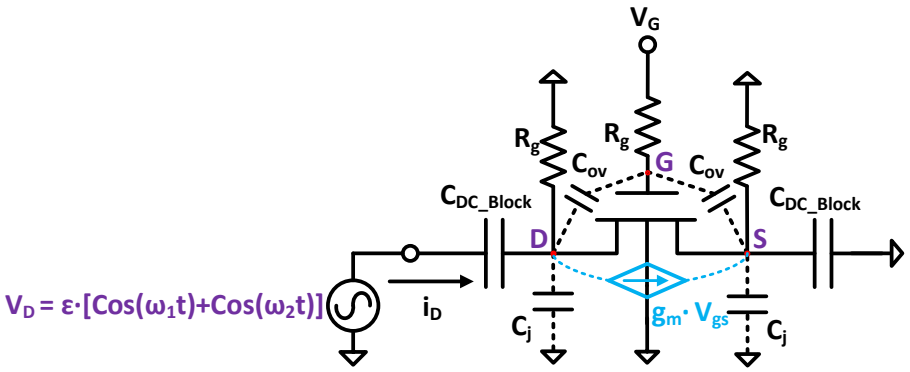

Figure 3.19 Analytical Method of Nonlinear Analysis for the RF Switches

According to KCL, we could write:

$$i_D = C_j \cdot \frac{dv_D}{dt} + g_m \cdot \left(V_G + \frac{1}{2}\varepsilon\right) + \frac{C_{ov}}{2} \cdot \frac{dv_D}{dt} \qquad \textit{Eq.3.18}$$

Since $v_D = \varepsilon \cdot [Cos(\omega_1 t) + Cos(\omega_2 t)]$, $\varepsilon \to 0$ therefore:

$$\frac{dv_D}{dt} = \varepsilon \cdot [\omega_1 \cdot Cos(\omega_1 t) + \omega_2 \cdot Cos(\omega_2 t)] \qquad \textit{Eq. 3.19}$$

$$i_D = g_m \cdot V_G + \varepsilon \cdot \left\{\left[\left(C_j + \frac{C_{ov}}{2}\right)[\omega_1 \cdot Cos(\omega_1 t) + \omega_2 \cdot Cos(\omega_2 t)]\right] + \frac{1}{2} g_m\right\} \qquad \textit{Eq. 3.20}$$

In order to solve for the weakly nonlinearity of the RF switch modeled around a quiescent point, (I_D, V_T), we could make use a Taylor series expansion by considering the two nonlinear elements of transconductance, g_m, and the junction capacitor, C_j. Here, we will again harness the intrinsic equation of drain current (Eq. 3.8) in the BSIM4 model to derive the transconductance:

Therefore,

$$g_m|_{v_{GS}=V_G} = -K \cdot (V_G - V_{th}) \qquad \textit{Eq. 3.21}$$

On the other hand, we will adopt the simple junction capacitor equation in this analysis.

$$C_j = \left.\frac{C_{jo}}{\left(1 + \frac{v_D}{\Phi_o}\right)^m}\right|_{v_D=\varepsilon+V_D} \qquad \textit{Eq. 3.22}$$

When the switch is on, $V_D = 0$. While when the switch is on, $V_D = V_{DD}$

The Taylor series of the nonlinear drain current is equal to:

$$i_D = I_D + G_1 \cdot v_D + G_2 \cdot v_D^2 + G_3 \cdot v_D^3 + \cdots \qquad \textit{Eq. 3.23}$$

Here, the set of G_n is the systematic coefficient of the nonlinear model of i_D. By applying the small-signal analysis, we are capable of writing the coefficients via extracting from the n^{th} order device's derivative:

$$G_n = \frac{1}{n!} \cdot \frac{d^n G}{dv_D^n}\bigg|_{v_D = \varepsilon + V_D} \qquad \textit{Eq. 3.24}$$

In the following derivation, the Taylor coefficients are calculated under different operation conditions of the switch. When the switch is ON, $v_{GS} = \frac{1}{2}\varepsilon + V_G = \frac{1}{2}\varepsilon + V_{DD}$, $v_{DS} = \varepsilon + V_D = \varepsilon$. Whilst, when the switch is OFF, $v_{GS} = \frac{1}{2}\varepsilon$, $v_{DS} = \varepsilon + V_D = \varepsilon + V_{DD}$.

$$\begin{cases} G_1^{ON}(\varepsilon = 0) = \left[C_{jo} + \dfrac{C_{ov}}{2}\right](\omega_1 + \omega_2) - K \cdot (V_{DD} - V_{th}) \\ G_1^{OFF}(\varepsilon = 0) = \left[\dfrac{C_{jo}}{\left(1 + \frac{V_{DD}}{\Phi_o}\right)^m} + \dfrac{C_{ov}}{2}\right](\omega_1 + \omega_2) + KV_{th} \end{cases} \qquad \textit{Eq. 3.25 a) \& b)}$$

we could calculate G_3 as:

$$\begin{cases} G_3^{ON}(\varepsilon = 0) = \dfrac{K[3(A-1)E_{sat}L - 6(V_{DD} - V_{th})]}{6E_{sat}^2 L^2} + \dfrac{C_{jo}m(1+m)}{6{\Phi_o}^2}(\omega_1 + \omega_2) \\ G_3^{OFF}(\varepsilon = 0) = \dfrac{K[3(A-1)E_{sat}L + 6V_{th}]}{6E_{sat}^2 L^2} + \dfrac{C_{jo}m(1+m)}{6(\Phi_o + V_{DD})^2\left(1 + \frac{V_{DD}}{\Phi_o}\right)^m}(\omega_1 + \omega_2) \end{cases} \qquad \textit{Eq. 3.26 a) \& b)}$$

Eventually, in order to find the $IP3$ of the switch, we could make use of the Eq. 3.25 a) & b) and Eq. 3.26 a) & b). The $IP3$ will equal to:

$$\begin{cases} IIP3^{ON}(\varepsilon = 0) = \sqrt{\frac{4}{3}\left|\frac{G_1^{ON}(\varepsilon = 0)}{G_3^{ON}(\varepsilon = 0)}\right|}\,(dBm) \\ IIP3^{OFF}(\varepsilon = 0) = \sqrt{\frac{4}{3}\left|\frac{G_1^{OFF}(\varepsilon = 0)}{G_3^{OFF}(\varepsilon = 0)}\right|}\,(dBm) \end{cases}$$

Eq. 3.27 a) & b)

To the Eq. 3.27 a) & b) above, we could make three important observations. First, this $IP3$ calculation is based on current. The reason is that as the Z_{bal} network matching the impedance of the antenna, it behaves as if a $50(\Omega)$ resistor, and its input power is $P_{TX} - 3(dBm)$. Namely, the magnitude of the current that flows through the Z_{bal} network will be a constant. Thus, an $IP3$ calculation based on the current approach is a reference as satisfactory as the one based on voltage. Second, the equations derived from Taylor expansion are merely valid when the RF switch is under the weakly nonlinear circumstance, which means the AC swing on the two terminals of the switches has to be small enough. Last but not least, you might have noticed that the internal resistance of the input source is missing from the derivation. The reason here for doing so is out of simplicity. By forcing the voltage on the switch to be a linear value, all the harmonic voltage coefficients can be ignored in the calculation. Nevertheless, the trending of the resulting solution manifests mightily meaningful. If we take a closer look at Eq. 3.27 a) & b), we could realize that an increment in the supply voltage, V_{DD} could boost up the $IP3$ performance. We could confirm this conclusion by inspecting the curves of the junction capacitors and on-resistance of the MOSFET with respect to the voltage drop because these two components are the main contributors to the nonlinear distortion arising from the switch.

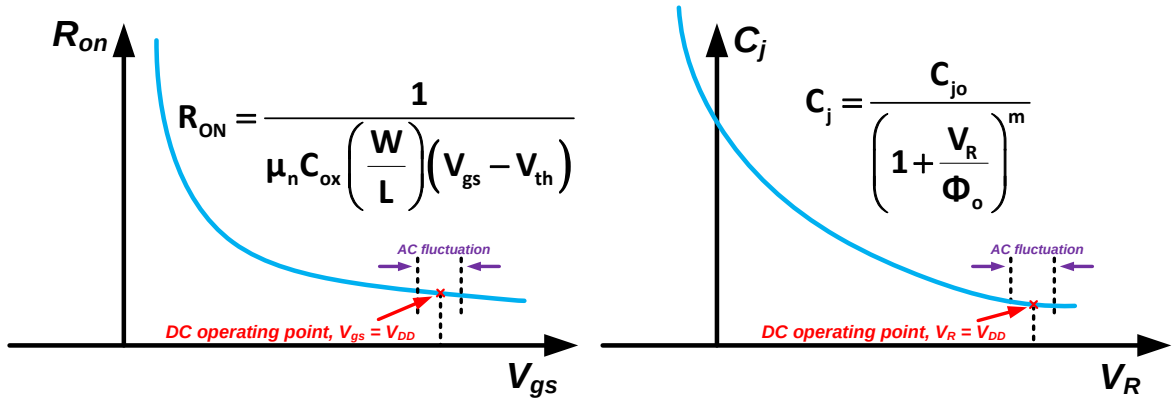

Figure 3.20 Curves of Junction Capacitor and On-Resistance.

By observing the slopes of the two curves in Figure 3.20, when the V_{DD} is increasing, the curves have a smoother changes of slope. Therefore, the nonlinearity of the switch will be mitigated.

3.5 Post-Layout Simulation and Specification

In this section, we will present the simulation results of the new design high-linear Z_{bal} network.

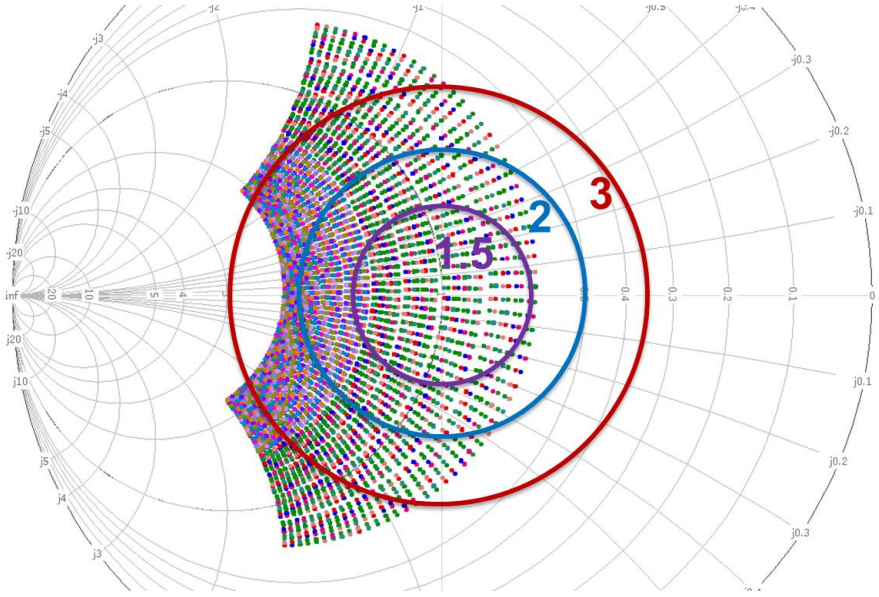

Figure 3.21 Simulated Admittance Coverage on Smith Chart

In Figure 3.21, the overall admittance features a 1.5:1 VSWR, especially the tunable imaginary part covers shows wide range beyond 3:1. Another feature about this tuning range is the uniformity for each impedance point. In future work, the tunable range can be further extended by utilizing a technology that has a higher substrate isolation.

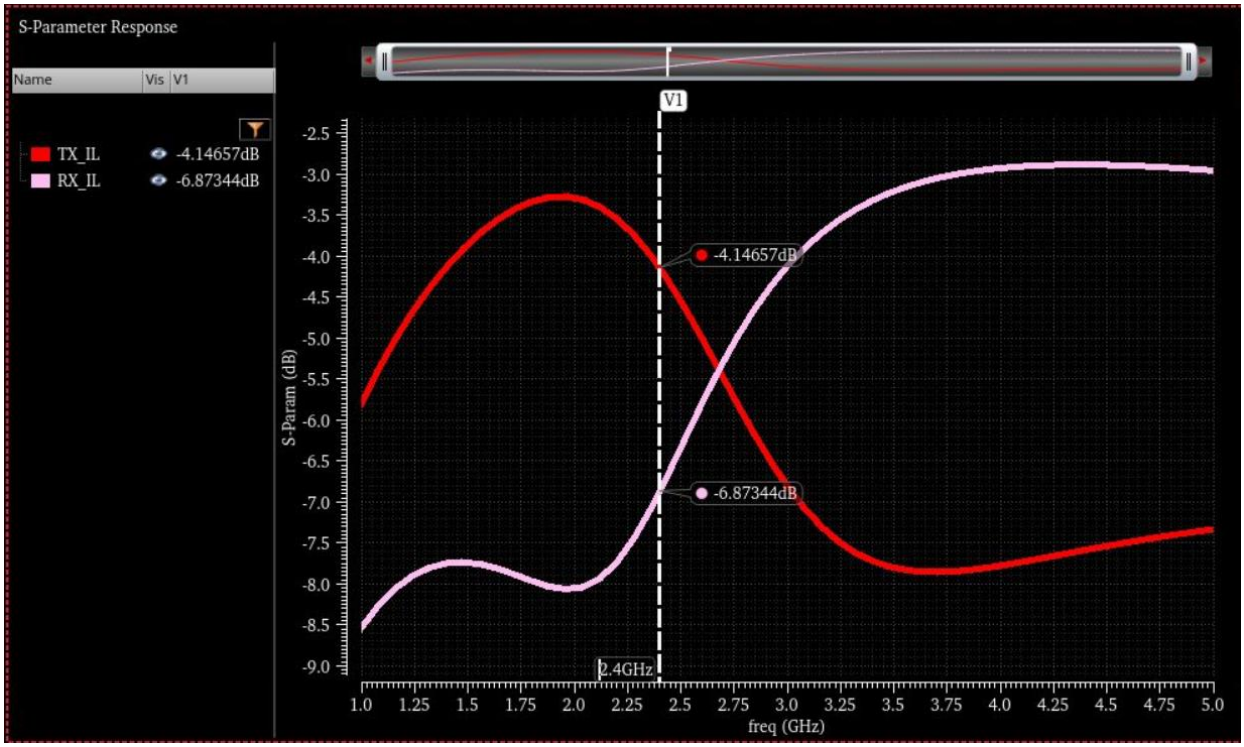

Figure 3.22 Simulated S Parameters of RX_IL & TX_IL

The Figure 3.22 above illustrates the insertion losses for RX & TX terminals. Since the input impedance of the TX is roughly $25(\Omega)$, the insertion loss is close to the ideal value of $3(dB)$. Conversely, the input impedance of the RX terminal is essentially dominated by the EBD, which basically has only an imaginary part. Thus, the RX insertion loss is greatly restrained. In future work, a matching network should be in place before the EBD, to mitigate the insertion loss.

4 A Complete Review of the New FD Chip

4.1 Layout

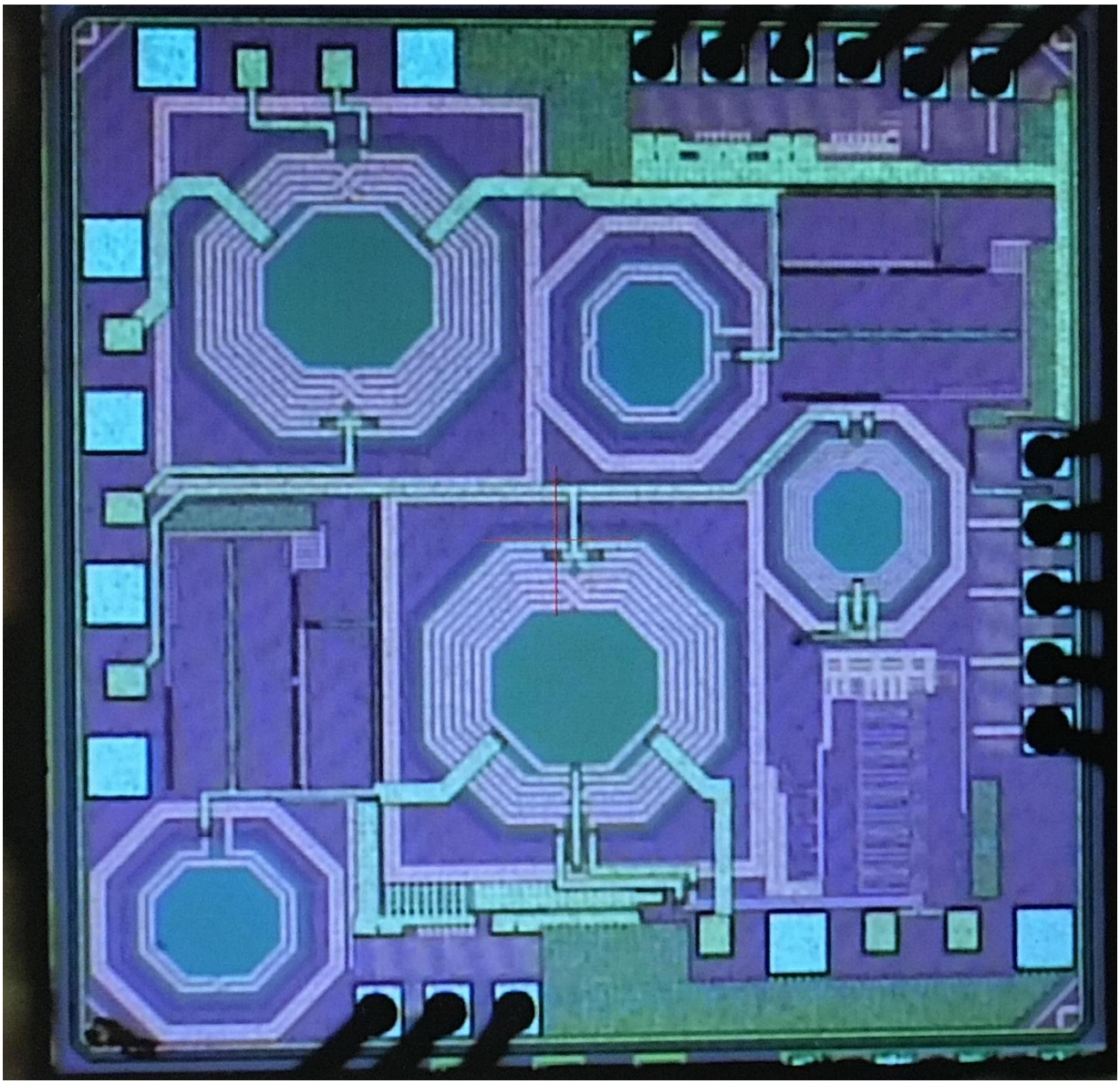

Figure 4.1 Layout of the Work

The complete layout of this work is shown in Figure 4.1. The dimension of the die is $1.2\ (\mu m) \times 1.2\ (\mu m)$ including two sets of the system, one of which has an auxiliary feedforward canceler to further suppress the self-interference. At the right bottom corner of the die, a set of

unconventional pads is positioned. The reason for designing these particular pads is to reduce the total die area.

4.2 Comparison of Simulation and Measurement Results

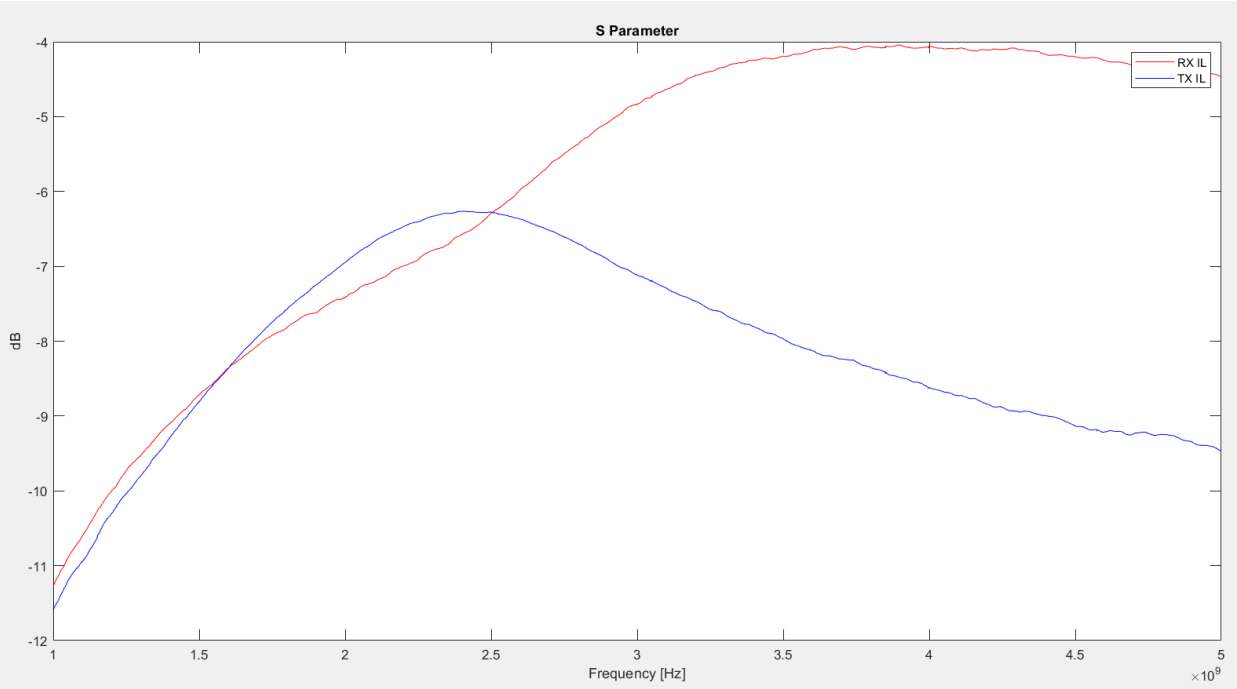

Figure 4.2 Measured S Parameters of RX IL & TX IL

The Figure 4.2 above demonstrates the RX & TX insertion losses from measurement. Both of the results are surpassed by the simulation results. The reason for that stems from the cables connected in between the vector network analyzer and the probe. As a result, both insertion losses are approximately $7(dB)$.

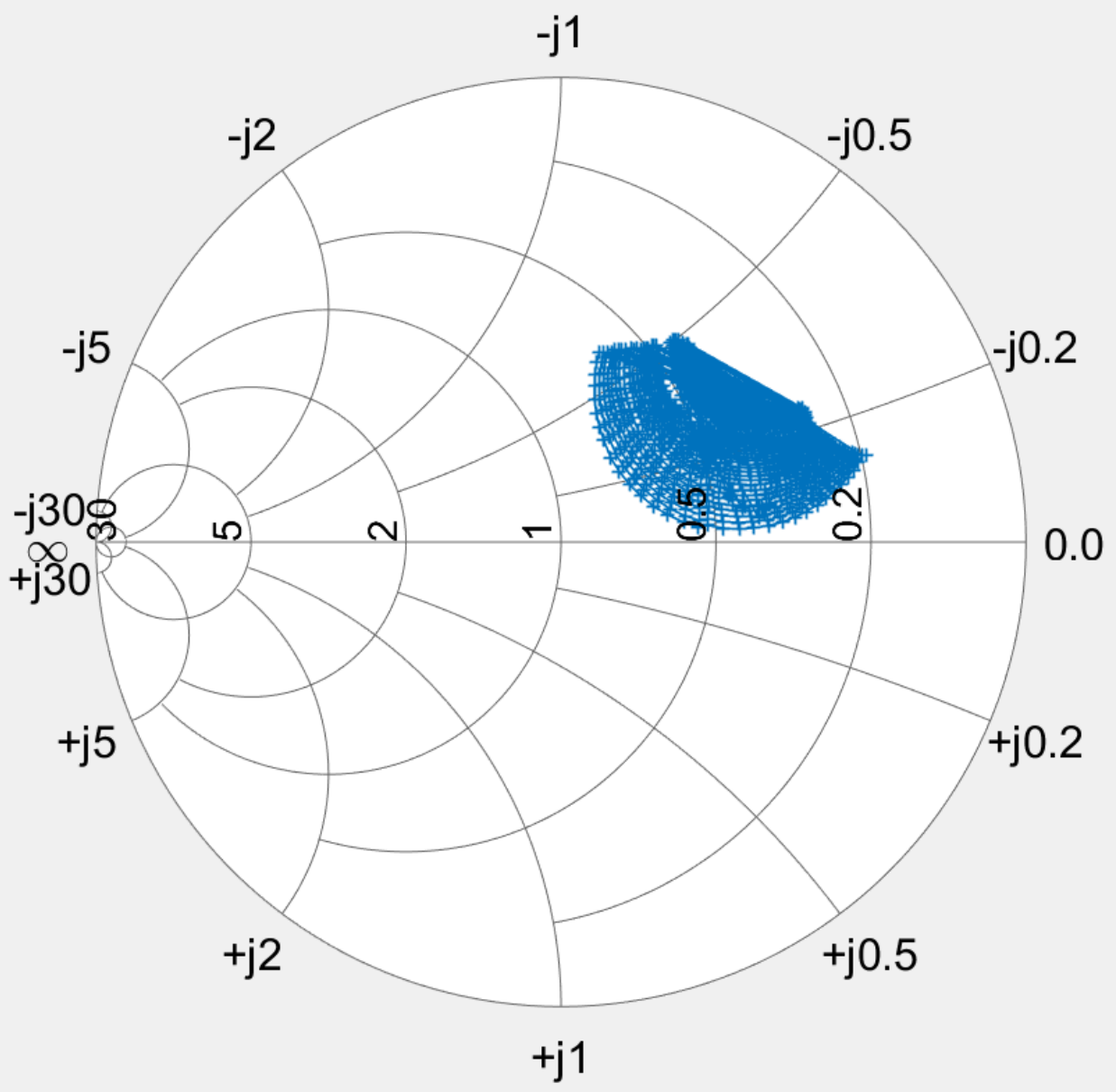

Figure 4.3 Measured Admittance Coverage on Smith Chart

Finally, the tunable range of the impedance matching network on a Smith Chart is presented in Figure 4.3. One notes the contour of the impedance is quite different from the simulation result. Yet, if we take a closer look, we will realize the coverage is shifted by the series resistance introduced by the cable, the probe, and the pad, and the series inductance from the EBD. Therefore, the measurement result is decent and agrees with the simulation.

5 Conclusion

The high linearity electrical balanced duplexer has been designed, taped out, and measured to validate the ideal of a high IP3 impedance matching which suppresses transmitter self-interference. Utilizing an advanced technology node of 10 layers CMOS effectively improves the loss mechanism of the EBD, which better serves the purpose of isolating the TX signal from the RX terminals. In the meantime, investigating the design of the impedance matching network in depth promises a high IP3 value that helps suppress the TX leakage signal.

The design techniques described in this thesis have significantly reduced the occupied area of the EBD as compared to prior art, Table 3. This was done without sacrificing the tuning range performance.

Items	Larson's [7f]	Craninckx's [14]	Long's [8]	This Work
Technology	90nm	0.18um	0.18um	40nm
Frequency Range (GHz)	3GPP bands 1,2,3,9	1.9-2.2	1.8-2	2.2-2.6
Bandwidth (MHz)	190 @40dB TX-RX Iso	229 @45dB TX-RX Iso	20 @50dB TX-RX Iso	200 @30dB TX-RX Iso
TX Insertion Loss (dB)	4.7	3.7	3	3.3
RX Insertion Loss (dB)	N/A	3.9	11	3.9
TX-to-RX IIP3 (dBm)	>16	70	49	>60 (w/o FFC) >38 (w/ FFC)
Tuning Range	N/A	1.5:1 VSWR	<1.5:1 VSWR	Imag: >3:1 VSWR Real: >1.5:1 VSWR
Area(mm^2)	0.6	1.75	0.67	0.14

Table 3 Comparison of Specifications

For the further development, the impedance matching network should consider the stabilization of the group delay to improve the TX SI suppression depth by adding more poles in transformer function. Additionally, the design of the RF switch should use BSIM6 model if possible to better capture the signal distortion.

6 Reference

[1] K. Chu, M. Katanbaf, T. Zhang, C. Su, and J. C. Rudell, “A Broadband and Deep-TX Self-Interference Cancellation Technique for FullDuplex and Frequency-Domain-Duplex Transceiver Applications,” presented at the 2018 IEEE International Solid-State Circuits Conference (ISSCC), 2018.

[2] X. J. Li, Y.P. Zhang, “Flipping the CMOS Switch,” IEEE Microwave Magazine, Volume: 11, Issue: 1, February 2010.

[3] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, (2010) “Achieving single channel, full duplex wireless communication,” in Proc. of ACM Mobicom, pp. 1–12.

[4] A. A. A. Apriyana, Y. P. Zhang, “A 60-GHz SPST switch in 65-nm CMOS technology,” Radio-Frequency Integration Technology (RFIT) 2014 IEEE International Symposium on, pp. 1-3, 2014.

[5] S. Abdelhalem, P. S. Gudem, and L. E. Larson, “A Tunable Differential Duplexer in 90nm CMOS,” Proc. IEEE RFIC, June, 2012, pp. 237-241.

[6] M. Mikhemar, H. Darabi, and A. Abidi, “A Multiband RF Antenna Duplexer on CMOS: Design and Performance,” IEEE J. of Solid-State Circuits, vol. 48, no. 9, pp. 2067- 2077, Sept. 2013

[7] S. H. Abdelhalem, P. S. Gudem, and L. E. Larson, “Tunable CMOS Integrated Duplexer with Antenna Impedance Tracking and High Isolation in the Transmit and Receive Bands,” in IEEE Trans. on MTT, pp. 2092–2104, Sep 2014.

[8] B. van Liempd, R. Singh, S. Malotaux, P. Reynaert, J. R. Long, and J. Craninckx, “A dual-notch +27dBm Tx-power electrical-balance duplexer,” in Proc. IEEE Eur. Solid-State Circuits Conf., Sep. 2014, pp. 463–466.

[9] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. K. Ramakrishnan, C. W. Rice, and N. K. Shankaranarayanan, “Design and characterization of a full-duplex multi-antenna system for Wi-Fi networks,” IEEE Trans. Vehic. Tech., 63 (3), 1160–1177, 2014.

[10] E. Sartori, “Hybrid transformers,” IEEE Transactions on Parts, Materials and Packaging, vol. 4, Issue 3, pp. 59-66, 1968.

[11] L. Anttila, D. Korpi, E. Antonio-Rodríguez, R. Wichman, and M. Valkama, “Modeling and efficient cancellation of nonlinear self-interference in MIMO full-duplex transceivers,” Proc. IEEE Glob. Telecom. Conf., pp. 777–783, 2014.

[12] S.C. Cripps, “RF Power Amplifiers for Wireless Communications,” Artech House, 1999.

[13] S. H. Abdelhalem, P. S. Gudem, L. E. Larson, “Tunable CMOS Integrated Duplexer With Antenna Impedance Tracking and High Isolation in the Transmit and Receive Bands,” IEEE Trans. on Micr. Theory and Techniques, vol. 62, no. 9, pp. 2092-2104, Sep 2014

[14] B. van Liempd, B. Hershberg, S. Ariumi, K. Raczkowski, K. F. Bink, U. Karthaus, E. Martens, P. Wambacq, and J. Craninckx, “A +70-dBm IIP3 Electrical-Balance Duplexer for Highly Integrated Tunable FrontEnds,” IEEE Transactions on Microwave Theory and Techniques, vol. 64, no. 12, pp. 4274–4286, Dec 2016.

[15] B.-K. Kim, T. Lee, D. Im, D.-K. Im, B. Kim, and K. Lee, “Design methodology of tunable impedance matching circuit with SOI CMOS tunable capacitor array for RF FEM,” Proc. Asia–Pacific Microw. Conf. (APMC), Nov. 2013, pp. 7–9.

[16] A. Tombak, M. S. Carroll, D. C. Kerr, J.-B. Pierres, and E. Spears, “Design of High-Order Switches for Multimode Applications on a Silicon-on-Insulator Technology,” IEEE Trans. Microw. Theory Tech., vol. 61, no.10, pp.3639-3649, 2013.

[17] K. Banerjee and C. Hu., “High Current Effect in Silicide Films for sub-0.25um VLSI Technologies,” in Proc. 36th IRPS, pp. 284- 292, 1998.

[18] X. Xi, K. Cao, J. He, H. Wan, M. Chan, C. Hu, “Symmetry realization of BSIM model with dynamic reference method for circuit simulation,” 60th DRC. Conference Digest Device Research Conference, 2002.

[19] E. Aryafar, M. A. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang, “MIDU: enabling MIMO full duplex,” in Proc. of ACM Mobicom, 2012, pp. 257–268.

[20] A. S. Sedra, P. O. Brackett, “Filter Theory and Design: Active and Passive,” Matrix: Champaign, IL, USA, 1978.

[21] J. K. Skwirzynski, “Design Theory and Data for Electrical Filters,” Van Nostrand: London, UK, 1965.

[22] L. Wanhammar, Analog Filters Using Matlab; Springer: Dordrecht, The Netherlands, 2009.

[23] R. D. Koller, B. M. Wilamowski, “A ladder prototype synthesis algorithm. In Proceedings of the IEEE 35th International Midwest Symposium on Circuits and Systems,” Washington, DC, USA, 1992.

[24] O. Nastov, R. Telichevesky, K. Kundert, and J. White, “Fundamentals of fast simulation algorithms for RF circuits,” IEEE Proc., pages 600–621, 2007.

[25] C. Nwankwo, L. Zhang, A. Quddus, M. Imran, R. Tafazolli, “A Survey of Self-Interference Management Techniques for Single Frequency Full Duplex Systems”

[26] M. Albreem, "5G wireless communication systems Vision and challenges"