Title
RF and Microwave Amplifier Design With ESD Protection

Author
Lin, Lin

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

RF and Microwave Amplifier Design With ESD Protection

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Lin Lin

December 2010

Dissertation Committee:

Dr. Albert Wang, Chairperson

Dr. Sheldon Tan

Dr. Daniel Xu

The Dissertation of Lin Lin is approved.

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENT

I would like to thank my advisor, Professor Albert Wang, for his encouragement and advice throughout my graduate career at University of California Riverside. His patience in education and passion for scientific training has left a supremely important and beneficial impression upon me. I am extremely grateful for the guidance, motivation, and opportunity that I have received from him.

I am indebted to Professor Tan and Professor Xu for being my committee and for their help and valuable advice over my graduate course. I am also very grateful for the talented members of the Albert Wang group from whom I have learned a tremendous amount over the past four years. I have been lucky to work with many exceptional people in the Wang group. These include Mr. Xin Wang, Mr. He Tang, Dr. Siqiang Fan, Mr. Jian Liu, Mr. Hui Zhao, Mr. Qiang Fang, and Mr. Zitao Shi. Lastly, I would like to thank my parents. Their ambition and passion for learning have inspired me in my own studies and future career. Their constant and unwavering confidence in me has encouraged me to persevere over the course of my degree. I owe them everything and cannot express enough my gratitude to them.

ABSTRACT OF THE DISSERTATION

RF and Microwave Amplifier Design With ESD Protection

by

Lin Lin

Doctor of Philosophy, Graduate Program in Electrical Engineering

University of California, Riverside, December 2010

Dr. Albert Wang, Chairperson

This dissertation is organized as follows. First, the concept and background of ultra wideband (UWB), orthogonal frequency-division multiplexing (OFDM), and Electrostatic discharge are reviewed. Second, the integration of OFDM-UWB and ESD protection and related unit architectures are discussed. Third, unit structures of the design, such as LNA, power amplifier, and ESD protection devices are also discussed and theoretical models are described for all the structures. A novel ESD design based on nano-phase-switching is developed and the ESD performance has been measured. Integration of ESD protection circuits with RF amplifier has been explored and discussed.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1
Introduction

This dissertation is organized as follows. First, the concept and background of ultra wideband (UWB), orthogonal frequency-division multiplexing (OFDM), and Electrostatic discharge are reviewed. Second, the integration of OFDM-UWB and ESD protection and related unit architectures are discussed. Third, unit structures of the design, such as LNA, power amplifier, and ESD protection devices are also discussed and theoretical models are described for all the structures.

1. Ultra wideband (UWB)

Ultra wideband (UWB) is a recently approved wireless communication standard that has been growing worldwide interest. The main reasons why this standard has been creating so much impact are its proposed very high data transfer speeds (480 Mb/s), its low power consumption while transmitting at these high speeds, and its spatial capacity. The spatial capacity (bits/sec/square-meter) is really where UWB outperforms the competition. UWB is projected at having upwards of 6 devices working simultaneously at 480 Mb/s within a range of 10 m, which is unheard of in today's wireless communications. UWB has the potential to revolutionize the consumer electronics industry.

A. UWB Definition

On February 14, 2002 the FCC released a report that officially allocated spectral space for UWB technology. This allocation was strictly to define and restrict the radio frequency (RF) emissions of this technology and bandwidth to allow coexistence. The minimum bandwidth of UWB as defined by the FCC must follow one of the two constraints listed below [1].

• The minimum bandwidth must occupy more than 20% of the center frequency.

• The minimum bandwidth must exceed 500 MHz.

The total bandwidth that could be occupied as defined by the FCC is from 3.1 GHz to 10.6 GHz. This covers a total span of 7.5 GHz. The power regulations for this technology were also strictly defined to allow coexistence. The max power emitted must be under −41.3 dBm/MHz. This is equivalent to 0.5 mW of average continuous power transmission across the full 7.5 GHz bandwidth (3.1−10.6 GHz) [2].

B. UWB History

UWB is not a new technology. There are recorded experiments performed by Heinrich Hertz in 1865 that involved transmitting electromagnetic (EM) waves. The transmission of those EM waves can be considered UWB. These experiments led Guglielmo Marconi to invent the Morse - code telegraphy in 1901 [3]. Then between the 1900s and 1950, wireless technology went mainly narrowband or tuned wireless transmission. Such inventions included telephones, amplitude modulation (AM), frequency modulation (FM), and various other narrowband technologies that are taken for granted today [4].

Around the 1950s is when UWB technology started to reappear. During this time period, the main research done on UWB was for communications, radar, and various other applications [4]. Around this same time period, the first patents for UWB type technologies start appearing. In 1952, Louis de Rosa was granted a patent (U.S. Patent No. 2671896) for his random impulse system [1]. Another UWB type patent was granted

to Conrad H. Hoeppner in 1961 for his work with a pulse communication system that reduced jamming capabilities [1].

Next, the UWB impulse radio was invented in the 1970s using digital techniques. This led to the first commercial UWB systems to be invented and sold around the 1980s and 90s [4]. One such company that was selling these products was PulsON or Time Domain Corporation. Subsequent to these important advancements, the FCC released its Report & Order on February 14, 2002 that allowed UWB to be used under Part 15, or commercial unlicensed use [5]. In 2003 there was a push by the Institute of Electrical and Electronics Engineers (IEEE) 802 groups to create a standard for UWB technology.

In January 2003, the IEEE 802.15.3a task group was created to standardize UWB. The first step taken was to call for all the proposals for the different technologies that could be classified as UWB. This led to 21 proposals that were submitted to the task group. Since this number was very large, a few companies tried to merge or consolidate their proposals into one as many of the proposals had only slight differences. Then in May 2003, the task group had a meeting where the proposals were narrowed down from 21 to 2. The two proposals were Multiband Orthogonal Frequency Division Multiplexing (MB‐OFDM) and Direct Sequence technique (DS). Through politics and voting loop‐holes the IEEE 802.15.3a task group was held up indefinitely. In January 2006, the supporters of both proposals shut down the IEEE 802.15.3a task group, as it had been stalled for 3 years. This task group was ended without conclusion [6].

While the task group was in constant deliberation, outside supporters of the MB‐OFDM proposal formed a group of their own to continue pursuing their idea. The Multiband OFDM Alliance (MBOA), now known as the WiMedia Alliance was created. This group continued to work while the IEEE task group was stalled. The WiMedia Alliance decided to release the WiMedia specifications and to get their proposal standardized [6].

WiMedia then searched for an alternative to IEEE to handle the standardization. ECMA International was used because of its strict rules about removing politics and business issues from deliberations.

ECMA created its own technical committee TC20 [6]. In December 2005, ECMA International released the ECMA 368 [11] and 369 UWB technical standards using MB‐OFDM.

WiMedia is currently working on getting the ECMA standards approved around the world. They believe that a new technology will grow to a successful state only if the following conditions are met [6]:

• The technology is not fragmented among multiple incompatible standards.

• The technology is standardized openly with no individual intellectual property holders.

• The standards are global to assure global trade of products with the new technology.

2. Main UWB Techniques

The only positive conclusion that came from the IEEE 802.15.3a task group was the elimination of all proposals for UWB techniques except for two. Those two include direct sequence and MB‐OFDM. These two techniques use very different methods to fill up the same allocated bandwidth and power spectral densities laid out by the FCC in 2002.

A. Direct Sequence Technique (DS)

This technique was the first to fall within the scope of UWB. It was first implemented in the late 1960s after key advancements in measurements technology. Murray Nicolson was the inventor of the tunnel diode constant false alarm rate receiver (CFAR), still in use today [1].

DS uses short impulses to send information. The shorter the pulses are in the time domain, the wider the bandwidth that these signals occupy in the frequency domain. This is due to the relationship defined by the Fourier Transform. In general, the greatest performance is obtained when these pulses occupy the most bandwidth.

B. Multiband Orthogonal Frequency Division Multiplexing (MBOFDM)

This technique differs from DS UWB in that instead of using short pulses to fill the bandwidth, a series of subcarriers are used to create a frame of data that is transmitted. Before OFDM, a majority of communications, such as AM or FM, involved taking a stream of data (or a serial data stream) and doing some baseband modulations in the time domain. Then taking that time domain signal and up converting it to higher frequencies for transmission. Instead, OFDM treats incoming streaming data (inherently

in the time domain) as if it is already in the frequency domain by grouping the data in a parallel stream. Then this parallel stream of data is orthogonally placed on a frequency spectrum to create an OFDM frame, hence the name orthogonal frequency division. Multiband-OFDM occurs when OFDM frames are interleaved in time to different frequencies (Figure 1.1).

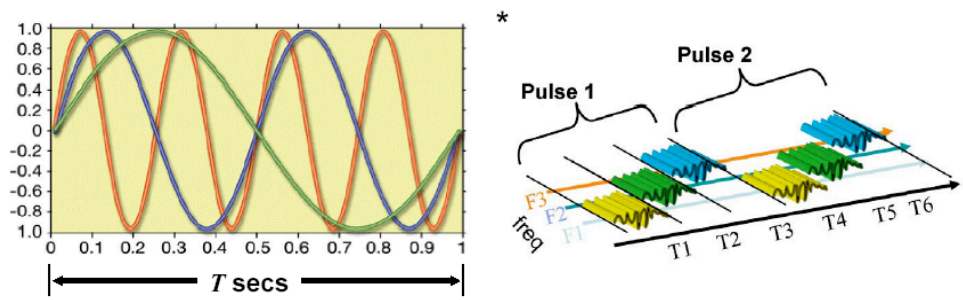


Figure 1.1 Schematic demonstration of multiband-OFDM. (Image courtesy: Intel)

3. The technique specifications of OFDM-UWB

Orthogonal Frequency-Division Multiplexing (OFDM) — essentially identical to Coded OFDM (COFDM) — is a digital multi-carrier modulation scheme, which uses a large number of closely-spaced orthogonal sub-carriers to carry data. These sub-carriers typically overlap in frequency, but are designed not to interfere with each other as would be the case with traditional FDM, and may be efficiently separated using a Fast Fourier Transform (FFT) algorithm. Each sub-carrier is modulated with a conventional modulation scheme (such as quadrature amplitude modulation) (Figure 1.2) at a low symbol rate, maintaining data rates similar to conventional single-carrier modulation schemes in the same bandwidth. OFDM has developed into a popular scheme for wideband digital communication, whether wireless or over copper wires, used in

applications such as digital television and audio broadcasting, wireless networking and broadband internet access.

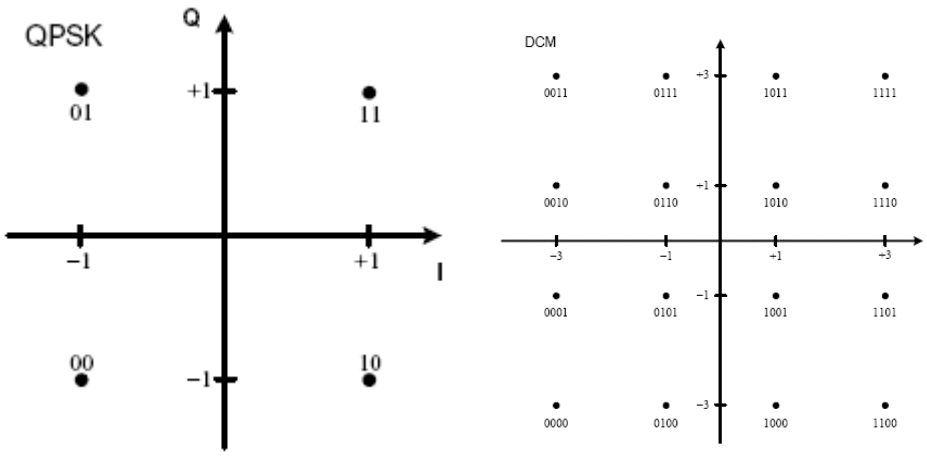


Figure 1.2 QAM Constellation (or π/4 QPSK Gray Coding) and 16 QAM (or DCM) Constellation

The primary advantage of OFDM over single-carrier schemes is its ability to cope with severe channel conditions — for example, attenuation of high frequencies in a long copper wire, narrowband interference and frequency-selective fading due to multipath — without complex equalization filters. Channel equalization is simplified because OFDM may be viewed as using many slowly-modulated narrowband signals rather than one rapidly-modulated wideband signal. The low symbol rate makes the use of a guard interval between symbols affordable, making it possible to handle time-spreading and eliminate intersymbol interference (ISI). This mechanism also facilitates the design of single-frequency networks, as the signals from multiple distant transmitters may be combined constructively, rather than interfering as would typically occur in a traditional single-carrier system.

An OFDM-based physical layer is one of the promising options for UWB devices due to its capability to capture multipath energy and eliminate inter-symbol interference and the fact that OFDM is considered by the industry as a mature and reliable technology [3]. For these reasons, our focus will be on the OFDM-based modulation scheme for UWB communications. Despite the merits mentioned above, the extremely short range, e.g., 10 meters for a data rate of 110 Mbps, puts UWB at an obvious disadvantage when compared to other competitive technologies, such as the soon coming IEEE 802.11n standard, which supports a data rate of 200 Mbps for at least 40 meters in indoor environments. Hence, in order to push UWB as an attractive option for Wireless Personal Area Network (WPAN) applications, it is crucial to improve the range limit of UWB devices. In a multiband-OFDM UWB system, the spectrum is divided into several sub-bands of bandwidth 528 MHz each. The system operates in one sub-band and then switches to another sub-band after a short time. The transmitted symbols aretime interleaved across the sub-bands to utilize the spectral diversity to improve the reliability of transmission. In each sub-band, OFDM modulation is used to transmit data symbols [5]. At the transmitter, the bits from information sources are first whitened by the scrambler and then encoded by the convolutional encoder. In order to exploit time-frequency diversity and combat multipath fading, the coded bits are further interleaved according to some preferred time-frequency pattern, and the resulting bit sequence is mapped into constellation symbols and then converted into a block of N symbols $x[0], \ldots, x[N-1]$ by the serial-to-parallel converter. The N symbols are the frequency components to be transmitted using the N subcarriers of the OFDM modulator, and are

converted to OFDM symbols X[0], . . . , X[N − 1] by the unitaryinverse Fast Fourier Transform (IFFT). After adding a cyclic prefix of length P, the resulting N +P time-domain symbols are converted into a continuous-time signal x(t) for transmission.

Short-range communication systems (so-called wireless personal area network (WPAN) systems) with ranges of up to 10m are becoming popular in replacing cables and in enabling new consumer applications. Examples such as Bluetooth and ZigBee, which operate in the 2.4 GHz ISM band, have however a limited data rate, typically about 1 Mbps, which is insufficient for many applications like fast transfer of large files (e.g., wireless USB) and high-quality video streaming. In order to increase the data rate to several hundreds of Mbps, a higher bandwidth is preferred over a larger SNR. This became possible at the moment the FCC released spectrum for UWB in the US spanning from 3.1 to 10.6 GHz with an average transmit power level of only -41.3 dBm/MHz [1, 2]. Several proposals have since then been presented to realize a short-range high data rate communication link. At present, both direct-sequence impulse communication and multiband OFDM UWB systems are under consideration as a standard.

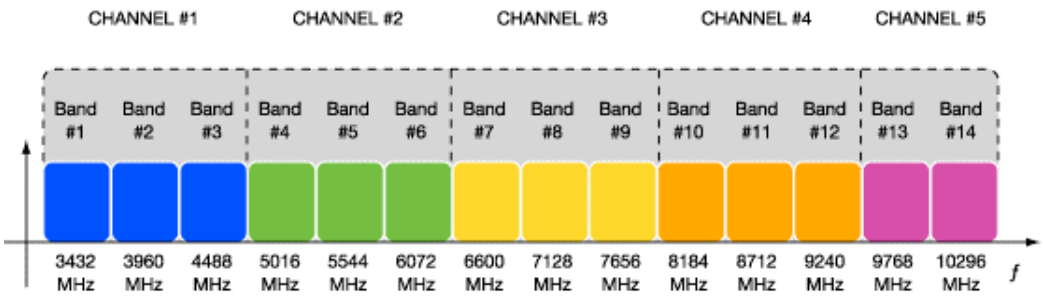


Figure 1.3 UWB frequency bands and their partitioning

The standard proposed by the multiband OFDM alliance (MBOA) is based on subdivision of the large available bandwidth in subbands of 528MHz (see Figure 1.3) [2]. The data is QPSK-OFDM modulated on 128 subcarriers. Various modes are defined with data rates up to 480 Mbps. In the mandatory mode of operation (Mode 1), a frequency-hopping scheme in the three lower bands is implemented. Using only the three lower bands allows the use of a bandpass prefilter to reduce the interferer level of the 5GHz ISM band. After each symbol period of 312.5 ns, a 9.5 ns guard time is available for hopping to the next band.

4. OFDM-UWB Transceiver Specifications

UWB receiver design is challenging, as it simultaneously requires a low noise density in a large bandwidth and a high linearity since large interferers can be present close to the used frequency band. An interferer scenario is required to determine the amount of filtering needed. On the transmit side, the challenge is in achieving a tunable, flat gain response over a 1.584 GHz bandwidth. Probably the most challenging block is the synthesizer due to the fast-hopping requirement.

A. Receiver requirements

For the receiver, the noise figure (NF) can be obtained from the system [NF.sub.system] according to NF = [NF.sub.system] - [IL.sub.prefilter] with [IL.sub.prefilter] the insertion loss of the prefilter. For a three-band system, the MBOA proposal assumes an [NF.sub.system] equal to 6 dB. For the 55 Mbps mode, this reflects a sensitivity level of -83.5 dBm with an SNR of -5 dB. For the highest data rate of 480 Mbps, the SNR is 6 dB and therefore the sensitivity level is increased to -72 dB. To

achieve graceful coexistence with other wireless technologies such as 802.11 WLAN and Bluetooth, an interferer robust receiver is needed. The MBOA interference scenario recommendation is depicted in Figure 1.4, indicating that even when a realistic 20 dB of prefiltering is taken into account, linearity requirements are severe. Most UWB systems target an input IP2 (iIP2) requirement above +20dBm and an iIP3 requirement in the order of -9 dBm.

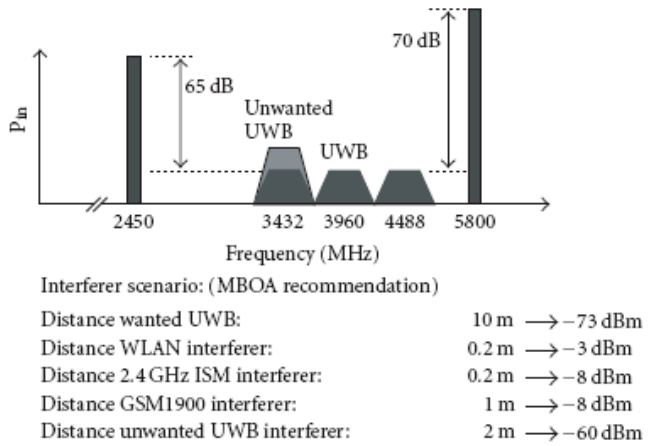


Figure 1.4 Interferer scenario. Indicated are received interferer powers.

Due to the strong interferers, there are severe filter requirements at IF as well. Consider the case where the closest 802.11a interferer is located only 398 MHz away from the edge of subband #3 centered at 4.488 GHz (5.15GHz - 4.752 GHz) at a distance of 0.2 m while the wanted UWB signal is transmitted from 10m distance. In such a case the filter has to provide more than 35 dB of attenuation relative to DC at 662 MHz offset. In a similar way, for the upper band of 802.11a an attenuation of 46 dB is required at an offset of 1.3G Hz.

B. Transmitter requirements

A key requirement for a UWB transmitter is that the spectral density limit of -41.3 dBm/MHz must be met. Based on this emission mask and the frequency hopping specification, the maximum transmit power can be calculated as -9.5 dBm. Assuming a power loss of about 2.5 dB between antenna and PA, the power that needs to be generated is -7.0 dBm. Study on the effect of nonlinearity on OFDM signals indicates that a 2-4 dB ensures acceptable degradation [3].

C. Synthesizer requirements

As the radio has to cover at least the lower three bands as defined in the MBOA and since most likely a zero-IF architecture is used, the synthesizer needs to provide quadrature signals at the center frequencies of the bands at 3432 MHz, 3960 MHz, and 4488 MHz. The I/Q mismatch must be lower than -30 dBc. In the MBOA proposal, frequency hopping between two subbands occurs once every symbol period of 312.5 ns. This period contains a 60.6 ns suffix, which is followed by a 9.5 ns guard interval in which the frequency hopping should be accomplished. The demands on the purity of the generated carriers are also stringent due to the presence of strong interferer signals. All spurious tones in the 5GHz range must be below -50 dBc to avoid down-conversion of strong out-of-band WLAN interferers into the wanted bands. For the same reason, the spurious tones in the 2GHz range should be below -45 dBc to allow co-existence with the systems operating in the 2.4 GHz ISM band, such as 802.11 b/g and Bluetooth. Finally, to ensure that the system SNR will not degrade by more than 0.1 dB due to intercarrier modulation, the overall integrated phase noise should not exceed 3.5 degrees rms. This

can be recalculated to a phase noise requirement of -100 dBc/Hz at 1MHz offset from the carrier.

D. RF Receiver building blocks

In addition to the receiver requirements, the low-noise amplifier (LNA) must provide broadband input matching and a broadband transfer. Several design options have been proposed in literature.

One possibility is to use a bandpass filter at the input in combination with an inductively degenerated (cascode) stage. In this way the reactive part of the input impedance will be cancelled over a wide frequency band [4, 5]. Implemented in a 0.18 [micro]m SiGe BiCMOS process, the LNA achieves an NF below 3 dB and an insertion gain above 20 dB. Distributed amplifiers also achieve wideband behavior. Where in mm-wave design coplanar wave guides or striplines are used to implement the transmission lines, silicon implementations use integrated inductors and capacitors as the lumped element replacement circuits for the transmission line.

An alternative CMOS LNA topology is presented in [9]. Here a common-gate input stage is loaded with three switched cascode devices with tanks resonating at the center frequency of each of the three bands. Note that the load switching must occur with the same speed as the hopping across the bands, that is, 9.5 ns. Noise figures between 5-7 dB and gains above 20 dB can be obtained.

Current feedback by means of a feedback resistor is also a quite commonly used method to broaden the bandwidth of the input match. In [10] a cascode topology including resistive feedback and a tuned load achieves an NF of 4 dB and a gain of 16 dB

in a 0.18 [micro]m CMOS process. Current feedback together with voltage feedback using an integrated transformer is demonstrated in [11]. This LNA consists of a cascode input stage (Q1 and Q2), followed by a voltage buffer (Q3 and Q4) known as a white emitter follower. There is voltage feedback by means of a transformer, formed by merging the collector coil and emitter degeneration coil of the input stage. In addition, there is current feedback formed by R1 and C1. This compound feedback mechanism gives high linearity, and also allows for matching of the input impedance to 50 [OMEGA] over the lower three bands, without the need for additional external matching components.

Due to the channel width of 528 MHz, most receivers apply a zero-IF architecture to relax the bandwidth requirements for the baseband filters and converters. In such an architecture, the LNA is in most cases directly followed by a (Gilbert) down-mixer. The subsequent mixer contains a combined common-emitter/common-base lower-stage, which is a well-known active balun structure [12]. It is highly degenerated by emitter resistors to obtain the required linearity. A fully balanced eight-transistor switching core has been used, which creates both the I and Q baseband signals. Noise caused by cross-conduction is reduced to a minimum by appropriate shaping of the LO drive signals. These signals should ideally be sinusoidal signals, but as they are the output of frequency dividers, they also contain higher-order harmonics.

A mixer with a variable gain range is demonstrated in [3, 9]. Here, the load resistor is decomposed into binary weighted segments so as to create dB steps in the gain. Implemented in 0.13 μm CMOS, a 30 dB gain is obtained over a large output bandwidth.

High-order filtering at IF/baseband is needed to achieve sufficient attenuation. The large bandwidth in combination with high linearity involves a careful distribution of gain, filtering, and noise. In [11], the baseband filter/VGA has been implemented as a fifth-order Chebyshev-like filter. The gain can be varied between 16 dB and 46 dB with 6 dB steps, and the bandwidth can be tuned in a range of 232 MHz to 254 MHz. At 662 MHz offset an attenuation of -57 dBr has been achieved. A fourth-order Sallen-and-Key filter has been used in [9], while in [13] a fifth-order elliptic filter has been used. In the latter case, the on-chip filter is a passive LC filter and, therefore, it is perfectly linear.

E. RF Transmitter building blocks

A crucial aspect of a UWB transmitter is the need of power control to ensure that the transmitted level does not exceed the -41.3 dBm/MHz limit (-14 dBm across 528 MHz). Furthermore, as with WLAN systems, RF impairments (e.g., I/Q mismatch, phase noise, carrier feed-through) must be kept to a minimum.

The RF power amplifier (PA) is in most cases based on an inductively loaded (cascode) transistor. An example is shown in Figure 1.5, where transistors M1 to M3 are used to implement a differential to single-ended structure [9]. Transistor M4 delivers an output level of -10 dBm. A straightforward approach has been used in [14], resulting in a -7 dBm output power level. By varying the bias, the gain of the amplifier can be varied with 6 dB.

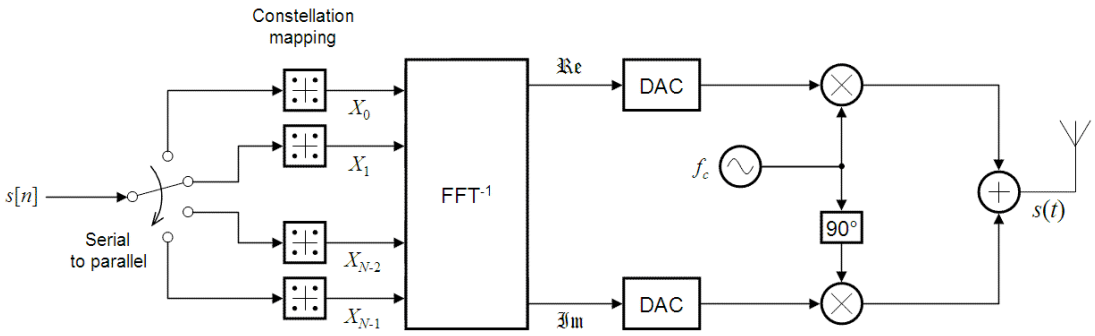


Figure 1.5 Scheme of a transmitter design.

Again, also the distributed amplifier has been proposed. In [15] a four-stage amplifier has been implemented in a 0.13 [micro]m CMOS process, resulting in a compression point of +3.5 dBm. In this case the transmission lines are implemented as micro-striplines.

Where in [9] an up-conversion circuit has been used based on resistively degenerated passive mixers along with a current feedback amplifier, two single-side-band Gilbert mixers have been used in [14]. The needed voltage-to-current converter as under stage for the Gilbert mixer core also implements a gain variation mechanism.

F. Fast-hopping synthesizer

A particularly challenging building block of the UWB receiver is the frequency synthesizer. A classical integer-N PLL with programmable loop divider ratio is unable to perform hopping within 9.5 ns, because such a PLL would require a loop bandwidth in the order of at least several hundreds of MHz and a reference frequency of several GHz. The high reference frequency contrasts the frequency resolution of 528 MHz. The high loop bandwidth, apart from being impractical, is in conflict with the phase noise demand

[17]. The same argument holds for a fractional-N PLL synthesizer, where the required high loop bandwidth is also hard to combine with the stringent spurious tone demands.

A straightforward frequency synthesizer architecture would be to use three separate PLLs (each generating one of the three required carrier frequencies) in combination with an output multiplexer. This is only practical in those cases where RC ring oscillators can fulfill the requirements. Three LC-oscillators-based PLLs will raise issues with respect to frequency pulling and occupation of die area. The option of using ring oscillators has been used in [9] for a three-band UWB system in a 0.13 [micro]m CMOS process, where each PLL consumes 15 mW from a 1.5 V supply voltage.

Most other proposed synthesizer concepts are based on frequency translation, where two frequencies can be added or subtracted by means of a single-sideband (SSB) mixer. Synthesizers using this method are also known as multi-tone generators.

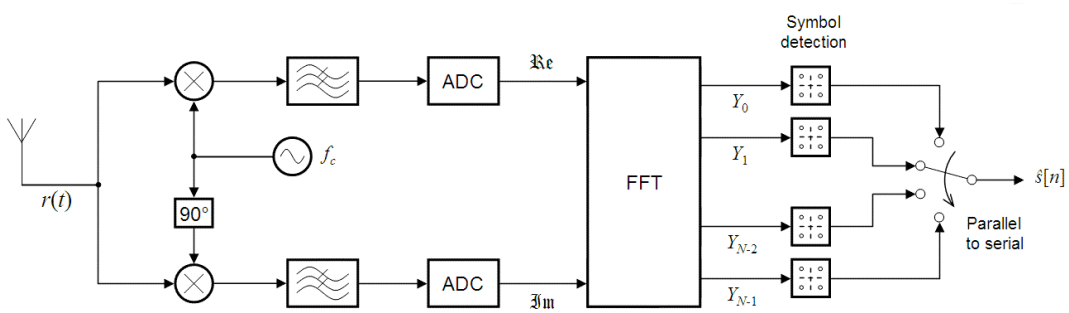


Figure 1.6 Scheme of a receiver design.

The problem of SSB mixing lies in the inherently generated spurious tones, for example, due to nonlinear behavior of the mixer. In this scheme the third harmonic of the 528 MHz signal (at 1584 MHz) is particularly troublesome because, after mixing with 3960 MHz, this harmonic will cause a spur at either 3960 + 1584 MHz = 5544 MHz or at

3960-1584 MHz = 2376 MHz. Both spurs are close to possible strong interferer signals (5 GHz and 2.4 GHz ISM bands, resp.) and this may result in UWB signal corruption. Because the 528 MHz signal is the output of a static divide-by-two circuit in the implementation of Figure 1.6, its harmonic content will inevitably be strong. Due to the use of quadrature signals, the third harmonic of +528 MHz is located at -1584 MHz. In [16] an integrated notch filter at the divide-by-two output (Figure 1.5) was used to place a notch at this frequency. In this way, all spurious tones in the 5GHz band are below -50 dBc, as can be seen from Figure 1.7. The fully integrated synthesizer consumes 73 mW from a 2.7 V supply and achieves frequency hopping within 1 ns.

To eliminate the need for two PLLs, the 3960 MHz signal needs to be divided by 7.5 to derive a 528 MHz signal. The challenge lies in the design of this divider, especially because of the need for quadrature signals with a 50% duty cycle. In [18] this is accomplished by two modified versions of the Miller divider, one realizing +3 and the other +2.5. The regenerative loop naturally leads to quadrature outputs and 50% duty cycle. Realized in 0.18 μm CMOS, the image suppression of the divider is -20 dBc while consuming 18 mW from a 1.8 V supply

One other possibility is demonstrated in [19]. Division by 7.5 has been realized using a frequency divider by 1.5 and a subsequent divider by 5 with postprocessing to make clean quadrature signals (Figure 1.6). The single PLL, single SSB mixer concept consumes 52 mW from a 2.7 V supply. Again due to additional filtering, out-of-band spurious tones are below -50 dBc. The integrated phase noise is below 2 degrees rms and the measured hopping speed is well below the required 9.5 ns.

In literature several proposals have been published in case the higher frequency bands also must be covered. A seven-band synthesizer based on two PLLs and one inductively loaded SSB mixer has been published in [20]. Fabricated in a 0.18 [micro]m CMOS technology, it achieves a sideband rejection of 37 dB. Covering the same bands can also be achieved using a 16GHz VCO, 2 SSB mixers, and only divide-by-two blocks [13]. A 12-band architecture based on three PLLs and two SSB mixers has been proposed in [3]. Multiplexing and routing of all RF signals will be challenging in this concept.

G. RF Transceiver for MB-OFDM UWB

As said, due to the wide channel bandwidth, the receiver and transmitter signal paths of UWB systems naturally employ direct conversion, that is, zero-IF. Such a direct conversion 3-band OFDM UWB transceiver has been demonstrated in [9]. The receiver consists of an LNA, quadrature mixers, a fourth-order Sallen-and-Key filter, and a first-order low-pass stage. The LO frequencies are synthesized using three independent PLLs using a 66MHz reference frequency. This allows a wide PLL loop bandwidth to suppress VCO phase noise. It is important to note that the LNA and PA share the same pin connected to the antenna. Designed in a 0.13 μm CMOS technology, this transceiver provides a total gain in the range of 69 to 73 dB and an NF in the range of 5.5 to 8.4 dB across the three bands. The circuit consumes 105 mW from a 1.5 V supply.

Direct conversion architecture for seven-band OFDM UWB has been proposed in [13]. The seven carrier frequencies are generated from a single 16 GHz VCO. The circuit has been fabricated in a 0.18 μm SiGe BiCMOS process and achieves an NF of 3.3-4.1

dB and a conversion gain of 52 dB. The current consumption is 88 mA from a 2.7 V supply.

A fully integrated receiver front end has been integrated in a SiGe BiCMOS technology with an NPN-f_T of 70 GHz [21]. The block diagram and chip micrograph are shown in Figure 1.8. The chip with a total area of 4 mm^2 has been packaged in an HVQFN package and mounted on an FR4 board. Digital control blocks for tuning the VCOs and the IF filter as well as a bandgap unit have also been implemented. The measured performance indicates that low noise figures can be achieved for complete receivers. The transmit chain is published in [14] and features wideband elliptic baseband filters, a VGA with dynamic range of 12 dB, an up-conversion mixer, and an RF output stage with a power of -7 dBm. The current consumption is 43 mA at 2.7 V for the complete transmit path.

Finally, some interesting studies on low-power UWB transceiver architectures have been presented in [22, 23]. The architectures are based on the use of distributed design approaches in the LNA and down-mixer circuits.

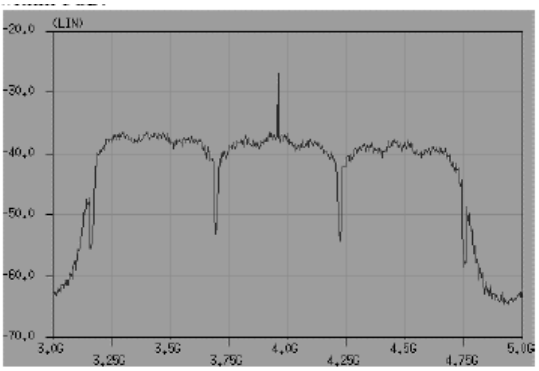


Figure 1.7 Measured output spectrum.

Chapter 2
RF Fundamentals and Ultra Wideband Radio

The goal of this chapter is to provide a review of the basic concepts of RF systems and circuits and a description of an ultra-wideband signal. At the end of this chapter, how to choose a proper ultra-wideband signal to ensure its power spectrum density (PSD) to meet FCC radiation power mask requirements will be discussed.

CMOS RF integrated circuit design resides at the convergence of two very different engineering traditions: the design of microwave circuits and systems, and the design of integrated circuits. As a consequence, a synthesis of these two traditions is required for CMOS RF engineer. In any RF systems, the matching, noise and linearity are the basic specifications of the system performance. Impedance matching is the major contributor to gain power efficiency. The noise characteristic affects the system sensitivity and bit error rate seriously. Linearity directly limits the system's working range and indirectly affects the system power consumption.

1. S-parameters and Smith Chart.

Systems can be characterized in numerous ways. To simplify analysis and perhaps elucidate important design criteria, it is often valuable to use macroscopic descriptions, which preserve input-output behavior but discard details of the internal structure of the system. At lower frequencies, the most common representations use impedance or admittance parameters, or perhaps a mixture of the two (called hybrid parameters). Impedance parameters allow one to express port voltages in terms of port currents. For the two-port shown in Figure 2.1, the relevant equations are:

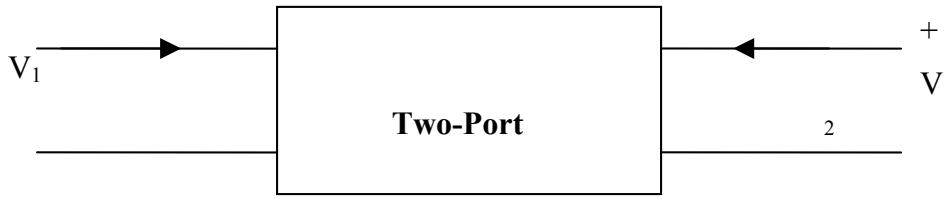


Figure 2.1. Port Variable Definitions

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \tag{2.1}$$

$$V_2 = Z_{21}I_1 + Z_{22}I_2 \tag{2.2}$$

It is very convenient to open-circuit the ports in succession to determine the various Z parameters experimentally, because various terms then become zero. For instance, determination of Z_{11} is easiest when the output port is open-circuited because the second term in Equation 2.1 is zero under that condition. Driving the input port with a current source and measuring the resulting voltage at the input allows direct computation of Z_{11}. Similarly, open-circuiting the input port, driving the output with a current source, and measuring V_1 allows determination of Z_{12}. Short-circuit conditions are used to determine admittance parameters, and a combination of open- and short-circuit conditions allows determination of hybrid parameters. The popularity of these representations for characterizing systems at low frequencies traces directly to the ease with which one may determine the parameters experimentally.

At high frequencies, however, it is quite difficult to provide adequate shorts or opens, particularly over a broad frequency range. Furthermore, active high-frequency circuits are frequently rather fussy about the impendence into which they operate, and may oscillate or even expire when terminated in open or short circuits. A different set of

two-port parameters is therefore required to evade these experimental problems. Called scattering parameters (or simply S-parameters), they exploit the fact that a line terminated in its characteristic impedance gives rise to no reflections. Interconnections between the instrumentation and the system under test can therefore be of a comfortable length. Since no short or open circuit needs to be provided, this greatly simplifies fixturing.

As implied earlier, terminating ports in open or short circuits are convenience for low-frequency two port descriptions because various terms then become zero, simplifying the math. Scattering parameters retain this desirable property by defining input and output variables in terms of incident and reflected (scattered) voltage waves, rather than port voltages or currents.

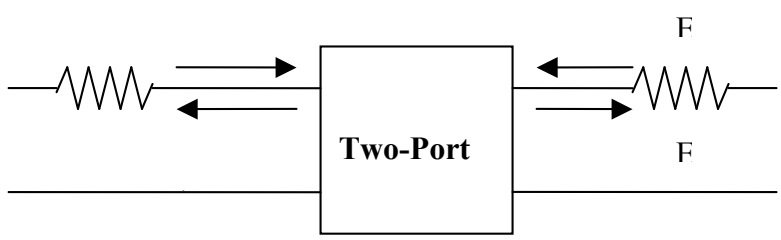


Figure 2.2. S-parameter Port Variable Definitions

As can be seen in Figure 2.2, the source and the load terminations are Z_0. With the input and output variables defined as shown, the two-port relations may be written as

$$b_1 = s_{11}a_1 + a_{12}a_2 \qquad (2.3)$$

$$b_2 = s_{21}a_1 + s_{22}a_2 \qquad (2.4)$$

where

$$a_1 = E_{i1} / \sqrt{Z_0} \tag{2.5}$$

$$a_2 = E_{i2} / \sqrt{Z_0} \tag{2.6}$$

$$b_1 = E_{r1} / \sqrt{Z_0} \tag{2.7}$$

$$b_2 = E_{r2} / \sqrt{Z_0} \tag{2.8}$$

The normalization by the square root of Z_0 is a convenience that makes the square of the magnitude of the various a_n and b_n equal to the power of the corresponding incident or reflected wave. Driving the input port with the output port terminated in Z_0 sets a_2 equal to zero, and allows us to determine the following parameters:

$$s_{11} = \frac{b_1}{a_1} = \frac{E_{r1}}{E_{i1}} = \Gamma_i \tag{2.9}$$

$$s_{21} = \frac{b_2}{a_1} = \frac{E_{r2}}{E_{i1}} \tag{2.10}$$

Thus, s_{11} is simply the input reflection coefficient, while s_{21} is a sort of gain since it relates an output wave to an input wave. Specifically, its magnitude squared is called the forward transducer power gain with Z_0 as source and load impedance. Similarly, terminating the input port and driving the output port yields

$$s_{22} = \frac{b_2}{a_2} = \frac{E_{r2}}{E_{i2}} = \Gamma_r \tag{2.11}$$

$$s_{12} = \frac{b_1}{a_2} = \frac{E_{r1}}{E_{i2}} \tag{2.12}$$

Here, we see that s_{22} is the reflection coefficient; s_{12} is the reverse transmission, whose magnitude squared is the reverse transducer power gain with Z_0 as source and load impedance. Once a two-port has been characterized with S-parameters, direct design of systems may proceed in principle without knowing anything about the internal workings of the two-port. For example, gain equations and stability criteria can be recast in terms of S-parameters. However, it is important to keep in mind that a macroscopic approach necessarily discards potentially important information, such as sensitivity to parameter or process variation. For this reason, S-parameter measurements are often used to derive element values for models whose topologies have been determined from first principles or physical reasoning.

To summarize, the reasons that S-parameters have become nearly universal in high-frequency work are that "zero"-length fixturing cables are unnecessary, there is no need to synthesize a short or open circuit, and terminating the two-port in Z_0 greatly reduces the potential for oscillation. Many graphical aids for S-parameter computation have been devised. Of these, the most generally useful has been one presented by Smith [7], which consists of loci of constant resistance and reactance plotted on a polar diagram in which radius corresponds to magnitude of reflection coefficient. Loci of constant resistance are circles and loci of constant reactance are circles orthogonal to those of constant resistance. The chart enables one to understand the behavior of complex impedance-matching techniques and to devise new ones. Here is an example for the chart's use. We define normalized impedance:

$$r + jx = \frac{Z_i}{Z_0} \qquad (2.13)$$

and a complex variable w equal to the reflection coefficient at the input port:

$$w = u + jv \qquad (2.13)$$

and then [8]

$$r + jx = \frac{1+w}{1-w} = \frac{1+(u+jv)}{1-(u+jv)} \qquad (2.14)$$

Equation 2.14 can be separated into real and imaginary parts as follows:

$$(u - \frac{r}{1+r})^2 + v^2 = \frac{1}{(1+r)^2} \qquad (2.15)$$

$$(u-1)^2 + (v - \frac{1}{x})^2 = \frac{1}{x^2} \qquad (2.16)$$

If we wish to plot the loci of constant resistance r on the w plane (u and v serving as rectangular coordinates), Equation 2.15 shows that they are circles with centers on the u axis at [$\frac{r}{1+r}$, 0] and with radii $\frac{1}{1+r}$. The curves for r=0, 1/2, 1, 2, ∞ are sketched in Figure 2.3. From Equation 2.16, similarly, we can draw the curves of constant x plotted on the w plane. The curves for x=0, $\pm 1/2, \pm 1, \pm 2, \infty$ are sketched in Figure 2.3 also.

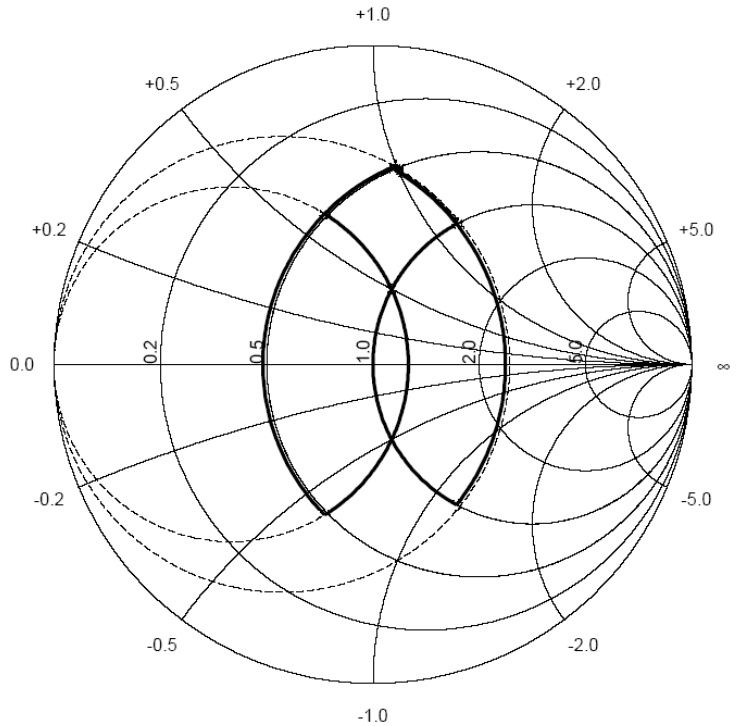


Figure 2.3. Basic Features of the Smith Chart

Several uses of the chart will be: Use as an admittance diagram; To find reflection coefficient given load impedance, and conversely; To find transform impedance along the transmission line; To find standing wave ratio and position of voltage maximum from a given impedance, and conversely; To Transform impedance along cascaded lines.

2. Noise.

The sensitivity of communication systems is limited by noise, and the system sensitivity can be denoted in terms of the minimum detectable signal (MDS) level at the antenna. To design a system with a good sensitivity, the noise characteristics must be analyzed, which is especially critical for UWB communication systems design. The broadest definition of noise as “everything except the desired signal”, for CMOS RFIC,

the noise is typically cataloged into 1. thermal noise, 2. shot noise, and 3. flicker noise. Thermal noise is the most important noise source in CMOS RFIC. The resistor, drain current, and gate thermal noise can be characterized respectively by:

$$\overline{e_{nR}^2} = 4kTR\Delta f \tag{2.17}$$

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f \tag{2.18}$$

$$\overline{i_{ng}^2} = 4kT\delta g_g\Delta f \tag{2.19}$$

where $g_g = \frac{\omega^2 C_{gs}^2}{5g_{d0}}$, where $\overline{e_{nR}}$ is the open-circuit rms noise voltage generated by the resistor R over the bandwidth Δf at a given temperature, k is Boltzmann's constant ($1.38\times10^{-23} J/K$), T is the absolute temperature in kelvins, and Δf is the noise bandwidth in hertz over which the measurement is made. g_{d0} is the drain-source conductance at zero V_{DS}. The parameter γ has a value of unity at zero V_{DS} and, in long devices, decreased toward a value of 2/3 in saturation, and δ has a value of 4/3 for the gate noise coefficient in long channel devices.

Shot noise is fundamental caused by the granular nature of the electronic charge, but how this granularity translates into noise is perhaps not as straightforward as one might think. In fact, shot noise can be expressed:

$$\overline{i_n^2} = 2qI_{DC}\Delta f \tag{2.20}$$

where $\overline{i_n^2}$ is the rms noise current, q is the electronic charge ($1.6\times10^{-19} C$), I_{DC} is the DC current in amperes, and Δf again the noise bandwidth in hertz. Flicker noise is the

most mysterious type of noise (also known as $1/f$ noise). Though no universal mechanism for flick noise has been identified, flick noise in resistors and MOSFET drain noise current is given by equation (2.21) and (2.22) respectively,

$$\overline{e_{nR}^2} = \frac{K_1}{f} \cdot \frac{R_o^2}{A} \cdot V^2 \Delta f \tag{2.21}$$

$$\overline{i_n^2} = \frac{K_2}{f} \cdot \frac{g_m^2}{WLC_{ox}^2} \cdot \Delta f \tag{2.22}$$

where A is the area of the resistor, R_o is the sheet resistivity, V is the voltage across the resistor, and K_1 is a material-specific parameter, and K_2 is a device-specific constant. W is the transistor width, L is the transistor length.

3. Signal Distortion and Dynamic Range.

The signal distortion and the dynamic range are other important performance of a system. If noise of the radio receiver sets its sensitivity, then the signal distortion that is created by a system's nonlinearity sets the maximum signal level of the system. Although most of the system design uses a linear approximation, that model is not valid in its non-linear region. This signal distortion can be specified by 1dB compression point. As the input signal approaches to the system's saturation region, the system signal gain begins to degrade. The point where the gain of the system deviated from its linear approximation by 1dB is called 1dB compression point. This 1dB compression point can be used to be defined the extent of the linear region. Therefore the system's dynamic range can be defined as the power difference from the noise level to the 1dB compression point. Figure

2.4 shows how the dynamic range is set from the relationship between input and output power.

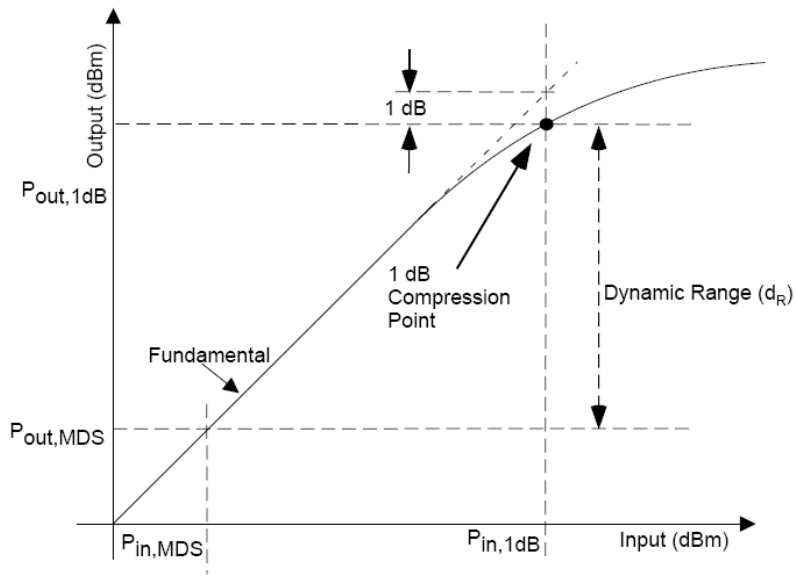


Figure 2.4. The Dynamic Range of a System from the Input and Output Power Relation

Two other characteristic of a narrow bandwidth system are the intermodulation distortion (IMD) and the third order intercept point (IIP_3). These characteristics come from the result of applying two unmodulated sinusoidal signals of slightly different frequencies to the input of a system. When two signals with different frequencies are applied to a nonlinear system, the output exhibits some components that are not harmonics of the input frequencies. Applying an input consisting of two closely-spaced sinusoidal components $V_{in} = A\cos(\omega_1 t) + A\cos(\omega_2 t)$, then the nonlinear system output comes as a power series $V_{out} = k_1 V_{in} + k_2 V_{in}^2 + k_3 V_{in}^3$ where k_1, k_2 and k_3 are gain. Expending this output, it contains several distortion products at frequencies $n\omega_1 + n\omega_2$ where n + m is

the order of the distortion product. The amplitude of each product varies as A^{n+m} so, second-order products vary in proportion to A^2 and third-order products in proportion to A^3 . Note that in a differential implementation, the even-order distortion is cancelled. Thus the third-order intermodulation distortion merits special attention. Figure 2.5 illustrates the behavior of the third-order intermodulation products with input amplitude. With the input and output amplitudes plotted on a log scale, the intermodulation product amplitudes follow straight line trajectories with slopes give by the order of the products. By extrapolating, intercept point can be found that serve as a characteristic for the linearity of a system.

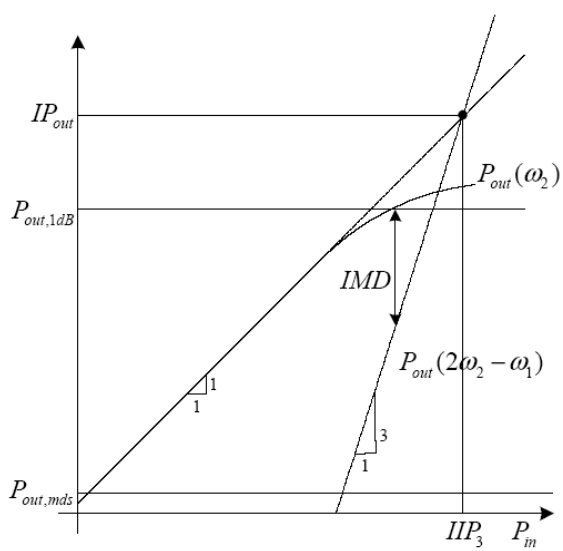


Figure 2.5. Intermodulation Behavior Based on Input-output Power Relation

In current communication systems, there are two different way to define a system's linearity. One is the dynamic range and the other is the IIP_3 as described above. Between these two definitions, the dynamic range, which is defined by the 1dB

compression point, is more general term for the linearity. Since the 1dB compression point is more general term for the linearity. Since the 1dB compression point can be determined by the system's gain saturation, this definition can be used both the narrow bandwidth and the wide bandwidth system. Therefore the dynamic range is more important in a none-carrier UWB system than the IIP_3.

Chapter 3
Power Amplifier Design

1. Distributed Amplifier

Distributed amplifiers are circuit designs that incorporate transmission line theory into traditional amplifier design to obtain a larger gain-bandwidth product than is realizable by conventional circuits.

The design of the distributed amplifiers is first formulated by William S. Percival in 1936. [1] In that year Percival proposed a design by which the transconductances of individual vacuum tubes could be added linearly without lumping their element capacitances at the input and output, thus arriving at a circuit that achieved a gain-bandwidth product greater than that of an individual tube. Percival's design did not gain widespread awareness however, until a publication on the subject was authored by Ginzton, Hewlett, Jasberg, and Noe in 1948. [2] It is to this later paper that the term distributed amplifier can actually be traced.

More recently, III-V semiconductor technologies, such as GaAs [3] [4] [5] and InP have been used. [6][7] These have superior performance resulting from higher band gaps (higher electron mobility), higher saturated electron velocity, higher breakdown voltages and higher-resistivity substrates. The latter contributes much to the availability of higher quality-factor (Q-factor or simply Q) integrated passive devices in the III-V semiconductor technologies.

To meet the marketplace demands on cost, size, and power consumption of monolithic microwave integrated circuits (MMICs), research continues in the development of mainstream digital bulk-CMOS processes for such purposes. The

continuous scaling of feature sizes in current IC technologies has enabled microwave and mm-wave CMOS circuits to directly benefit from the resulting increased unity-gain frequencies of the scaled technology. This device scaling, along with the advanced process control available in today's technologies, has recently made it possible to reach a transition frequency of 170 GHz and a maximum oscillation frequency of 240 GHz in a 90nm CMOS process. [8]

The operation of the DA can perhaps be most easily understood when explained in terms of the traveling-wave tube amplifier (TWTA). The DA consists of a pair of transmission lines with characteristic impedances of Z_0 independently connecting the inputs and outputs of several active devices. An RF signal is thus supplied to the section of transmission line connected to the input of the first device. As the input signal propagates down the input line, the individual devices respond to the forward traveling input step by inducing an amplified complementary forward traveling wave on the output line. This assumes the delays of the input and output lines are made equal through selection of propagation constants and lengths of the two lines and as such the output signals from each individual device sum in phase. Terminating resistors Z_g and Z_d are placed to minimize destructive reflections.

The transconductive gain of each device is gm and the output impedance seen by each transistor is half the characteristic impedance of the transmission line. So that the overall voltage gain of the DA is:

$$A_v = n * g_m * \frac{Z_0}{2} ,$$

Where n is the number of stages.

Neglecting losses, the gain demonstrates a linear dependence on the number of devices (stages). Unlike the multiplicative nature of a cascade of conventional amplifiers, the DA demonstrates an additive quality. It is this synergistic property of the DA architecture that makes it possible for it to provide gain at frequencies beyond that of the unity-gain frequency of the individual stages. In practice, the number of stages is limited by the diminishing input signal resulting from attenuation on the input line. Means of determining the optimal number of stages are discussed below. Bandwidth is typically limited by impedance mismatches brought about by frequency dependent device parasitic effects.

Delay lines are made of lumped elements of L and C. The parasitic L and the C from the transistors are used for this and usually some L is added to raise the line impedance. Due to the Miller effect in the common source amplifier the input and the output transmission line are coupled. For example for voltage inverting and current amplifying the input and the output form a shielded balanced line. Due to the current increasing in the output transmission line with every subsequent transistor, less and less L is added to keep the voltage constant and more and more extra C is added to keep the velocity constant. This C can come from parasitics of a second stage. These delay lines do not have a flat dispersion near their cut off, so it is important to use the same L-C periodicity in the input and the output. If inserting transmission lines, input and output will disperse away from each other.

For a distributed amplifier the input is feed in series into the amplifiers and parallel out of them. To avoid losses in the input, no input signal is allowed to leak through. This avoided by using a balanced input and output also known as push-pull amplifier. Then all signals which leak through the parasitic capacitances cancel. The output is combined in a delay line with decreasing impedance. For narrow band operation other methods of phase-matching are possible, which avoid feeding the signal through multiple coils and capacitors. This may be useful for power-amplifiers.

The single amplifiers can be of any class. There maybe some synergy between distributed class E/F amplifiers and some phase-matching methods. Only the fundamental frequency is used in the end, so this is the only frequency, which travels through the delay line version.

Due to this Miller effect a common source transistor acts as a capacitor (non inverting) at high frequencies and has an inverting transconductance at low frequencies. The channel of the transistor has three dimensions. One dimension, the width, is chosen depending on the current needed. The trouble is for a single transistor parasitic capacitance and gain both scale linearly with the width. For the distributed amplifier the capacitance - that is the width - of the single transistor is chosen based on the highest frequency and the width needed for the current is split across all transistors.

2. Power Amplifier

The term "power amplifier" is a relative term with respect to the amount of power delivered to the load and/or sourced by the supply circuit. In general a power amplifier is

designated as the last amplifier in a transmission chain (the *output stage*) and is the amplifier stage that typically requires most attention to power efficiency.

3. Power Amplifier Classes

The efficiency considerations lead to various classes of power amplifier. Power amplifier circuits (output stages) are classified as A, B, AB and C for analog designs, and class D and E for switching designs based upon the conduction angle or angle of flow, Θ, of the input signal through the (or each) output amplifying device, that is, the portion of the input signal cycle during which the amplifying device conducts. The image of the conduction angle is derived from amplifying a sinusoidal signal. (If the device is always on, Θ = 360°.) The angle of flow is closely related to the amplifier power efficiency. The various classes are introduced below, followed by more detailed discussion under individual headings later on.

Class A

100% of the input signal is used (conduction angle Θ = 360° or 2π); i.e., the active element remains conducting[5] (works in its "linear" range) all of the time. Where efficiency is not a consideration, most small signal linear amplifiers are designed as Class A. Class A amplifiers are typically more linear and less complex than other types, but are very inefficient. This type of amplifier is most commonly used in small-signal stages or for low-power applications (such as driving headphones). Subclass A2 is sometimes used to refer to vacuum tube Class A stages where the grid is allowed to be driven slightly positive on signal peaks, resulting in slightly more power than normal Class A (A1; where the grid is always negative[6]), but incurring more distortion.

Class B

50% of the input signal is used ($\Theta = 180°$ or π; i.e., the active element works in its linear range half of the time and is more or less turned off for the other half). In most Class B, there are two output devices (or sets of output devices), each of which conducts alternately (push–pull) for exactly 180° (or half cycle) of the input signal; selective RF amplifiers can also be implemented using a single active element.

These amplifiers are subject to crossover distortion if the transition from one active element to the other is not perfect, as when two complementary transistors (i.e., one PNP, one NPN) are connected as two emitter followers with their base and emitter terminals in common, requiring the base voltage to slew across the region where both devices are turned off.[7]

Class AB

Here the two active elements conduct more than half of the time as a means to reduce the cross-over distortions of Class B amplifiers. In the example of the complementary emitter followers a bias network allows for more or less quiescent current thus providing an operating point somewhere between Class A and Class B. Sometimes a figure is added (e.g., AB1 or AB2) for vacuum tube stages where the grid voltage is always negative with respect to the cathode (Class AB1) or may be slightly positive (hence drawing grid current, adding more distortion, but giving slightly higher output power) on signal peaks (Class AB2); another interpretation being higher figures implying a higher quiescent current and therefore more of the properties of Class A.

Class C

Less than 50% of the input signal is used (conduction angle $\Theta < 180°$). The advantage is potentially high efficiency, but a disadvantage is high distortion.

Class D

Main article: Switching amplifier

These use switching to achieve a very high power efficiency (more than 90% in modern designs). By allowing each output device to be either fully on or off, losses are minimized. The analog output is created by pulse-width modulation; i.e., the active element is switched on for shorter or longer intervals instead of modifying its resistance. There are more complicated switching schemes like sigma-delta modulation, to improve some performance aspects like lower distortions or better efficiency.

Additional classes

There are several other amplifier classes, although they are mainly variations of the previous classes. For example, Class G and Class H amplifiers are marked by variation of the supply rails (in discrete steps or in a continuous fashion, respectively) following the input signal. Wasted heat on the output devices can be reduced as excess voltage is kept to a minimum. The amplifier that is fed with these rails itself can be of any class. These kinds of amplifiers are more complex, and are mainly used for specialized applications, such as very high-power units. Also, Class E and Class F amplifiers are commonly described in literature for radio frequencies applications where efficiency of the traditional classes in are important, yet several aspects not covered

elsewhere (e.g.: amplifiers often simply said to have a gain of x dB - so what power gain?) deviate substantially from their ideal values. These classes use harmonic tuning of their output networks to achieve higher efficiency and can be considered a subset of Class C due to their conduction angle characteristics.

The classes can be most easily understood using the diagrams in each section below. For the sake of illustration, a bipolar junction transistor is shown as the amplifying device, but in practice this could be a MOSFET or vacuum tube device. In an analog amplifier (the most common kind), the signal is applied to the input terminal of the device (base, gate or grid), and this causes a proportional output drive current to flow out of the output terminal. The output drive current comes from the power supply.

The proposed Power Amplifier chooses to bias the amplifiers as class A mode. This is because class A amplifier has many advantages.

Amplifying devices operating in Class A conduct over the whole of the input cycle such that the output signal is an exact scaled-up replica of the input with no clipping. A Class A amplifier (or operational amplifier) is distinguished by the output stage (and perhaps the driver) device(s) being biased into Class A; even Class AB and B amplifiers normally have early stages operating in Class A. Class A is the usual means of implementing small-signal amplifiers, so the term Class A design applied to equipment such as preamplifiers (for example, in recording studios) implies not so much their use of Class A, but that their sound is top quality - good enough to be matched with top quality Class A power amplifiers.

Class A designs are simpler than other classes; for example Class AB and B designs require two devices (push-pull output) to handle both halves of the waveform, and circuitry to keep the quiescent bias optimal during temperature changes; Class A can use either single-ended or push-pull and bias is usually less critical.

The amplifying element is biased so the device is always conducting to some extent, normally implying the quiescent (small-signal) collector current (for transistors; drain current for FETs or anode/plate current for vacuum tubes) is close to the most linear portion (sometimes called the "sweet spot") of its characteristic curve (known as its transfer characteristic or transconductance curve), giving the least audio distortion.

Because the device is never shut off completely there is no "turn on" time, little problem with charge storage, and generally better high frequency performance and feedback loop stability (and usually fewer high-order harmonics).

The point at which the device comes closest to being cut off (and so significant change in gain, hence non-linearity) is not close to zero signal, so the problem of crossover distortion associated with Class AB and B designs is avoided, even in Class A double-ended stages.

4. Proposed Power Amplifier

A. Transistors biasing

Figure 3.1 shows the IV curve of the transistors which are biased in Class A mode and used in the proposed power amplifier. Equation 3.1. The transistors are biased in saturation region,

$$I_D = 50\% \times I_{DD}. \quad (3.1)$$

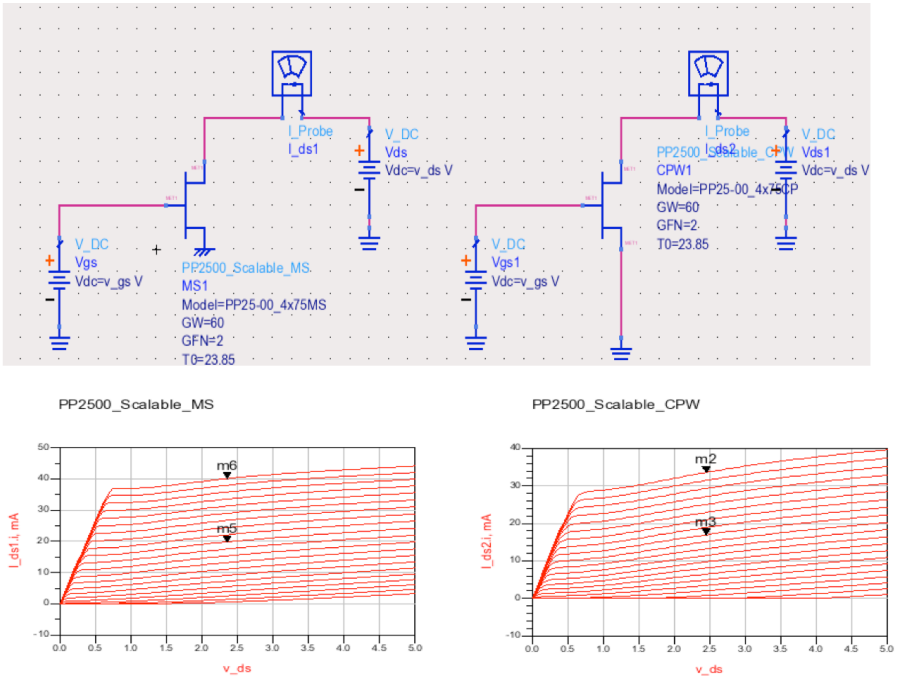


Figure 3.1 IV curve of the transistors biased in Class A mode

Figure 3.2 shows the schematic diagram of proposed Power Amplifier. The amplifier is formed by 10 stages. Each stage contains a conventional amplifier and is connected to each other by transmission lines through its drain and gate, which are so called Drain Line and Gate line correspondingly in the whole system. While an input signal is fed into the system, it will first go through the gate point of the first amplifier. At that point, part of the signal will be amplified in the first stage and sent to the drain line. The rest of the signal will then be transferred to the next stage. This part signal will be separated into two parts again. One of them will also be amplified in the second stage and sent to the drain line. The amplified signal will be added in to the signal coming from

previous stage along the drain line. The other part of the signal will be sent to the third stage and repeat the same routine. So as the signal propagated along the whole system, it will be amplified and added up. This will help to get higher gain bandwidth product.

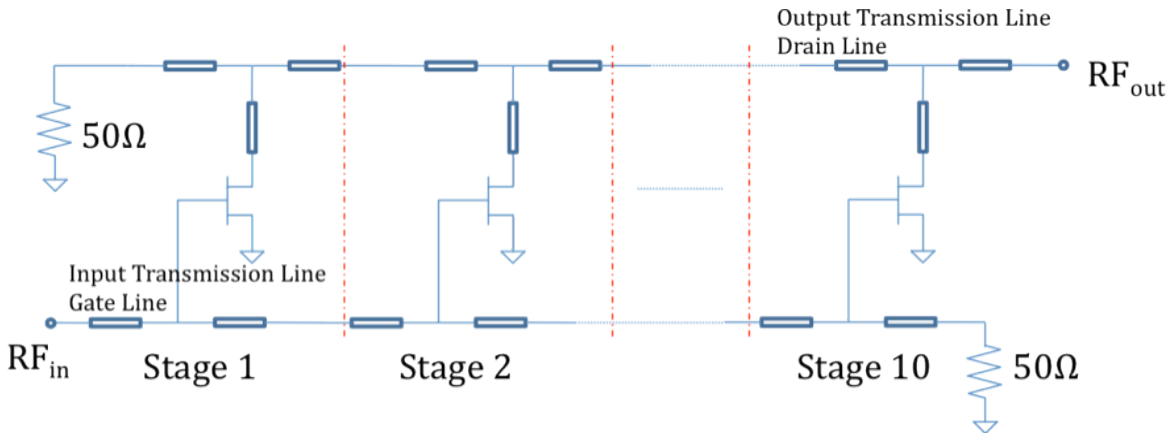


Figure 3.2 Schematic diagram of proposed Power Amplifier

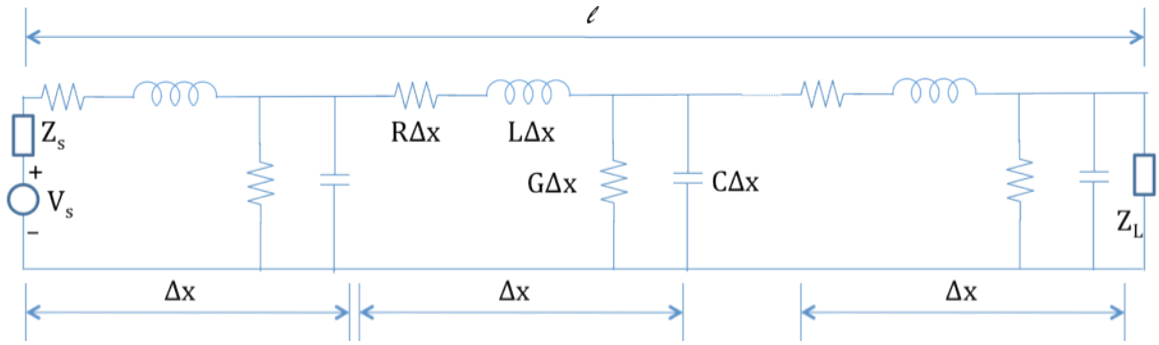


Figure 3.3 Equivalent circuit of a transmission line

B. Transmission lines

The transmission lines that used to connect each stage need to be designed carefully. Figure 3.3 shows the equivalent circuit of a transmission line. The electrical model of a transmission line is viewed as a two-port network, the transmission line receives power from the source at the input port (source end) and delivers power to the load at the output port (load end). The length *l* of the transmission line is divided into

many identical sections Δx. Each section Δx is modeled by a resistance R per unit length (R in Ω/m), an inductance L per unit length (L in H/m), a capacitance C per unit length (C in F/m), and a conductance G per unit length (G in S/m). These parameters are assumed to be constant along the transmission line. Figure 3.2.4 is the simplified equivalent circuit.

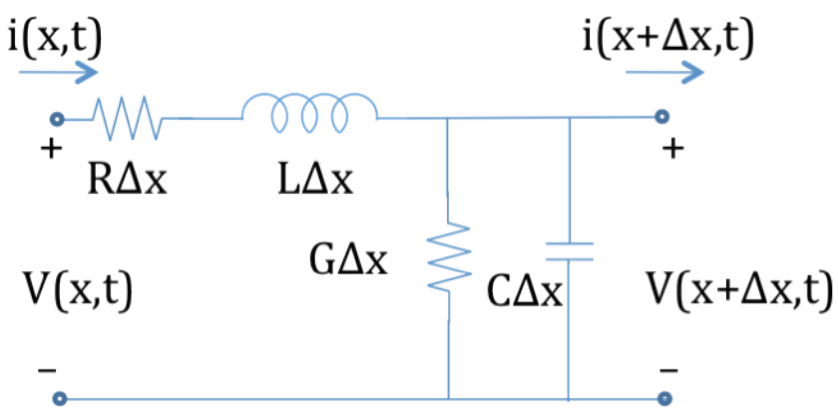


Figure 3.4 A section Δx of the transmission line and the voltage and currents at the input and output ports of the section Δx of transmission line

Figure 3.4 shows a section Δx of the transmission line and the voltage and currents at the input and output ports of the section Δx of transmission line. Observe that the voltages and currents along the transmission line are functions of position and time. At the input of the section Δx the voltage and current are $v(x, t)$, while at the output the voltage and current are $v(x+\Delta x, t)$ and $i(x+\Delta x, t)$.

Applying Kirchhoff's voltage law to the model in Figure 3.2.4 gives

$$v(x,t) - v(x+\Delta x,t) = R\Delta x i(x,t) + L\Delta x \frac{\partial i(x,t)}{\partial t}$$

Dividing by Δx we can write

$$\frac{v(x+\Delta x,t)-v(x,t)}{\Delta x} = -Ri(x,t) - L\frac{\partial i(x,t)}{\partial t}$$

As Δx approaches zero the left-hand side is recognized as the partial derivative of $v(x,t)$ with respect to x. Hence, taking the limit as $\Delta x \rightarrow 0$ we obtain

$$\frac{\partial v(x,t)}{\partial x} = -Ri(x,t) - L\frac{\partial i(x,t)}{\partial t} \qquad (3.2.1)$$

Similarly, applying Kirchhoff's current law to the model in Figure 3.2.4 gives

$$\frac{\partial i(x,t)}{\partial x} = -Gv(x,t) - C\frac{\partial v(x,t)}{\partial t} \qquad (3.2.2)$$

The partial differential equations (3.2.1) and (3.2.2) describe the voltages and currents along the transmission lines. Of particular practical interest in microwae electronics is the lossless transmission line – that is, a transmission line where R and G are negligible, or simply R = G = 0. A section Δx in a lossless transmission line is shown in Figure 3.5.

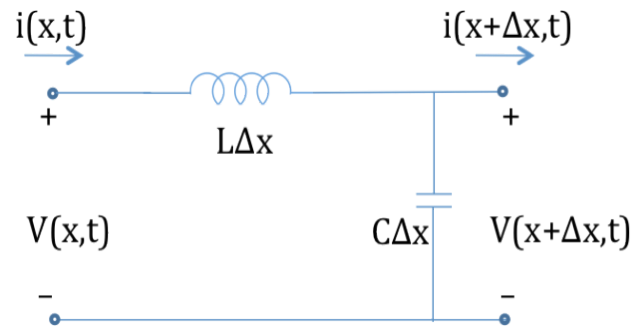


Figure 3.5 A section Δx in a lossless transmission line

In a lossless transmission line, R=G=0 and (3.2.1) and (3.2.2) reduce to

$$\frac{\partial v(x,t)}{\partial x} = -L\frac{\partial i(x,t)}{\partial t} \qquad (3.2.3)$$

$$\frac{\partial i(x,t)}{\partial x} = -C\frac{\partial v(x,t)}{\partial t} \qquad (3.2.4)$$

Define the phasor quatities

$$V(x) = f(x)e^{j\varphi(x)}$$

And

$$I(x) = g(x)e^{j\eta(x)}$$

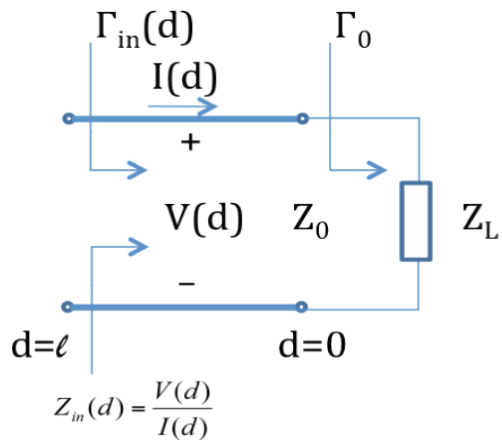


Figure 3.6 Input impedance of the transmission line

Referring to Figure 3.6, the input impedance of the transmission line at any position *d* is defined as

$$Z_{IN}(d) = \frac{V(d)}{I(d)} = Z_0 \frac{e^{j\beta d} + \Gamma_0 e^{-j\beta d}}{e^{j\beta d} - \Gamma_0 e^{-j\beta d}} \quad (3.2.5)$$

Where

$$\beta = \omega\sqrt{LC}$$

In equation 3.2.5 the constant Γ_0 can be evaluated using the boundary condition at the load, namely that the value of the input impedance at d=0 must be equal to Z_L. That is,

$$Z_{IN}(0) = Z_L$$

Then, from (3.2.5)

$$Z_{IN}(0) = Z_L = Z_0 \frac{1 + \Gamma_0}{1 - \Gamma_0}$$

Or

$$\Gamma_0 = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (3.2.6)$$

Substituting (3.2.6) into (3.2.5) gives

$$Z_{IN}(d) = Z_0 \frac{Z_L + jZ_0 tan\beta d}{Z_0 + jZ_L tan\beta d} \quad (3.2.5)$$

C. Cascode amplifiers

From Figure 3.2 we can see a distributed amplifier is obtained by several conventional amplifiers. These amplifier could be any type amplifier. In this design, cascode topology was choosen.

The cascode is a two-stage amplifier composed of a transconductance amplifier followed by a current buffer. Compared to a single amplifier stage, this combination may have one or more of the following advantages: higher input-output isolation, higher input impedance, higher output impedance, higher gain or higher bandwidth. In modern circuits, the cascode is often constructed from two transistors (BJTs or FETs), with one operating as a common emitter or common source and the other as a common base or common gate. The cascode improves input-output isolation (or reverse transmission) as there is no direct coupling from the output to input. This eliminates the Miller effect and thus contributes to a much higher bandwidth.

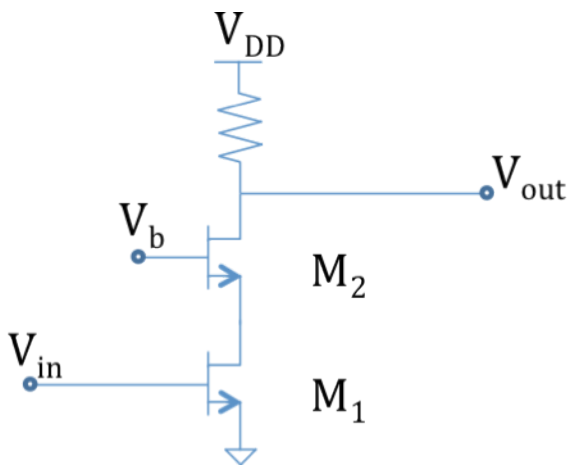


Figure 3.7 Cascode amplifier

Figure 3.7 shows an example of cascode amplifier with a common source amplifier as input stage driven by signal source Vin. This input stage drives a common gate amplifier as output stage, with output signal Vout. Figure 3.8 shows an equivalent circuit of a cascode amplifier.

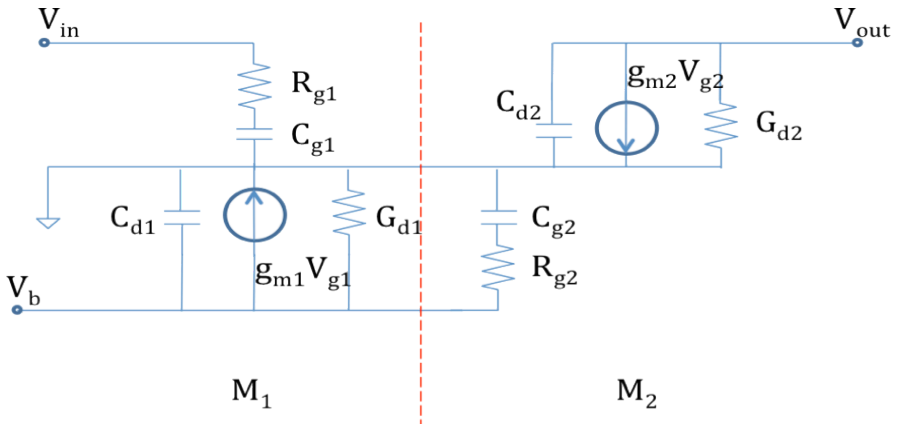


Figure 3.8 An equivalent circuit of a cascode amplifier

The major advantage of this circuit arrangement stems from the placement of the upper Field Effect Transistor (FET) as the load of the input (lower) FET's output terminal

(drain). Because at operating frequencies the upper FET's gate is effectively grounded, the upper FET's source voltage (and therefore the input transistor's drain) is held at nearly constant voltage during operation. In other words, the upper FET exhibits a low input resistance to the lower FET, making the voltage gain of the lower FET very small, which dramatically reduces the Miller feedback capacitance from the lower FET's drain to gate. This loss of voltage gain is recovered by the upper FET. Thus, the upper transistor permits the lower FET to operate with minimum negative (Miller) feedback, improving its bandwidth.

The upper FET gate is electrically grounded, so charge and discharge of stray capacitance Cdg between drain and gate is simply through RD and the output load (say Rout), and the frequency response is affected only for frequencies above the associated RC time constant: τ = Cdg RD//Rout, namely f = 1/(2$\pi\tau$), a rather high frequency because Cdg is small. That is, the upper FET gate does not suffer from Miller amplification of Cdg.

If the upper FET stage were operated alone using its source as input node (i.e. common-gate (CG) configuration), it would have good voltage gain and wide bandwidth. However, its low input impedance would limit its usefulness to very low impedance voltage drivers. Adding the lower FET results in a high input impedance, allowing the cascode stage to be driven by a high impedance source.

If one were to replace the upper FET with a typical inductive/resistive load, and take the output from the input transistor's drain (i.e. a common-emitter (CE) configuration), the CE configuration would offer the same input impedance as the

cascode, but the cascode configuration would offer a potentially greater gain and much greater bandwidth.

The cascode arrangement is also very stable. Its output is effectively isolated from the input both electrically and physically. The lower transistor has nearly constant voltage at both drain and source and thus there is essentially "nothing" to feed back into its gate. The upper transistor has nearly constant voltage at its gate and source. Thus, the only nodes with significant voltage on them are the input and output, and these are separated by the central connection of nearly constant voltage and by the physical distance of two transistors. Thus in practice there is little feedback from the output to the input. Metal shielding is both effective and easy to provide between the two transistors for even greater isolation when required. This would be difficult in one-transistor amplifier circuits, which at high frequencies would require neutralization.

As shown, the cascode circuit using two "stacked" FET's imposes some restrictions on the two FET's—namely, the upper FET must be biased so its source voltage is high enough (the lower FET drain voltage may swing too low, causing it to make saturation). Insurance of this condition for FET's requires careful selection for the pair, or special biasing of the upper FET gate, increasing cost.

The cascode circuit can also be built using bipolar transistors, or MOSFETs, or even one FET (or MOSFET) and one BJT. In the latter case, the BJT must be the upper transistor; otherwise, the (lower) BJT will always saturate (unless extraordinary steps are taken to bias it).

The cascode arrangement offers high gain, high slew rate, high stability, and high input impedance. The parts count is very low for a two-transistor circuit. The cascode circuit requires two transistors and requires a relatively high supply voltage. For the two-FET cascode, both transistors must be biased with ample VDS in operation, imposing a lower limit on the supply voltage.

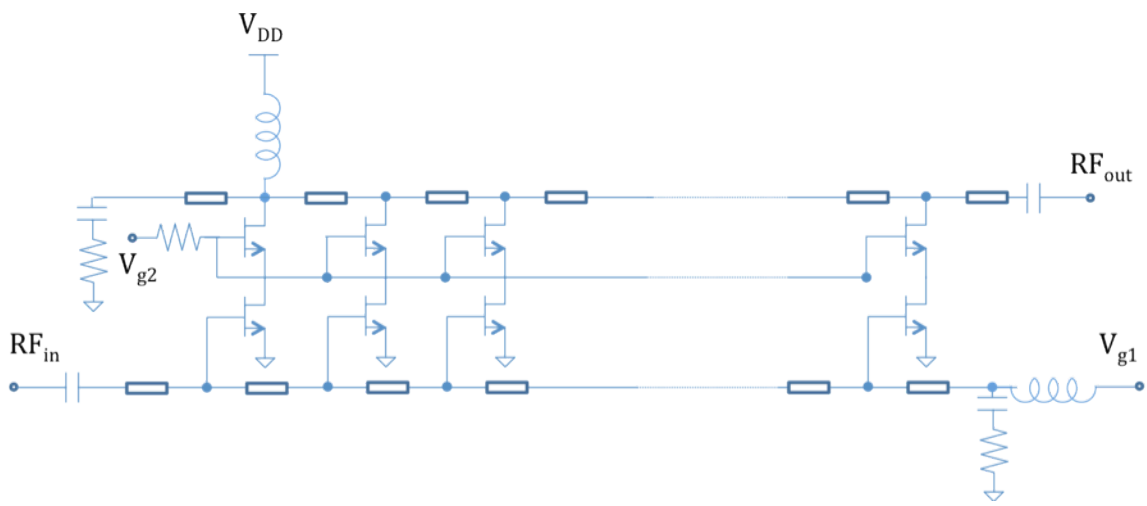


Figure 3.9 The propose power amplifier

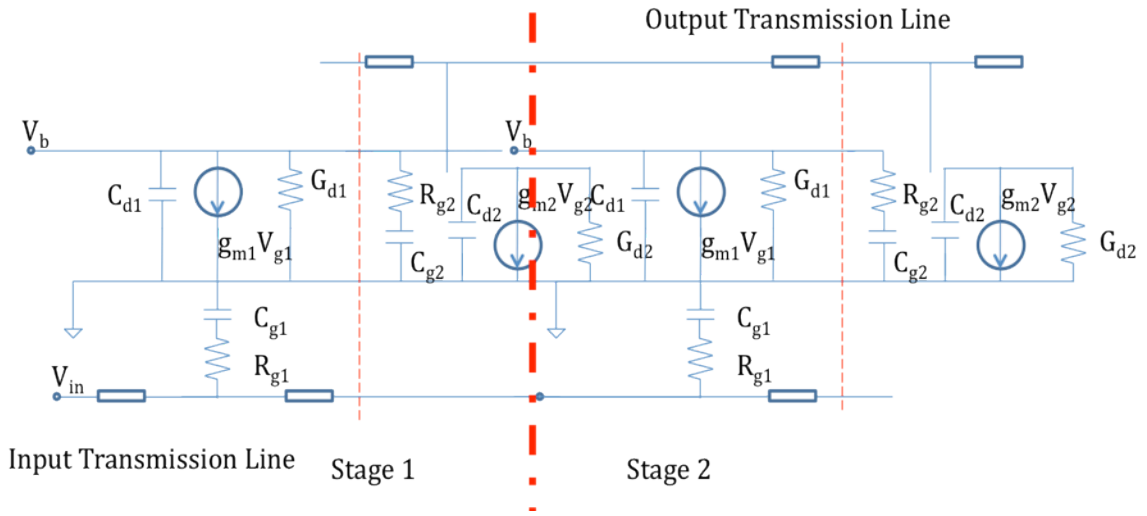


Figure 3.10 The equivalent circuit of 2 stages of figure 3.2.9.

Figure 3.9 shows the proposed power amplifier. Figure 3.10 is the equivalent circuit of 2 stages of Figure 3.9.

Figure 3.11 shows the simulated PA in ADS simulation schematic environment.

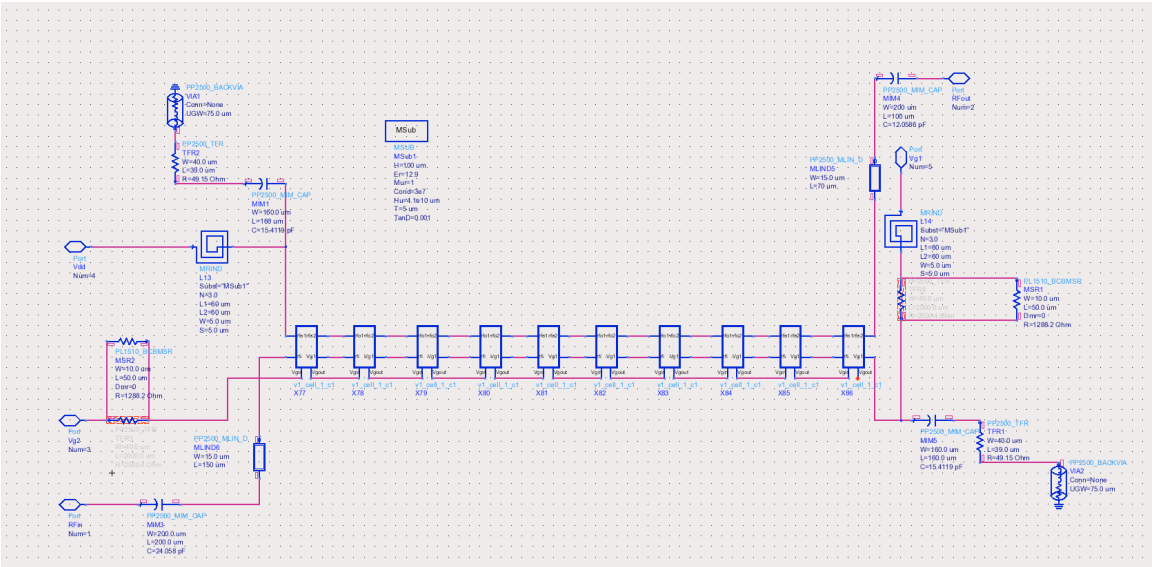


Figure 3.11 Simulated PA in ADS simulation schematic environment

As described in Chapter 2. In a RF circuits design, we used small signal test and large signal test to evaluate the circuit's performance. Both of the two testing environments are also simulated in ADS. Figure 3.12 shows the simulated small signal testing bench, and Figure 3.13 shows the large signal testing bench. Here I used inductors to simulate the bonding wires which are used to bond the on-chip pad to PCB board. Capacitors are used to simulate the bonding pads.

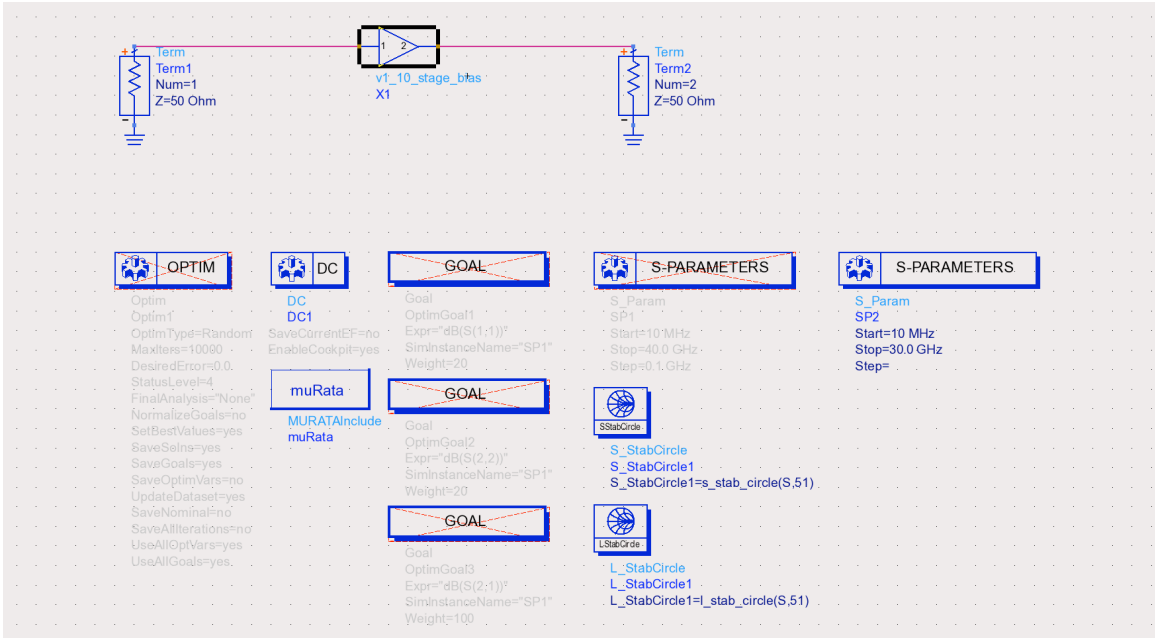


Figure 3.12 The simulated small signal testing bench

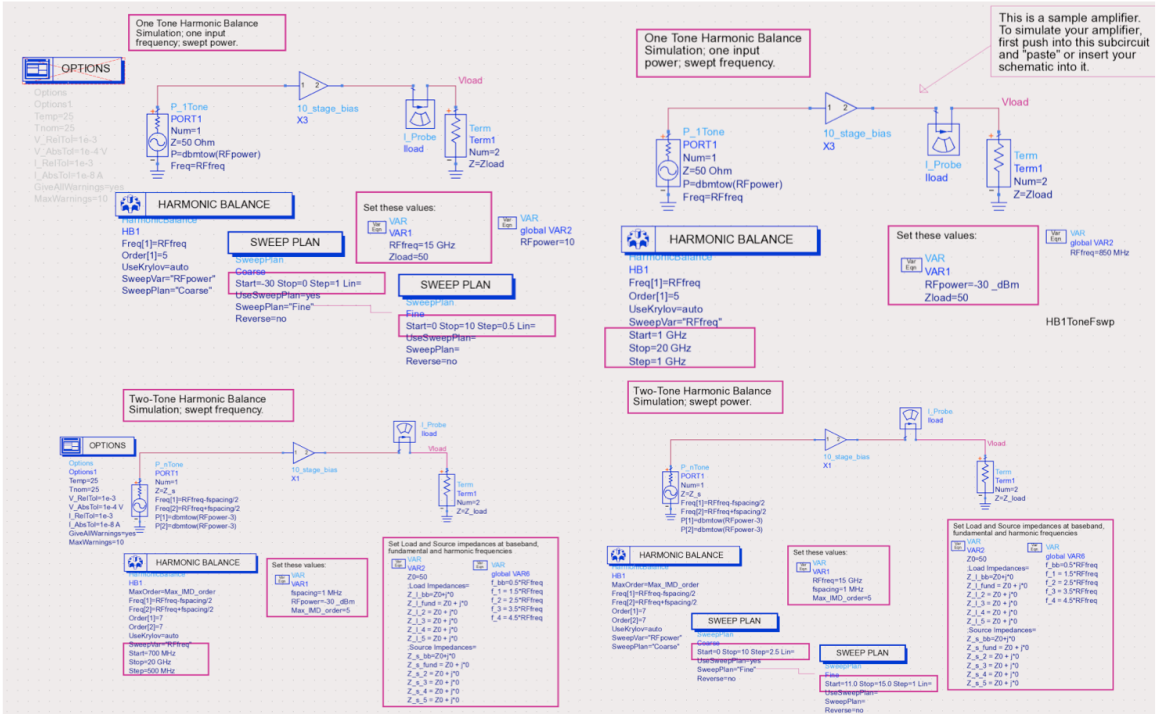


Figure 3.13 The large signal testing bench

Figure 3.14 shows the small signal gain of the proposed PA. Figure 3.15 shows the input return loss and output Return Loss of the PA.

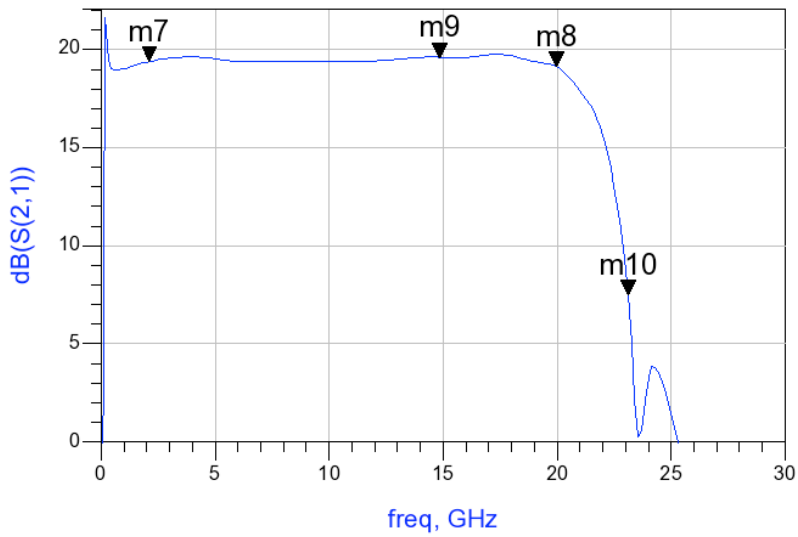


Figure 3.14 The small signal gain of the proposed PA

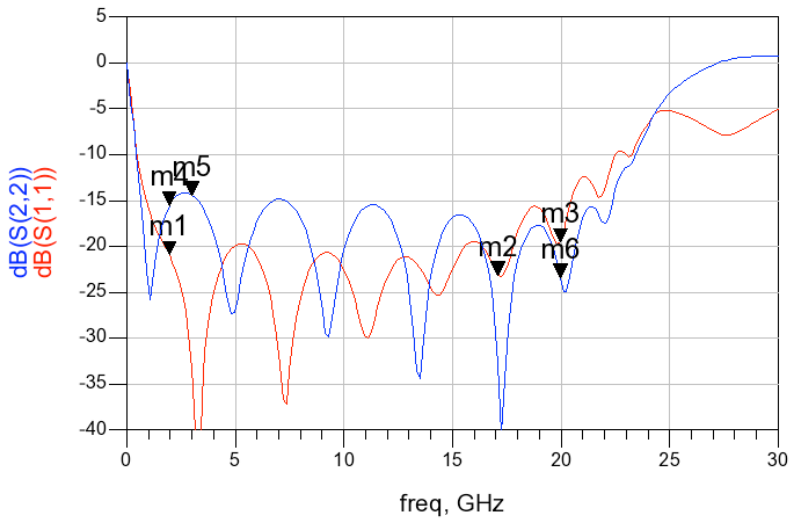


Figure 3. 15 The input return loss and output Return Loss of the PA

Figure 3.16 shows the OIP3 of the circuit. And Figure 3.17 shows the P-1dB of the PA.

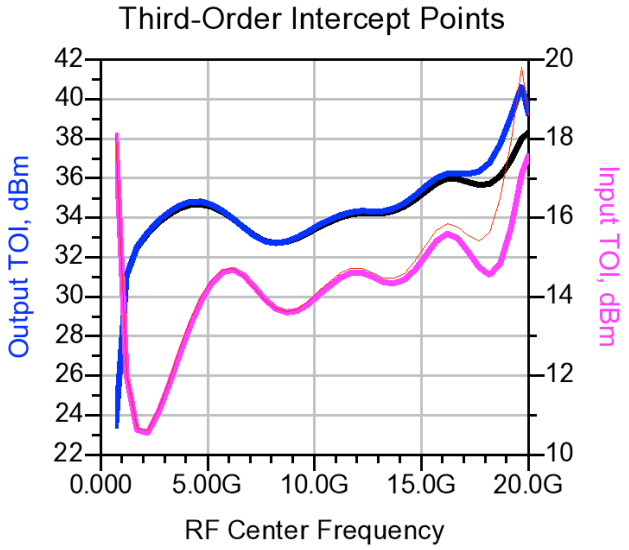


Figure 3.16 The OIP3 of the circuit.

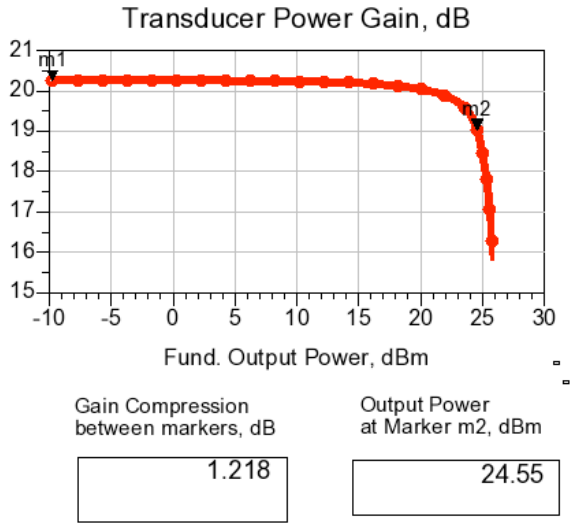


Figure 3.17 the P-1dB of the PA

Chapter 4
Low Noise Amplifier Design

1. Overview of Low Noise Amplifier

Low-noise amplifier (LNA) is an electronic amplifier used to amplify very weak signals (for example, captured by an antenna). It is usually located very close to the detection device to reduce losses in the feedline. This active antenna arrangement is frequently used in microwave systems like GPS, because coaxial cable feedline is very lossy at microwave frequencies.

An LNA is a key component which is placed at the front-end of a radio receiver circuit. Per Friis' formula, the overall noise figure of the receiver's front-end is dominated by the first few stages (or even the first stage only).

Equation 4.1

$$F_{total} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + \cdots$$

Using an LNA, the effect of noise from subsequent stages of the receive chain is reduced by the gain of the LNA, while the noise of the LNA itself is injected directly into the received signal. Thus, it is necessary for an LNA to boost the desired signal power while adding as little noise and distortion as possible, so that the retrieval of this signal is possible in the later stages in the system.

For low noise, the amplifier needs to have a high amplification in its first stage. Therefore JFETs and HEMTs are often used, and distributed amplifiers could be used.

They are driven in a high-current regime, which is not energy-efficient, but reduces the relative amount of shot noise. Input and output matching circuits for narrow-band circuits enhance the gain (see Gain-bandwidth product) and do not use resistors, as these would add noise. Biasing is done by large resistors, because energy efficiency is not needed, and a large resistor prevents leakage of the weak signal out of the signal path or noise into the signal path.

2. Low Noise Amplifier for UWB

Due to the unique natures of UWB radio, e.g., ultra wide bandwidth and extremely low signal power spectral density, the following challenges in designing UWB LNA must be addressed: sufficient voltage gain, good input matching, low noise figure, as well as minimized performance variation over the 3.1-10.6GHz spectrum. This thesis reports design and analysis of a single-chip 20GHz LNA for UWB radio.

Because of its ultra wideband nature, designing UWB LNA is very challenging compared with that for relatively narrow-band LNA for traditional frequency-domain RF transceivers. One of the critical tasks in broadband LNA design is to design proper resistive termination to realize across-ultra-wideband impedance matching between LNA input port and the output port of the driving source to achieve good input power and noise matching. However, because the input of a LNA circuit is typically a capacitive node, the varying parasitic capacitance of transistors makes it extremely difficult to realize good broadband impedance matching without degrading the noise performance and power delivery efficiency. Figure 4.1 illustrates four simple and typical LNA topology concepts based on their impedance matching methods: Figure 4.1(a) is a very simple resistive

termination topology, where the 50Ω resistance is directly connected to the input node of a common source amplifier. While reasonable broadband resistive matching is achieved, the terminating resistor would introduce too much thermal noise to the LNA circuit and attenuate the input signal to the LNA driving transistor port, resulting in an unacceptable high noise figure for the LNA circuit. Therefore, this resistive termination topology is not preferred in practical UWB LNA design. Figure 4.1(b) illustrates a $1/g_m$ LNA topology where the desired matching impedance is provided by the $1/g_m$ of the common-gate transistor. While this architecture is very simple, the required biasing conditions to deliver a $1/g_m$= 50Ω is usually different from that for optimum noise figure performance [23]. Figure 4.1(c) shows an inductive degeneration topology, where the desired resistive termination is realized at proper resonance frequency. This inductive LNA topology can achieve excellent noise figure [24], however, its narrow-band nature disqualified it for broadband UWB LNA applications. Figure 4.1(d) shows a closed-loop shunt-series feedback LNA topology. In contrast to the open-loop architectures, this topology breaks-up the troublesome global negative feedback and achieves proper trade-off between source impedance matching and noise figure performance. Hence, this shunt-series topology is attractive to broadband LNA design. In addition, the single-stage circuit ensures very low power dissipation compared to multi-stage topologies reported. One design challenge in using the shunt-series topology is to ensure circuit stability at the presence of the negative feedback.

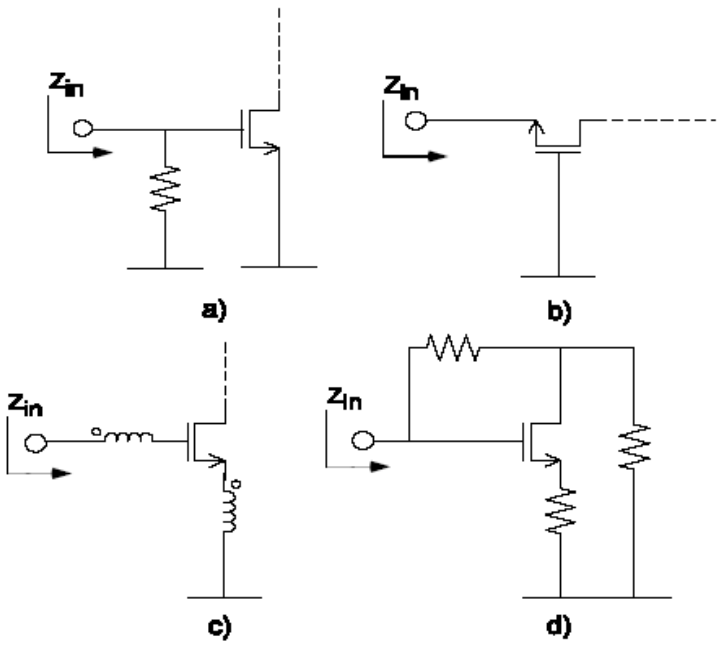


Figure 4.1 Simple LNA Topologies: (a) R-termination, (b) 1/gm -Termination, (c) Inductive Degeneration, and (d) Shunt-series Feedback.

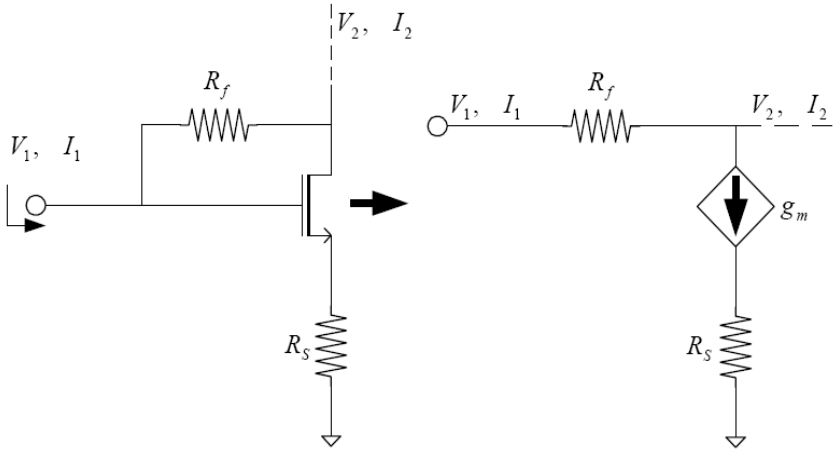


Figure 4.2 Small-signal Model for the Shunt-series Feedback Amplifier.

In this design, I used an improved shunt-series topology in each stage of the 20GHz UWB LNA for its simplicity and ability to achieve excellent full-band performance. Basic LNA circuit analysis follows: Using the small-signal equivalent circuit model in Figure 4.2, the two-port I-V transfer function for the LNA amplifier can be described by h-parameters as given in Equation 4.3,

$$\begin{pmatrix} V_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} I_1 \\ V_2 \end{pmatrix} = \begin{bmatrix} R_f & 1 \\ \dfrac{g_m R_f}{1+g_m R_s} - 1 & \dfrac{g_m}{1+g_m R_s} \end{bmatrix} \begin{pmatrix} I_1 \\ V_2 \end{pmatrix} \tag{4.3}$$

The h-parameter matrix can be further converted into the corresponding S-parameter format as in Equation 4.4,

$$[S] = \frac{1}{\Delta} \begin{bmatrix} \dfrac{R_f}{Z_0} - \dfrac{g_m Z_0}{1+g_m R_s} & 2 \\ 2(1 - \dfrac{g_m R_f}{1+g_m R_s}) & \dfrac{R_f}{Z_0} - \dfrac{g_m Z_0}{1+g_m R_s} \end{bmatrix} \tag{4.4}$$

Where the characteristic impedance is Z_0=50Ω and $\Delta = 2 + \dfrac{R_f}{Z_0} + \dfrac{g_m Z_0}{1+g_m R_s}$.

From the above S-parameter matrix, to achieve the ideal matching condition of S_{11}=S_{22}=0 for maximum signal power delivery, the series resistance R_s may be selected as,

$$R_s = \frac{Z_0^2}{R_f} - \frac{1}{g_m} \tag{4.5}$$

Substituting Equation 4.5 into 4.4 results in the following S-parameter expression,

$$[S] = \begin{bmatrix} 0 & \frac{Z_0}{R_f + Z_0} \\ 1 - \frac{R_f}{Z_0} & 0 \end{bmatrix} \tag{4.6}$$

Equations 4.4 and 4.6 reveal that the broadband forward gain of S_{21} can be flat and good matching can be achieved by choosing proper values for R_f and R_s.

3. LNA for ultra high frequency band

In addition to the desired performance uniformity of a broadband LNA, particularly for the 20GHz UWB bandwidth, it is hard, yet extremely critical to guarantee the stability of the ultra wideband LNA in design. To ensure the desired unconditional stability for the UWB LNA, the condition of $K_f > 1$, where K_f is given in Equation 4.7, must be satisfied in design.

$$K_f = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{21}| \cdot |S_{12}|} \tag{4.7}$$

Where, $\triangle = S_{11}S_{22} - S_{21}S_{12}$. Detailed analysis for the circuit stability will be given later.

Figure 4.3 and Figure 4.4 show the schematic and equivalent circuit models of each stage in the UWB LNA in this design. The previous analysis for the shunt-series feedback LNA circuit neglects all reactance effects in the circuit. In the real design, all the parasitic components (capacitance and inductance) must be taken into account, particularly for the full-band 3.1-10.6GHz UWB circuit, because such parasitic will adversely affect UWB LNA circuit behaviors, such as, gain, noise figure and stability, etc.

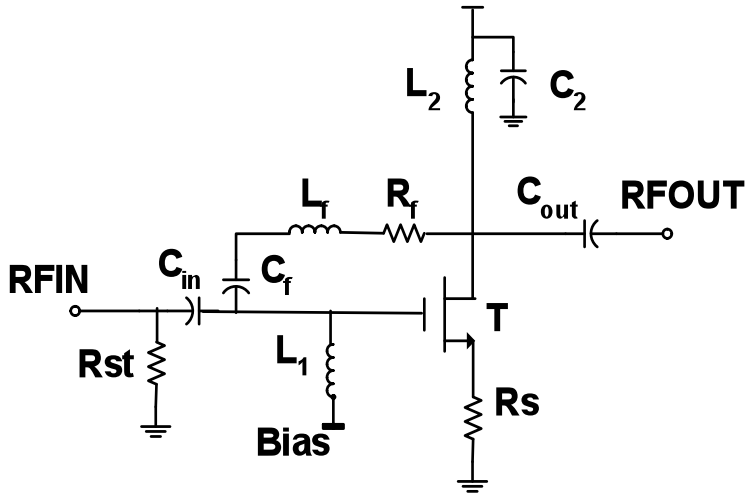
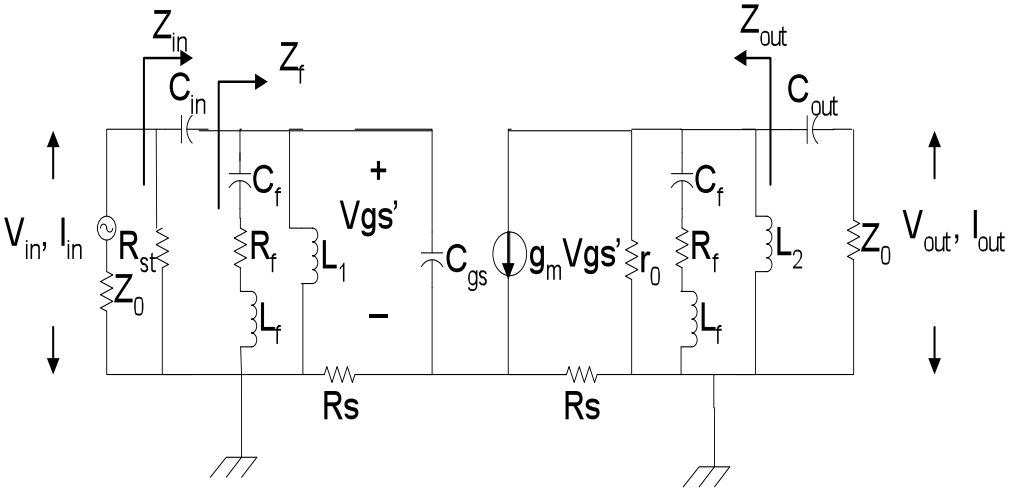
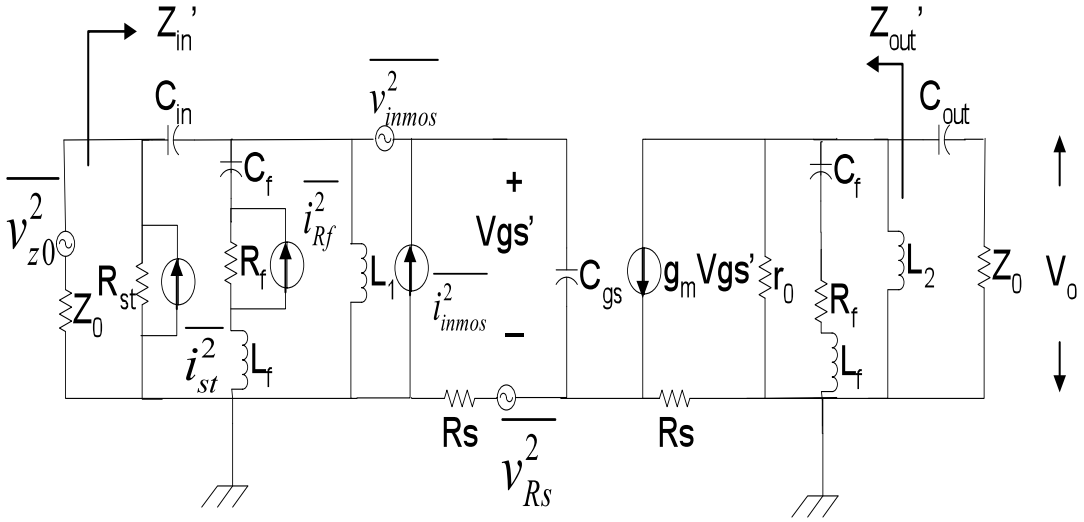


Figure 4.3. Improved Shunt-Series Feedback LNA Schematic in this Design.



(a)



(b)

Figure 4.4. UWB LNA Circuit Models: (a) Small-Signal Equivalent Circuit, (b) Equivalent Noise Circuit. Z_0–matching is assumed at Input & Output.

The R_S and R_f of the UWB LNA have to be selected very carefully to achieve a balanced maximum overall LNA circuit performance across the UWB bandwidth, instead

of one specific parameter. Mismatching-compensation technique is widely used at the input or output port devices of an UWB LNA circuit to address the frequency variation problem [25, 26], which leads to a frequency-dependent impedance at the input or output port. As a result, one would not get the ideal maximum power transfer of input signals over the desired ultra wideband frequency range. Hence, generally, the optimal condition given in Equation 4.5 for R_S needs not to be satisfied over the whole UWB frequency range since the maximum power matching is not always required in UWB LNA design. The ideal condition of Equation 4.5 is typically valid for low frequency, narrowband LNA region only where parasitic capacitor effect is not the most critical concern in the design. Accordingly, in this design, in order to achieve better noise performance for UWB LNA, an R_S=0 is selected for a well-balanced overall circuit performance, which, in the meantime, still ensures the best possible power matching for the LNA. With the R_S determined, the selection of R_f becomes critical to optimizing the UWB LNA circuit, where, again, one ought to consider a well-balanced overall circuit performance that includes noise figure, return loss, gain flatness, stability, as well as maximum power matching, which are tightly correlated with each other. Firstly, consider the influence of R_f on the minimum noise figure, NF_{min}. Figure 4.5 shows the simulated $NF_{min} \sim R_f$ relationship, which clearly shows that a close-to-zero R_f leads to minimum noise figure, while a middle-range R_f (several tens to a few hundreds ohms) results in the highest NF_{min}. A relatively large R_f (~ one thousand ohms) delivers reasonably good noise figure. A more comprehensive noise analysis for the UWB LNA circuit is provided in Section III. Secondly, consider the $S_{11} \sim R_f$ as given in Figure 4.6, it is readily observed that too

small a R_f (close to zero) induces too much return loss, a middle-range R_f ensures very low return loss, while a relatively large R_f would achieve a fairly reasonable S_{11}. Thirdly, considering the feedback behavior, too large an R_f will reduce the desired feedback effect significantly, which leads to substantial degradation in the voltage gain flatness over the ultra wide frequency spectrum. Next, one has to consider the important forward gain, S_{21}. Equation 4.6 clearly suggests that the magnitude of S_{21} is directly determined by the feedback resistance R_f, which prefers a large R_f for a desired high LNA gain. Lastly, yet very importantly, one must consider the unconditional circuit stability. A large R_f reduces the feedback effects, which in turn increases the LNA stability. Apparently, R_f is a critical parameter in UWB LNA design optimization. Based upon the above analysis and to ensure a well-balance overall optimum UWB LNA circuit performance, a moderately large feedback R_f of 700Ω is selected for optimized LNA circuit performance in terms of noise figure, gain, gain flatness and power delivery over the full 3.1-10.6GHz frequency range. Such an UWB LNA circuit design consideration is verified by LNA measurement results to be discussed in later Section.

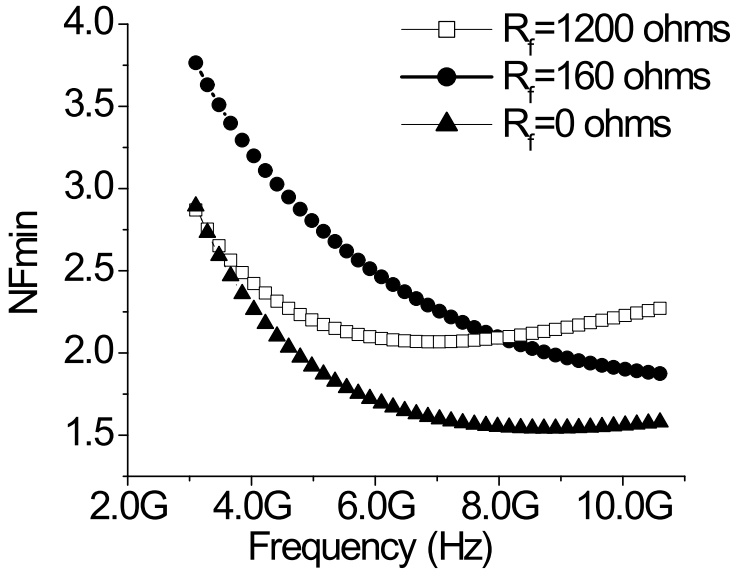


Figure 4.5. An Optimal R_f is Carefully Selected According to the NF_{min} Versus R_f Relationship in the UWB LNA Circuit Design.

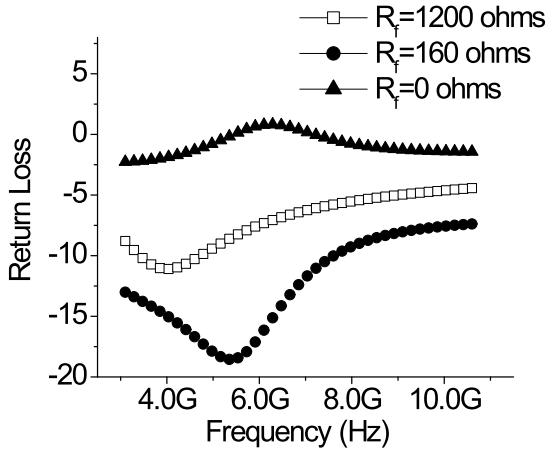


Figure 4.6. An Optimal R_f is Carefully Selected According to the S_{11} Versus R_f Relationship in the UWB LNA Circuit Design.

4. Other design considerations

Design consideration in selecting other circuit components for the UWB LNA shown in Figure 4.3 follows. Use of an inductor L_f in the feedback path helps relaxing the need for too large a feedback resistor R_f, since it helps to improve the gain and the gain flatness at high frequency end. The load inductance L_2 is used to replace the resistive load in the original shunt-series feedback configuration. Since the magnitude of the equivalent impedance of the inductor increases as frequency increases, it serves to compensate for the amplifier gain degradation that occurs in higher frequency range for UWB application. With this inductor load, the resonant frequency must be set outside of the operating frequency range to avoid the LNA becoming a narrow-bandwidth amplifier. One cannot simply use a very large inductor for large high-frequency gain compensation to achieve the desired broadband UWB gain flatness, because large inductance would result in a quite low self-resonance frequency. Adequate high-frequency gain compensation is realized by introducing the feedback inductor L_f into the circuit. Another important component in the circuit is the DC blocking capacitance C_f, which blocks the DC current coming from the output node. This blocking is necessary to separate the input and output voltages, hence, makes the input bias tunable so that it is possible to simultaneously achieve the optimal transistor biasing and maximum transistor g_m. This helps to reduce circuit power consumption while increases the gain.

Properly design of the DC operating point is equivalently important to the UWB LNA circuit, for which an optimal transistor gate-to-source bias, *i.e.,* V_{GS} as shown in Figure 4.4, is critical to achieving high transistor gain and low noise figure

simultaneously [27]. To further comprehend this point, the LNA circuit gain is derived as following,

$$A(s) = g_m Z_0 \frac{\frac{1}{sc_{gs}} \times Z_{in} \times Z_{out}}{(R_s + \frac{1}{sc_{gs}}) \times Z_{in} \times (Z_{out} + Z_0 + \frac{1}{sc_{out}})} \tag{4.8}$$

Where the FET transconductance is given by $g_m = \mu C_{ox} \frac{W}{L}(V_{GS} - V_{TH})$, Z_{in} is the equivalent input impedance looking rightward from shunt stability resistor R_{st}, Z_f is the equivalent input impedance looking into the feedback connection node at the import port, and Z_{out} is the equivalent output impedance. Clearly, a higher V_{GS} leads to a higher g_m and results in a larger circuit gain in the saturation region. However, too high a V_{GS} would force the transistor into high current mode that might significantly degrades the transconductance efficiency, noise figure and linearity performance in the high frequency region, which should be avoid in design. With the above consideration, an optimal V_{GS} of ~0.8V is selected for the UWB LNA circuit to achieve the best trade-off among g_m, gain, and noise figure and linearity range in this design.

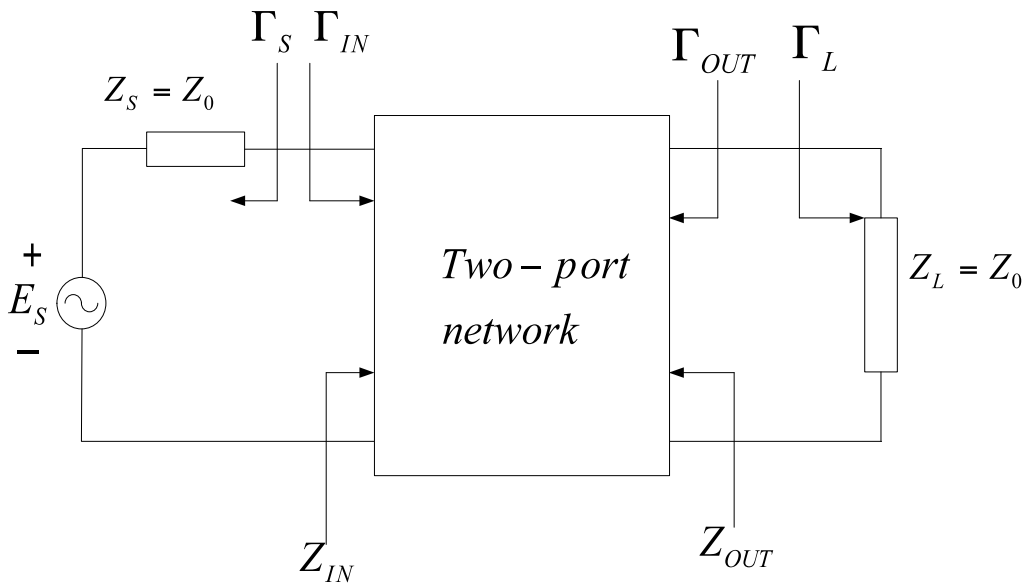


Figure 4.7. A Two-port Network for Circuit Stability Analysis.

Stability consideration is one of the most important, yet challenging issues in ultra broadband circuit design. In general, as is stated in the previous Section, the condition of K_f >1 must be met to ensure unconditional stability for the UWB LNA circuit. A comprehensive circuit stability analysis is provided in this Section. A typical two-port network for stability analysis is illustrated in Figure 4.7, where Γ_S is the source reflection coefficient, Γ_L is the load reflection coefficient, Γ_{IN} the input reflection coefficient, and Γ_{OUT} is the output reflection coefficient. Z_S, Z_L, Z_{IN}, and Z_{OUT} are the equivalent source impendence, equivalent load impedance, equivalent input impendence, and equivalent output impedance of the two-port network, respectively. In theory, a linear two-port network is to be unconditionally stable if, for all passive source and load impedance/admittance, the real parts of the input and output impedances are greater than

zero, *i.e.,* $\mathrm{Re}(Z_S + Z_{IN}) > 0$ and $\mathrm{Re}(Z_L + Z_{OUT}) > 0$. In terms of reflection coefficients, the unconditionally stable conditions are equivalent to $|\Gamma_S| < 1$, $|\Gamma_L| < 1$, $|\Gamma_{IN}| < 1$ and $|\Gamma_{OUT}| < 1$, respectively [28]. Such stability conditions can be appreciated as following. First the $|\Gamma_L| < 1$ and $|\Gamma_{IN}| < 1$ requirements are to be satisfied. the Γ_{IN} can be derived as,

$$|\Gamma_{IN}| = |S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}| = |\frac{1}{S_{22}}(\Delta + \frac{S_{12}S_{21}}{1 - S_{22}\Gamma_L})| < 1 \tag{4.9}$$

Where $\triangle = S_{11}S_{22} - S_{21}S_{12}$, and $|\Gamma_{IN}| < 1$ condition can be met for any Γ_L values with $|\Gamma_L| < 1$. This condition set is equivalent to,

$$|\frac{1}{S_{22}}(\Delta + \frac{S_{12}S_{21}}{1 - |S_{22}|^2})| + \frac{|S_{12}S_{21}|}{1 - |S_{22}|^2} < 1, \tag{4.10}$$

and

$$0 < 1 - \frac{|S_{12}S_{21}|}{1 - |S_{22}|^2} \ . \tag{4.11}$$

Equation 4.10 can be converted to,

$$|\frac{S_{11} - \Delta S_{22}^*}{1 - |S_{22}|^2})|^2 < (1 - \frac{|S_{12}S_{21}|}{1 - |S_{22}|^2})^2, \tag{4.12}$$

Which is equivalent to,

$$| S_{11} - \Delta S_{22}^{*} |^2 < (1 - | S_{22} |^2 - | S_{12} S_{21} |^2)^2 \tag{4.13}$$

Where S_{22}^{*} is the conjugate of S_{22}.

Rewriting Equation 4.13 into,

$$| S_{11} - \Delta S_{22}^{*} |^2 = | S_{12} S_{21} |^2 + (1 - | S_{22} |^2)(| S_{11} |^2 - | \Delta |^2), \tag{4.14}$$

One can readily reach to,

$$K_f = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{21}| \cdot |S_{12}|} > 1 \quad . \tag{4.15}$$

Next, similar derivation applied to the condition set of $| \Gamma_S | < 1$ and $| \Gamma_{OUT} | < 1$, the former leads to,

$$0 < 1 - \frac{| S_{12} S_{21} |}{1 - | S_{11} |^2} \quad , \tag{4.16}$$

While the latter gives,

$$| \Gamma_{OUT} | = | S_{22} + \frac{S_{12} S_{21} \Gamma_S}{1 - S_{11} \Gamma_S} | < 1 \tag{4.17}$$

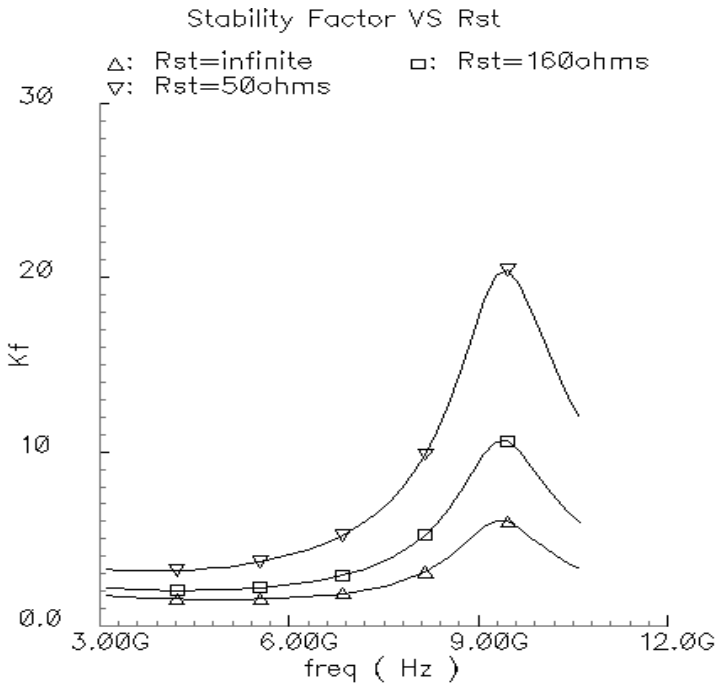
That reaches to the same inequality of Equation 4.15. The combined Equation 4.11, 4.15 and 4.16 inequalities set the unconditional stability conditions, called Rollett condition, which should be satisfied in circuit design, where $K_f > 1$ must be guaranteed by selecting proper circuit parameters [28].

Next, consider the practical stability design issue. In principle, a simple and effective way to stabilize an active circuit is to add a series resistance or a shunt conductance to its input port, *e.g.*, the stability resistor, R_{st}, shown in Figure 4.3, to ensure the real part of the equivalent input impendence, Z_{IN}, big enough. The stability condition of $\mathrm{Re}(Z_S + Z_{IN}) > 0$ can be rewritten into its admittance format as $\mathrm{Re}(Y_S + Y_{IN}) > 0$, and

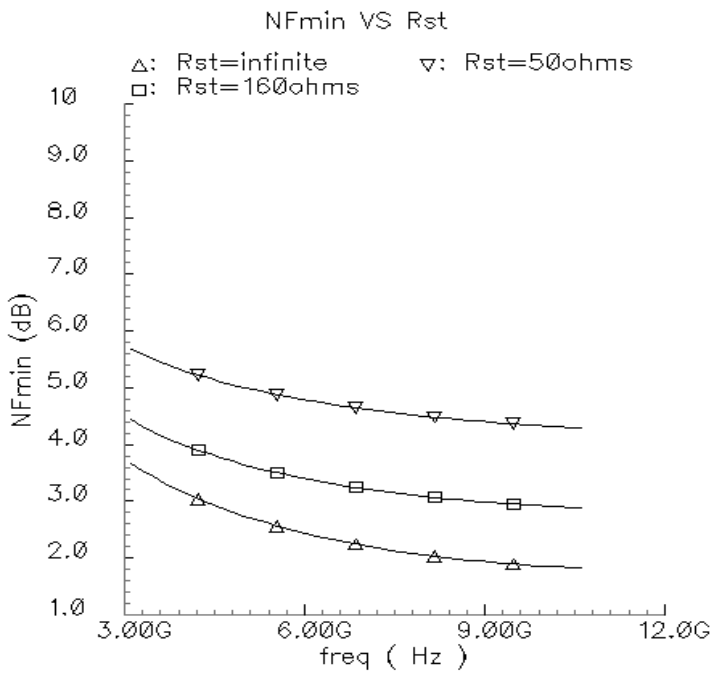
$$\begin{aligned} \mathrm{Re}(Y_S + Y_{IN}) &= Y_0 + 1/R_{st} + \mathrm{Re}(Y_{IN}^{'}) \\ &= Y_0 + 1/R_{st} + G(w) \end{aligned} \tag{4.18}$$

Where Y_0=1/50Ω is the characteristic admittance, $Y_{IN}^{'} = G(w) + jI(w)$ is the equivalent input admittance at the node before C_{in} , for which $G(w)$ is the real part and $I(w)$ is the imaginary part. Since R_{st} in Equation 4.18 is a positive value only, a smaller R_{st} is required to guarantee $\mathrm{Re}(Y_S + Y_{IN}) > 0$, hence, to ensure the unconditional stability condition of $\mathrm{Re}(Z_S + Z_{IN}) > 0$. The proper selection of the critical stability resistor, R_{st}, in this design is conducted by investigating the influences of R_{st} on both the stability and noise performance for the UWB LNA circuit in the simulation design phase with the results shown in Figure 4.8. The $K_f \sim R_{st}$ given in Figure 4.8(a) shows clearly that a smaller R_{st} delivers a larger K_f factor value, which is preferred for circuit stability.

However, the $NF_{min} \sim R_{st}$ relationship in Figure 4.8(b) indicates that too low an R_{st} would seriously degrade the noise performance of the circuit. Again, for a balanced overall circuit performance in practical design, an optimal R_{st} should be selected for the best possible circuit performance trade-off. Based upon the above analysis, an R_{st} of infinity is used to achieve the lowest possible NF_{min}, while still ensures the unconditional circuit stability with $K_f > 1$. The measurement data to be discussed in the next Section confirms this design trade-off concept. The final unconditional stability condition for the UWB LNA circuit is satisfied by an optimal selection of proper R_{st}, R_f and L_f in this work.



(a)



(b)

Figure 4.8. Influences of the Stability Resistor, R_{st}, on the Stability Factor (a) and NF_{min} (b) For the UWB LNA Circuit Designed.

Noise performance is certainly one of the most critical factors in LNA circuit design. For the UWB LNA designed in this work, the selection of the important resistive components, *e.g.,* R_{st}, R_s and R_f is carefully conducted for optimized noise behavior along with a balanced consideration for gain, power matching and stability ,etc for the circuit. Noise performance of the UWB LNA circuit can be analyzed using the equivalent small-signal noise circuit shown in Figure 4.4, where the total input-referred two-port noise voltage and current sources for the LNA circuit are modeled by $\overline{v_{ni}^2}$ and $\overline{i_{ni}^2}$. Since all the R_{st}, R_s and R_f inevitably generate thermal noises; design optimization in selecting these resistors must be conducted with the total LNA noise figure considered. Using the

equivalent noise model circuit of Figure 4.4, the total input-referred noise current and voltage power, *i.e.,* $\overline{i_{ni}^2}$ and $\overline{v_{ni}^2}$, for the UWB LNA circuit with shunt feedback path can be expressed as [27, 29, 30],

$$\overline{v_{ni}^2} = 4kT(\delta \frac{1}{5g_{d0}} + R_S)\Delta f \tag{4.19}$$

and

$$\overline{i_{ni}^2} \approx (\overline{i_{st}^2} + \overline{i_{Rf}^2} + \frac{\omega^2 C_{gs}^2}{g_m^2} 4kT\gamma g_{d0})\Delta f$$

$$= (4kTG_{st} + 4kTG_{Rf} + \frac{\omega^2 C_{gs}^2}{g_m^2} 4kT\gamma g_{d0})\Delta f \tag{4.20}$$

Hence, the total noise figure for the UWB LNA circuit is given as,

$$NF = 1 + \frac{\overline{v_{ni}^2}}{4kTZ_0} + \frac{Z_0 \overline{i_{ni}^2}}{4kT}, \tag{4.21}$$

where NF is the noise figure, k is the Boltzmann constant, T is the Kelvin temperature, Δf is the bandwidth in Hertz, g_{d0} is the drain-source conductance at zero V_{DS} biasing, γ is a device noise factor, δ is the gate noise coefficient, G_{st} and G_{Rf} are the conductance of R_{st} and R_f, respectively. The obviously monotonous relationship between the NF and the R_{st}, R_s and R_f revealed in Equations 4.19 to 4.21 readily suggests that a very small R_s and very large R_{st} and R_f values should be selected to achieve a smaller NF for the LNA circuit. With this understanding, and taking into consideration of other key LNA performance parameters, such as, gain, power matching and stability, this design chose a $R_s = 0$, a R_{st} = infinite and a moderately large $R_f = 700\Omega$ as stated in the

previous Section. Such optimal values for R_{st}, R_s and R_f ensure a well-balanced overall optimum circuit performance for the UWB LNA in terms of all key circuit specifications. The optimized overall LNA circuit performance is further ensured by carefully selecting other component values, including L_f, L_1, L_2 and C_f, as well as an optimal biasing voltage of V_{GS}.

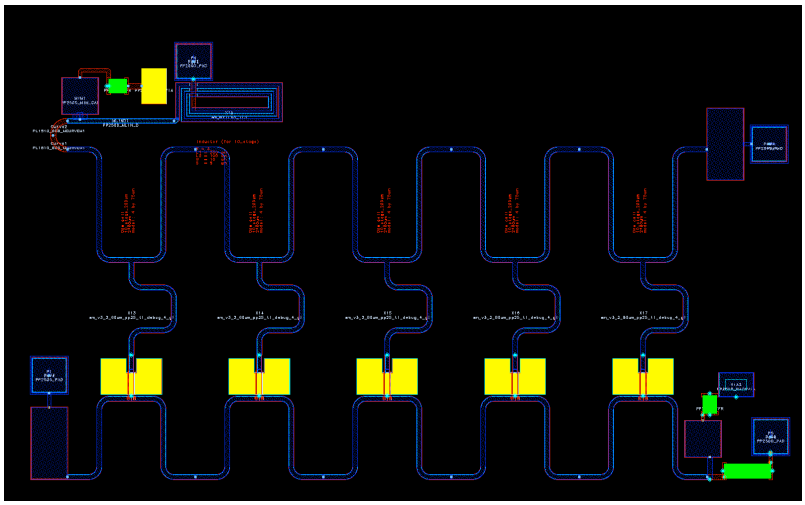


Figure 4.9 Layout for the LNA circuit

This fully-integrated 20GHz UWB LNA is designed and implemented in a commercial 0.25µm pHEMT technology. Figure 4.9 shows the layout for the LNA circuit. Parasitic effects of the bonding pads and transmission lines are considered in the LNA design and measurements. Figure 4.10 shows the simulated gain of 10dB and NF of 3.8dB over the full 20GHz UWB bandwidth at the optimal bias of V_{GS}=0.8V, which achieves the best reported gain flatness factor of 0.27 across the 3.1-10.6GHz frequency spectrum where the Gain Flatness Factor is defined as $\frac{(V_{\max} - V_{\min})(dB)}{BW(GHz)}$ with a smaller value representing a better flatness.

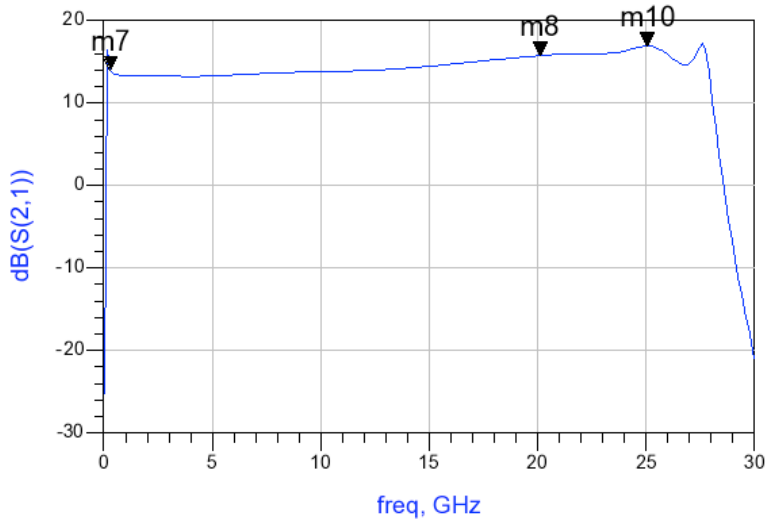


Figure 4.10 Simulated gain of 10dB and NF of 3.8dB over the full 20GHz UWB bandwidth at the optimal bias of V_{GS}=0.8V

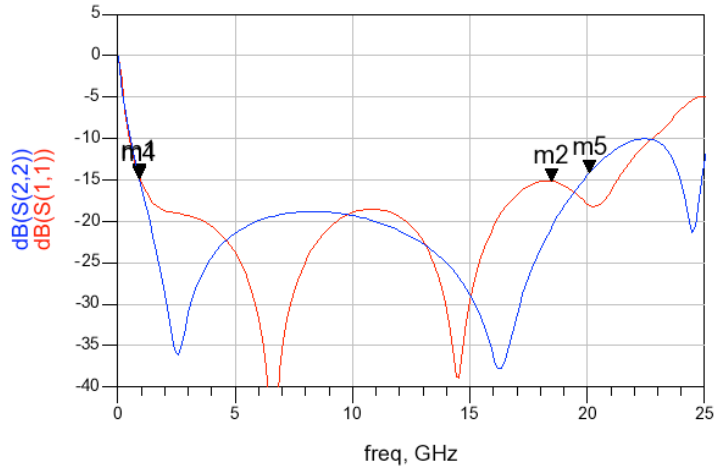


Figure 4.11 Simulated S_{11}, S_{22} and K_f factor

Figure 4.11 shows the simulated S_{11}, S_{22} and K_f factor. A low S_{11} of better than -9dB and a S_{22} of better than -15dB are achieved across the 3.1-10.6GHz range. The broadband small value S_{22} is important to maintain the LNA stability and also important for 50Ω silicon measurement. The simulated K_f >1 for the UWB LNA circuit indicates

that the circuit is unconditionally stable over the entire 20GHz UWB frequency spectrum. This unconditional LNA stability is achieved by comprehensive stability analysis and accurately selection of the key circuit component values as discussed previously. Figure 4.12 shows the simulated 1-dB compression point (CP), which achieves the best reported UWB LNA linearity of better than -0.8dBm over the full 20GHz UWB bandwidth. The whole UWB LNA circuit consumes only 9.8mW at 1.8V power supplies. Compared the key specifications of this UWB LNA with other reported designs[31, 32, 33, 34], this thesis reported LNA shows the best reported linearity and gain flatness across the full 7.5GHz bandwidth in this design [35], benefited by the improved broadband shunt-series feedback approach used and thorough consideration in achieving optimum overall circuit performance in this work.

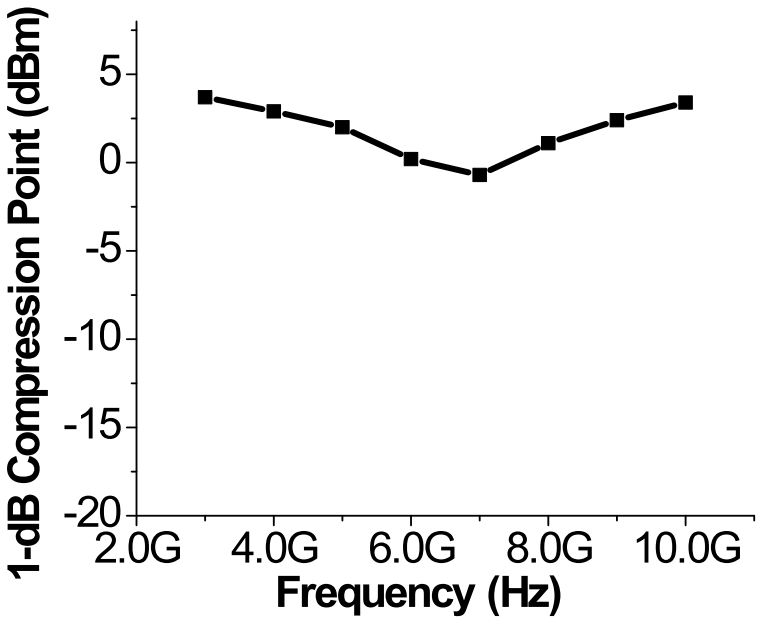


Figure 4.12 Measured 1-dB Compression Point (CP) for the LNA Circuit.

This section reports the design and analysis of a fully-integrated low-power 20GHz LNA circuit implemented in a commercial 0.20μm pHEMT for full-band OFDM UWB transceiver. The LNA circuit uses an improved shunt-series feedback topology to achieve the desired UWB matching, gain flatness, noise performance and linearity. Careful design trade-off is excised to ensure optimum overall LNA circuit performance. Measurement shows good circuit performance over the 20GHz spectrum including a flat broadband gain of 12dB, a minimum NF of 3.8dB, low return loss of S_{11} < -9dB and S_{22} < -15dB. The UWB LNA achieves the best reported gain flatness factor of <0.27 and linearity with a 1-dB compression point better than -0.8dBm in similar designs. The full UWB LNA consumes a low power consumption of 9.8mW at 1.8V.

Chapter 5
ESD Protection for RF circuits design

Electrostatic Discharge, or ESD, is a single-event, rapid transfer of electrostatic charge between two objects, usually resulting when two objects at different potentials come into direct contact with each other. ESD can also occur when a high electrostatic field develops between two objects in close proximity. With continuous scaling down at VLSI technology, ESD is one of the major causes of device failures in the semiconductor industry. The advancement of CMOS IC technologies presents increasing challenges in the design of reliable Electrostatic Discharge ESD) protection.

ESD protection design rapidly emerges as a big challenge to RF IC designers because ESD-induced parasitic effects are inevitable and become intolerable to parasitic-sensitive RF ICs for beyond GHz operations. Generally, RF ICs, typically used in wireless handheld devices, demand for more robust ESD protection, much higher than 2kV HBM (human body model) typically for digital ICs, which results in more parasitic that adversely affect RF circuit performance [1]-[8]. In principle, ESD protection structures are connected between I/O and supply/ground and/or between supply buses. In normal circuit operation, these ESD structures remain in OFF state without affecting core circuit operation. When ESD transients appear, ESD structure triggers off immediately to form low-impedance path that discharges ESD pulses safely without generating too much heat and clamps IC pads to a safely low level, hence protect IC against any ESD damage. Unfortunately, though remaining OFF in normal circuit operation, ESD structure inevitably introduces substantial parasitic effects such as parasitic capacitance (C_{ESD}), resistance (R_{ESD}), noise coupling and self-generated noises, etc, which, if not considered

in RF circuit designs, can severely corrupt RF I/O impedance matching, resulting in serious degradation of RF IC circuit performance including gain, reflection, linearity, power efficiency, bandwidth and noise figure, etc [7], [8]. Therefore, it is vitally critical to include ESD-induced parasitic into RF IC circuit design flow in order to achieve whole-chip ESD+RFIC design optimization. Unfortunately, lack of ESD-aware RF IC design techniques and accurate ESD device models have made it impossible to achieve this goal until recently [9]-[17]. Two main technical challenges exist in this regard: firstly, RF ESD protection structures must be optimized for minimum parasitic and accurately characterized for design integration. Secondly, such ESD-induced parasitic effects must be integrated into core RF IC circuit design flows. It is hence imperative to develop a new ESD-RFIC co-design method to address this huge emerging RF IC design challenge. In this paper, we report the first ESD-RFIC co-design methodology developed to accurately address the troublesome ESD-RFIC interactions and hence enable whole-chip design optimization of ESD-protected RF IC circuits. The new co-design flow includes design techniques for RF ESD protection design optimization, accurate RF ESD protection characterization, RF+ESD circuit co-simulation and S-parameter-based RF+ESD I/O re-matching, etc.

1. Typical ESD Failure

One of the causes of ESD events is static electricity. Electrostatic charge build-up occurs as a result of an imbalance of electrons on the surface of a material. Such a charge build-up develops an electric field that has measurable effects on other objects at a distance [1, 2]. The process of electron transfer as a result of two objects coming into

contact with each other and then separating is known as 'triboelectric charging'. Examples of tribocharging include walking on a rug, descending from a car, or removing some types of plastic packaging. In all these cases, the friction between two materials results in tribocharging, thus creating a difference of electrical potential that can lead to an ESD event.

Another cause of ESD damage is through electrostatic induction. This occurs when an electrically charged object is placed near a conductive object isolated from ground [3-5]. The presence of the charged object creates an electrostatic field that causes electrical charges on the surface of the other object to redistribute. Even though the net electrostatic charge of the object has not changed, it now has regions of excess positive and negative charges. An ESD event may occur when the object comes into contact with a conductive path. For example, charged regions on the surfaces of styrofoam cups or plastic bags can induce potential on nearby ESD sensitive components via electrostatic induction and an ESD event may occur if the component is touched with a metallic tool [4, 6].

2. Types of ESD.

The most spectacular form of ESD is the spark, which occurs when a strong electric field creates an ionized conductive channel in air. This can cause minor discomfort to people, severe damage to electronic equipment, and fires and explosions if the air contains combustible gases or particles. However, many ESD events occur without a visible or audible spark. A person carrying a relatively small electric charge may not actually feel a discharge that is still sufficient to damage sensitive electronic components

(some devices may be damaged by discharges as small as 20 volts). These invisible forms of ESD can cause device outright failures, or less obvious forms of degradation that may affect the long-term reliability and performance of electronic devices. The degradation in some devices may not become evident until well into the service life of some equipment.

3. Typical ESD failure in CMOS IC technology.

In the semiconductor industry, the potentially destructive nature of ESD in integrated circuits (IC) became more apparent as semiconductor devices became smaller and more complex. The high voltages result in large electric fields and high current densities in the small devices, which can lead to breakdown of insulators and thermal damage in the IC. The losses in the IC industry caused by ESD can be substantial if no efforts are made to understand and solve the problem. The largest category is that of electrical overstress (EOS), of which ESD is a subset. In many cases, failures classified as EOS could actually be due to ESD, which would make this percentage even higher. ESD can destroy a semiconductor device in many ways, resulting in observable signs of damage or failure attributes.

4. ESD Test Models

There are three predominant ESD models for IC's: 1) the *Human Body Model (HBM)*; 2) the *Charged Device Model (CDM)*; and 3) the *Machine Model (MM)*. The HBM simulates the ESD event when a person charged either to a positive or negative potential touches an IC that is at another potential [7]. The CDM simulates the ESD event wherein a device charges to a certain potential, and then gets into contact with a conductive surface at a different potential. The MM simulates the ESD event that occurs

when a part of an equipment or tool comes into contact with a device at a different potential. HBM and CDM are considered to be more 'real world' models than the MM.

A. The Human Body Model (HBM)

The Human Body Model simulates the ESD phenomenon wherein a charged body directly transfers its accumulated electrostatic charge to the ESD-sensitive device. A common example of this phenomenon, and from which the name of this model was derived, is when a person accumulates static charge by walking across a carpet and then transferring all of the charge to an ESD-sensitive device by touching it. Of course, other 'non-human' materials that accumulate and transfer charge in a similar manner are also covered by the HBM.

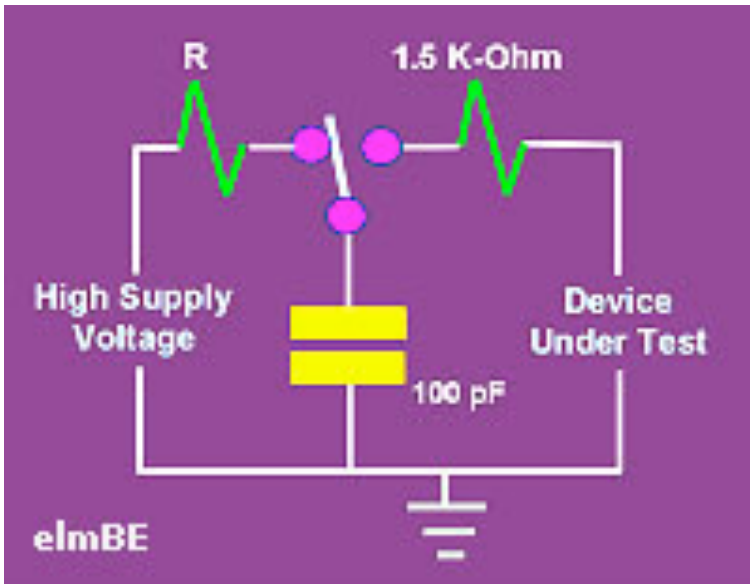


Figure 5.1. HBM Model. (Http://www.siliconfareast.com/esdmodels.htm)

Dating back to the 19th century when it was used to investigate gas explosions in mines, the HBM is the oldest and most commonly used model for testing the ESD sensitivity of a device. The general ESD testing set-up for this model consists of a 100 pF capacitor that can be charged to a certain voltage, and then discharged by a switching component into the device being tested through a 1.5 K-Ohm resistor. Figure 5.1 shows a basic HBM test circuit.

B. The Machine Model (MM).

Machine model (MM) describes an ESD pulsing event where charged machinery discharges when touching IC parts during testing. MM circuit is similar to HBM model with different parameters: C = 200 pF, R = 0 Ω and L ≈ 2.5 μH, leading to an oscillatory discharging waveform with higher peak current and shorter rise time.

Originated in Japan as a result of investigating worst-case scenarios of the HBM, the Machine Model simulates a more rapid and severe electrostatic discharge from a charged machine, fixture, or tool. The MM test circuit consists of charging up a 200 pF capacitor to a certain voltage and then discharging this capacitor directly into the device being tested through a 500 nH inductor with no series resistor. Figure 5.2 shows a basic MM test circuit.

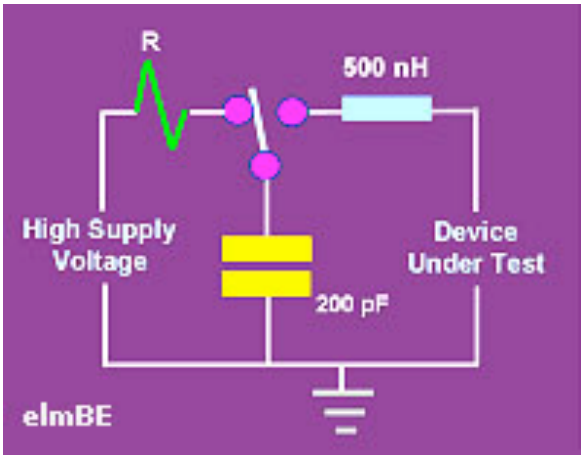


Figure 5.2. MM Model. (http://www.siliconfareast.com/esdmodels.htm)

C. The Charged Device Model (CDM).

Not all ESD events involve the transfer of charge *into* the device. Electrostatic discharge *from* a charged device to another body is also a form of ESD, and a quite commonly encountered one at that.

A device can accumulate charge in a variety of ways, especially in situations where they undergo movement while in contact with another object, such as when sliding down a track or feeder. If they come into contact with another conductive body that is at a lower potential, it discharges into that body. Such an ESD event is known as Charged Device Model ESD, which can even be more destructive than HBM ESD (despite its shorter pulse duration) because of its high current.

There are currently two widely used models for CDM testing: 1) the Socketted Discharge Model (SDM); and 2) the Real-world Charged Device Model (RCDM). SDM simulates a device inserted in a socket, then charged from a high voltage source, and then

discharged through a 1-ohm resistor. SDM is easy to conduct but is not always replicating real-world CDM ESD events.

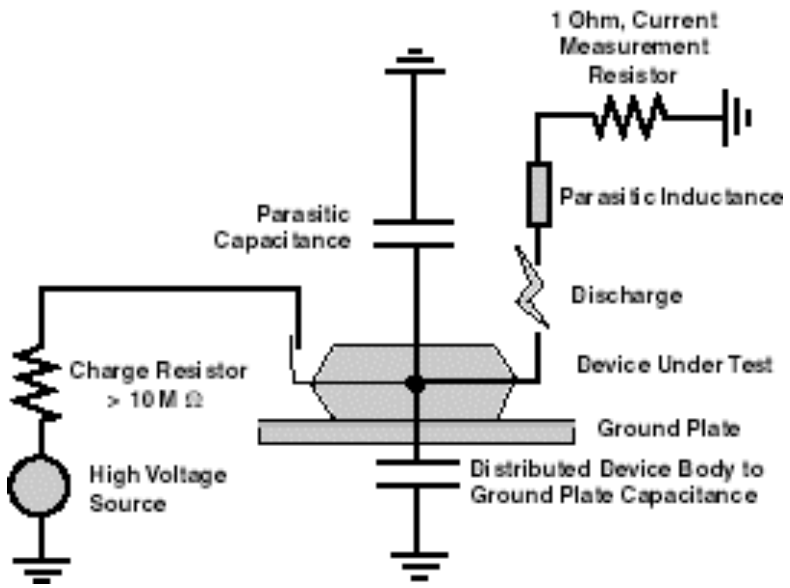


Figure 5.3. CDM Model. (http://www.siliconfareast.com/esdmodels.htm)

RCDM testing consists of putting the DUT in dead bug position on a thin dielectric (FR4), which is then placed over a ground plate. The DUT is then charged either directly by a charging probe or indirectly by field induction. Each pin is then discharged through a 1-ohm resistor to ground.

5. ESD protection Design

Electrostatic discharge (ESD) is one of the most important reliability problems in the integrated circuit (IC) industry [8]. Typically, one-third to one-half of all field failures (customer returns) are due to ESD and other failures known collectively as electrical overstress (EOS). As ESD damage has become more prevalent in newer technologies due

to the higher susceptibility of smaller circuit components, there has been a corresponding increase in efforts to understand ESD failures through modeling and failure analysis. This has resulted in a greater industry-wide knowledge of ESD mechanisms and thus a greater ability to design robust ICs which sustain fewer field failures. Despite these efforts, there are still ESD-related problems, which are not well understood, especially the phenomenon of "latent damage." There are two ways to reduce IC failures due to ESD. One is to ensure proper handling and grounding of personnel and equipment during manufacturing and usage of packaged chips, i.e., to prevent ESD events from occurring. The other approach is to connect protection circuits (almost always on-chip) to the pins of a packaged IC, which will divert high currents away from the internal circuitry and clamp high voltages during an ESD stress. A chip manufacturer has limited control over a customer's handling of its product, so incorporating effective protection circuitry is essential. Since the spectrum of stresses under the label of EOS/ESD is broad and the amplitude of stress is virtually unlimited, it is not possible to guarantee total EOS/ESD immunity. However, through the proper design of protection circuitry the threshold of sustainable stress can be significantly increased, resulting in improved reliability of ICs.

Designing ESD protection circuits becomes more challenging as device dimensions shrink, particularly in MOS technologies [9]. As ICs become smaller and faster, susceptibility of the protection circuits to damage increases due to higher current densities and lower voltage tolerances. Use of lightly doped drains (LDDs) and silicide in newer technologies exacerbates these problems. If the LDD diffusions are shallower than the source/drain diffusions, then for a given current level there is a greater current density

in the LDD region, which means there is more localized heating and therefore a greater chance of damage during an ESD stress. This effect has been verified with simulations as well as through failure analysis. Similarly, silicided source/drain diffusions lead to current localization by concentrating current flow at the surface of devices as well as reducing the ballasting resistance needed to distribute the current. Finally, the thinner gate oxides of newer MOS processes are more susceptible to high-field stress, i.e., dielectric breakdown.

Typically the design of ESD protection is an empirical, trial-and-error procedure in which several variations of a circuit or types of circuits are laid out, processed, packaged, and tested on a simple pass/fail basis. This approach is time consuming and does not facilitate the evolution of protection circuits in future technologies. A better design methodology includes a more complex testing technique and modeling of ESD circuit behavior in order to provide understanding of the functionality of the transistors, diodes, and lumped capacitors and resistors making up the circuit as well as to extract critical parameters of the circuit. In conjunction with a relatively small array of test structures, proper modeling can be used to design an optimum protection circuit as well as predict the performance of similar circuits in next-generation technologies. Recent advances in two-dimensional numerical device simulation have made possible the modeling of ESD events [6, 7]. These simulations predict the device's current-voltage response to an ESD stress and provide analysis capabilities, which suggest how and where a protection device will fail.

The focus of this thesis is on modeling, simulating, and design of ESD protection devices in state-of-the-art silicon CMOS technology using advanced testing techniques and numerical simulation. An outline of the thesis is presented at the end of the chapter.

A. Protecting Integrated Circuits from ESD.

A protection circuit serves two main purposes: providing a current path during a high-stress event and clamping the voltage at the stressed pin below the gate-oxide breakdown level. Additionally, the protection circuit itself should not become severely damaged during an ESD event. Although the odds of having the same pair of pins stressed more than once is small, it is important that the protection circuit not become leaky and degrade chip performance. Also, in the case of output-protection circuits which double as output drivers, long-term reliability may be reduced if damage is incurred. For example, it has been shown that MOSFETs driven deep into snapback during an ESD stress may suffer hole trapping in the gate oxide as well as interface-state generation, leading to a shift in the threshold voltage [1, 10, 11]. The hole trapping can increase the susceptibility to time-dependent dielectric breakdown of the gate oxide. To avoid being damaged, protection circuits should minimize self-heating by keeping current densities and electric fields in the silicon low and prevent dielectric breakdown of the gate oxides in the protection circuit by minimizing the electric fields across the oxides.

B. Design Methodology.

Another approach to designing and optimizing protection circuits is to create models for transient I-V and failure parameters using statistical design of experiments. By characterizing a set of protection transistors with variations in layout, models can be

developed to describe the TLP I-V parameters, TLP withstand current, and HBM withstand voltage as functions of transistor width, gate length, contact-to-gate spacing, number of poly-gate fingers, and other layout parameters. (The withstand current or voltage is defined as the maximum TLP current or HBM voltage, respectively, a structure can withstand without incurring damage [1, 9, 12]. Thus, the withstand level is always slightly lower than the failure level.) A statistical design-of-experiments program is useful for determining the minimum number of test structures needed and for extracting the model equations. Once models are developed for a given technology, the performance of any ESD circuit designed in that technology can be determined. (Figure 5.4)

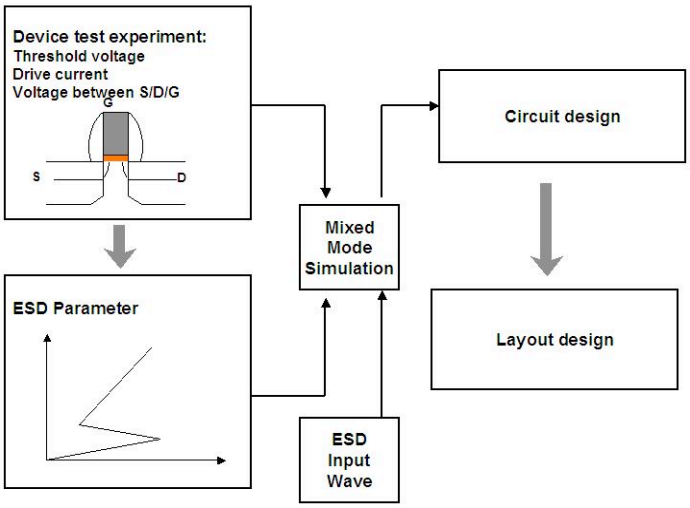


Figure 5.4. Scheme of ESD design using mixed mode simulation.

6. ESD Protection Circuits' Effects on RF ICs

The inevitable ESD-to-Circuit influences are all-scale, including RC delay and RF I/O impedance matching corruption due to C_{ESD} and R_{ESD}, substrate-to-I/O noise coupling due to C_{ESD} and ESD self-generated noises, which affect clock speed, signal integrity, RF

gain, linearity, power efficiency, bandwidth and noise figure (NF) of RF ICs, etc. Unfortunately, the foundry provided ESD device SPICE models are over-simplified that treat an ESD device as an ideal-OFF device; hence, it cannot be used to simulate the ESD effects on RF IC performance. To argue this point, an ESD-protected fully-integrated 5GHz low-noise amplifier (LNA) was designed in commercial 0.18μm RFCMOS as shown in Figure 5.5, where a typical foundry-provided 5.7kV N+/P-substrate diode ESD structure is used to protect its input pad.

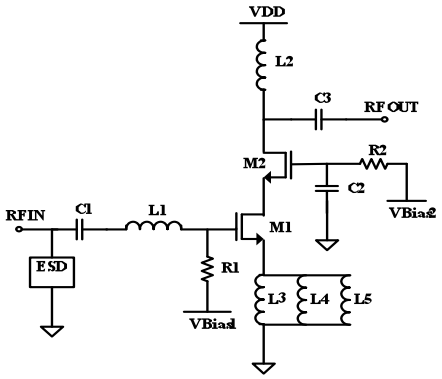
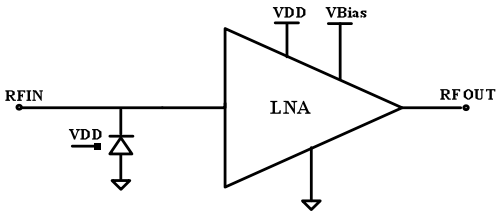
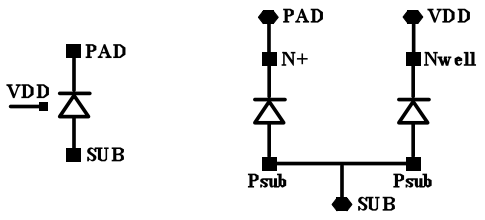


Figure 5.5. Schematic for an ESD-protected 5GHz LNA circuit where the ESD protection device is connected at input port.

In design, the foundry-provided ESD SPICE model, which is a common diode sub-circuit model consisting of a N+/Psub diode in parallel with a Psub/Nwell guard-ring diode shown in Figure 5.6, was used to simulate the LNA circuit.



(a)

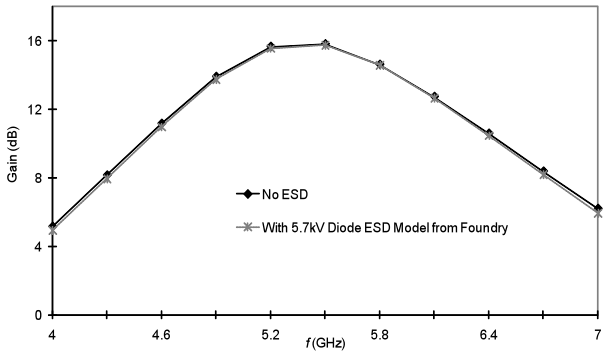


(b)

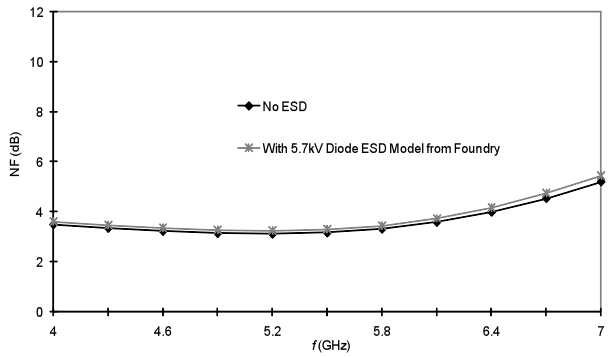
Figure 5.6. Simulation setup for ESD-protected LNA using foundry-provided diode ESD structure (a) and its normal ESD sub-circuit model (b).

Figure 5.7 is simulated LNA gain and NF, showing almost no visible degradation: 0.08dB (-0.5%) drop in gain and 0.13dB (+4%) increase in NF, which contradicts to the testing results. Obviously, traditional simulation method using foundry ESD models do not work for evaluating ESD impacts on RF ICs.

To show actual ESD impact on RF IC performance, the same LNA in Figure 1 was designed in two splits, with and without a custom-optimized 5kV diode ESD structure, with its die photo shown in Figure 5.8. The 5kV ESD structure is optimized by mixed-mode ESD simulation-design method for accurate ESD protection and minimum parasitic [14]. The LNA circuit,



(a)



(b)

Figure 5.7. Simulation results using foundry-provided ESD circuit model show negligible LNA performance degradation for both gain (a) and NF (b), which conflict with measurements.

It is designed for power-constrained noise optimization, uses a fine-tuned 50Ω input matching network (C1, L1, L3, L4 and L5). The non-ESD LNA features a gain of15.82dB and NF of 3.12dB at center frequency, and draws a 6.4mA current from a 1.8V supply.

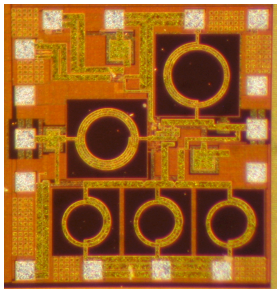


Figure 5.8. Die photo for the ESD-protected LNA circuit designed in 0.18μm RFCMOS.

Figure 5.9 is measured NF showing a 0.6dB increase (+12.56%) due to ESD device. Figure 5.10 is measured LNA gain showing a 0.25dB drop (-1.7%). The LNA performance degradation is attributed to ESD-induced input matching corruption, which is confirmed by Figure 5.11 where the measured input reflection ($|S_{11}|$) is affected

severely by the ESD device, i.e., ~6.6dB shift. Clearly, even using a parasitic-optimized ESD structure, serious LNA performance degradation due to ESD parasitic effects is readily observed, which must be addressed in real-world RF IC designs.

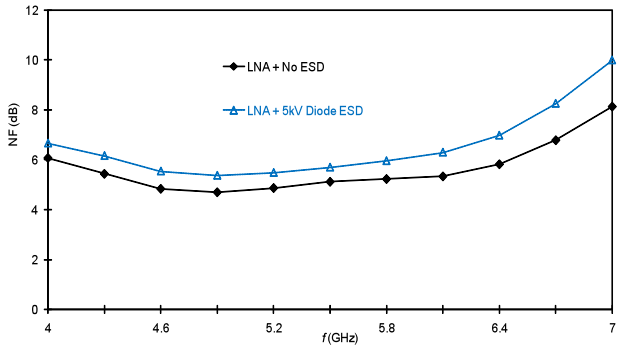


Figure 5.9. Measured LNA NF curves show significant degradation due to ESD protection noise effect.

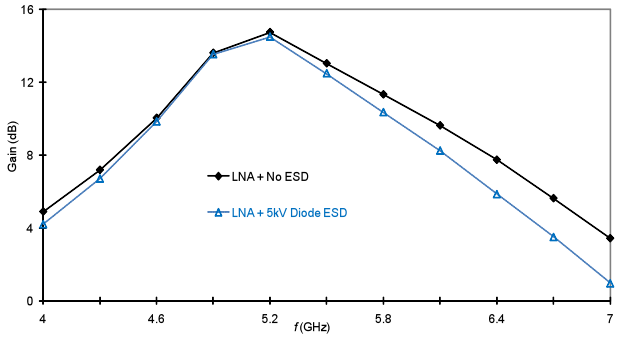


Figure 5.10. Measured LNA gain curves shows substantial degradation due to ESD protection induced parasitic effects.

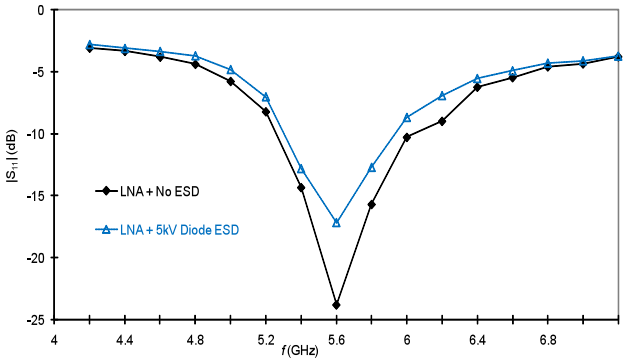


Figure 5.11. Measured LNA S_{11} data reveal significant input port reflection due to matching corruption induced by ESD protection parasitic.

Consider serious ESD-to-Circuit influences, it is hence imperative to optimize RF ESD protection designs and accurately characterize RF ESD parasitic parameters. To ensure minimum ESD-induced parasitic effect, RF ESD protection structures can be designed and optimized using a mixed-mode ESD simulation-design method [14]-[17]. In a separate study, a group of commonly used ESD structures, e.g., diode strings with 1-5 diodes (Dx1, Dx2 Dx3, Dx4 and Dx5), grounded-gate NMOS (ggNMOS), silicon-controlled rectifier (SCR) and dual-direction SCR (dSCR), are first designed and optimized using the mixed-mode ESD design method and implemented in a commercial 0.35µm BiCMOS, which were then accurately characterized using a new RF characterization method recently developed where ESD parasitic C_{ESD}, R_{ESD} and NF are measured and extracted using a two-port network model [15]. To ensure accuracy, Ground-Signal-Ground (GSG) pattern, as well as "short" and "through" calibration patterns are used for ESD testing structures to eliminate any parasitic effects associated with pads and metal lines. All testing was done across a 9GHz spectrum. Figure 5.12 shows the extracted C_{ESD} for all the ESD devices studied.

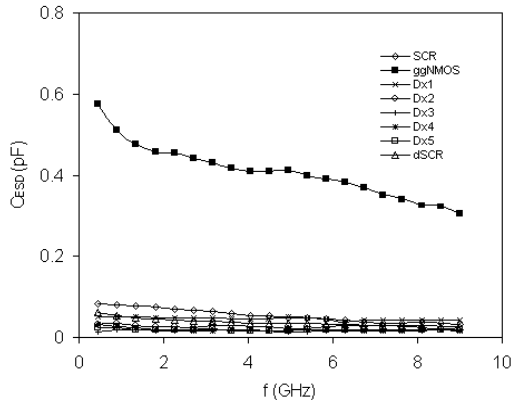


Figure 5.12. Measured ESD parasitic C_{ESD} for various ESD protection structures shows that ggNMOS ESD structure is poor for RF ESD protection.

Clearly, the ggNMOS ESD structure introduces the highest parasitic C_{ESD} and is not recommended for RF IC circuits. To further explore the details, Figure 5.13 shows the C_{ESD} data without ggNMOS.

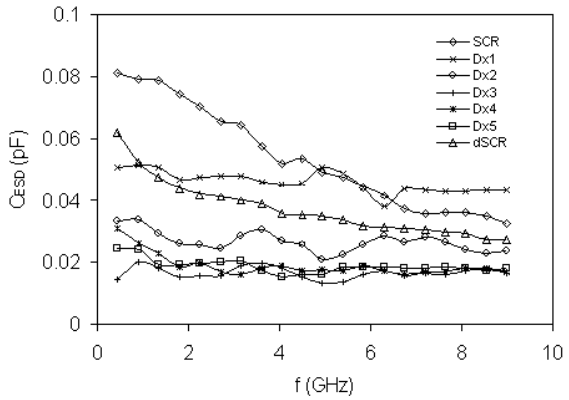


Figure 5.13. Measured ESD parasitic C_{ESD} for various ESD protection structures excluding the ggNMOS ESD structure reveal design insights in selecting proper RF ESD protection structure.

It is readily observed that SCR structure has very low parasitic C_{ESD} while dSCR further reduced the C_{ESD}. It is also interesting to notice that, contrary to common belief

that adding diodes in a series diode string would monolithically reduced the total C_{ESD}, this C_{ESD}-reduction trend actually saturates for a diode string with more than three diodes, mainly due to extra stray capacitances associated the diode structures. However, in practical ESD designs, more factors, such as layout sizes and NF, should be considered in order to achieve overall RF ESD design optimization. To enable this overall RF ESD design optimization concept, a new figure-of-merit, entitled F-factor, is introduced to evaluate overall merit of RF ESD structures, which is given in (1).

$$F = \frac{kV(ESDprotection)}{Area \times C_{ESD} \times NF} \qquad (1)$$

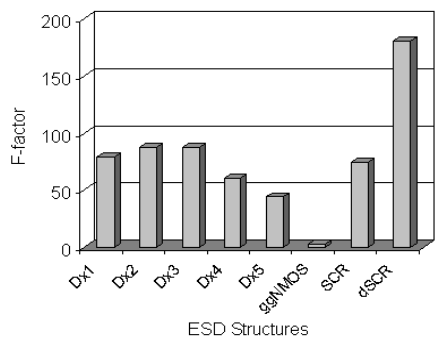


Figure 5.14. Extracted F-factors for the ESD protection structures studied in this work clearly reveal the relative merits of different designs in terms of overall performance.

Clearly, a higher F-factor is preferred for an overall optimized RF ESD protection structure. Figure 5.14 is the extracted F-factors for all ESD structures studied, which readily shows that ggNMOS is not a good option for RF ESD protection while SCR type

structure is good for RF ESD protection. It is also important to notice that a diode string with about 2-3 diodes may be optimized RF ESD protection solution.

7. Mixed Model ESD Methodology

Since ESD structures normally consume large amount of chip area and inevitably produce parasitic capacitance, they may negatively influence performance of the circuits protected. Such ESD-induced parasitics become intolerant issues to high-speed and VDSM ICs and may affect such critical specifications as timing, stability, and bandwidth, etc. It is therefore desirable to optimize on-chip ESD protection design for both silicon structures and metal interconnects. The former can be addressed by traditional CAD work, while the latter is mainly guided by some empirical rules-of-thumb. For example, virtually all IC design rules suggest 20um of ESD metal line width or "better wide". The challenge is that there is no practical simulation approach that can simulate both semiconductors and metals altogether and simultaneously, leaving design of ESD metals largely with the experiences [6, 7]. This shortcoming certainly prohibits IC designers from optimizing ESD design, minimizing its impacts on circuits, and taking full advantages of some new technologies, *e.g.*, copper interconnects.

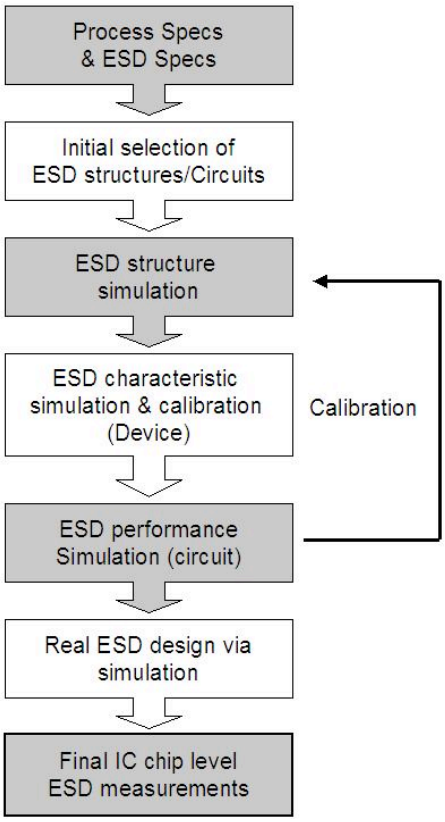


Figure 5.15 Mixed mode ESD design route.

Our lab has developed a new simulation-design approach that intends to address such challenge [1, 2, 7, 10, 12]. The new approach can not only predict the design performance via coupled thermal-electro and device-circuit simulation, but also optimize the design by accurate sizing. The simulation results demonstrate that significant size reduction can be realized in ESD design using this new approach.

Figure 5.15 shows a low chart of the mix-mode ESD protection design methodology. This methodology is called mixed-mode because it involves process simulation, device characteristic simulation and circuit performance simulation. In device

characteristic simulation, this methodology couples electron models with thermal behavior models, to address ESD thermal failure issues. Prediction in ESD protection design is hence made possible by using this mixed-mode methodology. Furthermore, by following this procedure, designers could also optimize their devices and circuits to alleviate the trouble some interactions between ESD protection structures and the circuits being protected.

With the understanding of the substantial ESD effects on RF IC performance, a new ESD-RFIC co-design method is then developed aiming to achieve whole-chip ESD-protected RF IC design optimization by integrating any ESD-introduced parasitics into RF IC simulation design flow. The new ESD-RFIC co-design method has three steps: Firstly, any RF ESD structure ought to be designed and optimized using the mixed-mode ESD simulation-design method discussed previously to achieve best overall ESD performance with minimum parasitic effects [14]. Secondly, accurate RF ESD characterization will be conducted for the ESD structures used to obtain accurate parasitic C_{ESD} and NF, etc. Thirdly, the measured ESD parasitic C_{ESD} and NF parameters will be used for ESD-RFIC co-design that is described by Figure 5.16. In co-design simulation, the RF IC circuit is designed first with its I/O impedance matching networks tuned up for its specs. Next, the extracted ESD parasitic parameters are included in the I/O matching network for re-simulation to evaluate ESD influences. Inevitably, the pre-designed RF IC I/O matching will be corrupted by ESD parasitics, resulting in degradation of RF IC circuit performance as expected. In next step, an I/O re-matching technique is used to re-adjust the I/O matching networks in order to recover any I/O matching corruption

induced by ESD structures. Currently, since no applicable SPICE circuit model is available for ESD parasitic modeling yet, we introduce a new S-parameter-based I/O re-matching technique that uses a direct S-parameter insertion technique to insert the extracted ESD s-parameter data directly into the I/O ports in simulation. It proves that this new S-parameter insertion technique ensures accuracy of RF IC re-design including ESD-induced parasitic, which is not possible when using any existing ESD SPICE model yet. At last, post simulation for the whole chip will be conducted to ensure overall design optimization. The goal of the new I/O re-matching design is to recover almost all RF IC circuit performance degradation induced by ESD protection parasitic effects, which was verified experimentally.

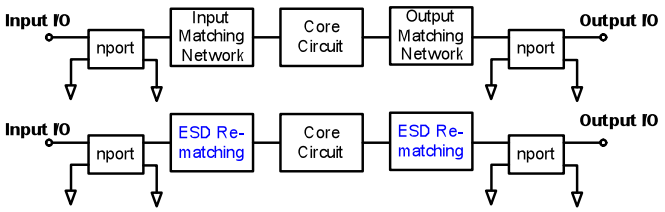
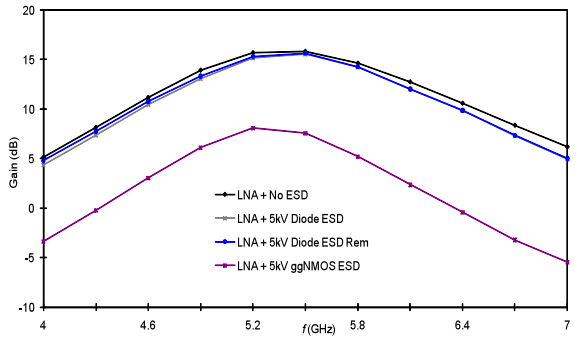


Figure 5.16. Flow chart for the new ESD-RFIC co-design method.

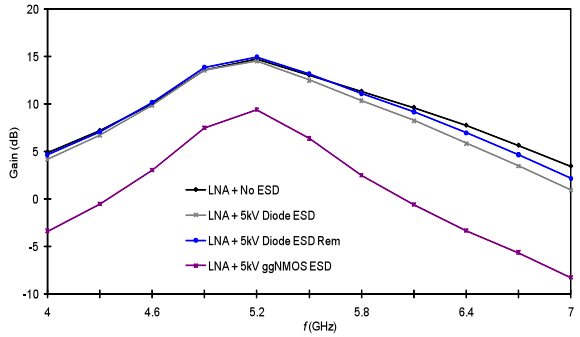
In the following discussions, several practical design examples are presented to demonstrate the usefulness of the new ESD-RFIC co-design method, including a 5GHz LNA, a 2.4GHz power amplifier (PA) and a 2.4GHz mixer circuit, all are part of a dual-band WLAN transceiver chip. To simulate the practical application and for uniform comparison, 5kV ESD protection structures (i.e., N+/PW diode, ggNMOS and SCR) are used to protect the input of LNA, output of PA and RF and LO pins of mixer circuit. There is no ESD protection for any internal pin due to transceiver integration

consideration. All the ESD structures used were carefully optimized for minimum parasitic effects, though such minimized ESD parasitics still need to be considered in ESD-RFIC co-design optimization at whole chip level. The extracted C_{ESD} is 90fF for the diode ESD (lowest reported for 5kV ESD protection) and 1071fF for ggNMOS ESD (highest in this group, yet still lower than any published data). All the designs were implemented in a foundry 0.18μm RFCMOS technology. For fair comparison, the RF IC block design splits include the ones without ESD protection, with ESD structures and the ones after I/O re-matching co-design (labeled as "-REM").

The schematic for the 5GHz LNA circuit is shown in Figure 1 with its specs given before. Figure 5.17 shows the post-simulation and measured gain results for the LNA circuits. It is clear that the ggNMOS ESD structure severely affects the LNA circuit performance, while the diode ESD also causes visible gain degradation (-14.3% to -4.2% drop from its nominal gain). After I/O re-matching co-design, the I/O matching corruption encountered, hence the gain degradation observed, was almost recovered completely (by 76%). It is interesting to notice that using the new co-design method, post-simulation result agrees with measurement very well, which is critical for ESD-RFIC co-design optimization at full-chip level. Such good agreements are observed for all other design examples, hence validated the accuracy of the new co-design method.

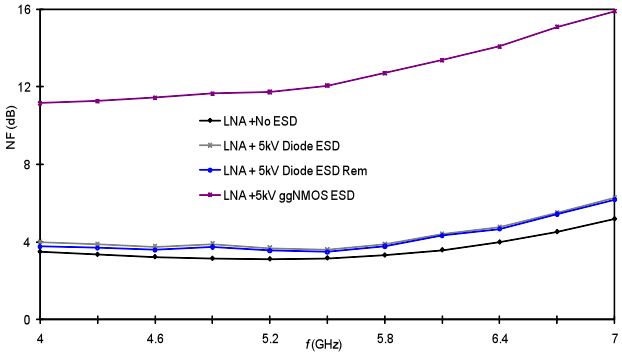


(a)

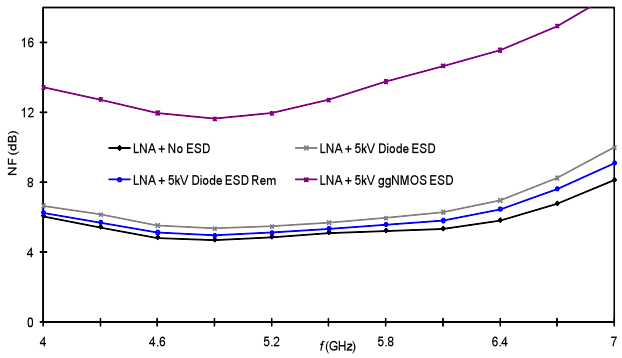


(b)

Figure 5.17. Gain for the LNA different ESD protection structures by post simulation co-design (a) and measurement (b) clearly reveal the ESD impacts on LNA performance.



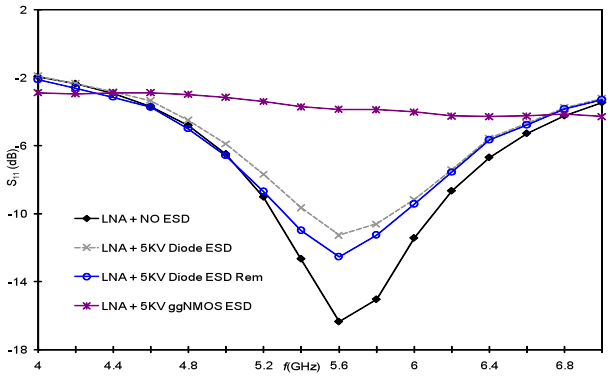
(a)



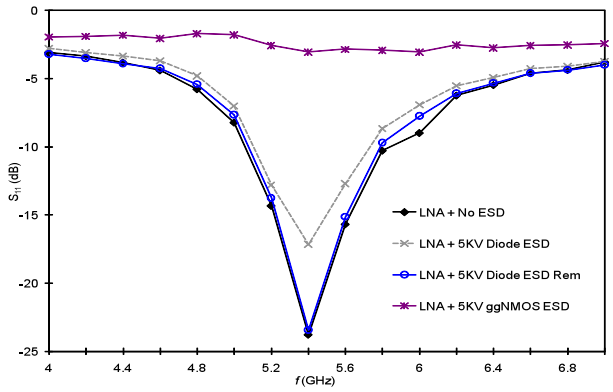
(b)

Figure 5.18. NF for LNA with different ESD protection structures by post simulation co-design (a) and measurement (b) readily shows the influences of ESD protection on LNA performance.

Figure 5.18 shows the simulated and measured NF results, which also reveal that the ggNMOS ESD structure introduces too much extra noises and the diode ESD structure affects LNA noise performance noticeably (+11.3% to +18%) that can be recovered by I/O re-matching co-design technique (by 82.7%). The LNA performance degradation is attributed to the I/O matching network corrupted by ESD-induced parasitics, which can be better understood by the measured S_{11} data in Figure 14 that clearly shows the strong influences of ESD structures on input port reflection, hence the I/O matching and circuit performance. The performance recovery using the new I/O re-matching co-design technique can be readily observed in the S_{11} data chart.



(a)



(b)

Figure 5.19. S_{11} data for LNA with different ESD protection structures by post simulation co-design (a) and measurement (b) suggest that ESD-induced parasitics corrupt the I/O matching network without co-design.

To further illustrate the ESD impacts on RF circuit, Figure 5.20 shows the measured Smith chart results for the LNA with 5kV diode and ggNMOS ESD structures. It clearly reveals that the ggNMOS ESD structure strongly affects LNA performance while the optimized diode ESD structure has much lower impact that can be further suppressed by using the new co-design method.

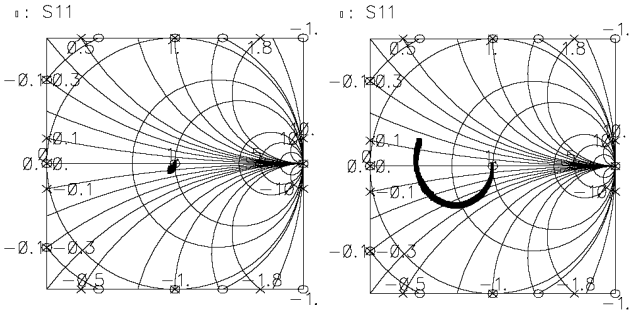


Figure 5.20 Measured S_{11} results on Smith Chart for LNA with diode (a) and ggNMOS (b) ESD protection structures clearly shows the relatively impacts of ESD parasitic on LNA performance.

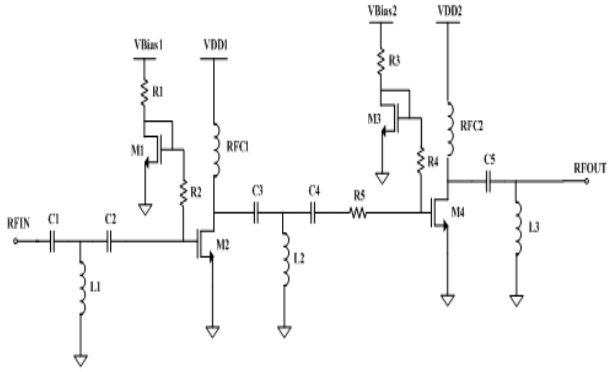


Figure 5.21. Schematic for the 5kV ESD-protected 2.4GHz PA designed in 0.18µm RFCMOS.

The second design example is a 2.4GHz PA circuit with different ESD protection structures. Figure 5.21 shows the schematic for the two-stage fully-integrated PA circuit with its die photo shown in Figure 5.22, where output L-matching network is fine-tuned using L3, C5, RFC2 and ESD parasitics. The PA design splits consist of non-ESD PA, PA+diode ESD, PA+ggNMOS ESD, PA+SCR ESD, PA+diode ESD re-matched and PA+SCR ESD re-matched (co-design optimization, noted with “-Rem”).

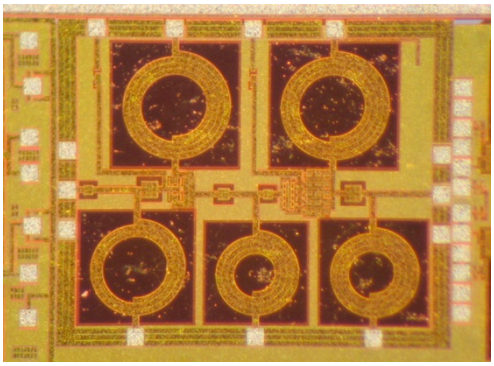
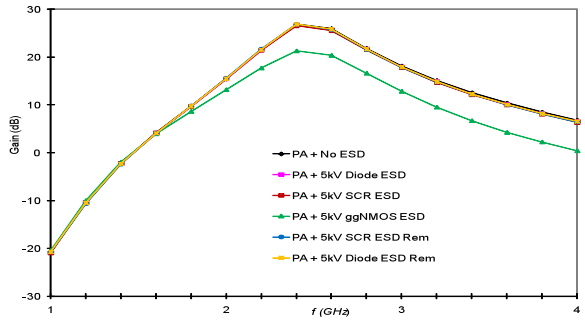
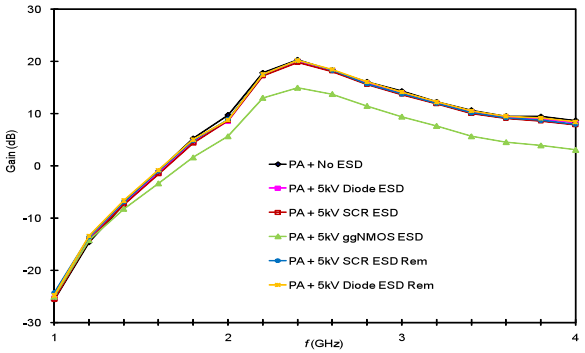


Figure 5.22. Die photo for the ESD-protected PA circuit fabricated in a foundry 0.18μm RFCMOS.

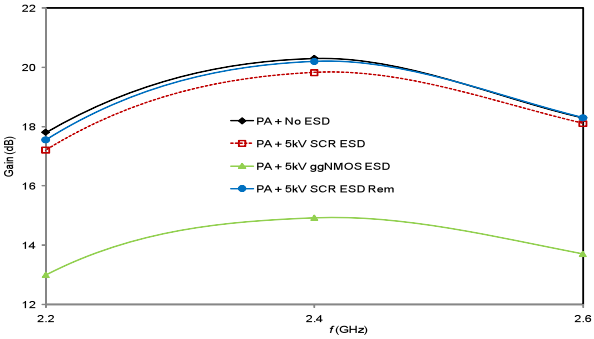
The designed nominal specs for the PA include a gain of 27.65 dB at 2.45GHz, a maximum linear output power of 18.64dBm, a P_{-1dB} of -8dBm and a PAE of 22.8%. Figure 5.23 gives the gain results for the PA with different ESD protection by post simulation co-design and measurement. It is readily observed that ggNOMS is again a poor ESD protection solution, causing up to -42.26% gain reduction, while diode and SCR ESD structures are optimized for low parasitic effects. Figure 5.23(c) shows that even the optimized SCR ESD protection causes sizable PA gain drop, which, however, is almost fully recovered by using the new I/O re-matching co-design technique (~99%). The observed gain difference between post simulation and measurement is mainly due to bias variation in Si induced by series resistance from a layout problem.



(a)



(b)



(c)

Figure 5.23. Gain data for the PA with different ESD protection structure: by post simulation co-design (a) and measurement (b) show the impacts of ESD parasitic on PA performance. (c) reveals the detailed gain recovery using new co-design method.

In the third design example, a 2.4GHz single-balanced direct-conversion mixer is designed with ESD protection at its RF and LO ports. Figure 5.24 shows the schematic for the mixer featuring an IF of 100MHz, a conversion gain of 0.1dB and NF of 20dB.

The following figures give the reflection S_{11} results at the input port that shows that ggNMOS ESD causes significant more reflection (~+83.3%), while both the optimized diode and SCR.

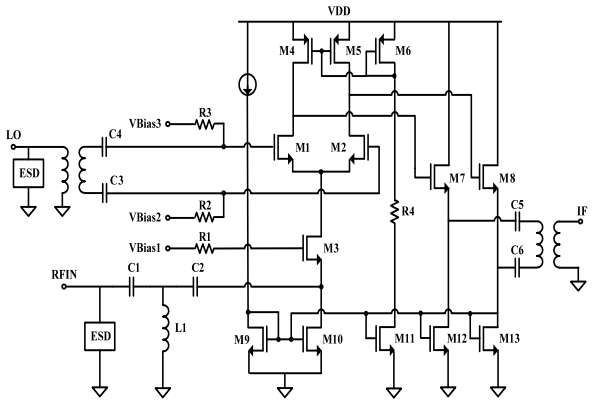


Figure 5.24 An ESD-protected 2.4GHz mixer circuit designed in this work has the ESD devices connected at its RF and LO ports.

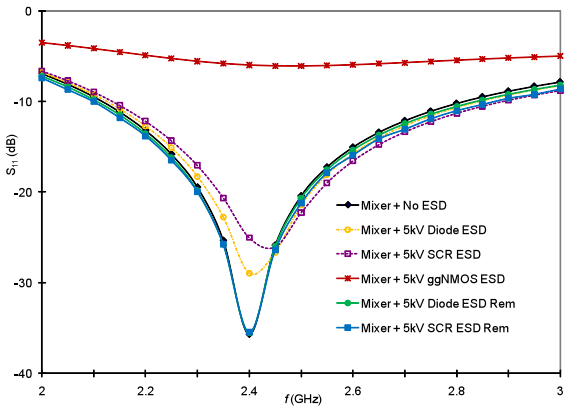
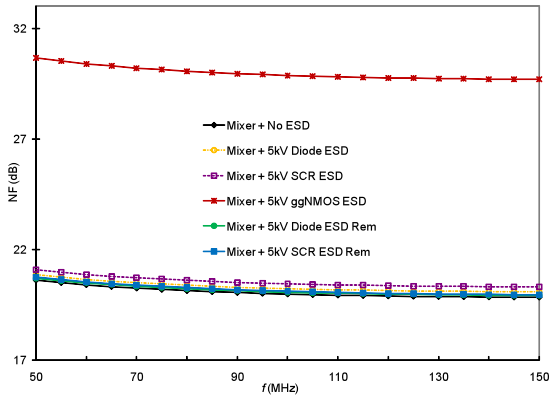
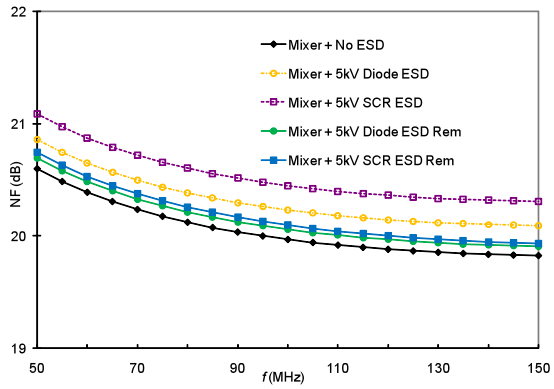


Figure 5.25. S_{11} results for the ESD-protected mixer clearly reveal negative influences of different ESD protection devices on mixer performance.



(a)



(b)

Figure 5.25. NF results for the ESD-protected mixer show ESD impacts on mixer performance, post simulation (a) and measurement (b).

ESD still affect S_{11} performance, which can be recovered by re-matching co-design (by almost 99.4%). Figure 5.26 is the NF simulation and measurement results for the mixer with different ESD protection, which again reveal that ggNMOS is a poor RF ESD protection solution (up to +33.25% increase in NF), and diode and SCR ESD structures introduced much lower extra noises to the mixer. Figure 5.26 (b) clearly shows that after re-matching co-design, the mixer circuit noise performance can be recovered almost completely (by ~98%).

All the above three ESD-protected RF IC design examples strongly support the argument that any ESD protection structure may introduce significant parasitic effect that adversely affect RF IC circuit performance. The new ESD-RFIC co-design method can be used to achieve ESD-protected RF IC circuit design optimization at whole chip level.

Chapter 6
New methodologies for the design of ESD protection strutures

The rapid evolution of the semiconductor industry makes it increasingly difficult to maintain technology- and application- adapted on-chip ESD protection, since existing ESD solutions may need to be redesigned or even redefined within short time cycles. The design of protection circuits is commonly accomplished following an experimental approach. In the best scenario, this procedure may lead to a fast solution, but more often results in a long cycle of experiments and numerous test devices that may or may not lead to the required ESD protection solution. This uncertainty in the design of ICs' ESD protections is exacerbated in new technologies- and circuit- applications, in which the window for the protection devices is narrower, the devices are required to respond fast, incorporate minimum parasitic components, and occupy the minimum silicon area [1, 2, 7, 10].

As the semiconductor industry evolves, changes are not only observed in the conception of the technology, but also in the companies' target and corporate strategies pursued to gain markets and establish benchmarks in specific sectors. These strategies have resulted in expansion of fabrication outsourcing to specialized foundries, and consequently, less number of IC design companies running in-house fabrication facilities. Using foundry processes, however, the designers are restricted to the specific technology and a pre-established process. Furthermore, they have also limited information to reformulate design rules or generate custom structures that may be necessary for implementation of different ESD protection designs.

Even though the implementation of application-specific ESD solutions at the process development phase provides a better possibility of reliable ESD design libraries, each condition for the designers and process type requires its own ESD design methodology. The approaches can be either empirical, more systematic through the use of TCAD simulations, or alternatively combining experimental procedures and TCAD simulations. In this chapter, the design methodologies that combine the previous approaches are discussed.

1. ESD Protection Design Following an Experimental Approach

Similar to the methodology already established in the industry to optimize the technology performance and reach certain design goals, the efficient implementation of ESD protection devices also requires a disciplined characterization of the specific process, and pursues design strategies to comply with the design-specific requirements. In an effort to address the strategic ESD planning and provide a methodology to estimate ESD technology benchmarking, the SEMATECH working group has proposed standardized structures, testing and equipment that allow for ESD technology roadmap assessment [13, 14].

SEMATECH working group's ESD assessment identifies standard structures provided in the process, evaluates ESD robustness of CMOS technology, and provides insights about possible strategies to improve the ESD performance. Besides this standard experimental method to evaluate and project the ESD capability of the technology, there are also other subsets of nonstandard structures embedded in the processes. These structures can be also optimized to provide custom ESD protection, thereby increasing

the amount of circuit applications that can be implemented in the specific technology. Evaluation and design of such custom ESD protection structures following an experimental approach, however, commonly require long development cycles. A comprehensive flow diagram summarizing the FA (failure analysis) and redesign cycles of ESD development following an experimental approach is shown.

As described in chapter 5, custom ESD protection devices need to depict specific conducting characteristics depending on the design window set by the technology- and the circuit application- requirements [3, 11, 15]. Commonly, the ESD design library does not provide devices optimized for each circuit application or specific customers' demands. To address this limitation, structures are investigated by building matrices of test devices such as the one shown in Figure 6.1. In this case, a well-defined experimental matrix of devices can provide insights about possible solutions and failure limits, but it may also result in useless results.

Considering the elevated cost of processing and the necessary time for testing, a purely experimental approach for the design of ESD devices is accompanied by some disadvantages. The uncertainty and drawbacks of the experimental approach lead to the search of alternative ways to obtain relevant design information. As it is described in the next section, TCAD simulations can play an important role in reducing the uncertainty of the experimental approach trial-and-error characteristic, and thus support ESD protection design and expedite the development time.

2. ESD Protection Design Based on TCAD Simulations and Measurements

The use of TCAD tools for process assessment, or to gain insights about the performance of device structures without the need of costly fabrication runs, is vital to lead innovations and new developments [1, 5, 8, 16]. Even though this premise is followed in different areas of technology development, it is still not the case when it comes to the use of TCAD for ESD protection design and evaluation. The most common practice is to achieve ESD robust IC design by destructive testing and physical failure analysis (FA).

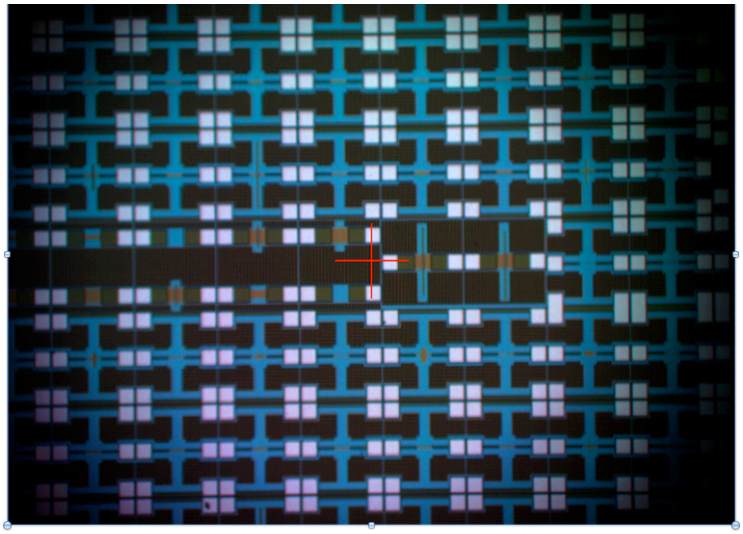


Figure 6.1. Layout Top-View Of an ESD Experimental Matrix.

The design of ESD protection structures using TCAD simulation tools is a systematic method to evaluate in the pre-silicon phase 1) device's I-V characteristics, and 2) critical physical properties, such as current density, electric field, temperature distribution, and impact ionization rate, which are properties that cannot be directly assessed experimentally. Nowadays, there are different industry-standard TCAD tools

that may provide different sets of models, numerical methods and flexibility to accomplish the simulation. In general, the TCAD device simulator uses a set of discrete fundamental equations, which correlate the electrostatic potential and carrier density within a finite element grid. In this study, the set of TCAD simulation and visualization tools are from Silvaco [3, 4].

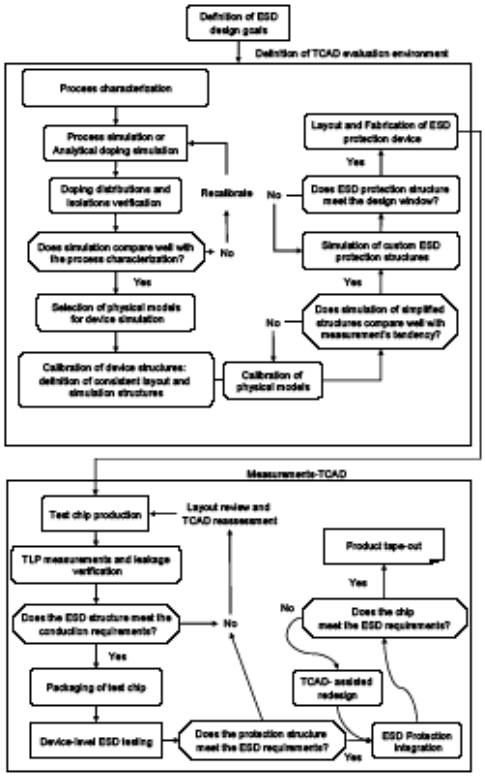


Figure 6.2. Flow Diagram for TCAD-Guided ESD Design

TCAD-assisted ESD design is affected by the limitation in predictability of the TCAD tool, as well as in the measurement setups necessary to realize reliable calibration of the simulation. On the other hand, electro-dynamics, solid-state physics, and thermodynamics models are also necessary in the simulation environment, resulting in complex differential equations systems for which a general solution is not possible.

Instead, a discretization with a finite number of elements is used, and solutions are approximated by stable numerical solvers. As a result, the mesh size, number of grid points, convergence and approximations establish a tradeoff in the utilization of the TCAD tool for ESD design. In the next sections and chapters, different methods are presented based on steady-state and transient simulations of the ESD device to assist in the design of custom protection structures, regardless of the complexity of the device structures or intrinsic limitations previously pointed out.

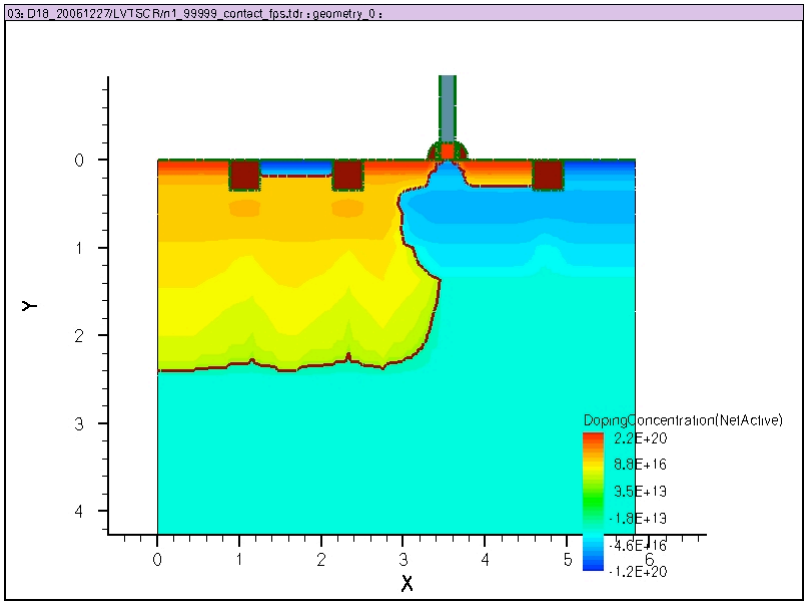


Figure 6.3. The Cross-Sectional View of a Structure Generated in One of the Fabrication processes utilized.

Figure 6.2 shows a flow diagram for a TCAD-guided ESD design, which also incorporates a reduced number of fabrication experiments [9, 14, 17]. The TCAD methodology starts from the definition of the simulation environment, including: 1) process characterization, 2) simulation/calibration of the technology, 3)

selection/calibration of the models for the device simulation, 4) simulation of the custom ESD protection device, 5) layout and fabrication, and 6) TCAD-assisted redesign. During the calibration, the focus of attention lies on the choice of physical models and appropriate parameters, which are obtained by fitting the measurement results. Calibration of the process simulation can be accomplished by defining the detailed flow of the technology, or by analytically approaching the different doping and isolating regions defined in the process.

Figure 6.3 shows the cross-sectional view of a structure generated in one of the fabrication processes utilized in this study, using the simulation of the actual fabrication process. doping profile can be obtained by cut line in the device of Figure 5.3. In this case, the analytical simulation closely resembles the results of the process simulation and actual data measured via SIMS. Different alternative methods can be used for gathering information about the cross-sections, doping distributions and isolations characteristics. This information can be subsequently utilized to define the input deck of the TCAD environment, even for those cases where foundry processes are being used and the process information is not available.

3. Physical Models Incorporated in the Device Simulation

In the previous section, simulations have been included in the design flow to take the place of large numbers of process runs and layout structures, thereby the time and cost of ESD technology development can be reduced. Due to the complexity of the device structures introduced in this study for the custom design of ESD protection systems, incorporation of electro-thermal numerical simulation is necessary to assess the

device behavior and critical physical phenomena in the design of ESD protection devices. The set of physical models used along with the thermal equation in device simulation are described below.

The classic heat flow equation is given by equation 6.1 :

$$\rho \cdot C \cdot \frac{\partial T}{\partial t} = H + \nabla(\kappa(T) \cdot \nabla T) \tag{6.1}$$

where ρ is the density (g/cm3), *C* is the specific heat (J/g·K), κ is the thermal conductivity of the material (W/cm·K), and *H* is the heat generation term (W/ cm3) [115], [118]. In the electrothermal simulation, the heat generation term is modeled as:

$$H = \bar{J}_n \cdot E + J_p \cdot E + H_{GR} \tag{6.2}$$

where *n J* , *p J* are electron and hole current densities, *E* is the electric field, and *HGR* is the generation recombination contribution expressed by:

$$H_{GR} = -GR \cdot E_g \tag{6.3}$$

where *GR* is the generation-recombination rate, which will be presented later, and *Eg* is the band gap. Related to the band gap definition, the simulation also incorporates the standard set of equations associated with the theory of carrier statistics, i.e., Fermic-Dirac/Boltzmann- statistics, effective density of states, intrinsic carrier concentration, and the bandgap narrowing.

The Poisson Equation is derived from Maxwell's equations and relates the space charge density to the electrostatic potential. Since the lattice temperature is no longer spatially constant, the Poisson equation is written as:

$$\nabla \cdot \nabla(\Psi - \theta) = -\frac{q}{\varepsilon(\bar{r})} \cdot \rho(x) - \rho_f = -\frac{q}{\varepsilon(\bar{r})} \cdot \left(p(x) - n(x) + N_D^{+} - N_A^{-}\right) - \rho_f \tag{6.4}$$

where q is the electron elementary charge, $\varepsilon$is the permittivity of the material, Ψ is the electrostatic potential, the electric field relates to the electrostatic potential through $\Psi - \nabla = E$, $n(x)$ and $p(x)$ are electron and hole densities, corresponds to the net ionized impurity concentration,

$$N_D^{+} - N_A^{-} \tag{6.5}$$

$f\rho$ is the fix charge density, and θ the band structure parameter given by:

$$\theta = \chi + \frac{E_g}{2q} + \frac{kT}{2q} \ln\left(\frac{N_C}{N_V}\right) \tag{6.6}$$

where χ is the electron affinity, k is Boltzman's constant, T is the local lattice temperature, and NC and NV are the conduction band and valence band density of states.

The Continuity Equation states the net conservation of charge principle in any possible volume of the semiconductor:

$$\frac{\partial n}{\partial t} = GR + \frac{1}{q}\nabla \cdot \bar{J}_n$$

$$\frac{\partial p}{\partial t} = GR + \frac{1}{q}\nabla \cdot \bar{J}_p \tag{5.7}$$

where *GR* represents the net generation/recombination rate. The parameters of the generation/recombination rate are also temperature dependent. It can be expressed as:

$$GR = G - R_{SRH} + R_{AUG} \tag{5.8}$$

where *G* is the generation rate mainly dominated by impact ionization, and it is modeled by:

$$G = \frac{1}{q}\left(\alpha_n \bar{J}_n + \alpha_p \bar{J}_p\right) \tag{5.9}$$

in which *n* α and *p* α are the electron and hole ionization rates. To describe the dependence of the ionization rate on field and temperature, the simulation incorporates the model proposed by Selberherr [13]. and are the Shockley-Read-Hall- and Auger-recombination rates, respectively. These recombination rates are described by the general equations:

$$R_{SRH} = \frac{np - n_i^2}{\tau_p(n + n_1) + \tau_n(p + p_1)}$$

$$R_{AUG} = \left(C_n n + C_p p\right)\left(np - n_i^2\right) \tag{5.10}$$

where $n\tau$ and $p\tau$ are minority electron and hole lifetimes, is the intrinsic concentration, and constants defined by the n/p concentrations associated to the trap level, and C_n, Cp are constant parameters. Additionally, the Shockley-Read-Hall recombination also incorporates the concentration dependent lifetime models as described.

Current Density Equations describes the transport of charge. To consider changes in the spatial lattice temperature, additional thermal-diffusion terms are placed in the current density equations.

$$\bar{J}_n = qn\mu_n E + k\mu_n\left(T\overline{\nabla}n + n\overline{\nabla}T\right)$$

$$\bar{J}_p = qp\mu_p E - k\mu_p\left(T\overline{\nabla}p + p\overline{\nabla}T\right)$$

(5.11)

where $n\mu$ and $p\mu$ are the electron- and hole- mobility, respectively.

The electron- and hole- mobility model incorporated in the simulation [63] considers the doping, temperature, parallel electric field and perpendicular electric field effects. In this mobility model, the field-, doping-, and temperature- dependent part of the mobility are represented by three components that are combined using the Mathiessen's rule:

$$\mu_T^{-1} = \mu_{SR}^{-1} + \mu_{AP}^{-1} + \mu_{OIP}^{-1}$$

(5.12)

where *SR* μ accounts for the surface roughness scattering, *AP* μ accounts for the acousticalphonon scattering, and *OIP* μ accounts for the optical inter-valley phonons scattering. These mobility terms are functions of different parameters, i.e., *SR* μ is a function of the doping and electric field, *AP* μ is a function of the doping, electric field and temperature, while *OIP* μ is a function of the temperature and electric field. In the case of a high electric field, the velocity saturates and the Caughey-Thomas expression is used to provide smooth transition between low field and high field behavior:

$$\mu_{n,p} = \mu_{Ln,p} \cdot \left(1 + \left(\frac{1}{1 + \left(\frac{\mu_{Ln,p} \cdot E'}{v_{sat}} \right)^{\beta}} \right) \right)^{\beta^{-1}} \tag{5.13}$$

where *p Ln*, μ is the low field mobility, βis a fitting parameter, commonly 1 for electrons and 2 for holes, and *E'* is the electric field in the direction of the current flow. This mobility model is implicitly dependent on the lattice temperature *T* (K) through the temperature-dependent saturation velocity:

$$v_{sat} = \frac{2.4 \cdot 10^{7}}{1 + 0.8 \cdot \exp\left(\frac{T}{600} \right)}. \tag{5.14}$$

The two-dimensional (2D) process and device simulators allow the user to create a 2D (two dimensional) cross-section of a semiconductor device, including definition of

silicon and oxide regions, doping profiles, and electrodes, and then simulate the I-V characteristics of the device. Coefficients for various mobility models, impact-ionization models, and material parameters can be slightly adjusted to calibrate simulations to experimental data, but even non-calibrated simulations can offer the necessary qualitative understanding and tendencies in the device performance with changes in the structure. The numerical analysis using the previous set of physical models, allows the user to examine many physical properties at all the locations in the ESD device for any simulated I-V point and thermal border condition.

4. Steady-State and Transient Simulation.

In addition to predicting I-V curves, simulations can identify the point of device failure by monitoring the electric field, temperature, and other properties throughout the device. Transient simulations can be used to approach tests such as the human-body model, charged-device model, and transmission-line pulsing, by using simplified waveforms (see Figure 5.4), while steady-state I-V sweeps are useful in predicting junction breakdown voltages, or even S-type I-V characteristics of ESD protection devices.

A curve-tracing technique is used to automate the steady-state simulation of complex I-V curves. Automation of complex simulations, such as latchup or snapback, saves the time and effort needed to manually change simulation boundary conditions any time there is a sharp turn in the I-V curve. This algorithm is discussed earlier. Figure 6.5 shows a schematic example of the simulation setup, and example I-V steady-state simulation results for one of the devices studied in this work. The 2D contours of lattice

temperature, conduction current density and impact generation rate at the two points indicated in Figure 6.5(b) are depicted in Figure 6.6. In this case, the snapback characteristics are predicted, along with the changes in the 2D contours at different operating conditions. Similar contours can be obtained for the other physical parameters of the device previously discussed, and tendencies with changes in the device cross-section can be evaluated.

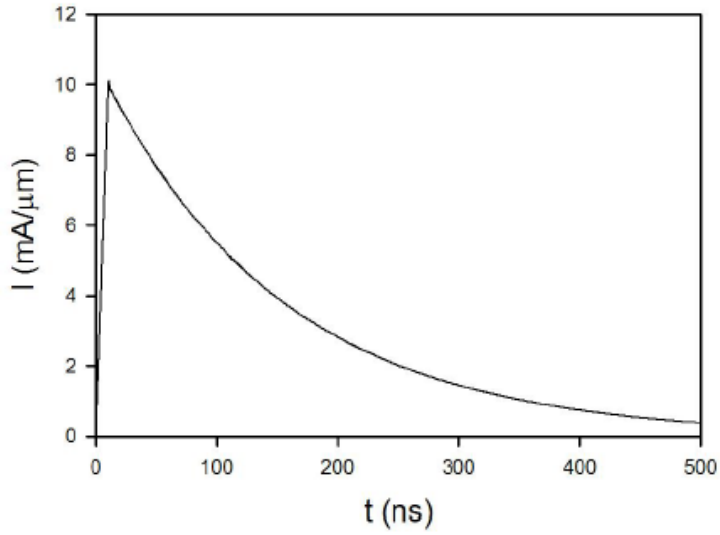
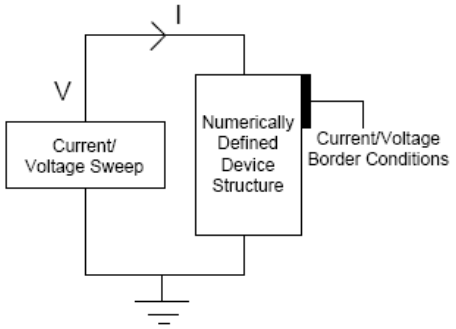
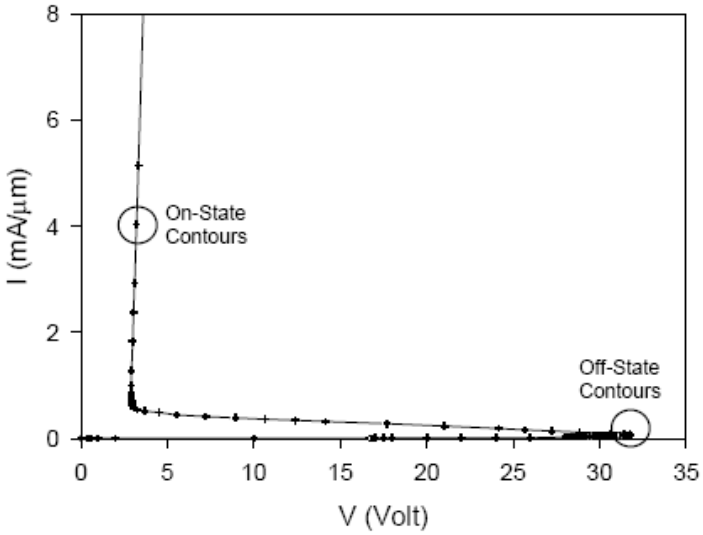


Figure 6.4. TCAD-simulated transient waveform using 10 ns rise time and 150 ns decay time. This is the approximate transient characteristic of an HBM event.



(a)



(b)

Figure 6.5. (a) General Schematic of the Device Simulation, and (b) Example of Simulated S-type IV Characteristics and Points where Contours are Taken.

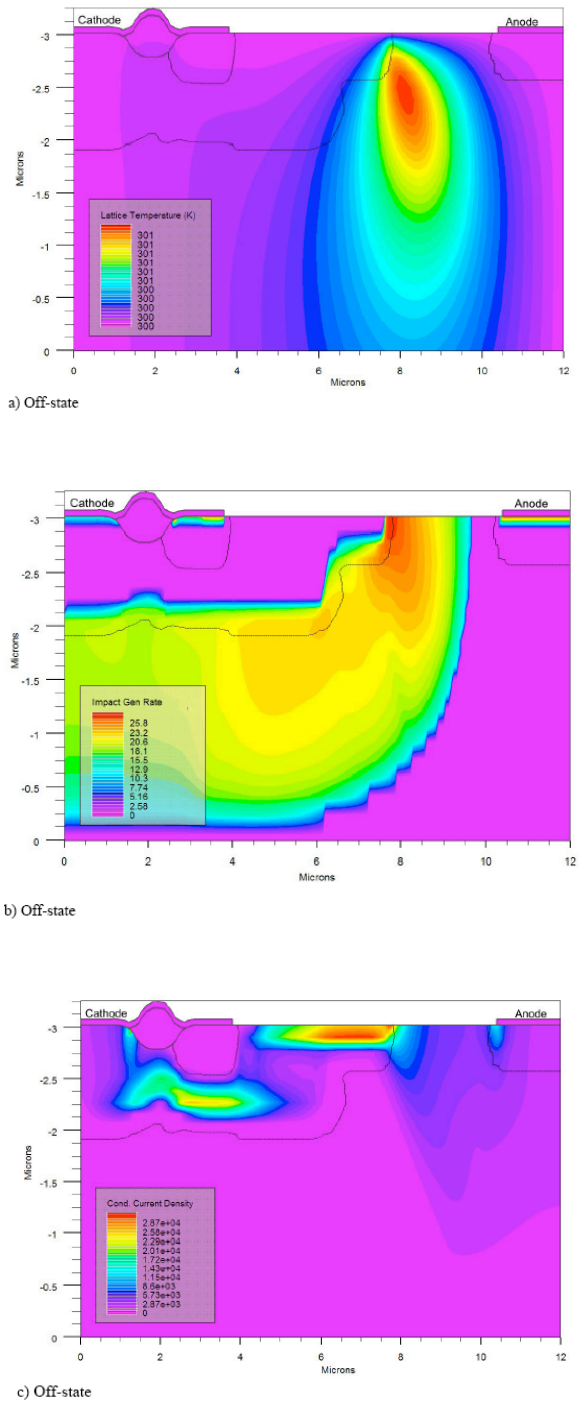


Figure 6.6. Comparison of On- and Off-state 2D Current Density Contours of (a) LatticemTemperature in (K), (b) Impact Generation Rate and (c) Current Density (A/cm).

The contours shown in Figure 2.6 give important insights that are used along this study in the implementation of tailored ESD devices. For instance, the lattice temperature contour shows the optimum location for the hot spot deep in the device during the on-state. Likewise, the current density evolves from a narrow conduction area, when the device is operating close to the trigger point, to a well-spread conduction along the device during the on-state. Additionally, the impact generation rate gives results that can be used for calibrating the trigger voltage.

The time required for the transient simulation previously described can be considerably more than that required for the steady state simulations. However, in the transient simulation the device response can be also evaluated and optimized at different rise-times, which can be useful when designing for protection against CDM.

5. Mixed-Mode Simulation.

The TCAD mixed-mode simulation creates the capability of embedding one or more numerically simulated devices in a SPICE-like circuit that may include lumped resistors, capacitors, and inductors as well as voltage sources, current sources, and compact models for diodes, MOSFETs, or BJTs. The total circuit is normally solved in a coupled manner, in which the semiconductor equations previously discussed are still incorporated in the simulation of the 2D devices, as well as the current and voltage Kirchhoff circuit equations given below:

$$\sum_{x=1}^{m} i_x = 0\,,\ m = \text{ number of paths converging in one node,}$$

$$\sum_{y=1}^{n} v_y = 0\,,\ n = \text{ number of branches in a close loop.} \tag{2.15}$$

Mixed-mode simulations can be used for transient characterization of ESD tests, such as the ESD standards discussed in chapter 1. This type of simulation can also be used to generate the I-V points of the snapback curve typically obtained in ESD protection devices. Figure 5.7 shows a schematic representation of a mixed-mode simulation incorporating a numerically-described ESD protection device.

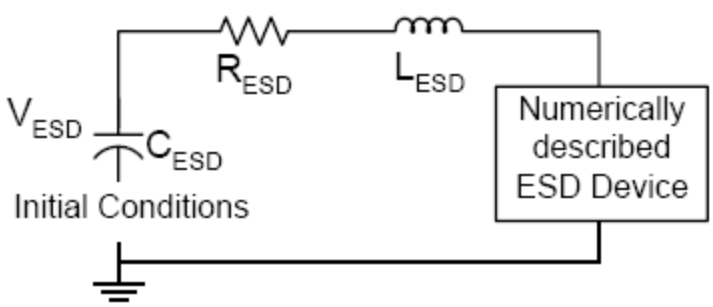


Figure 6.7. Schematic Representation of Mixed-mode Simulation Using Simplified ESD Model Lumped Circuit and Numerically-Simulated device.

Applications of the mixed mode simulation have been demonstrated for assessing the I-V characteristics, and device performance, under events closer to those obtained from the actual ESD tester [7, 11, 12]. This type of simulation, however, is time consuming and convergence problems are difficult to solve when the device structure is complex. Thus, even though for some analysis mixed-mode simulation it may be

necessary in the assessment of ESD protection designs, there are intrinsic limitations regardless of the approach that is followed. Therefore, reevaluation of the TCAD strategy is also necessary for selecting the most efficient way to obtain the information.

Chapter 7
ESD Protection design for 0.18μm/0.13μm CMOS technology

Based on the mixed-mode ESD protection simulation-design methodology developed before, ESD devices that are expected to protect 0.18μm and 0.13μm CMOS IC chips have been designed and simulated. The design of the individual devices is discussed in this chapter and the corresponding Sentaurus simulation will be discussed in the Chapter.

1. Basic requirements for ESD devices

The requirement for ESD protection circuits to carry currents way beyond the levels for which the elements were initially designed results in regions of high thermal dissipation and high electric fields [1, 3, 4, 6-12, 14, 17, 18]. As discussed in Chapter 4, a good protection circuit needs to be able to withstand the heating effects, sink the large currents during the ESD event, and not be damaged by the ensuing high electric fields.

The capability to meet these requirements is critically dependent on the specific design and layout of the protection circuits and the individual elements as well as the circuits that are being protected. Hence, ESD design is considered to be a very layout-intensive activity. Furthermore, while the design concepts have to change with the technology, the layout techniques must be revised to make them compatible.

In this chapter, we present and discuss the design of ESD devices for input/output pins. The focus will be on circuits designed in advanced CMOS processes, but we will also present some typical approaches for bipolar and BiCMOS circuits. Specific examples will be given to illustrate approaches to protection design and layout.

The first requirement for a good protection circuit design methodology is the choice of the appropriate type of protection device that is compatible and/or suitable for the technology [6]. Next, the chip function must be considered, and the operating requirements for the circuit being protected must be compatible with the choice of protection circuit. This includes the operating voltage conditions, capacitance and resistance loading, and area requirements. Finally, the customer and product engineering requirements for ESD and handling capability must be weighed against the possible impact of the protection circuit on the performance of the IC. For the protection circuit designs described here, each of these aspects is considered in detail.

The protection devices mentioned previously are described in the following sections. The effectiveness of the total protection concept is realized only when all of the components are harmonized to work together during an ESD event.

2. ESD protection design for a 0.18μm CMOS

The goal of this project is to design full-chip ESD protection solution for a 0.18μm 1.8V & 3.3 CMOS technologies that consume as little chip are as possible. According to given data, the initial proposed ESD protection structures are: ggNMOS, ggPMOS, SCR, MVTSCR, LVTSCR; the proposed full-chip ESD protection scheme are power-clamp and ESD bus.

After defining the ESD specifications, the next step is to select initial ESD protection structures according to core circuit specifications and process features. There are many possible ESD protection structures, i.e., diode, GGNMOS/GGPMOS, SCR, etc. However, not all of them may be applicable in a specific design. One has to carefully

choose the most possible structures based upon specific process technologies and circuit features.

Next, the selected possible ESD protection structures are created by process simulation using TCAD tool, TSUPREiM4**,** which solves process model equations at each node of the device mesh and produces detailed impurity doping profiles. It is imperative for one to start the design from process simulation because many advanced process steps have negative impacts on ESD robustness. These types of considerations can be simulated in design phase.

A. Initial Calibrations.

Calibration is one of the most complicated yet critical steps in mixed-mode ESD simulation [7], which includes both process calibration and device characteristic calibration. Process calibrations focus on the parameters related to IC processes, e.g., doping profiles and voltage threshold values. Device characteristic calibrations are made by comparing steady-static and transient simulation results with related testing data performed by curve tracer, TLP (transmission line pulsing) and ESD zapping tests. The main calibration covers mobility model parameters, material parameters and thermal coefficients. Impact ionization calibration is a critical step. Physical layout design follows the mixed-mode ESD simulation.

A variety of test patterns for both ESD protection structures and core circuits will be generated in layout for testing and calibration purpose. Generally, these ESD test structures should be as simple as possible in order to avoid any confusion.

B. Individual Structure Design and Simulation.

Any I/O protection circuit should provide a low-impedance path from input to supply during an ESD event to absorb current but provide a very high impedance during normal operating conditions so as not to affect circuit performance, e.g., through increased leakage current or parasitic capacitance. Additionally, an ESD circuit should clamp input voltages at a safe level, i.e., below the dielectric breakdown voltage of a thin gate transistor [1, 10]. The dielectric threshold electric field is actually time dependent: it must be held across an oxide for a certain length of time before the oxide breaks down, as measured by leakage current. The time to breakdown is lower for a higher stress field. Although the consequences of this time dependence on ESD protection ability are important, for simplicity we will assume that the voltage across a thin gate must not exceed some critical level for any amount of time.

When designing ESD protection circuits, there are some important differences to consider between input protection and output protection. While the high-impedance input pads of a CMOS chip are connected to the thin gates of the input buffer transistors, the low impedance outputs are connected to the drains of output-buffer transistors. Design of output protection is thus more restricted than that of input protection because of low output-impedance requirements. For example, a well resistor may be placed between an input pad and the protection MOSFET to reduce the rise time of an ESD pulse, but such a resistor cannot be placed on an output pad because the increased impedance would exceed circuit specifications. Also, since the output-protection transistors often double as the CMOS output buffer, they must meet certain chip-performance specifications. As a

result, output protection relies more on the proper layout of one or two transistors than on the use of creative circuit designs.

The work was started from ggNMOS structure parameters selection by running DC sweep. The DC sweep performed here would optimize the parameters by investigating the influence on thermal breakdown point. In order to minimize the device area, the minimum gate length L=0.18μm was applied [7, 14].

3. ESD Protection Devices

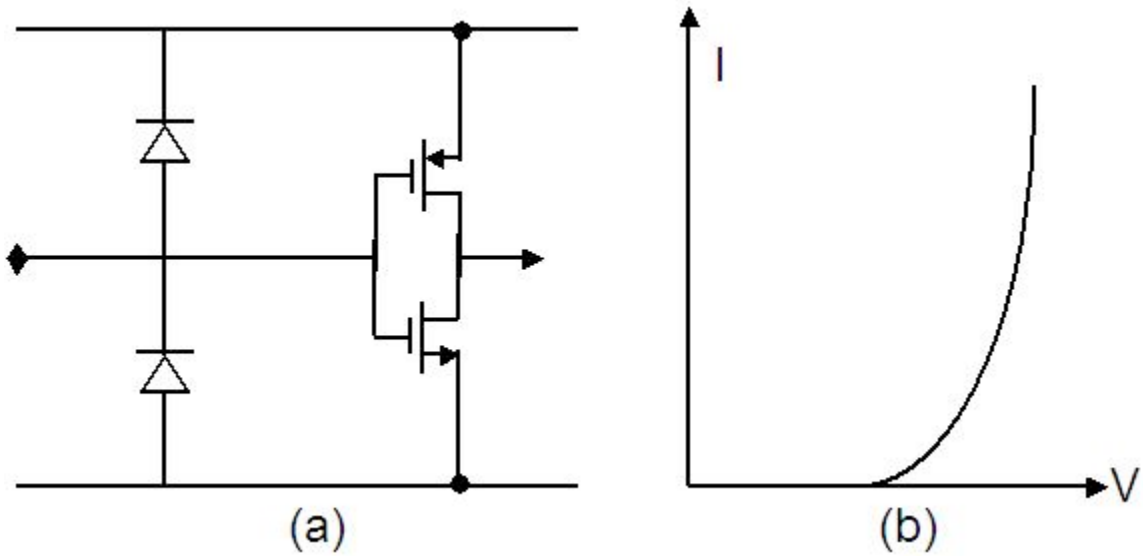


Figure 7.1. (a) Diode Protection Scheme. (b) Typical I-V Curve of a Turn-on Diode.

A. Diode.

In early ESD protection design, reverse-connected diodes are heavily used for the purpose of prevention breakage. (Figure 7.1a) Its IV curve follows the standard turn-on scheme of a typical diode without a snapback, which is graphed in Figure 3.1b. The cross

section of a typical diode is shown in Figure 7.2. In particular, Zerner diodes are often employed for this kind of protection design. This solution can simply simulate by SPICE.

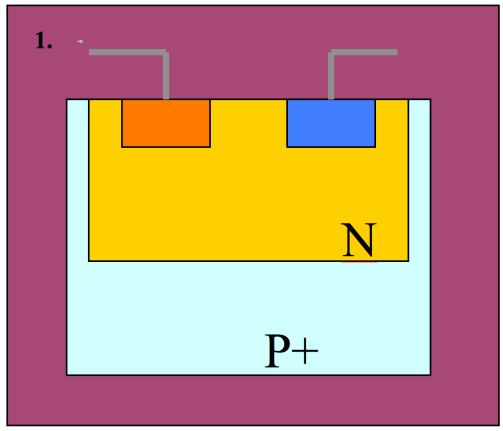


Figure 7.2. Cross-Sectional View of the Diode for ESD Protection

Depending on ESD protection operation, the diode could work under either forward or reverse bias mode. Since ESD is a high current event, the formula is given under high injection level.

In forward mode, current can be defined as the following equation

$$I = \frac{2qAD(T)n_i(T)}{L_d}\exp(\frac{qV}{2kT}) \qquad (7.1)$$

L_d Diffusion length

A Junction Area

D(T) Minority carrier diffusion coefficient

$n_i(T)$ Intrinsic carrier concentration of semiconductor (cm^{-3})

k Boltzmann's constant= $1.38 * 10^{-23}$ J/K

T Temperature (K)

V Voltage across the junctions (V)

In the reverese mode, current can be defined the following equation.

$$I_R = \frac{qADN_C N_V}{L_d N_B} \exp(\frac{-E_g}{kT}) + \frac{qW}{\tau_e} \sqrt{N_C N_V} \exp(\frac{-E_g}{2kT}) \quad (7.2)$$

N_C Density of states in the conduction band

N_V Density of states in the valence band

N_B Minority carrier doping concentration

W Width of depletion region

τ_e Effective carrier lifetime

There are some drawbacks of using diodes as ESD protection design. The fixed forward turn-on voltage of the diode restricts its application in varied-VDD cases. The problem could be improved by using multiple diodes strings (forward or reverse). Yet, the on-resistance add-up prohibits the current handling capacity of the diodes strings.

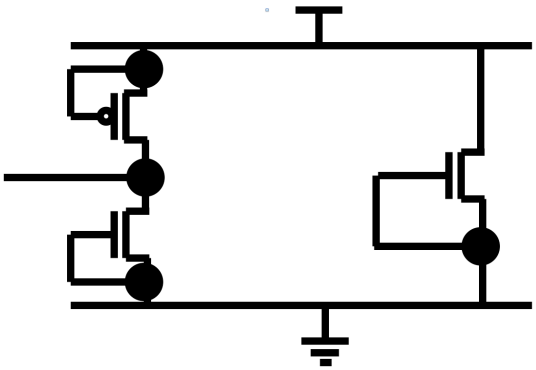


Figure 7.3. Shows a GGNMOS ESD Protection Scheme

B. Ground-Gate NMOS Transistors (GGNMOS).

The GGNMOS device is chosen since it has been the most popular protection device in use up to now. Utilizing the extracted equivalent circuit, we examine the effects of the parasitic of the NMOS protection device on the characteristics of a low noise amplifier (LNA), which is one of the important RF circuits. Figure 7.3 shows a GGNMOS ESD protection scheme, where the drain is connected to an I/O pad.

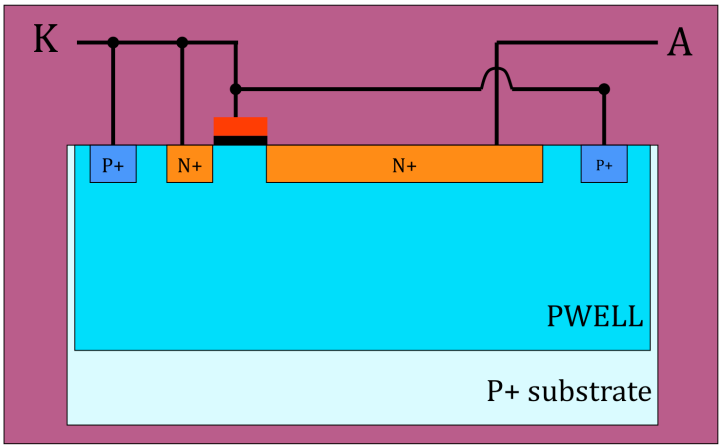


Figure 7.4. The Cross-Section of GGNMOS

C. Gate-Coupled NMOS Transistors (GCNMOS).

In most applications, the thin oxide device is used as a protection device with its gate grounded. This will always ensure that the protection device, while being robust for ESD protection, will not cause any extra leakage at the pin. However, the thin oxide device can be a more robust protection element if its gate is coupled high during an ESD event. The effect of gate bias on the ESD performance of nMOS devices has been reported. As illustrated in Figure 6.5, if the gate voltage is about 1V, the nMOS trigger is lowered to less than the onset for avalanche breakdown ($Vt1\_ < Vt1$) and is therefore ideal for improved ESD protection. It was also noted that if the gate voltage goes above 5V, the $It2$ value (defined in Figure 7.5) decreases to give reduced failure threshold voltage ($It2\_ < It2$). Hence, a gate voltage of between 1 and 2V is typically needed for best ESD performance.

In a multifinger structure, the gate coupling improves the uniform turn-on of all the fingers. This is not always possible in a grounded gate device as the first *npn* device to turn on discharges all the ESD current potentially preventing the other fingers from turning on. In such a device, each of the other fingers has a chance to turn on as the voltage increases again toward $Vt2$. That is, after one finger begins *npn* conduction and clamps at the snapback voltage, the pad voltage builds up again because of the snapback resistance. The GCNMOS can have applications for input, output, and input/output pins. For inputs it can be directly connected to the buffer gates. However, for maximum efficiency and good protection for CDM, the recommended approach is to isolate the protection device with a secondary clamp, as shown in Figure 7.5.

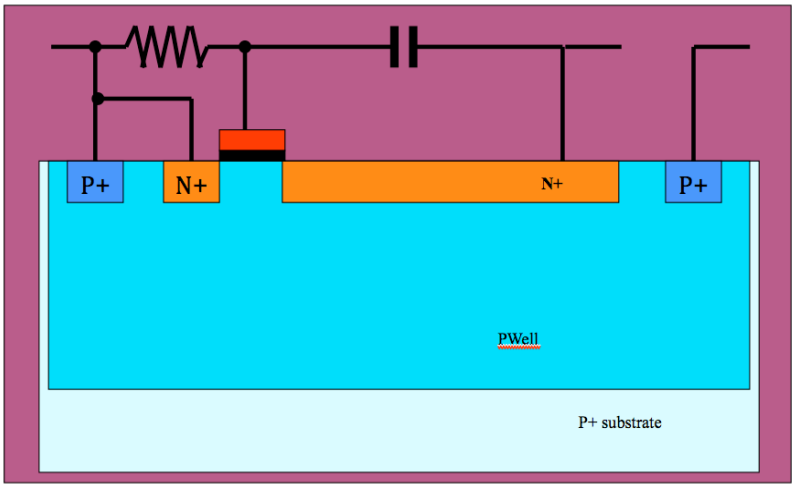


Figure 7.5. GCNMOS

Finally, the GCNMOS concept can also be applied for *V*dd protection. However, it must be noted that owing to the large capacitance associated with the *V*dd pin the transient voltage at the pad slows down and discharges the gate to be below the transistor *Vt* in a short time. In other words, the gate potential could be close to 0V when the drain potential reaches *Vt*1, making the protection device trigger nonuniform. This can be simply overcome by making the booting capacitor Cc about 10 times larger.

D. Silicon Controlled Rectifiers (SCR) protection device.

The SCR is the most efficient of all protection devices in terms of ESD performance per unit area. The basic SCR is a *pnpn* device, as shown in Figure 3.6. The device shown in this figure is also referred to as a lateral SCR or LSCR.

The adjacent *n*+ and *p*+ diffusions in the *n*-well are connected to the input terminal. A vertical *pnp* device is formed with the *p*-substrate as the collector, *n*-well as the base, and input *p*+ diffusion as the emitter. The *n*+ diffusion in the *p*-well is

connected to the ground or substrate bus and forms the emitter of the *npn* transistor. The base of the *npn* is formed by the *p*-substrate and the collector is the *n*-well and the *n*-well contact. During normal circuit operation, CMOS latchup should not be a problem as the emitter and base of the *pnp* are at the same potential. During an ESD stress pulse the collector–base junction of the *npn* goes into avalanche breakdown generating the electron current in the *n*-well which forward biases the emitter–base junction of the *pnp*. The turn on of the *pnp* occurs in less than 1 ns and this leads to the regenerative *pnpn* action. Once the SCR is turned on the device is in a low impedance state and the anode to cathode clamping voltage is of the order of 1–2V in a submicron process. This dramatically reduces the power dissipation and results in an improved ESD performance. The nature of the device operation means that it is not strongly influenced by salicidation, which is a big advantage in advanced CMOS processes.

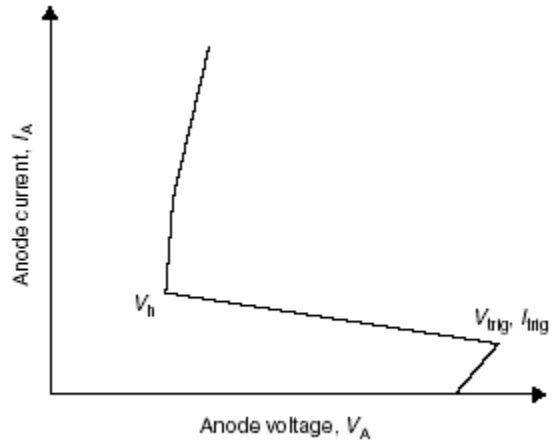


Figure 7.6. SCR Device Cross-Section and Simulated I-V Performance.

The avalanche breakdown voltage of the *n*-well to substrate is about 20V in an advanced CMOS process. In order to make the SCR a good ESD protection circuit the

trigger voltage must be reduced. This is accomplished by using additional *n*+ diffusion at the *n*-well edge. The breakdown voltage is now reduced to that of the *n*+ to substrate, which is about 15V in a submicron CMOS process. A further reduction of the trigger voltage is achieved by using a gated diode at the *n*-well edge. The gate of the nMOS transistor that forms the gated diode is connected to the cathode. The trigger voltages for these devices are between 6 and 10V in submicron processes.

E. Low Voltage Triggered SCR.

The LVTSCR essentially uses an MOS device in parallel with the SCR. Triggering occurs after the drain junction of the MOS transistor begins avalanching. (Figure 6.7) The avalanche-generated hole current in the *p*-substrate turns on the lateral *npn* and then the vertical *pnp*, followed by eventual regenerative SCR action. The low trigger voltage of the LVTSCR means that it can be used as an ESD protection device for CMOS output buffers. However, one still needs to ensure that the output device does not trigger before the SCR. This can be solved by either making the channel length of the MOS device in the LVTSCR shorter than the output device or by placing an isolation resistor between the output device and the SCR protection device. The latter option has been successfully implemented in large submicron circuits, but requires that the performance degradation of the output buffers caused by the addition of the resistor is compensated in the circuit design. An alternative is to reduce the trigger voltage of the SCR further by techniques similar to those used for the GCNMOS. By raising the gate voltage of the LVTSCR the voltage at which the SCR turns on can be significantly reduced [Diaz94]. This technique requires careful tuning to ensure that the trigger circuit

does provide the correct triggering for the SCR. For the advanced submicron technologies the MLSCR trigger voltage is around 10–12V and the LVTSCR trigger voltage is about 8–9 V. However, the MLSCR may not be reliable unless design optimization is properly done.

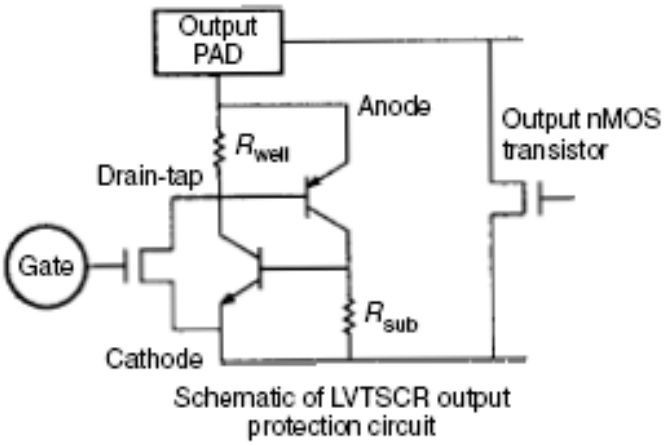


Figure 7.7. LVTSCR

Chapter 8
Non-Traditional Nano-Phase-Switching ESD Protection

This chapter summarizes a new non-traditional *nano phase switching* ESD protection mechanism and *nano crossbar* ESD structures. Prototype Nano-crossbar $Cu/Si_xO_yN_z/W$ devices show excellent ON/OFF switching and ESD protection, i.e., fast response of 100pS, ultra low leakage of 0.26pA and ESD protection of >267V/μm^2. New *dispersed local ESD tunneling model* is proposed. Heterogeneous integration into CMOS makes it an alternative solution to traditional ESD protection.

1. Introduction of ESD protection design

ITRS cites ESD protection as an emerging design challenge in sub-100nm era, particularly for RF and mixed-signal ICs. Traditional ESD protection uses PN junction based active conduction (diode, BJT, FET, SCR, etc.) to discharge ESD pulses [1]. As CMOS continues scale down, such traditional ESD structures are facing un-resolvable barriers inherent to PN-junction ESD protection. Firstly, similar to high leakage in sub-100nm CMOS devices, e.g., ~100nA/μm reported for 32nm MOSFET [2], ESD junction leakage (I_{leak}) can easily be tens of nA, intolerable to ICs. ESD device scaling-down is not possible due to reduced ESD robustness and cannot reduce leakage. Secondly, conventional PN turn-on ESD triggering may cause serious mis-triggering for GHz+ RF ICs. Thirdly, multi-supply ICs require special ESD designs with flexible triggering voltage (V_{t1}). All these problems urgently call for revolutionary non-traditional ESD protection mechanisms and structures. We report a new non-*traditional nano phase switching* ESD protection mechanism and design as a solution.

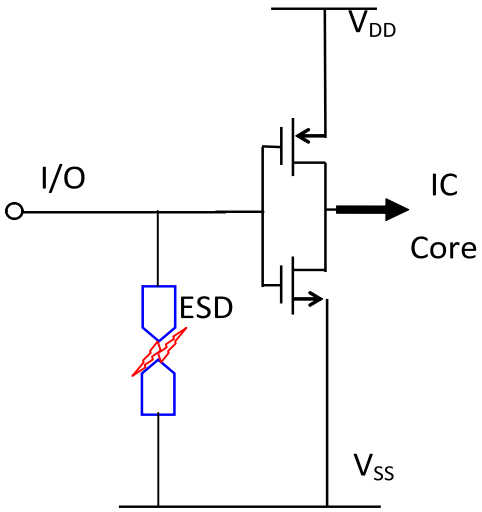


Figure 8.1 Nano-Crossbar ESD Protection Structure

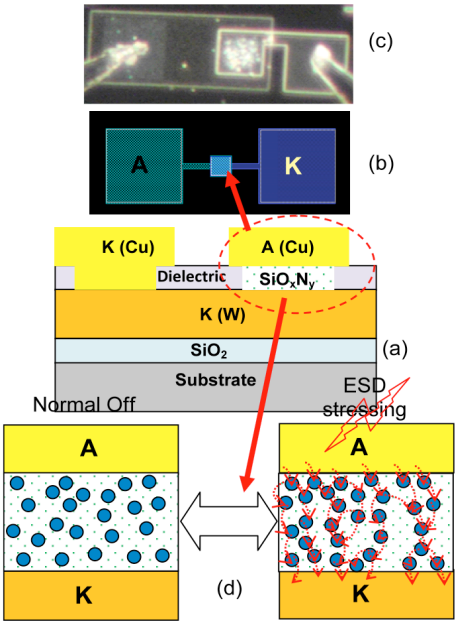


Figure 8.2 Nano phase switching ESD protection: a) x-section, b) layout for one crossbar node, c) photo, d) dispersed local ESD tunneling model.

Figure 8.1 shows common ESD protection scheme where an ESD device, connected between I/O and GND, is turned on by an ESD pulse to form an active conducting path to discharge ESD surge. Figure 8.2 depicts the new nano-phase-

switching ESD protection concept where a nanowire crossbar array is formed. Each node in the array is a sandwiched nano-crossbar device consists of an insulator and two metal electrodes (A & K), which are connected to I/O and GND. Unlike PN-based ESD device, the insulator of the new ESD structure ensures no leakage, ideally. When an ESD surge appears at I/O, electric field instantly triggers phase switching and turn on the device to discharge ESD current. After ESD pulse, it returns OFF state automatically. The key is to find magic dielectrics to enable nano phase switching for nano-crossbar ESD structure. Organic (CuTCNQ, etc.) and inorganic (SiO, etc.) materials were reported for resistive switching under DC biasing [3, 4]. We identified $Cu/Si_xO_yN_z$ as new nano phase switching structure suitable for transient ESD ON/OFF switching.

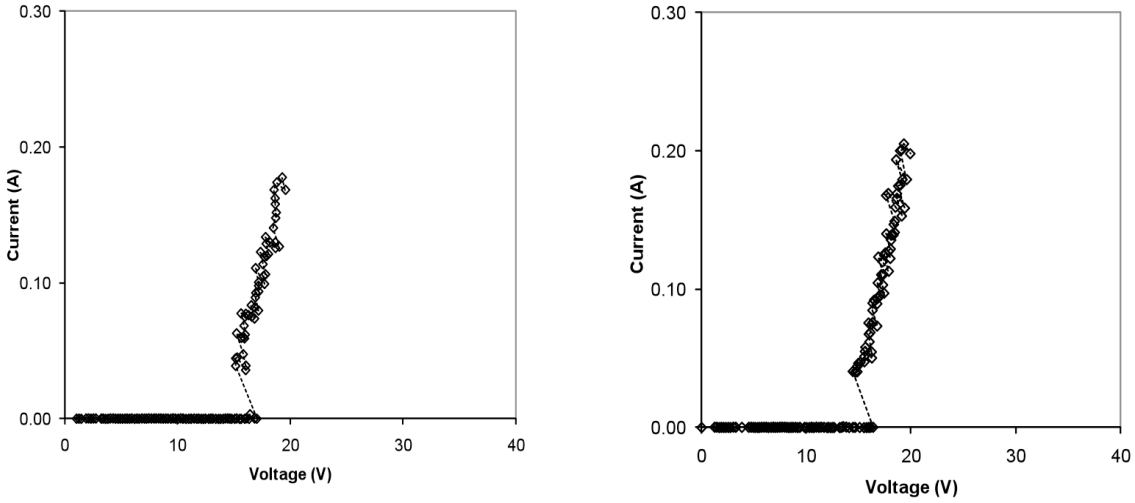


Figure 8.3 Measured I-V curves by TLP: (L) 1μX1μ, (R) 10μX10μ devices.

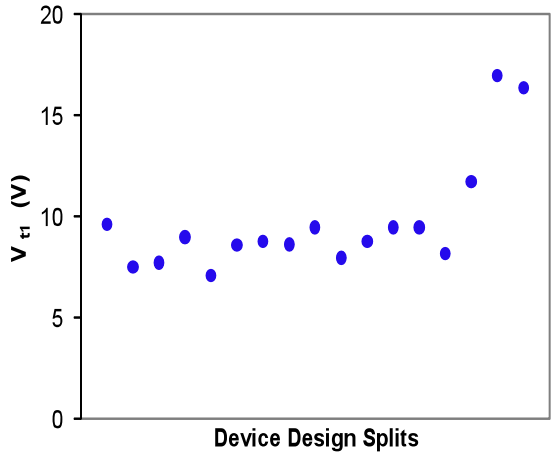


Figure 8.4 Adjustable V_{t1} for different devices.

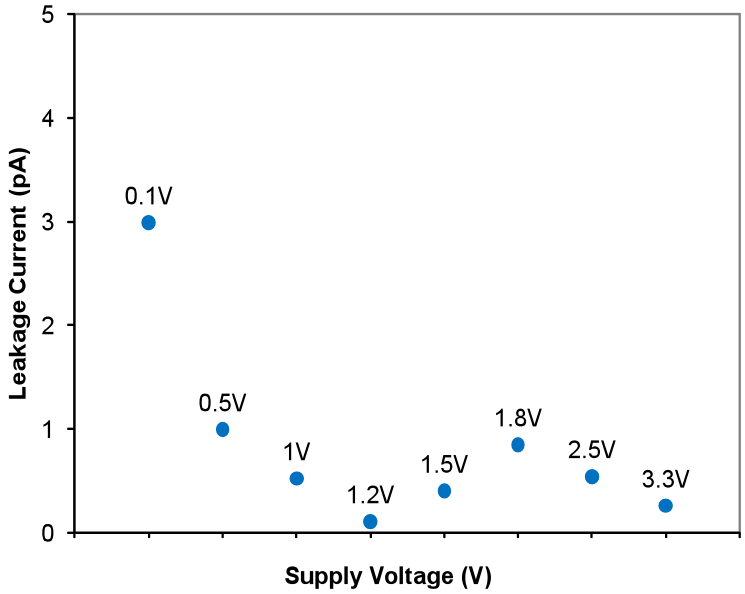


Figure 8.5 Measured ESD leakage

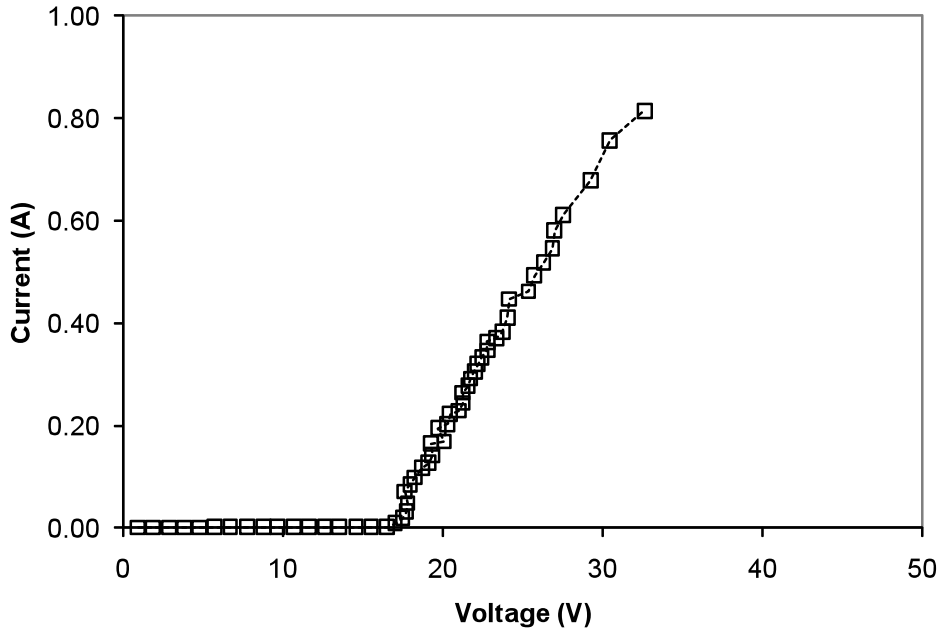


Figure 8.6 Measured fast ESD I-V by VG-TLP for a 40μX40μ device

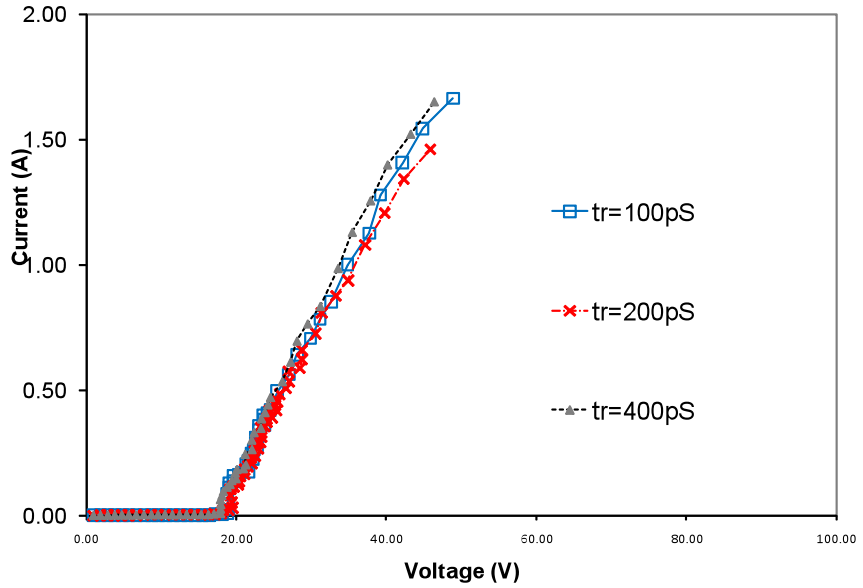


Figure 8.7 Measured I-V *vs.* pulse t_r by VF-TLP for a 60μX60μ devices.

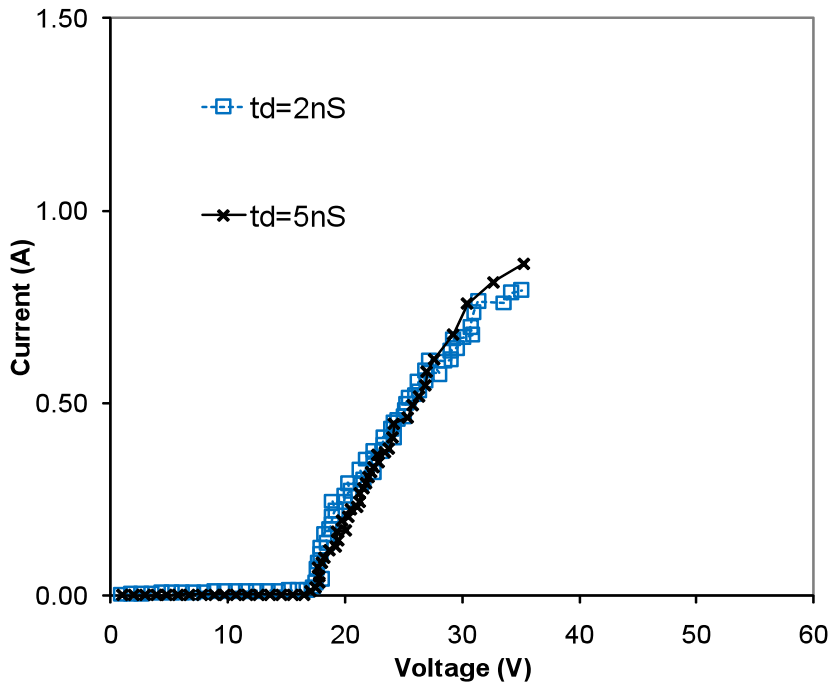


Figure 8.8 Measured I-V *vs.* t_d by VF-TLP for a 40μX40μ devices.

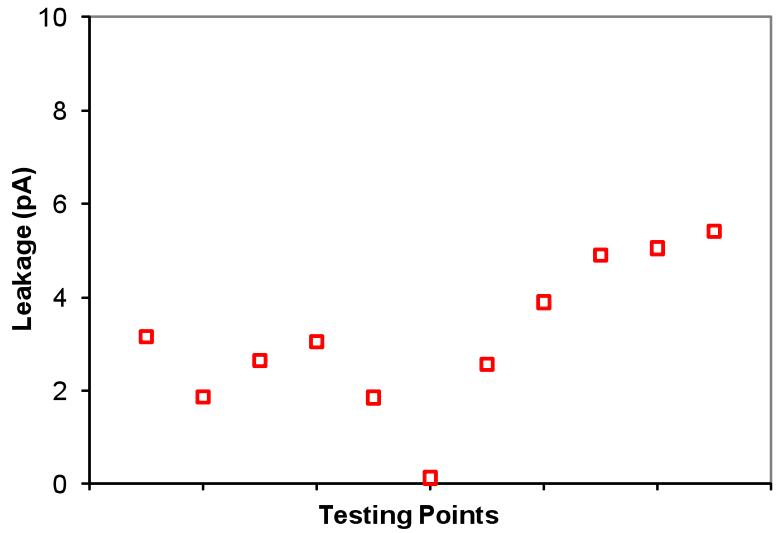


Figure 8.9 Measured leakages for a 60μX60μ device after repeating ESD stresses.

TABLE I $V_{t1} \sim t_r$ COMPARISON

VF-TLP t_r (pS)	V_{t1} (V)
100	18.833
200	17.966
400	17.883

TABLE II $V_{t1} \sim Si_xO_yN_z$ COMPOSITION

Devices	Si:O:N	V_{t1} (V)
I	1:0.95:0.54	39
II	1:0.66:0.45	16
III	1:0.42:0.39	9
IV	1:1.93:0.08	20

Figure 8.10 Tables to summarize the performance

2. Testing bed design and ESD performance measurement

New CMOS-compatible process was developed to fabricate the nano-crossbar ESD devices: A 100nm-thick W layer was deposited on SiO_2/Si substrate, followed by lithography and RIE etching to define electrode K. Next, a 50nm $Si_xO_yN_z$ layer was deposited by PECVD and tuned by reactive gas ratio of N_2O/SiH_4. A via is formed for bottom connection. Finally, Cu film was deposited by PVD followed by lift-off to define device size and pads (A & K). A large set of Cu/$Si_xO_yN_z$/W nano crossbar structures were formed with varying sizes (1μmX1μm to 80μmX80μm), $Si_xO_yN_z$ thickness and composition ratio for characterization and optimization. AES analysis confirms $Si_xO_yN_z$ atomic ratios as designed. Device photo is given in Figure 2. Complete ESD testing was conducted. A new *dispersed local ESD tunneling model* is proposed for the new nano phase switching ESD protection mechanism as depicted in Figure 2d. Porous $Si_xO_yN_z$ is formed for easy Cu diffusion. Low-T annealing pre-diffuses and disperses Cu ions into $Si_xO_yN_z$ where Cu is trapped by O/N throughout $Si_xO_yN_z$, yet the hi-R OFF state remains. When ESD transient comes, strong ESD electric field causes local electron tunneling between neighbor Cu ions dispersed throughout $Si_xO_yN_z$, resulting in low-R conduction to discharge ESD surges. After ESD stressing is over, the tunneling stops and the device

returns to OFF state. The new mechanism is reasonably verified by measurements. TLP test, for HBM ESD testing (rising time t_r~10nS & pulse duration t_d~100nS), clearly shows ESD triggering as depicted in Figure 8.3. Figure 8.4 shows desirable varying ESD triggering V_{t1} ≅7-17V for different devices. Figure 8.5 shows ultra low ESD leakage, I_{leak}≅260fA~2.99pA, desired for low-V ICs. Fast CDM ESD function (t_r~400pS & t_d~ a few nS) is verified by very fast VF-TLP testing (t_r=100, 200 & 400pS and t_d=1, 2 & 5nS). VF-TLP result in Figure 8.6 clearly shows ultra fast ESD triggering at V_{t1}≅17.5V under t_r~100pS and t_d~5nS. Figures 8.7 & 8.8 and Table I show the V_{t1}~t_r and V_{t1}~t_d results suggesting no major impact of t_r and t_d on V_{t1} in fast CDM ESD modes. It is believed that ESD triggering may be controlled by the $Si_xO_yN_z$ atomic composition because O and N play roles in trapping the diffused Cu ions inside $Si_xO_yN_z$. VF-TLP testing results given in Table II suggest that excess O/N increases ESD V_{t1} (Device I, II & III) and N may play a more sensitive role in Cu trapping (Device IV). To examine the nano phase switching and ESD ON/OFF repeatability, new ESD devices were stressed repeatedly and leakage current was tested after each TLP stressing. Figure 8. 9 gives statistics for leakage after repeating ESD pulses and clearly shows that ESD devices return to OFF state very low I_{leak}≅0.13~5.42pA after each ESD stressing. This ESD-triggered ON/OFF switching can be explained by the new dispersed local tunneling model, which is also supported by the extremely fast ESD response time of t_r~100pS in VF-TLP testing. Figure 8.3L shows that a 1μmX1μm ESD device can survive I>178mA, i.e., HBM 267V/μm^2, indicating a fairly positive ESD protection capacity. However, since prototype devices show metal damage (not SiON) under ESD stressing, we cannot find ESD protection capacity accurately for

these new ESD devices yet. While our results readily verified the new non-traditional nano phase switching ESD protection mechanism and structures, work is underway to further confirm the new ESD tunneling model and optimize ESD designs. It is expected that a nanowire crossbar array of the reported single crossbar devices shall be used in real CMOS IC designs for better ESD performance in future.

REFERENCE

Chapter 1:

[1] http://www.ieee802.org/15/pub/TG3a.html.

[2] "'Multi-band OFDM Physical Layer Proposal for IEEE 802.15 Task Group 3a'," IEEE P802.15 Working Group for Wireless Personal Area Networks, March 2004.

[3] B. Razavi, H.-C. Kang, C.-C. Hsu, and C.-C. Lee, "Multiband UWB transceivers," in Proceedings of the IEEE Custom Integrated Circuits Conference (CICC '05), pp. 140-147, San Jose, Calif, USA, September 2005.

[4] A. Ismail and A. Abidi, "A 3 to 10 GHz LNA using a wideband LC-ladder matching network," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '04), vol. 1, pp. 384-534, San Francisco, Calif, USA, February 2004.

[5] A. Bevilacqua and A. M. Niknejad, "An ultra-wideband CMOS LNA for 3.1 to 10.6 GHz wireless receivers," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '04), vol. 1, pp. 382-533, San Francisco, Calif, USA, February 2004.

[6] P. Heydari, D. Lin, A. Shameli, and A. Yazdi, "Design of CMOS distributed circuits for multiband UWB wireless receivers [LNA and mixer]," in Proceedings of IEEE Radio Frequency Integrated Circuits (RFIC '05), pp. 695-698, Long Beach, Colo, USA, June 2005.

[7] P. Heydari and D. Lin, "A performance optimized CMOS distributed LNA for UWB receivers," in Proceedings of the IEEE Custom Integrated Circuits Conference (CICC '05), pp. 330-333, San Jose, Calif, USA, September 2005.

[8] M. Tsai, K.-Y. Lin, and H. Wang, "A 5.4 mW LNA using a 0.35[micro]m SiGe BiCMOS technology for 3.1-10.6GHz UWB wireless receivers," in Proceedings of the IEEE Custom Integrated Circuits Conference (CICC '05), pp. 337-340, San Jose, Calif, USA, September 2005.

[9] B. Razavi, T. Aytur, F.-R. Yang, et al., "A 0.13 [micro]m CMOS UWB transceiver," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '05), vol. 1, pp. 216-594, San Francisco, Calif, USA, February 2005.

[10] S. Iida, K. Tanaka, H. Suzuki, et al., "A 3.1 to 5 GHz CMOS DSSS UWB transceiver for WPANs," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '05), vol. 1, pp. 214-594, San Francisco, Calif, USA, February 2005.

[11] J. Bergervoet, K. Harish, G. van der Weide, et al., "An interference robust receive chain for UWB radio in SiGe BiCMOS," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '05), vol. 1, pp. 200-593, San Francisco, Calif, USA, February 2005.

[12] G. Chevallier and E. F. Stikvoort, Transformer Circuit, Double-Balanced Mixer, US patent 5825231.

[13] A. Ismail and A. Abidi, "A 3.1 to 8.2 GHz direct conversion receiver for MB-OFDM UWB communications," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '05), vol. 1, pp. 208-593, San Francisco, Calif, USA, February 2005.

[14] S. Aggarwal, D. M. W. Leenaerts, R. van de Beek, et al., "A low power implementation for the transmit path of a UWB transceiver," in Proceedings of the IEEE Custom Integrated Circuits Conference (CICC '05), pp. 148-151, San Jose, Calif, USA, September 2005.

[15] C. Grewing, K. Winterberg, S. van Waasen, et al., "Fully integrated distributed power amplifier in CMOS technology, optimized for UWB transmitters," in Proceedings of IEEE Radio Frequency Integrated Circuits Symposium (RFIC '04), pp. 87-90, FortWorth, Tex, USA, June 2004.

[16] D. M. W. Leenaerts, R. van de Beek, G. van der Weide, et al., "A SiGe BiCMOS 1ns fast hopping frequency synthesizer for UWB radio," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '05), vol. 1, pp. 202-593, San Francisco, Calif, USA, February 2005.

[17] D. M.W. Leenaerts, J. van der Tang, and C. S. Vaucher, Circuit Design for RF Transceivers, Kluwer Academic, Dordrecht, The Netherlands, 2001.

[18] C.-C. Lin and C.-K. Wang, "A regenerative semi-dynamic frequency divider for mode-1 MB-OFDM UWB hopping carrier generation," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '05), vol. 1, pp. 206-207, San Francisco, Calif, USA, February 2005.

[19] R. van de Beek, D. M. W. Leenaerts, and G. van der Weid, "A fast-hopping single-PLL 3-band UWB synthesizer in 0.25 [micro]m SiGe BiCMOS," in Proceedings of the 31st European Solid-State Circuits Conference (ESSCIRC '05), pp. 173-176, Grenoble, France, September 2005.

[20] J. Lee and D. Chiu, "A 7-band 3-8 GHz frequency synthesizer with 1 ns band-switching time in 0.18 [micro]m CMOS technology," in Proceedings of IEEE International Solid-State Circuits Conference (ISSCC '05), vol. 1, pp. 204-593, San Francisco, Calif, USA, February 2005.

[21] R. Roovers, D. M. W. Leenaerts, J. Bergervoet, et al., "An interference-robust receiver for ultra-wideband radio in SiGe BiCMOS technology," IEEE Journal of Solid-State Circuits, vol. 40, no. 12, pp. 2563-2572, 2005.

[22] P. Heydari, "Design considerations for low-power ultra wideband receivers," in Proceedings of IEEE 6th International Symposium on Quality of Electronic Design (ISQED '05), pp. 668-673, San Jose, Calif, USA, March 2005.

[23] P. Heydari, "A study of low-power ultra wideband radio transceiver architectures," in Proceedings of Wireless Communications and Networking Conference (WCNC '05), vol. 2, pp. 758-763, New Orleans, La, USA, March 2005.

Chapter 2:

[1] P.H. Smith. *Electronics,* 12 (1937): 29-31.

[2] Ramo, Simon, and John R. Whinnery, et al. *Fields and Waves in Communication Electronics.* 3rd ed. Wiley, 1994

[3] Doh, H, and Y. Jeong, et al. "Design of CMOS UWB Low Noise Amplifier with Cascode Feedback", *IEEE MWSCAS*, 2 (2004): 641-644.

[4] Oppermann, Ian, and M. Hamalainen, et al. *UWB Theory and Applications*. Wiley Press, 2004.

[5] Choi, Seok, and Jong K. Park, et al. "A New Ultra-wideband Antenna for UWB Applications." *Microwave and Optical Technology Letters*, 40.5 (March 2004): 399-401.

Chapter 3:

[1] W. S. Percival, "Thermionic Valve Circuits," British Patent Specification no. 460,562, filed 24 July 1936, granted January 1937.

[2] E. L. Ginzton, W. R. Hewlett, J. H. Jasberg, and J. D. Noe, "Distributed Amplification," Proc. IRE, pp. 956-69, August 1948.

[3] E. W. Strid and K. R. Gleason, "A DC-12 GHz Monolithic GaAsFET Distributed Amplifier," IEEE Trans. Microwave Theory and Techniques, vol. MTT-30, no. 7, pp. 969-975, July 1982.

[4] Y. Ayasli, R. L. Mozzi, J. L. Vorhaus, L. D. Reynolds, and R. A. Pucel, "A Monolithic GaAs 1-13-GHz Traveling-Wave Amplifier," IEEE Trans. Microwave Theory and Techniques, vol. MTT-30, no. 7, pp. 976-981, July 1982.

[5] K. B. Niclas, W. T. Wilser, T. R. Kritzer, and R. R. Pereira, "On Theory and Performance of Solid-State Microwave Distributed Amplifiers," IEEE Trans. Microwave Theory and Techniques, vol. MTT-31, no. 6, pp. 447-456, June 1983.

[6] R. Majidi-Ahy, C. K. Nishimoto, M. Riaziat, M. Glenn, S. Silverman, S.-L. Weng, Y.-C. Pao, G. A. Zdasiuk, S. G. Bandy, and Z. C. H. Tan, "5-100 GHz InP

Coplanar Waveguide MMIC Distributed Amplifier," IEEE Trans. Microwave Theory and Techniques, vol. MTT-38, no. 12, December 1990.

[7] S. Kimura, Y. Imai, Y. Umeda, and T. Enoki, "Loss-compensated Distributed Baseband Amplifier for Optical Transmission Systems," IEEE Trans. Microwave Theory and Techniques, vol. MTT-44, no. 10, pp. 1688-1693, October 1996.

[8] D. Linten, S. Thijs, W. Jeamsaksiri, J. Ramos, A. Mercha, M. I. Natarajan, P. Wambacq, A. J. Scholten, and S. Decoutere, "An Integrated 5 GHz Low-Noise Amplifier with 5.5 kV HBM ESD protection in 90 nm RF CMOS," 2005 Symp. on VLSI Circuits Digest of Technical Papers, pp. 86-89, July 16-18 2005.

[9] Robert Boylestad and Louis Nashelsky (1996). *Electronic Devices and Circuit Theory, 7th Edition*. Prentice Hall College Division. ISBN 978-0133757347. http://books.google.com/books?id=TRBTAAAAMAAJ&q=Electronic+Devices+and+Circuit+Theory&dq=Electronic+Devices+and+Circuit+Theory&pgis=1.

[10] It is a curiosity to note that this table is a "Zwicky box"; in particular, it encompasses all possibilities. See Fritz Zwicky.

[11] John Everett (1992). *Vsats: Very Small Aperture Terminals*. IET. ISBN 0863412009. http://books.google.com/books?id=MWuXmf4V4NwC&pg=PA84&dq=amplifier+impedance+transmission-line++dB&as_brr=3&ei=nsyFR5GqJ6eGtgPX4rXQBQ&sig=Lwap3QvlOSI2ClwZlO4MM4BcD_A#PPA81,M1.

[12] Lee, Thomas (2004). *The Design of CMOS Radio-Frequency Integrated Circuits*. New York, NY: Cambridge University Press. p. 8. ISBN 9780521835398.

[13] RCA Receiving Tube Manual, RC-14 (1940) p 12

[14] ARRL Handbook, 1968; page 65

[15] Paul Horowitz, Winfield Hill (1989). *The Art of Electronics*. Cambridge University Press. ISBN 0521370957. http://books.google.com/books?id=bkOMDgwFA28C&pg=PA411&dq=crossover-distortion+class-b&lr=&as_brr=0&ei=v7nlR4CCDZaytgOT7IzZBA&sig=Dj6XntJelOmRA80c0IMp9gcG01s.

[16] Ask the Doctors: Tube vs. Solid-State Harmonics

[17] Volume cranked up in amp debate

[18] A.P. Malvino, *Electronic Principles* (2nd Ed.1979. ISBN 0-07-039867-4) p.299.

[19] N. O. Sokal and A. D. Sokal, "Class E — A New Class of High-Efficiency Tuned Single-Ended Switching Power Amplifiers", *IEEE Journal of Solid-State Circuits*, vol. SC-10, pp. 168–176, June 1975. HVK

[20] US patent 2210028, William H. Doherty, "Amplifier", issued 1940-08-06, assigned to Bell Telephone Laboratories

[21] US patent 3314034, Joseph B. Sainton, "High Efficiency Amplifier and Push-Pull Modulator", issued 1967-04-11, assigned to Continental Electronics Manufacturing Company

[22] Class Z Direct Digital Feedback Amplifiers, Zetex Semiconductors, 2006.

Chapter 4:

[1] Kobayashi, K, and A. Oki, et al. "A Low-Noise Baseband 5GHz Direct-Coupled HBT Amplifier with Common-base Active Input Match," *IEEE Microwave and Guided Wave Letters*, 4.11, (Nov. 1994): 373-375.

[2] Huang, Q, and F. Piazza, et al. "The Impact of Scaling Down to Deep Submicron on CMOS RF Circuits," *IEEE Journal of Solid State Circuits*, 33.7, (July 1998): 1023-1036.

[3] Lee, S. "Design and analysis ultra-wideband impulse radio receiver," *Ph.D. Thesis*, Electrical Engineering Dept, University of South California, California, USA, 2002.

[4] Lee, F, and A. Chandrakasan, et al. "A BiCMOS Ultra-Wideband 3.1–10.6-GHz Front-End," *IEEE Journal of Solid-State Circuits*, 41.8, (Aug. 2006): 1784-1791.

[5] Gary, P. *Analysis and Design of Analog Integrated Circuits*. Wiley & Sons, 2000.

[6] Gonzalez, G. *Microwave Transistor Amplifiers Analysis and Design*. Prentice Hall, 1996.

[7] Lee, T. *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 1998.

[8] Razavi, B. *RF Microelectronics*. Prentice Hall, 1997.

[9] Kim, C. "An Ultra-wideband CMOS Low Noise Amplifier for 3-5GHz UWB System." *IEEE Journal of Solid-State Circuits*, 40.2 (February 2005): 544-547.

[10] Gharpurey, R. "A Broadband Low-noise Front-end Amplifier for Ultra Wideband in 0.13um CMOS." *IEEE Journal of Solid-State Circuits*, 40.9 (September 2005): 1983-1986.

[11] Barras, D. "A Low Supply Voltage SiGe LNA for Ultra-wideband Frontends." *IEEE Microwave and Wireless Components Letters*, 14.10 (October 2004): 468-471.

[12] Ismail, A. and A. Abidi, et al. "A 3 to 10 GHz LNA Using a Wideband LC-ladder Matching Network." *Proc. IEEE International Solid State Circuits Conference*, 1 (February 2004): 384-385.

[13] [Xie, Haolu, and Xin Wang, et al. “A Fully-Integrated Low-Power 3.1-10.6GHz UWB LNA in 0.18µm CMOS.” *IEEE Radio & Wireless Symposium*, (January 2007).

Chapter 5:

[1] A. Z. H. Wang, H. Feng, R. Zhan, H. Xie, G. Chen, Q. Wu, X. Guan, Z. Wang, and C. Zhang, "A Review on RF ESD Protection Design," *IEEE Trans. Electron Devices*, 52:1304-1310, 2005.

[2] A.Wang, "On-Chip ESD Protection for Integrated Circuits: An IC Design Perspective," *Norwell, MA: Kluwer*, 2002.

[3] C. Richier, P. Salome, G. Mabboux, I. Zaza, A. Juge, and P. Mortini, "Investigation on different ESD protection strategies devoted to 3.3VRF applications (2 GHz) in a 0.18 _m CMOS process," In *Proc. Electrical Overstress Electrostatic Discharge Symp.*, pages 251-259, 2000.

[4] B. Kleveland, T. Maloney, I. Morgan, L. Madden, T. Lee, and S. Wong, "Distributed ESD protection for high-speed integrated circuits," *IEEE Electron Device Lett.*, 21:390-392, 2000.

[5] J. Barth, K. Verhaege, L. Henry, and J. Richner, "TLP calibration, correlation, standards and new techniques," *IEEE Trans. Electron. Packag. Manuf.*, 24:99-108, 2001.

[6] M. Xu, D. S. D. Shaeffer, T. Lee, and A.Wooley., "Measuring and modeling the effects of substrate noise on the LNA for CMOS GPS receiver," *IEEE J. Solid-State Circuits*, 36:473-485, 2001.

[7] H. Feng, G. Chen, R. Zhan, Q. Wu, X. Guan, H. Xie, and A. W. R. Gafiteanu, "A mixed-mode ESD protection circuit simulation-design methodology," *IEEE J. Solid-State Circuits*, 38:995-1006, 2003.

[8] P. Yue, S.Wong, D. Su, and W. McFarland, "System for providing electrostatic discharge protection for high-speed integrated circuits," *U.S. Patent 6 597 227*, 2003.

[9] S. Galal and B. Razavi, "Broadband ESD protection circuits in CMOS technology," *IEEE J. Solid-State Circuits*, 38:2334-2340, 2003.

[10] A. Wang and C. Tsay, "On a dual-direction on-Chip electrostatic discharge protection structure," *IEEE Trans. Electron Devices*, 48:978-984, 2001.

[11] R. V. P. d. Vreede and P. Woerlee, "Diode network used as ESD protection in RF applications," In *Proc. Electrical Overstress Electrostatic Discharge Symp*, 2001.

[12] A. Wang, H. Feng, R. Zhan, G. Chen, Q. Wu, and . "ESD protection design for RF integrated circuits: New challenges," *Proc. IEEE Custom Integrated Circuits Conf.*, (2002):411-418.

Chapter 7

[1] P. Leroux, M. Steyaert, "High-Performance 5.2 GHz LNA with on-Chip Inductor to Provide ESD Protection", *Electronics Letters*, *Vol. 37, No. 7*, 2001, pp. 467-469.

[2] D. Linten, X. Sun, S. Thijs, M.I. Natarajan, A. Mercha, G. Carchon, P. Wambacq, T. Nakaie and S. Decoutere, "Low-Power Low-Noise Highly ESD Robust LNA and VCO Design Using Above-IC Inductors", *Proc. IEEE Custom Integrated Circuits Conference*, 2005, pp. 497–500.

[3] V. Chandrasekhar, C.M. Hung, Y.C. Ho and K. Mayaram, "A Packaged 2.4 GHz LNA in a 0.15µm CMOS Process with 2kV HBM ESD Protection", *Proc. IEEE European Solid-State Circuits Conference*, 2002, pp. 347–350.

[4] D. Linten, S. Thijs, M. I. Natarajan, P. Wambacq, W. Jeamsaksiri, J. Ramos, A. Mercha, S. Jenei, S. Donnay, S. Decoutere, "A 5GHz Fully Integrated ESD-Protected Low Noise Amplifier in 90nm RFCMOS", *IEEE Journal of Solid-State Circuits, Vol. 40, No.7,*2005, pp. 1434-1442.

[5] M. Liu, J. Craninckx, N. M. Iyer, M. Kuijk, A. R. F. Barel, "A 6.5-kV ESD-Protected 3-5GHz Ultra-Wideband BiCMOS Low-Noise Amplifier Using Interstage Gain Roll-Off Compensation"**,** *IEEE Transactions on Microwave Theory and Techniques, Vol. 54, No. 4,* 2006, pp. 1698-1706.

[6] A. Wang, H. Feng, R. Zhan, G. Chen and Q. Wu, "ESD Protection Design for RF Integrated Circuits: New Challenges", *Proc. IEEE Custom Integrated Circuits Con.,* 2002, pp. 411-418.

[7] H. Feng, H. Xie, A. Wang, Y. Cheng and S. Lloyd, "A Full-Monolithic LNA in 0.18um SiGe: Performance Variation Due to ESD Protection", *Proc. IEEE International Conference on Solid-State and Integrated Circuits Technology*, 2004, pp. 1226 – 1229.

[8] A. Wang, H. Feng, R. Zhan, H. Xie, G. Chen, Q. Wu, X. Guan, Z. Wang, C. Zhang, "A Review on RF ESD Protection Design", *IEEE Trans. Electron Devices*, *Vol. 52, No. 7,* 2005, pp. 1304 – 1311.

[9] J. Li, S. Joshi, R. Barnes, E. Rosenbaum, "Compact Modeling of on-Chip ESD Protection Devices Using Verilog-A", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, *Vol. 25, No. 6*, 2006, pp. 1047 – 1063.

[10] X. Gao, J. Liou, J. Bernier, G. Croft, A. Ortiz-Conde, "Implementation of a Comprehensive and Robust MOSFET Model in Cadence SPICE for ESD Applications", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, *Vol 21, No. 12*, 2002, pp.1497 – 1502.

[11] R. Zhan, H. Feng, Q. Wu and A. Wang, "ESDExtractor: A New Technology-Independent CAD Tool for Arbitrary ESD Protection Device Extraction", *IEEE Trans on CAD of Integrated Circuits and Systems, Vol. 22, No. 10*, 2003, pp1362-1370.

[12] R. Zhan, H. Feng, Q. Wu, H. Xie, X. Guan, G. Chen and A. Wang, "ESDInspector: A New Layout-level ESD Protection Circuitry Design Verification Tool Using A Smart-Parametric Checking Mechanism", *IEEE Trans on CAD of Integrated Circuits and Systems, Vol. 23, No. 10,* 2004, pp. 1421-1428.

[13] H. Xie, R. Zhan, A. Wang and R. Gafiteanu, "Real 3D Electro-Thermal Simulation and Analysis for ESD Protection Structures", *Proc. of the 5th IEEE International Caracas Conference on Device, Circuits and Systems*, 2004, pp. 61-64.

[14] H. Feng, G. Chen, R. Zhan, Q. Wu, X. Guan, H. Xie, A. Wang and R. Gafiteanu, "A Mixed-Mode ESD Protection Circuit Simulation-Design Methodology", *IEEE Journal of Solid-State Circuits, Vol. 38, No. 6*, 2003, pp. 995-1006.

[15] G. Chen, H. Feng, H. Xie, R. Zhan, Q. Wu, X. Guan, A. Wang, K. Takasuka, S. Tamura, Z. Wang and C. Zhang, "RF Characterization of ESD Protection Structures", *Proc. IEEE Radio-Frequency Integrated Circuits Symp.*, 2004, pp. 379-382.

[16] G. Chen, "Design and Characterization of ESD Protection for RF ICs", *MS Thesis*, Illinois Institute of Technology, 2003.

[17] L. Lin, "Design and Analysis of ESD Protection Circuit in CMOS", *MS Thesis*, Illinois Institute of Technology, 2007.

Chapter 8

[1] A. Wang, *On-Chip ESD Protection for ICs,* Kluwer, 2002.

[2] P. Sivasubramani, *et al, IEDM,* pp.543-546, 2007.

[3] S. Liu, *et al*, *Solid thin Films, Elsevier, v280,* pp.271-277, 1996.

[4] M. N. Kozicki, *et al*, *IEEE Trans. Nanotech.,* pp. 331, May 2005.