

mm-Wave Phase Shifters and Switches

Ehsan Adabi Firouzjaei

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2010-163

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-163.html>

December 16, 2010



Copyright © 2010, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Acknowledgement

There are many people who I am grateful for knowing and working with them during my stay at Berkeley. First of all, my advisor professor Ali Niknejad who I am thankful for all the time and energy that he dedicated. I am also thankful of my qualification exam and thesis committee members professors Robert Meyer, Jan Rabaey and Paul Wright. I thank professors Khoramabadi, Yablonovitch, Nikolic, Carmena, Bahai , Kuroda and for courses they offered and I had the opportunity to take. I also would like to thank my team mates at the RF research group and all researchers, staffs and directors at BWRC (Berkeley Wireless Research Center) who I had the opportunity to work with, exchange ideas and enjoy the productive and memorable environment of BWRC.

mm-Wave Phase Shifters and Switches

by

Ehsan Adabi Firouzjaei

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

in

Engineering—Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ali M. Niknejad, Chair
Professor Jan M. Rabaey
Professor Paul K. Wright

Fall 2010

mm-Wave Phase Shifters and Switches

Copyright Fall 2010
by
Ehsan Adabi Firouzjaei

Abstract

mm-Wave Phase Shifters and Switches

by

Ehsan Adabi Firouzjaei

Doctor of Philosophy in Engineering—Electrical Engineering and Computer Science

University of California, Berkeley

Professor Ali M. Niknejad, Chair

The ever increasing speed of transistors in mainstream silicon-based technologies made the mm-wave domain open to consumer electronic applications. Solutions that previously had to be implemented in advanced compound (III-V) technologies and were limited to high-end systems due to cost purposes, are now entering the market of low-cost consumer electronic products. Emerging mm-wave market contains various applications from extremely high data rate transceivers in “Personal Digital Assistant (PDA)” devices to automotive radar modules and to point to point links for replacing the fiber connectivity in sparse areas. Chapter one highlights the specific requirements of each application that makes it more compatible with a certain type of technology.

To have a complete mm-wave system suitable for low cost applications, a single chip or a single package solution is preferred. To achieve this goal integrated low loss transmit / receive switching structures that are highly linear should be employed. A miniature transformer-based shunt T/R switch is introduced and implemented in a standard 90nm CMOS technology. Design equations and trade-offs for such a structure are described in this thesis.

Due to a much higher free space path loss of a 60GHz signal compared to its low frequency counterparts (30dB higher loss than WiFi), and lower performances extractable from devices at these high frequencies, phased antenna array structures should be exploited to add passive antenna gains to the transceiver and help meet the link budget requirement. Fundamentals of phased antenna array structures are described and two different implementations, one through true time delay elements and the other one employing phase shifters are presented.

For wideband applications and for very large arrays intended to have a wide field of view, true time delay elements should be employed to steer the array pointing beam. This work investigates true time delay elements, and an “inductance tuning technique” is introduced which enhances the delay tunability of a synthesized transmission line while keeping its characteristic impedance constant. In most mm-wave applications, delay cells in antenna array structures can be approximated and replaced with phase shifters. Hence different types of phase shifters are studied and an active I-Q interpolating phase shifter in the RF-path is designed and implemented at 60GHz.

Contents

List of Figures	iii
------------------------	------------

List of Tables	vii
-----------------------	------------

1 mm-Wave opportunities and choice of technologies	1
1.1 mm-Wave opportunities	2
1.1.1 60GHz connectivity	2
1.1.2 Automotive radar	4
1.1.3 Imaging	5
1.1.4 Point to point links, Gigabit ethernet	6
1.2 Standard digital CMOS vs. SiGe technology	8
1.2.1 Cost of the process	8
1.2.2 Performance	9
1.2.3 Quantitative comparison	13
1.2.4 Choice of technology	14
2 Transmit/Receive Switching	16
2.1 Antenna array reuse via T/R switching	17
2.1.1 On-chip or On-package antennas	17
2.1.2 Size and loss trade-off in a transmit/receive array	18
2.2 T/R switching configurations	20
2.2.1 Series vs Shunt switching	20
2.2.2 The transformer based shunt switching structure	22
2.3 Design equations of a transformer based T/R switch	24
2.3.1 Equivalent shunt loading	24
2.3.2 Center frequency and matching	26
2.3.3 Insertion Loss	27
2.3.4 Leakage	27
2.3.5 Isolation	28
2.4 Transformer based switch design example in 90nm CMOS	29
2.4.1 MOS transistors performance in switch mode	29

2.4.2	Prototype design	32
2.5	Prototype Measurement Results	34
3	Phased Array Structures	37
3.1	Fundamentals of phased arrays	37
3.2	Advantages of phased arrays	38
3.3	Different phase shifting schemes	42
3.3.1	Digital phased arrays	42
3.3.2	LO domain phase shifting scheme	42
3.3.3	RF domain phased arrays	43
3.4	Automotive radar link budget and the required size of the array	44
3.4.1	Radar Equation	45
3.4.2	Array size	48
3.5	Phase shifters or delay elements	48
4	True Time Delay Elements	52
4.1	Tunable delay structures	52
4.1.1	Slow wave transmission lines	53
4.1.2	Synthesized transmission lines	55
4.2	Inductance tuning technique	57
4.3	Passive mode implementation of delay elements employing inductance multiplication technique	60
4.4	Active mode implementation of delay elements employing inductance multiplication technique (variable delay amplifiers)	63
4.5	Conclusion	68
5	Phase Shifting Structures	69
5.1	Passive phase shifters	69
5.1.1	Varactor-based passive phase shifting structures	69
5.1.2	Butler mixer and high/low pass filters	74
5.2	Active phase shifter : I-Q interpolation	75
5.2.1	I-Q Generation	76
5.2.2	Variable Gain Amplifiers	81
5.2.3	Signal dividers / combiners	83
5.3	mm-Wave implementation of the active phase shifter	86
5.3.1	Transformer matching networks	86
5.3.2	Measurement results	88
6	Conclusion	95
	Bibliography	98

List of Figures

1.1	2007 ITRS roadmap failed to predict current mm-wave design trend [1]	2
1.2	60GHz wireless connectivity for video streaming and WPAN applications	3
1.3	Long range(left) and short range(right) automotive radar solutions	4
1.4	mm-wave imaging for medical and security applications.	5
1.5	Point to point links are preferred to operate at frequency bands where mm-wave signals free space absorption loss is minimum[2]	7
1.6	At each process node SiGe has more number of mask layers, however in advanced CMOS technologies the difference is diminishing[3].	9
1.7	Cost per die size of each CMOS process node equals to that of a SiGe technology which is two generations older [4].	10
1.8	Digital intensive solutions get shrunk considerably as they are migrated to newer process nodes and the die cost will go down.	11
1.9	Comparing f_t and NF_{min} of 65nm CMOS and 130nm SiGe technologies	12
2.1	A fully integrated transceiver with an on-chip antenna requires an integrated T/R switch to save die area.	17
2.2	In case of not including T/R switches, two separate RX and TX arrays are required (B), which correspondingly lengthens interconnection routings and increases their associated loss	18
2.3	Graph of the maximum achievable communication range in two configurations of shared or separate arrays (for both small and large arrays). For larger arrays the routing loss is more detrimental than the T/R switch loss and higher quality packaging should be employed.	19
2.4	Measured S_{21} of an NMOS transistor acting as a series switch in a 50Ω environment.	20
2.5	Combinations of series and shunt switches in a T (A) or π (B) or L -shape configuration (C)[33] decrease the amount of off-state leakage at the price of degrading the on-state insertion loss	21
2.6	Traditional shunt switches occupy a large footprint as a result of employing quarter wavelength transmission lines	22
2.7	Schematic diagram of a transformer-based shunt switch and its equivalent circuit model	23

2.8	Simplified model of the SPDT network including loading effects	24
2.9	Calculating the shunt equivalent loading network	25
2.10	Parallel tank equivalent network of the switch for calculating the center frequency	26
2.11	Equivalent circuit for calculating the isolation between two thru ports	28
2.12	Source network is transferred to the load side (A) and then the structure is converted to a parallel tank configuration (B)	28
2.13	Higher biasing impedance at the gate not only improves the insertion loss but also results in less distortion to the output signal and a higher linearity number for the T/R switch	30
2.14	Layout example of a MOS transistor to be used as a switch	31
2.15	Die microphotograph of the miniaturized shunt switch employing a transformer	32
2.16	Measured insertion loss, leakage(input to the off-thru port) and isolation (between the on-thru and off-thru ports)	33
2.17	Measured gain, input and output return loss curves	34
2.18	Large signal measurement ($P_{-1dB} = 14dBm$)	35
3.1	The incoming wavefront reaches each antenna element at a different time and corresponding delay element are needed in each path to compensate for that.	38
3.2	The passive gain achieved through array directivity increases the EIRP and relaxes PA design requirements.	39
3.3	A simple triangular gain windowing significantly decreases side lobe levels at the price of less main lobe gain and lower main lobe resolution (wider half-power beam-width).	40
3.4	In a phased array receiver, signals add up in voltage domain whereas uncorrelated noises add up in power domain. This increases the overall sensitivity of the receiver array.	41
3.5	Accomplishing the phase shifting and signal combining in the digital domain has the advantage of ultimate flexibility and programability. It results in the highest component count and area consumption as well as higher required dynamic range for the entire chain	43
3.6	A LO phase shifting scheme has less component count with respect to digital phase shifting. Phase shifter non-idealities are not directly in the RF path. A fully symmetric LO distribution network is necessary	44
3.7	The RF phase shifting scheme has the least component count and area/power consumption. Phase shifter noise and non-linearity are directly in the signal path.	45
3.8	Bulding block diagram of an FMCW radar	46
3.9	In a narrowband system, true time delay elements (linear phase response vs. frequency) can be approximated with phase shifters (constant phase response vs. frequency).	49
3.10	The nonlinear relation of equation 3.16 leads to erroneous estimation of direction of arrival for end-fire angles (frequency deviation is swept up to 20% in 5% steps)	50

3.11	As the array gets larger, the HPBW shrinks and the phase shift approximation of delay elements fails for incident angles closer to the end-fire angle even for narrowband signals	51
4.1	Electric field pattern in a normal (a) and a slow wave (b) transmission line.	53
4.2	Measurement comparison of a slow wave structure (blue) with a normal transmission (red) line in terms of the dielectric constant(ϵ), propagation constant(β), characteristic impedance(Z_0), and resonant quality factor($Q_{res} = \frac{\beta}{2\alpha}$).	54
4.3	A π -section of an artificial transmission line synthesized out of lumped component inductors and capacitors instead of infinitesimal distributed inductance and capacitance of a classic transmission line	55
4.4	(a) Switched transmission lines (large footprint). (b) Varactor loaded artificial transmission line (Z_0 variation). (c) Modifying the effective series reactance by adding a varactor in series with the inductor (narrowband network). (d) Broadband solution for tunable synthesized transmission lines.	56
4.5	Inductance tunability, net magnetic flux crossing a loop is altered via the flux generated by another loop.	57
4.6	Four different ways to realize inductance tuning: (a) Using a current amplifier, (b) varying the mutual inductance, or by rerouting the current in the secondary in a (c) single-ended or (d) differential manner.	59
4.7	Group delay, return loss (input/output) and phase response of a single delay cell and a cascade of two delay cells are demonstrated in (a) and (b) respectively.	61
4.8	Die Microphotograph	62
4.9	A variable delay amplifier implemented in CMOS technology	63
4.10	A variable delay amplifier in BiCMOS SiGe technology with $V_{DD} = 3.3V$ that can be used to stack more devices	65
4.11	A 4-way phased array receiver comprising of 4 low noise VDAs, 8 VDAs and a 4 to 1 combining network implemented in 130nm SiGe technology	66
4.12	(a) Gain response of each channel for different delay settings, (b) phase response of each channel for different delay settings (c) group delay $\tau = -\frac{d\phi}{d\omega}$	67
5.1	The reactance of a variable tank provides 180° <i>phaseshift</i> , but due to the variation of its normalized impedance ($\frac{Z}{R_p}$) it can not act as an all-pass filter on its own.	70
5.2	A quadrature hybrid structure in conjunction with two variable tanks form a reflective type phase shifting structure (RTPS).	71
5.3	Increasing the channel length of a MOS varactor increases $\frac{C_{max}}{C_{min}}$ and hence the frequency tunability $((\frac{C_{max}}{C_{min}})^{\frac{1}{4}} - (\frac{C_{min}}{C_{max}})^{\frac{1}{4}})$ at the price of lowering the varactor quality factor.	73
5.4	Two types of passive phase shifters that do not use varactors to achieve phase variation (Butler mixer and high-pass/low-pass networks).	74
5.5	Schematic diagram of an I-Q interpolating active phase shifter.	75

5.6	A voltage domain poly phase filter (left) and its gain degradation as loading capacitances being presented output nodes.	77
5.7	Adding the poly phase filter lowers the input impedance of a MOS amplifier which results in lower matching network Q and consequently lower gain through the chain	78
5.8	A current mode poly phase filter (left) and its gain degradation as a nonzero load impedance being presented at the output.	79
5.9	a) A transmission line coupler, b) A transformer can be exploited for the lumped version of a transmission line coupler, c) 1.25 turn loops comprising a transformer coupler resemble a quadrature hybrid.	80
5.10	Gain variation capability can be achieved through changing bias conditions of an NMOS transistor.	82
5.11	A phase inverting building block with input and output tuning inductances.	83
5.12	Schematic of core transistors for a current commuting type VGA (a) and a fixed gain cascode amplifier (b).	84
5.13	Traditional microwave method of signal dividing and combining.	85
5.14	Two lumped component version for dividing and combining signals in current (a) or voltage (b) mode	85
5.15	A transformer acts as a wideband matching network	86
5.16	(a)-Simulated Self inductances of a 2 : 1 transformer, (b)- quality factors of primary and secondary windings, (c)-transformer insertion loss, (d)- coupling between transformers	87
5.17	Schematic of the complete active I-Q interpolating phase shifter	89
5.18	Die microphotograph of the I/Q interpolating active phase shifter	90
5.19	Measurement data for gain, phase and return losses at input and output and their comparison with simulation prediction	91
5.20	Measured phase responses for different phase shift settings	92
5.21	$V_{g,tune}$ can be varied from V_{DD} down to $V_{DD}/2$ in order to adjust the gain	93
5.22	Large signal measurements for two extreme cases of I+Q and only-I settings and their comparison with simulation data	94

List of Tables

- 1.1 For comparable f_t and current dissipation, CMOS technology has less intrinsic parasitics which makes it more vulnerable to layout added parasitics and their variations. 13
- 2.1 Comparison to recently published mm-wave T/R switches 36

Acknowledgments

There are many people who I am grateful for knowing and working with them during my stay at Berkeley. First of all, my advisor professor Ali Niknejad who I am thankful for all the time and energy that he dedicated. For the directions that he showed and schools of thoughts that he made me familiar with. For his patience to go through my lengthy reports and give me his feedbacks and for classes he taught (EE142-EE217-EE242) that covered cutting edge "Radio Frequency Integrated Circuit" topics.

I am also thankful of my thesis committee members professor Jan Rabaey and professor Paul Wright. I had the pleasure to have them on my qualification exam committee as well and not only I had the opportunity to have their brilliant comments during the qualification exam and the thesis review, but also I benefited from their visionary discussions and presentations at the "Berkeley Wireless Research Center(BWRC)" and its retreats. I am also thankful of professor Robert Meyer who kindly accepted to be in my qualification committee and let me learn from his comments and ideas. I thank professors Yablonovitch, Nikolic, Carmenta, Bahai, Kuroda and Dr. Shana for courses they offered and I had the opportunity to take.

I learned a lot in EE247 from professor Haideh Khoramabadi, but her kindness and caring character in the years afterwards made me think of her like my mom and I am greatly thankful of her. I bugged Mary Burns, Ruth Gredje and Patrick Hernan a lot during my PhD program and I thank them all for their help and patience.

My group-mates Babak Heydari and Mounir Bohsali who I was co-working with for three years made me long lasting good memories of doing research at BWRC. Discussions that I had with Amin Arbabian and Bagher Afshar whether they were about research topics or class projects made me realize that there are always different angles to look at something. I also thank the mm-wave group members Alex Pai, Omar Bakr, Ashkan Borna, Steven Callendar, Jiashu Chen, Wei-hung Chen, Debo Chowdhury, Zhiming Deng, Mohan Dunga, Shinwon Kang, Cristian Marcu, Nuntachai Poobuapheun, Maryam Tabesh and my other colleagues at BWRC along with BWRC staffs and directors for making a productive environment and an enjoyable research center.

I would like to thank my co-workers Alfred G. Besoli and Michael Boers at Broadcom Corp. where I did an internship working on phase shifting structures. Attending UC Berkeley gave me the opportunity to live at the lively Berkeley community, benefit from its intellectual people, the "International House" and much more. I will be always proud of being part of the Berkeley community.

Chapter 1

mm-Wave opportunities and choice of technologies

The RF revolution has opened up a huge market for personalized devices that transmit and receive wireless data. The main characteristic of a product targeted for a consumer electronic market is its aim to be as cost effective as possible with respect to other competitive solutions. Therefore readily available and less expensive technologies that are advanced enough to meet the system requirements have a clear chance of victory for such applications. Nowadays highly scaled and advanced CMOS technologies have established themselves as primary solutions for many RF circuits and systems design at multi-GHz frequency band. mm-Wave solutions that deal with circuits and systems functioning at multi-ten gigahertz range of frequencies are gaining momentum to enter the consumer electronic market. Different technologies are suitable for different types of applications due to their technical requirements and the corresponding size of the market.

2007 ITRS roadmap (Fig. 1.1) [1] had suggested that CMOS technology would not be used for multi-ten gigahertz applications and even BiCMOS (SiGe) technologies wouldn't be deployed in applications beyond the frequency range of 60GHz. Silicon technologies have lower speeds compared to their more advanced compound (III-V) technologies and their lower breakdown voltages pose challenges on the transmit power of silicon-based solutions. However feasibility studies over the past few years ([44] [48] [53]) showed that advanced silicon technologies are capable of handling mm-wave applications. To create the communication link for different types of applications, different link budgets and overall system requirements are needed to be met which demands for a case by case study of respective applications. The next few years will be decisive to figure out the direction of silicon-based products in the consumer electronic market. In next section, various mm-wave opportunities and their criteria for being implemented efficiently will be covered.

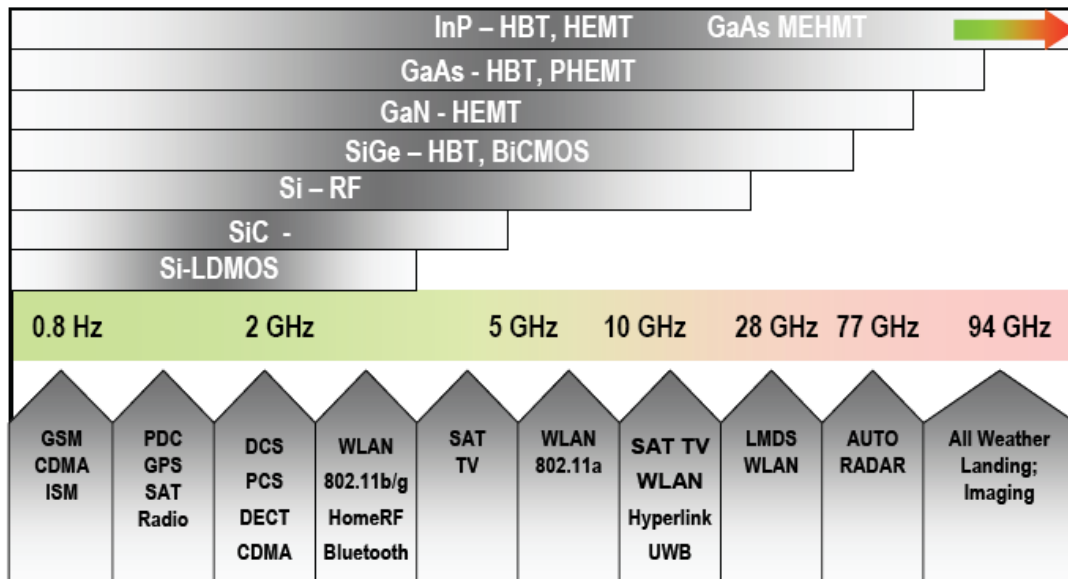


Figure 1.1: 2007 ITRS roadmap failed to predict current mm-wave design trend [1]

1.1 mm-Wave opportunities

1.1.1 60GHz connectivity

The universal availability of 7GHz of unlicensed bandwidth around 60GHz opens up opportunities for high data rate short range communications. Wirelessly streaming HD video that replaces the HDMI cable has a great potential as most portable media player devices will have HD capabilities in the future. 1600x1200 pixels at 24bits (true color) per pixel with a 120 Hz refresh rate lead to 10Gbit/s of uncompressed data-rate. Transmitting and receiving uncompressed video helps to remove the latency and cut the power dissipation associated with the video processing unit. Larger bandwidth available around 60GHz and much more relaxed limits on the transmit power levels are key advantages of 60GHz over WiFi and UWB solutions respectively.

In WPAN realm, as the capacity of storage devices is steadily increasing, more time is needed to synchronize the data between different devices. High speed wireless connectivity between personal computers, external hard drives, HD cameras and USB devices is highly desired in the future of WPAN applications. 60GHz can be used to connect PDAs, smart phones, portable media players and other personal devices while enhancing the overall efficiency (Joule/bit number) of communication.

60GHz can also be used for very high speed Gbps WLAN connectivity. 7GHz of bandwidth, opportunity to transmit high power levels and high path losses through walls and obstacles make 60GHz communication an option for high data rate and secure indoor communication where out-



Figure 1.2: 60GHz wireless connectivity for video streaming and WPAN applications

sider interferences are greatly attenuated. Small wavelengths at 60GHz make a complete phased array solution with relatively large number of antennas for such applications (8-64 antennas) realizable in either a SOC (system on a chip) or SOP (system on a package) configuration. With antenna arrays being embedded on chip or on the package, the phased array solution enhances the flexibility of the system in addressing multiple devices at the same time by exploiting the spatial multiplexing on top of time and frequency domain multiplexing.

Since most of the above mentioned applications are handheld and battery operated, the complete mm-wave system should be delivered with a reasonable power consumption, small footprint and a low cost of on the order of a dollar including testing and packaging costs. However due to the large volume of these products ($\sim 10B$ per year), the estimate market size would be around a 10B\$ revenue annually which is large and a great incentive for employing a routine and inexpensive technology such as standard digital CMOS technologies in the design.

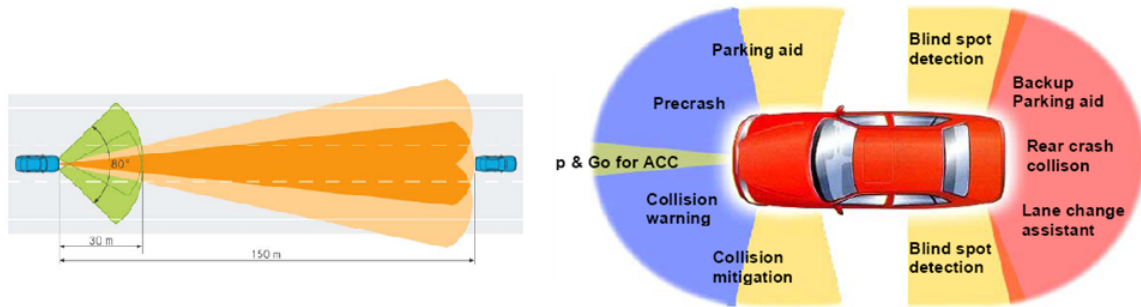


Figure 1.3: Long range(left) and short range(right) automotive radar solutions

1.1.2 Automotive radar

It is well known that motor vehicle accidents are one of the leading causes of death. 90 percent of fatal accidents involve driver errors. Therefore having some kind of system that recognizes situations where drivers are not responding accordingly and takes action that results in prevention or mitigation of the accident would be much beneficial. A key part of such a system is an automotive radar to detect objects surrounding a vehicle. Automotive radar solutions have been around for more than a decade now. Due to stringent requirements, all automotive radars were implemented exploiting advanced and expensive III-V MMIC modules. Using MMIC modules in discrete designs sets the price level of the complete solution to a high level which is out of reach for being installed on vehicles other than high-end cars.

Demanding requirements of detection range, range and angular resolutions, and target discrimination in heavily cluttered areas make the single chip solution of an automotive radar system in silicon technologies challenging. However, the ever increasing speed of transistors in emerging Si process nodes due to the scaling, and the luxury of being able to implement a microprocessor residing next to a radar chip are bridging the gap to make an automotive radar commercially viable for the purpose of being included in every single vehicle as a standard option for added safety and driving comfort. mm-Wave silicon technologies can directly revolutionize the automotive radar industry in terms of cost, size and power consumption. Having a single chip fully integrated car-radar module or a single package solution with all the antennas being included, decreases the size and hence the cost of the system considerably. Being implemented on an standard silicon technology makes it possible to have the digital processing part in the same chip that includes the RF part, which yields in a highly flexible single chip solution with the capability of calibrating and programming various parameters as well as incorporating built in self test (BIST) techniques. Upcoming novel solutions in the car industry such as smart electric cars open up a great opportunities to design short to high range radars that are compatible and can be easily embedded in electric car systems.

Right now the 24GHz band is allocated for pulse based UWB communication for short range

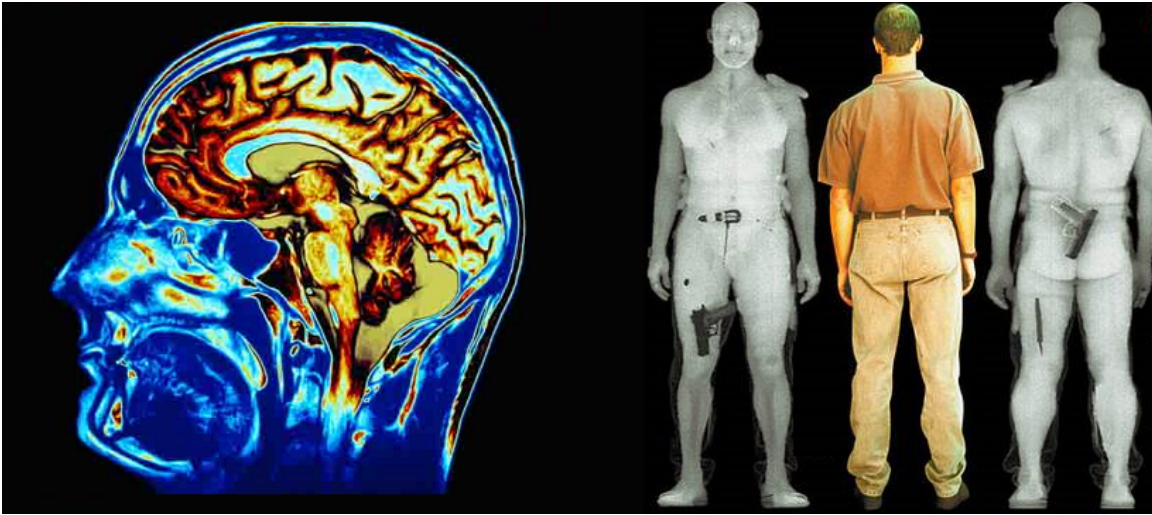


Figure 1.4: mm-wave imaging for medical and security applications.

radar(SRR) applications that help with parking, blind spot detection, lane detection, stop and go in traffic and collision avoidance/mitigation. By 2013, 24GHz band for UWB short range radars will be moved to 79GHz. 77GHz band is already dedicated to FMCW long range radar(LRR) solutions. 77GHz FMCW radar aids with the adaptive cruise control (ACC) that is primarily intended for highway driving. ACC allows target detection from a distance of few meters up to 150m. The data can be provided to the driver as an assistance or as an ultimate goal, with help of digital processing units it could be used for autonomous driving.

As depicted in figure. 1.3, 8 of car radar modules is needed per vehicle. Taking into account that 50 million cars are being made worldwide each year and the estimated unit price of an automotive radar module to be around \$10, it is clear that mm-wave automotive radar is a market with few billion dollars(~\$4B) of revenue each year. After 60GHz connectivity, automotive radar market seems to be the next domain that silicon technology will spread into and replace compound technologies.

1.1.3 Imaging

Basic principles of microwave imaging have been understood for decades. Extending these imaging techniques to higher carrier frequencies where larger bandwidths are also available enabling for higher image resolutions both in depth and lateral spacing. Two main areas that can benefit from this enhanced imaging technology are medical diagnosis and security (figure 1.4).

There are two types of imaging systems, passive and active. Passive mm-wave imaging exploits an array of very low noise receivers to reconstruct a high resolution image of an object. All objects at a temperature greater than absolute zero emit blackbody radiation. This energy emission occurs at a broad range of spectrum with emission peaks at infrared frequencies. The amount of radiation

at mm-wave frequencies is 10^8 times smaller than that emitted at the infrared range. However state of the art mm-wave receivers have at least 10^5 times better sensitivity with respect to infrared detectors and the temperature contrast can recover the remaining factor of 1000. This makes mm-wave passive imaging solutions comparable to infrared imaging systems. For optimal detection the imaging windows are chosen where the atmospheric loss is minimum, namely at 35GHz, 94GHz, 140GHz, 220GHz. Passive imaging can be used in various applications such as airport safety, weather radar, remote sensing for environmental and geological explorations and non-invasive and non-destructive in-situ testings.

In active imagers, an ultra high bandwidth pulsed based system reconstructs an image of the object based on the scattered components of the radiated short-duration pulses. Strength of each scattered component determines the intensity of the corresponding pixel in the image. Pulse based active mm-wave imaging can be used in medical diagnosis since reflection properties of various tissues and substances are different. It can be as well used in security systems for detection of drugs and weapons. Because of the fact that mm-wave signals can easily pass through clothing but can not do so through the body and will get reflected through metallic / liquid objects.

As silicon technologies are capable of integrating large arrays of transceivers, the mm-wave imaging technology could be a low cost competitor to existing technologies such as MRI, CAT scan and infrared imaging systems. However silicon based mm-Wave imaging is an active area of research and there are still parts with challenging requirements that pose a great challenge for the complete system to be implemented in a monolithic fashion. Since their target applications are mainly medical diagnosis and security, they are anticipated to have a unit price in the range of \$100-1000 and the annual demand would be 10-100K parts. So an annual revenue of about \$50M is estimated for the silicon based mm-wave imaging market.

1.1.4 Point to point links, Gigabit ethernet

The cost of realizing fiber optic links in remote areas that are not heavily populated can hamper the implementation of such links. In those circumstances having wireless point to point links between base stations (BTS) and network nodes is a key solution to decrease the implementation cost. Since point to point links are kilometer-range links, it is mostly desired that frequency bands where the free space path loss is minimum be allocated for such applications. Although the fiber optic infrastructure already exists in many urban areas for realizing LAN connectivity. However, hardware implementation of wired communication in the first/last mile to reach the end users is extremely costly. Replacing the fiber link with a high data rate mm-wave wireless link in the first/last mile is an option to reduce the overall cost and make it a more feasible solution.

As of now, for such applications frequency spectrum is allocated in the unlicensed 24.25-24.45GHz and 25.05-25.25 GHz bands. However, the 72.5-82.5GHz ISM band (E-band) is going to be used to extend LAN backbones. Communication at the E-band requires smaller antennas, resulting in a more directive antenna array being implemented at a given specified area devoted to radiation elements. E-band communication with the throughput in the range of multi-10Gbps is foreseen to be employed as a solution with carrier grade performance for achieving link dis-

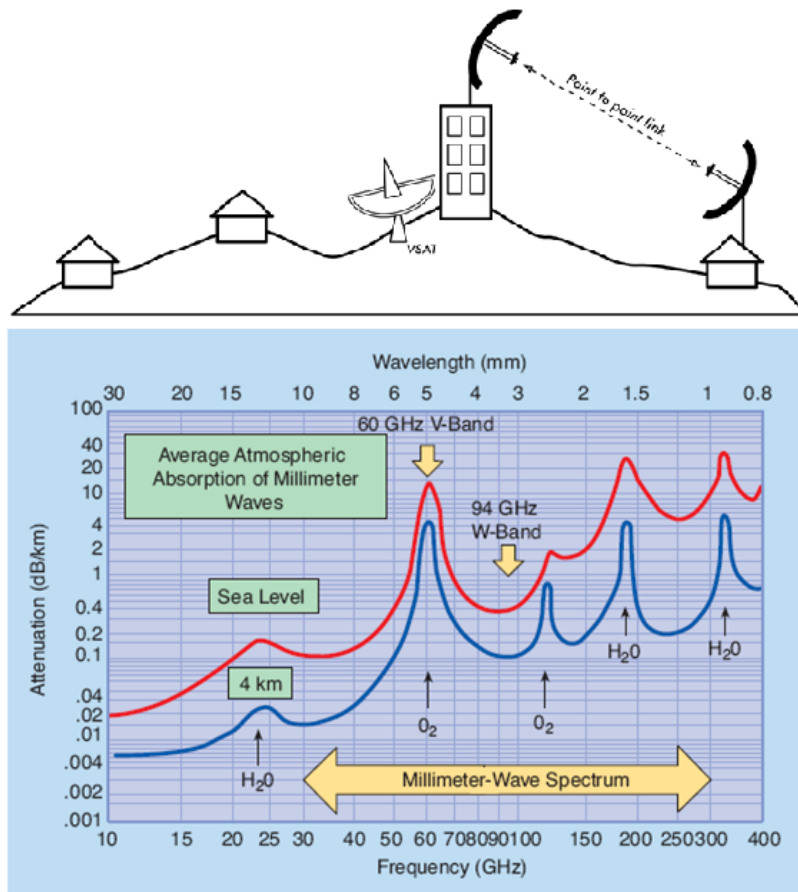


Figure 1.5: Point to point links are preferred to operate at frequency bands where mm-wave signals free space absorption loss is minimum[2]

tances over several kilometers. It will provide means for interconnection and backhaul of 3G and 4G (WiMAX and LTE) networks. Other applications of E-band communication will be gigabit ethernet access, fiber backup, path redundancy and network extension applications.

Due to the nature of these applications, each unit will be on the order of \$1000 and the total need is 10k parts per year. So the market(\$10M) is smaller than the market size for previously mentioned applications. Cost for these systems are less important with respect to issues like performance and reliability.

1.2 Standard digital CMOS vs. SiGe technology

Although III-V compound technologies have the best performance at the mm-wave regime and most MMIC based mm-wave solutions exploit these high performance technologies nowadays. Due to their high cost of implementation as a result of their demand for specialized substrates and low process yields, and because of their lack of flexibility which comes from the inability to integrate digital circuitry on the same die, they are not suitable for consumer electronic markets. To investigate the most suitable technologies for emerging high volume markets of mm-wave applications, we focus our study on two promising silicon technologies: standard digital CMOS technology and the BiCMOS (SiGe) technology. They are both capable of providing low cost solutions that exact requirements of different applications may favor one with respect to the other one. In next sections important criteria for a mm-wave system design will be investigated and advantages and disadvantages of CMOS and SiGe technologies with respect to each other will be pointed out. As it is the case in other situations, price and performance are the two main parameters that rule the choice of technology to be adopted for mm-wave design.

1.2.1 Cost of the process

Is CMOS really cheaper?

Ever increasing speed of transistors is the main scaling advantage for RF-Silicon technologies. However scaling below 100nm of feature length poses strong technical challenges in the fabrication which requires highly advanced and expensive tools (sources with smaller wavelengths) or new process techniques such as off-axis illumination and phase-shift masking which considerably adds to the complexity of the fabrication process. Therefore with the introduction of each new process node there is a hike in the cost per die size to account for more advanced tools and techniques employed in the fabrication process. At a given node, CMOS normally requires less masking steps than SiGe. This will result in faster fabrication cycle time and lower cost. However as depicted in figure 1.6, since masking steps increase with CMOS scaling, the gap between masking layer counts of CMOS and SiGe technologies is diminishing as new generations of scaling emerge [3].

Although at each process node the cost of manufacturing a SiGe chip is more than the cost per die area of a CMOS chip, due to higher performance of SiGe technologies a newer CMOS technology should be employed to compete with an older SiGe technology. In fact cost of each process node of CMOS technology is comparable to that of a SiGe technology which is two generations older (figure 1.7)[4]. If we look at the same transit frequency (f_t) as a figure of merit for performance, it can be seen that for each process node, SiGe technology has a better performance which is slightly more than twofold. Hence price per performance (f_t) for SiGe technology is marginally better than that of CMOS technologies if the cost of the die area is considered to be the dominant factor in determining the total system cost. Therefore for applications that the size is mostly ruled by passive component footprints and hence does not change with proceeding to newer process nodes, die area cost is almost the same for similar performance obtained out of CMOS and SiGe

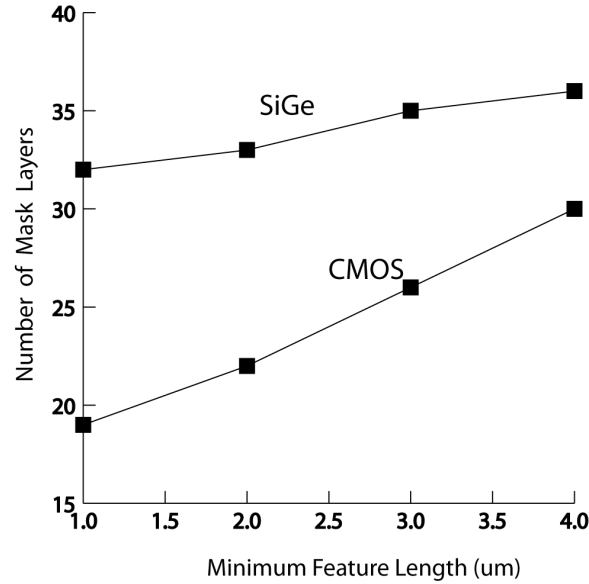


Figure 1.6: At each process node SiGe has more number of mask layers, however in advanced CMOS technologies the difference is diminishing[3].

technologies and therefore it is not the major key player for the choice of the technology. On the other hand for applications that are digitally intensive and a great portion of the chip is occupied by digital or analog/mixed signal blocks, their size will shrink by migrating to a newer generation of scaled technology and then CMOS technology would be more die area cost effective (figure 1.8).

mm-Wave solutions need a more specialized package since RF interfaces are at such high frequencies and are prone to parasitics which are considered negligible for other applications at lower frequency bands. In most cases, antennas will be embedded in the same chip or they will be placed on a substrate in the flip chip package configuration that the mm-wave chip can be directly mounted on top of it. These special cares make the cost of testing and packaging a noticeable portion of the overall cost of a complete mm-wave system. Again CMOS technology with the capability of fully exploiting the built-in self test (BIST) feature can profit from the scaling and reduce this portion of the cost. Moreover making use of a standard CMOS fab that offers no specialized options and relies only on standard options as offered in the fabrication of all-digital processors has the advantage of much less lead time that can help lower the price by lowering the time to the market.

1.2.2 Performance

Intrinsic gain

A figure of merit for amplifier design is the intrinsic gain of the device ($g_m \cdot r_o$). Due to the exponential I-V characteristic of bipolar transistors, SiGe offers a better transconductance(g_m)

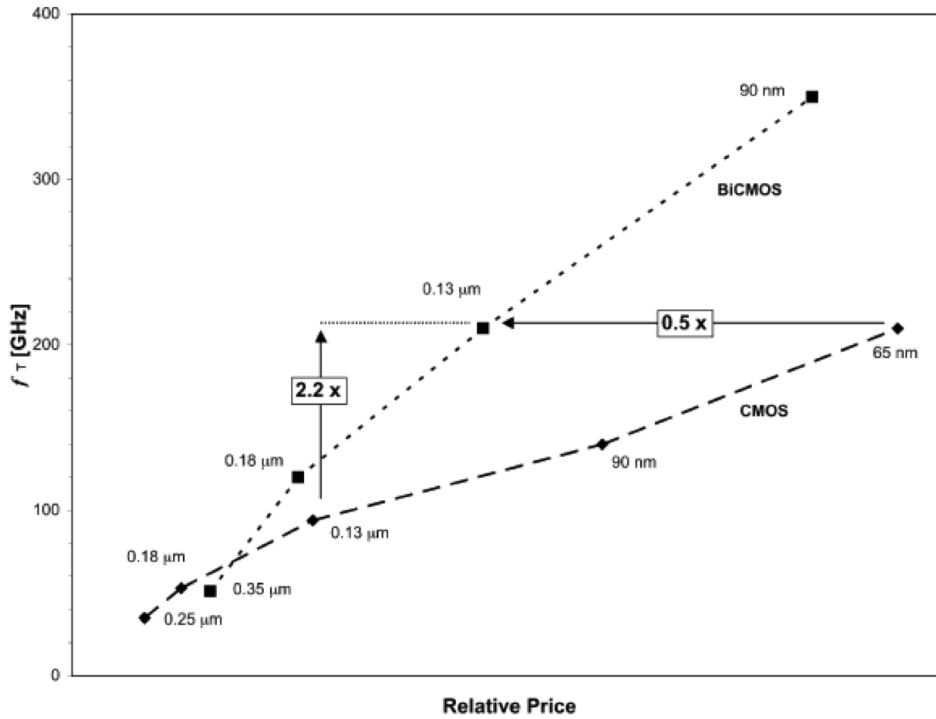


Figure 1.7: Cost per die size of each CMOS process node equals to that of a SiGe technology which is two generations older [4].

over CMOS where the output current has a quadratic relationship with input voltage (for highly scaled short channel devices, as a result of velocity saturation this relationship approaches toward a linear relationship and makes the situation even worse for the CMOS transconductance). SiGe BiCMOS technology provides a knob for the device designer to tweak the emitter germanium concentration and modify the electric field at the base, as a result of it one can improve the early voltage and the intrinsic output resistance of the device. In CMOS technology, in order to improve the reliability of analog and RF applications, a few extra processing steps are needed to increase the break-down voltage (eg. dual oxide technology), as well as additional implants to compensate for the short channel effects (eg. halo implants to control the punch through). These reliability precautions usually end up with lower output resistance for CMOS and hence considerably lower intrinsic gain. Although for many RF applications the quality factor of passive components which are tuning out the output capacitance determines the total loading at the output of the device, higher g_m of the SiGe technology eventually results in higher gain for SiGe designs of RF and mm-wave amplifiers.

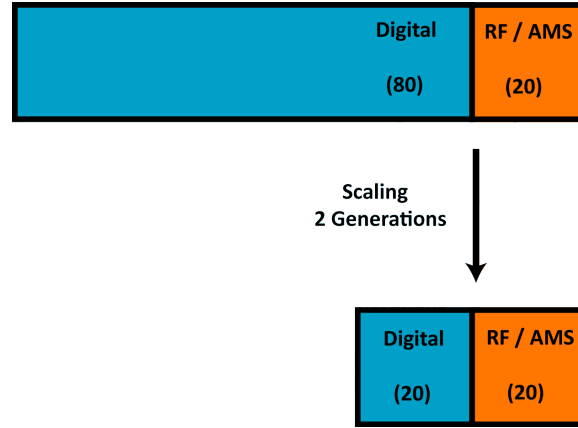


Figure 1.8: Digital intensive solutions get shrunk considerably as they are migrated to newer process nodes and the die cost will go down.

Matching

Matching is an important criteria for analog and RF circuits and systems design. For mm-wave application where there are multiple TX and RX chains to form a phased array solution, good matching between paths relaxes design requirements and simplifies algorithms that control the phased array operation. In the case of bipolar devices, the matching of short circuited base-emitter potential drop (V_{BES}) is determined by the doping profiles of p/n areas across the emitter/base junction. With each technology improvement these doping levels are increasing and better matching is obtained consequently. Matching between threshold voltages (V_{TH}) is playing the equivalent role in CMOS devices. The CMOS threshold voltage depends not only on the doping profile, but it also depends on the lateral parameters such as channel length and finger width of the device. These are much more process dependent and less stable which result in inferior matching performance of the CMOS technology with respect to bipolar SiGe counterparts.

Noise

For mm-wave amplifiers and buffers where there is no frequency translation, NF_{min} is sufficient to characterize a technology in terms of the noise performance. In VCOs, mixers and other building blocks where up-converting of low frequency noise takes place, flicker ($1/f$) noise should be considered as well.

FET and BJT transistors have different mechanisms in generating the $1/f$ noise. In bipolar technologies flicker noise is generated in the emitter-base junction. In CMOS technologies, defects at the interface of $Si-SiO_2$ is responsible for the trap and release of charges that leads to the noisy current flow. The quality of the emitter-base junction and $Si-SiO_2$ interface determines the power of the flicker noise. Present-day SiGe devices have a high-quality of $Si-SiGe$ interface which makes their flicker noise behavior superior to CMOS devices. Furthermore CMOS scaling moves

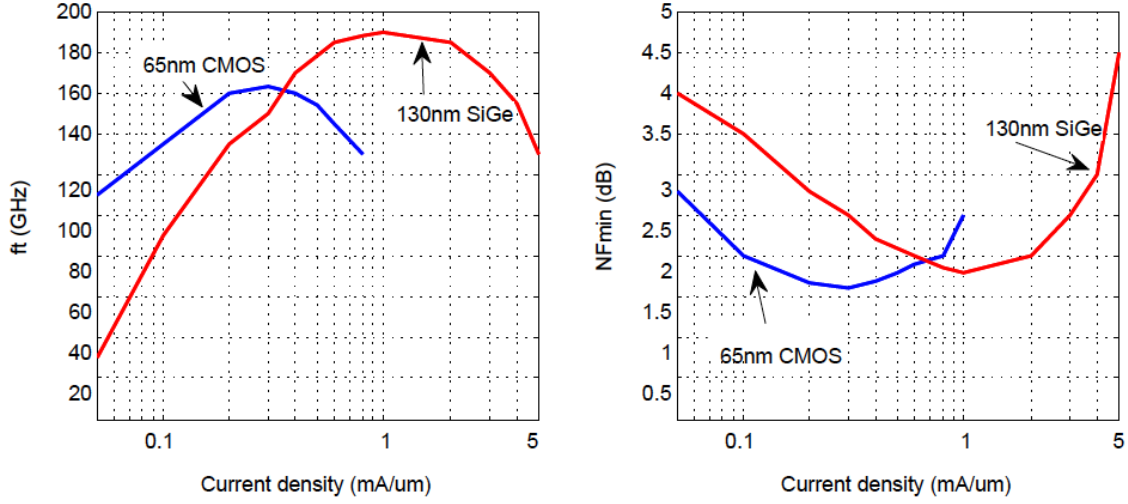


Figure 1.9: Comparing f_t and NF_{min} of 65nm CMOS and 130nm SiGe technologies

the oxide interface closer to the active channel which worsens the $1/f$ noise performance in each new generation of scaling.

In terms of NF_{min} , although in older technologies SiGe had a clear advantage over CMOS processes for high frequency noise, today CMOS technologies with higher source and drain dopings (lower parasitic R_S and R_D) and higher transconductances as a result of shorter effective gate lengths (which results in lower thermal noise generated by the channel), the NF_{min} is comparable to the minimum noise figure attainable via SiGe technologies (figure 1.9).

f_t and f_{max}

The most important FOMs for high frequency applications are current gain cut-of frequency (f_t) and maximum oscillation frequency (f_{max}). Maximum oscillation frequency is the highest frequency that an active device is capable of providing power gain and hence f_{max} sets a higher limit for amplifier and fundamental-frequency oscillator designs in each process node. With each generation of scaling, lateral and vertical dimensions shrink and these two FOMs increase. As depicted in Fig. 1.7, f_t of each CMOS process node is comparable to f_t of a SiGe technology which is two generations older in terms of scaling. A careful layout that minimizes external added parasitics associated with routings and interconnections will result in a f_{max} which is larger than f_t for both technologies. The quantitative relations are derived in [3] and [9]:

$$f_{max,CMOS} = \frac{f_{t,CMOS}}{2} \cdot \sqrt{\frac{C_{GS} + C_{GD}}{r_g \cdot (g_m C_{GD} + g_{ds} C_{GS} + g_{ds} C_{GD})}} \quad (1.1)$$

Table 1.1: For comparable f_t and current dissipation, CMOS technology has less intrinsic parasitics which makes it more vulnerable to layout added parasitics and their variations.

Technology	CMOS(65nm)	SiGe(130nm)
I_{dc}	14.5mA	15mA
f_t	152GHz	160GHz
g_m	40mS	300mS
c_{gs}/c_{π}	30fF	240fF
c_{gd}/c_{μ}	12fF	60fF

$$f_{max,SiGe} = \sqrt{\frac{f_{t,SiGe}}{8\pi \cdot R_B \cdot C_{\mu}}} \quad (1.2)$$

As seen from above equations, lowering loss and resistive parasitics in the layout as well as decreasing the amount of output signal fed back to the input through C_{GD}/C_{μ} , results in an f_{max} number higher than the f_t value which is a significant aid to amplifier design. An f_{max} number as high as three times the f_t has been reported in [48] which allows for amplifier and oscillator design beyond the f_t of the process.

1.2.3 Quantitative comparison

There are multiple figure of merits that characterize each process for the RF design. f_t and f_{max} set the limit for the operating frequency of signal generation and amplification. NF_{min} sets the minimum achievable noise figure and is very important for designing low noise amplifiers in sensitive receivers. Maximum current handling capability and break-down voltages are principal parameters determining maximum power that can be delivered to the output load in each technology. Robustness and parasitic tolerance are also important parameters for first-pass designs that depending on the situation can favor one technology to the other one.

As depicted in figure 1.9, 65nm CMOS and 130nm SiGe technologies are on par with each other in terms of peak f_{max} and minimum NF . Consequently designing for amplifier's gain and noise figure or oscillator's maximum allowable frequency have comparable challenges in these two technologies. On the other hand peak f_t and minimum NF happens at a higher current density for the SiGe technology. Being capable of handling higher currents in SiGe ($1.2mA/\mu m$) with respect to CMOS ($0.5mA/\mu m$) and also having higher breakdown voltages ($BV_{SiGe} = 4V$ compared to

$BV_{CMOS} = 2.4V$) make SiGe technologies more attractive for applications that output power delivery is the main concern. Higher supply voltages allow for more headroom at the output and higher linearity for building blocks designed in SiGe, therefore due to the higher gain and linearity the power figure of merit ($FOM_{PA} = P_{out} \cdot G \cdot PAE \cdot f^2$) of a SiGe power amplifier can be an order of magnitude higher than its CMOS counterpart.

One other aspect to compare CMOS and SiGe technologies is how insensitive these two technologies are with respect to layout parasitics and their variations, and hence how capable they are for a first pass design. According to Table 1.1, for two CMOS and SiGe transistors biased in a way that they have the same I_{DC} and f_t , parasitic capacitances are almost an order of magnitude larger for the SiGe technology. Hence routing, interconnection and unknown layout parasitics have less effects in a SiGe design which makes this technology less sensitive and the goal of first pass design more attainable. The much higher transconductance gain of a 130nm SiGe technology ($\sim 30mS/\mu m$) with respect to 65nm CMOS ($\sim 1.4mS/\mu m$) makes high speed off-chip capacitance driving easier for a SiGe technology, which is an important factor for two-chip solution where there is a need for a high-speed interface.

1.2.4 Choice of technology

As it was stated earlier, thanks to the scaling, both standard CMOS and SiGe technologies are capable of operating at mm-wave frequency regimes. Inherently a CMOS technology is optimized mainly for digital circuitry, and digital requirements are the drive behind the direction of each generation of scaling. SiGe technologies are optimized for high frequency and RF designs with a superior metallization stack and higher performance RF transistors. Noticeably higher power handling capability and faster switching of off-chip capacitances are advantages for the SiGe technology. However for mm-wave designs in which, the antenna array is implemented in an on-chip or on-package fashion(in order to minimize RF interface parasitics), there is a flexibility in terms of antenna impedance to make it closer to the optimum output impedance that the power amplifier desires to interface. Moreover due to the small wavelengths, a larger antenna array can be utilized to exploit the spatial power combining and bridge the power delivery gap between CMOS and SiGe technologies. In conclusion both technologies can handle mm-wave design and depending on the application whether it is digital intensive (high definition video transmission, WPAN applications, etc) or more high frequency oriented (radar and imaging systems, point to point links, etc) one can be favored with respect to the other one.

Considering the cost, even though the cost of CMOS and SiGe technologies with equivalent performance seems to be comparable in terms of the die area cost, since digitally intensive applications can benefit from the area reduction as a matter of scaling, CMOS technology can result in a more cost efficient solution. Furthermore since there are more foundries worldwide offering CMOS technologies with more flexible schedules, it is more suitable for high-volume products which usually require shorter lead times. Since the technical characteristics of different runs offered by multiple foundries are similar, re-tweaking the design and sending it to another foundry should be easy with a CMOS choice which again makes it the more flexible and more suitable

pick for high-volume products. Nonetheless application that typically have smaller markets but need the extra RF performance, can benefit from the SiGe technology which is optimized to have higher performance under those circumstances. More advanced III-V compound technologies will be limited to very specialized and customized solutions in the future of mm-wave systems and circuits design.

In conclusion CMOS has advantages of lower power, lower wafer cost at a comparable node and shorter fab cycles. SiGe offers advantages of high speed combined with high drive capability and low noise. The ultimate selection will be based on system specifications.

Chapter 2

Transmit/Receive Switching

Previous works [44]-[48] have demonstrated the feasibility of many key mm-wave transceiver building blocks in standard digital CMOS processes. On the other hand, there have been relatively few demonstrations of transmit/receive switches operating at mm-wave frequencies, which is a key building block in RF systems. Operating at mm-wave frequencies has many advantages. One key advantage is due to small wavelengths, which allows antennas to be realized on chip or on the package, further reducing the cost of a system. Moreover many antennas could be integrated with suitable phase shifters to create phased array systems that effectively increase the aperture size and directivity of transmit/receive antennas by a factor of N (number of antennas). Spatial filtering achieved by a phased antenna array system alleviates impairments such as delay spread and co-channel interference and further helps to extend the communication range and bandwidth. Without a transmit/receive (T/R) switch, two separate antennas should be employed for the receiver and transmitter, which translates to half the transmit/receive antenna gain and aperture size for a given area. Whereas if T/R switches with performances that pass the system requirements can be implemented in the same CMOS chip, a single antenna will be shared between the transmitter and receiver and twice the number of antennas (antenna array gain) will be realizable out of the same die area. Hence T/R switch can save a great deal of area, and thus lower the cost, since antennas are relatively large. Therefore designing CMOS T/R switches at mm-wave frequencies is an important goal toward the realization of a low cost mm-wave system.

The design of T/R switches is challenging since it resides before the LNA and after the PA. In order not to degrade the sensitivity of the receiver and the transmit power, switches should incur low insertion loss in the ON mode. To have enough directivity, insertion loss should be rather high when the switch is OFF. To keep the LNA and PA isolated, enough isolation between the two thru ports should be ensured and finally switches should be capable of handling high powers beyond the PA output compression point. As with all high frequency circuits, input and output ports should be matched to guarantee adequate return loss for incident signals.

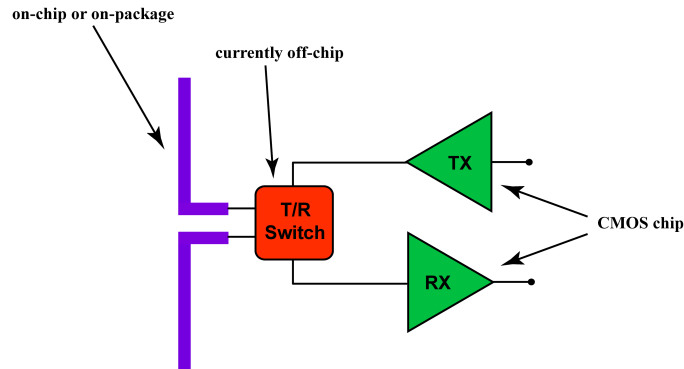


Figure 2.1: A fully integrated transceiver with an on-chip antenna requires an integrated T/R switch to save die area.

2.1 Antenna array reuse via T/R switching

2.1.1 On-chip or On-package antennas

Modern CMOS technologies are fabricated on a relatively conductive substrate (conductivity of $\sigma \sim 10S/m$). The high permittivity of silicon substrate ($\epsilon = 11.7$) tends to redirect the electromagnetic energy radiated by on-chip antennas to the lossy substrate. Surface and substrate modes are additional problems needed to be addressed with on-chip antenna design. There are techniques to mitigate these problems, one such technique is placing an electromagnetic hemispherical lens on the backside of the die ([30]). An electromagnetic lens on the backside of the wafer reshapes the medium underneath the antenna, collects surface and substrate mode radiations and converts them to a useful radiation from the back. Despite these solutions, having on-chip antennas incur a noticeable loss in advanced CMOS technologies due to conductive substrate and thin metal stacks. Moreover since the wavelength at 60GHz is $5mm$, accounting for ϵ of SiO_2 reduces the size of the antenna but still the spacing between antennas should be $\lambda_0/2$ where λ_0 is the wavelength of the signal in the air. The on-chip real estate that would be devoted to the antenna array would be substantial for large number of antennas in the array. Therefore having antennas on the package seems to be a viable solution according to these facts :

1. Higher performance antennas will be available on the package which has less conductive and dielectric losses compared to a silicon substrate.
2. Large arrays will be implemented on low-cost packages instead of a more expensive silicon wafer. This lowers the implementation cost.
3. Multi-layer boards with high-quality metallization not only provide good design environments for the antenna array, but they can also be employed to route signal, power and ground lines and furthermore act as a heat-sink to take the dissipated power off the chip.

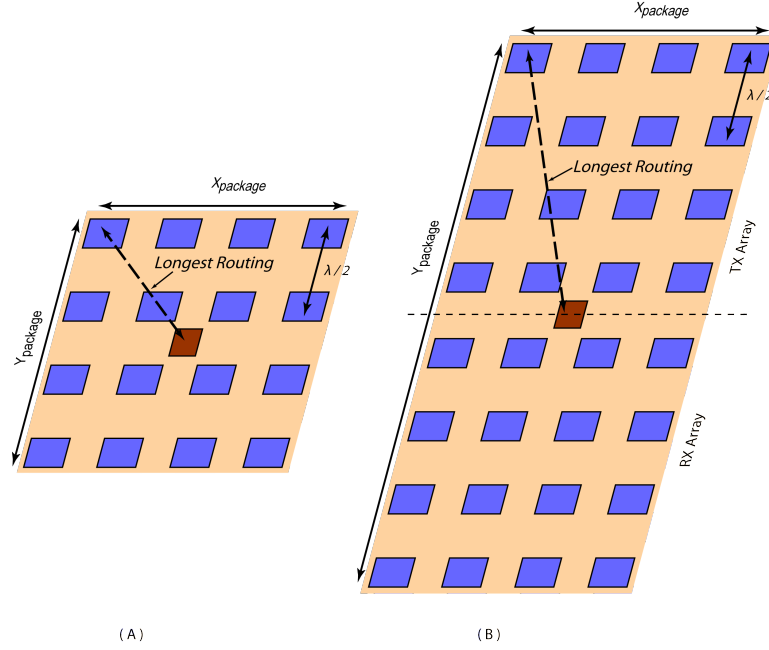


Figure 2.2: In case of not including T/R switches, two separate RX and TX arrays are required (B), which correspondingly lengthens interconnection routings and increases their associated loss

2.1.2 Size and loss trade-off in a transmit/receive array

Even when antennas are implemented on a package, having a smaller package makes the over-all solution more cost-effective and allows for easier adoption of the solution by WPAN applications such as smart phones, portable media players, etc. But whenever the package size is a secondary issue, then removing TR switches should result in a better over-all performance since there is no extra component after the PA to harm the TX linearity and output power nor is there a component before LNA which degrades the receiver noise figure and consequently sensitivity. However in practical cases, for solutions without the TR switch, bigger size of the package means longer on-package routings to connect front-ends residing on the chip to the antennas implemented on the package (figure 2.2). The routing loss is as important as the TR switch loss since it directly affects the transmit power and receive sensitivity. Depending on the array size and on-package transmission line loss characteristics, the resultant routing loss can offset the benefit of not including a T/R switch.

To have a quantitative comparison we consider the scenarios described in figure 2.2. In figure 2.2-a, one antenna array is shared between TX and RX via TR switches and In figure 2.2-b two separate arrays are used for RX and TX portions of the transceiver. To compare these two scenarios we compare the maximum communication range achievable by each case while assuming nominal values of transmitter $P_{out}=10\text{dBm}$, receiver $NF=10\text{dB}$ and a required SNR of 10dB for

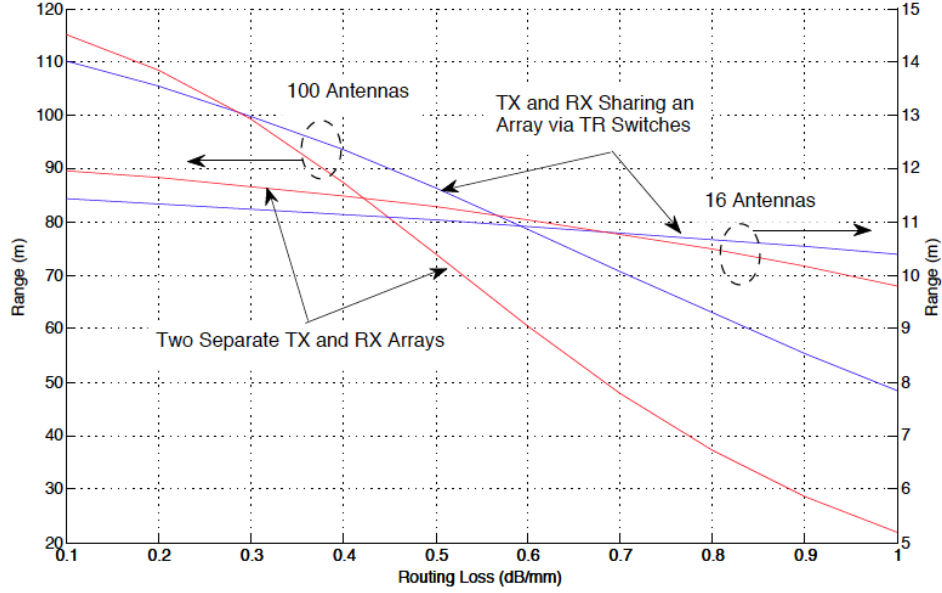


Figure 2.3: Graph of the maximum achievable communication range in two configurations of shared or separate arrays (for both small and large arrays). For larger arrays the routing loss is more detrimental than the T/R switch loss and higher quality packaging should be employed.

maintaining the communication link at the frequency band of $57 - 64GHz$.

According to the Friis equation [12](if we account for the transmit and receive array directivities in the TX effective radiated power (EIRP) and RX sensitivity and not in the free space path loss), the maximum achievable communication link is related to the maximum tolerable path loss as :

$$Path Loss = \left(\frac{\lambda}{4\pi R}\right)^2 \Rightarrow R = \frac{\lambda}{4\pi\sqrt{Path Loss}} \quad (2.1)$$

To calculate the maximum allowable path loss we need to know the effective isotropic radiated power at the transmitter and the minimum detectable signal at the receiver:

$$Path Loss = EIRP - P_{Noise} - NF - SNR_{min} \quad (2.2)$$

The effective isotropic radiated power for an array of N omnidirectional antenna elements will be :

$$EIRP = P_{PA} + 20\log(N) - Loss_{TR} - Loss_{Routing} \quad (2.3)$$

In a N -element phased array receiver since signals add up in the voltage domain and noises get combined in the power domain, as it will be derived in the next chapter, the total noise factor of the array will be reduced to:

$$F_{Array} = \frac{1}{N}(F_{element}) \quad (2.4)$$

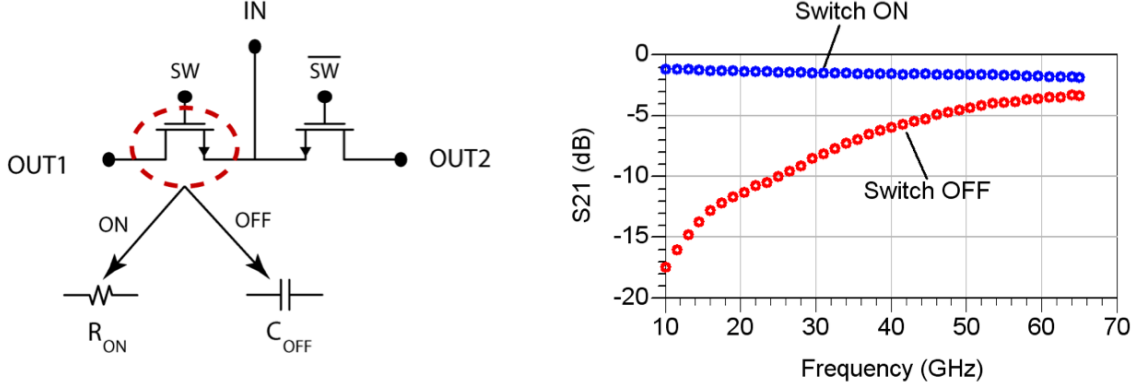


Figure 2.4: Measured S_{21} of an NMOS transistor acting as a series switch in a 50Ω environment.

Where $F_{element}$ is the noise factor of each path at the antenna :

$$F_{element} = 1 + \frac{F_{receiver} - 1}{Loss_{routing} + Loss_{TR}} \quad (2.5)$$

Routing loss depends on the dies size, array size/shape and the loss characteristic of the package. For the scenario depicted in figure 2.2 , the maximum routing length will be calculated as :

$$Routing\ Length = \sqrt{\left(\frac{X_{package} - X_{Chip}}{2}\right)^2 + \left(\frac{Y_{package} - Y_{Chip}}{2}\right)^2} \quad (2.6)$$

Assuming inter-element spacing of half the wavelength and plugging previously mentioned numbers for these two array configurations, figure 2.3 depicts the maximum achievable communication range as a function of the routing transmission line loss. As can be seen from figure 2.3, for larger arrays (100 element array in this case) transmission line routing loss is more pronounced and unless very low loss routing is realizable on package, including the T/R switch will provide better overall performance. For smaller arrays (16 antenna element in figure 2.3) more routing loss is tolerable. However for ultimate single-chip solutions where the antenna array is realized on chip, due to higher on-chip routing losses ($\sim 1dB/mm$) a solution with a shared array between TX and RX through a T/R switch looks more promising.

2.2 T/R switching configurations

2.2.1 Series vs Shunt switching

By its nature, when the gate terminal of a MOS transistor is driven by a rail to rail voltage, it acts as a switch. In fact this single transistor switch is sufficient for many digital and analog

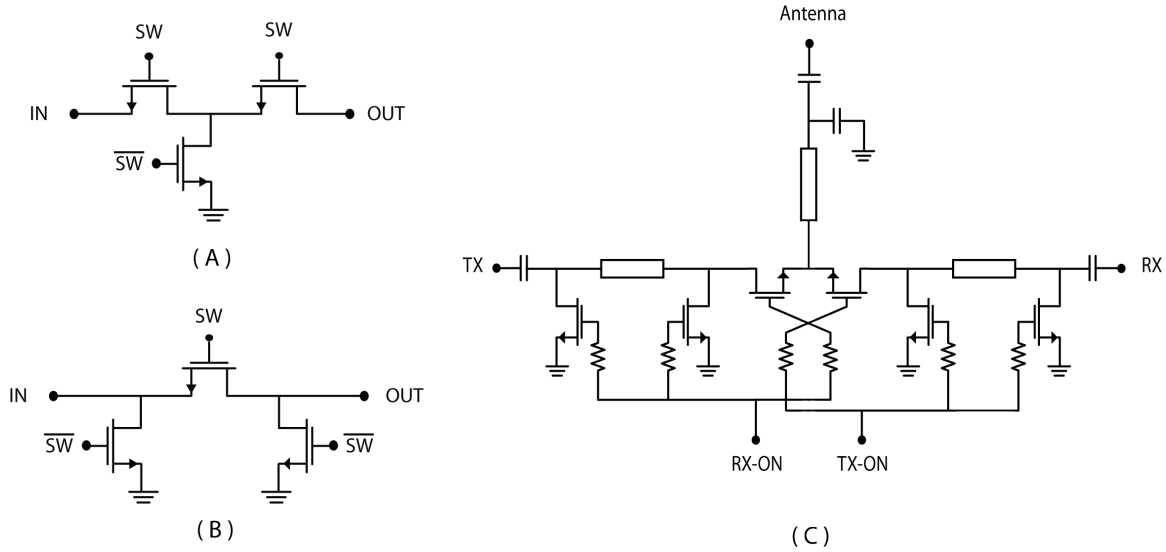


Figure 2.5: Combinations of series and shunt switches in a T (A) or π (B) or L -shape configuration (C)[33] decrease the amount of off-state leakage at the price of degrading the on-state insertion loss

applications. When the switch is on, the transistor is in the triode region and a low resistive channel connects the MOS source and drain terminals. To the first order, the on-mode resistance of a MOS switch (R_{ON}) is equal to $1/g_m$ where the transconductance is calculated at the edge of saturation. When the switch is off, there is no conducting channel between the source and drain terminals and the resistive path that connects source and drain nodes of the transistor exhibits a very large R_{OFF} . However in the off mode there is a feedthrough path through parasitic capacitors between source and drain terminals. This path is through both $C_{gd} - C_{gs}$ and $C_{db} - C_{sb}$ networks. At low frequencies, when the transistor is operating at frequencies much lower than its f_t , the capacitive feedthrough path introduces a much higher off-mode reactance than the resistance of the channel in the on-mode. But at mm-wave frequencies, when the device is operating close to the f_t limit, the impedance of the switch in the off state is comparable to when it is on. Figure 2.4 depicts the measured S_{21} of a $40\mu\text{m}$ NMOS transistor acting as a switch in a 50Ω environment. As can be seen in this figure, S_{21} for on and off states at 60GHz are comparable.

There are techniques to diminish the feedthrough path by cascading multiple switches (decreasing the effective series capacitance) and adding a shunt switch in between two series switches with an inverted gate signal with respect to the series switches (shorting out the feedthrough signal). These techniques (figure 2.5-a,b) are mostly applicable to lower RF frequency systems since at mm-wave frequencies having multiple transistors in the signal path introduces substantial insertion loss. Switches cannot be made too large due to parasitic capacitance limitations. In mm-wave regime, a switching structure with the least number of transistors is desirable to reduce parasitics. Furthermore, any transistors in series with the signal path can incur a noticeable loss and a structure

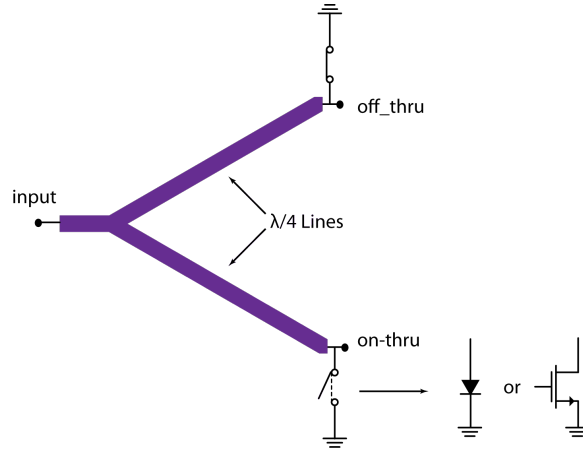


Figure 2.6: Traditional shunt switches occupy a large footprint as a result of employing quarter wavelength transmission lines

that removes series transistors is more suitable. This is the reason that in [33] an L-configuration with corresponding matching networks as depicted in figure 2.5-c is exploited with only one series transistor and two transistors in shunt with the signal path. An ultimate shunt SPDT switch at mm-wave frequencies has the capability to provide lower losses due to the elimination of transistors in series with the signal path.

The shunt SPDT switch configuration is demonstrated in figure 2.6. Traditionally shunt switches were designed exploiting two quarter-wave transmission line sections (figure 2.6) and shunt switches at each end. When a switch is closed, the corresponding thru-port is grounded and all the incident power reflects with negligible amount coupling through. The $\lambda/4$ line converts this low impedance of the short from the off-thru port to a high impedance (open circuit) at the common input port which results in directing all the input power toward the on-port. The impedance of the on-port load is matched to the quarter-wave line characteristic impedance which is in turn matched to the input impedance. The off-state switch at the on-thru port has a very high shunt loading resistance and its parasitic capacitance is absorbed into the quarter wavelength transmission line design. Even at mm-wave frequencies, $\lambda/4$ lines are too bulky and take too much area to be implemented on chip. Absorption of parasitic capacitances added by the interfacing building blocks helps to reduce the transmission line size but still the footprint can be huge. Therefore devising a lumped component counterpart for the shunt SPDT switch is highly valuable. In this work we introduce a transformer based shunt switch employing a transformer and shunt NMOS switches as shown in figure 2.7 .

2.2.2 The transformer based shunt switching structure

A transformer based shunt switching structure is demonstrated in figure 2.7. Comparing it with the traditional shunt SPDT structure (figure 2.6) reveals that the two quarter wavelength arms

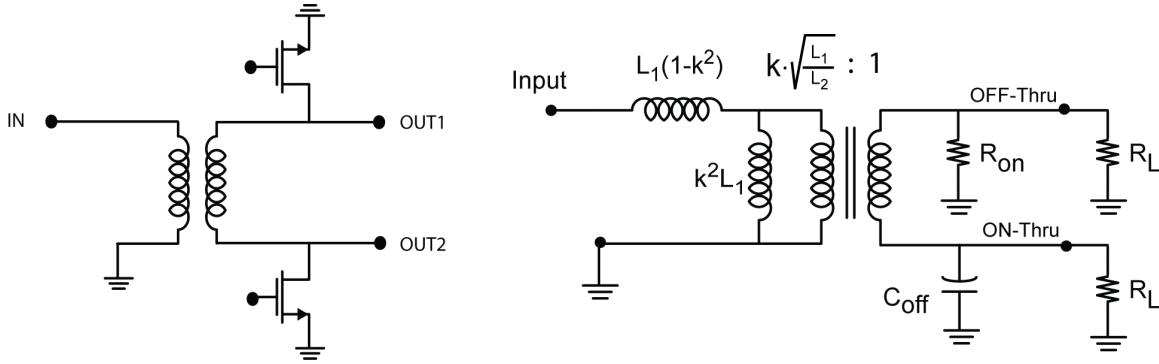


Figure 2.7: Schematic diagram of a transformer-based shunt switch and its equivalent circuit model

are replaced with a miniature transformer that saves a considerable amount of the die area. In a transformer based shunt switch, the input is connected across the primary winding and each end of the secondary winding is connected to an output with an NMOS transistor shunted to ground. Input current flowing in the primary winding produces magnetic flux that passes through the secondary winding and induces a voltage across the secondary. For each switching state, one transistor is on and the other one is off. When a transistor is on, it introduces a low impedance at its port which quenches any voltage swing on it and the induced voltage will majorly appear at the other end of the secondary winding. As shown in figure 2.7, all parasitic capacitances in this structure are shunt capacitors to ground and do not provide a feedthrough path between the input and output terminals. Moreover, these capacitances will be resonated out with the equivalent inductance appearing at the transformer terminals at the desired frequency and will not be seen by load and source impedances.

Unlike the traditional distributed switch, this structure introduces a 180° phase difference between the two outputs. In a TR switch, this 180° phase shift is of minor concern since this corresponds to moving the transceiver a distance of $\lambda/2$ and most circuits that are sensitive to RF phase use carrier locking techniques. Shunt switches shown in figure 2.6 are reflective type of switches which means that when the input is not connected to an output port, that port sees a short circuit and hence it is not matched. A mismatch at the input of LNA and output of PA can be problematic and cause oscillation. Since these switches are used in a time division duplexing (TDD) scheme, whenever one output is disabled, the interfacing building block can be turned off. Shutting down the block not only eliminates the potential oscillation problems as a result of a low impedance loading the off-thru port, but also it has the advantage of saving power dissipation at the expense of recovery time needed for those blocks to be turned on to reach their steady state condition required for the system operation.

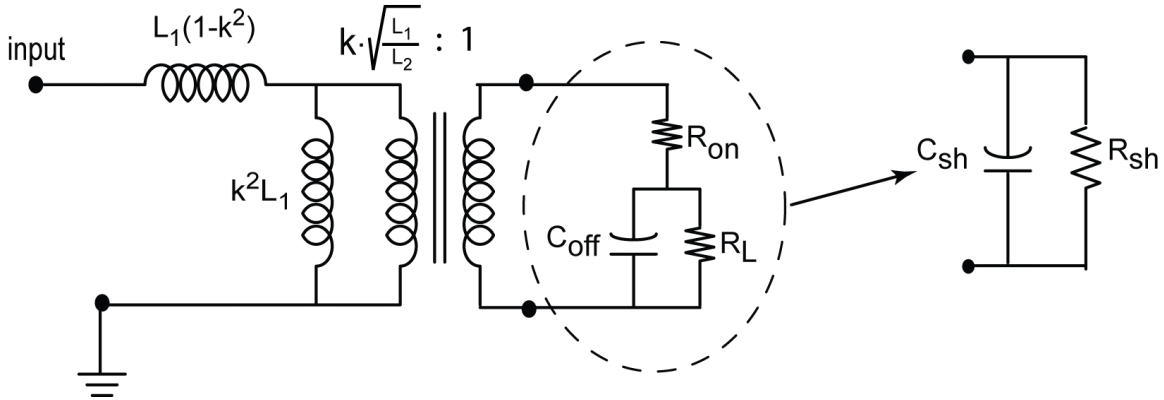


Figure 2.8: Simplified model of the SPDT network including loading effects

2.3 Design equations of a transformer based T/R switch

As depicted in figure 2.7. In a transformer based switch the input(antenna port) is connected to the primary loop of the transformer and outputs are taken from each terminal of the secondary loop. Shunt transistors are also connected to each terminal of the secondary in parallel with each output. These two transistors are driven at the gate by inverted control signals with respect to each other. so whenever the switching signal is high, the transistor acts as a low impedance path between its source and drain terminals.

Since the on-resistance of the transistor is much smaller than the load resistance ($R_{ON} \ll R_L = Z_0$) and for operating frequencies(ω) adequately less than the technology's transit frequency (ω_t) the on-resistance is much lower than the shunt reactance of parasitic capacitances of the transistor (C_{drain}), then we can assume that the loading at the off-thru port is only R_{ON} of the transistor. At the on-thru port, the transistor is off and the total loading at that port will be the load resistance(R_L) in parallel with the drain capacitance of the transistor at the off mode (C_{off}) as depicted in figure 2.8.

2.3.1 Equivalent shunt loading

To derive the transfer function, the loading network depicted in figure 2.9-a is transformed into its equivalent shunt impedance (figure 2.9-c). Two quality factors are defined in order to accomplish this. The first one is the quality factor of the shunt network of R_L and C_{off} which is the loading network quality factor and can be written as:

$$Q_L = R_L C_{off} \omega \quad (2.7)$$

And the second one is the quality factor for the series network of R_{on} and C_{off} which is the transistor-only quality factor and is determined by the choice of the technology. It Increases as

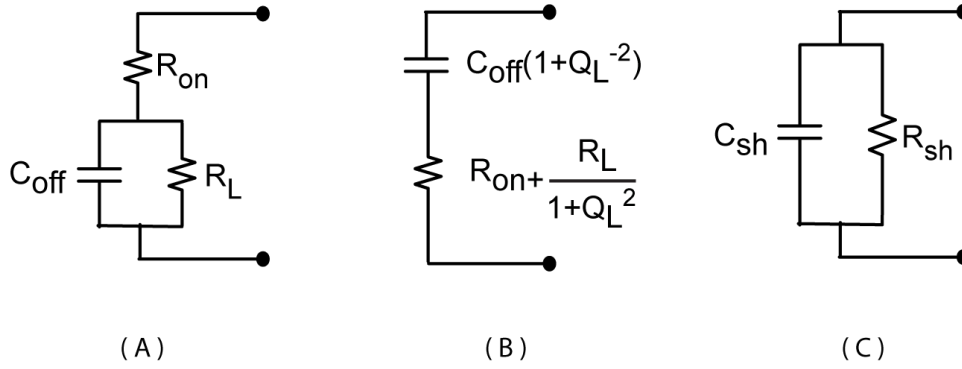


Figure 2.9: Calculating the shunt equivalent loading network

more advanced technologies are employed and it can be formulated as:

$$Q_0 = \frac{1}{R_{on}C_{off}\omega} \quad (2.8)$$

In order to reach to the shunt equivalent network, first the parallel combination of R_L and C_{off} is transformed to its equivalent series network to obtain the intermediate series network depicted in figure 2.9 as :

$$R_S = R_{on} + \frac{R_L}{1 + Q_L^2}, \quad C_S = C_{off}(1 + Q_L^{-2}) \quad (2.9)$$

To transform this intermediate series network to the equivalent shunt network we should calculate its quality factor(Q_{int}):

$$\frac{1}{Q_{int}} = \omega C_{ser} R_{ser} = \omega \left(R_{ON} + \frac{R_L}{1 + Q_L^2} \right) C_{off}(1 + Q_L^{-2}) \quad (2.10)$$

Expanding and regrouping the above equation yields to:

$$\frac{1}{Q_{int}} = \frac{1}{Q_0} + \frac{Q_L(1 + Q_L^{-2})}{1 + Q_L^2} \quad (2.11)$$

For large values of the load quality factor ($Q_L^2 \gg 1$) the intermediate quality factor can be simplified as:

$$\frac{1}{Q_{int}} = \frac{1}{Q_0} + \frac{1}{Q_L} \Rightarrow Q_{int} = Q_0 || Q_L \quad (2.12)$$

Finally with the aid of the intermediate quality factor (Q_{int}), the series network depicted in figure 2.9-b can be transformed into its shunt equivalent network (figure 2.9-b). This shunt loading impedance can be characterized as a resistance in parallel with a capacitance with values of :

$$R_{sh} = R_{ON}(1 + (Q_0 || Q_L)^2) + R_L \frac{1 + (Q_L || Q_0)^2}{1 + Q_L^2} \quad (2.13)$$

$$C_{sh} = C_{off} \frac{1 + Q_L^{-2}}{1 + (Q_L \| Q_0)^{-2}} \quad (2.14)$$

This equivalent loading is used in the next section to calculate the center frequency.

2.3.2 Center frequency and matching

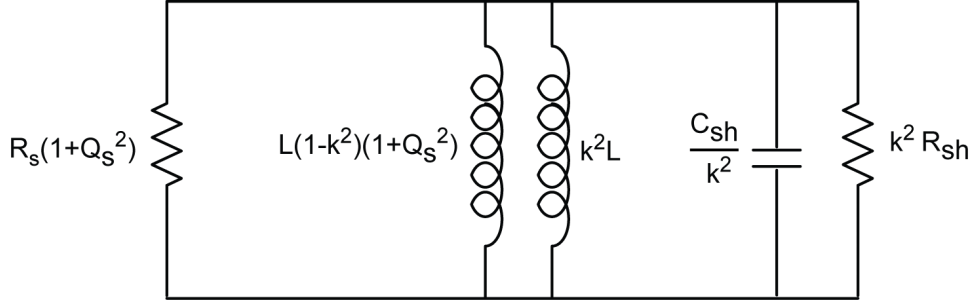


Figure 2.10: Parallel tank equivalent network of the switch for calculating the center frequency

Since a lower desired loss requires a lower on-resistance and consequently a larger transistor with more capacitance loading, to operate at mm-wave frequencies the smallest realizable inductance is preferred. So from now on we assume that the transformer is a 1:1 structure with self inductance of L for each loop. As depicted in figure 2.8 there is one remaining series structure that if it is transformed to its parallel equivalent network, the structure will resemble a parallel tank and calculation of the center frequency that the switch is operating at would be handy. The series network of the source impedance (R_S) and the leakage inductance ($L(1 - k^2)$) have a quality factor of $Q_S = \frac{L(1-k^2)\omega}{R_S}$. For a good transformer (tight coupling) the leakage inductance is small and the reactance associated with it at mm-wave frequencies is still much smaller than the source impedance ($Q_S \ll 1$). This small Q makes the equivalent shunt impedance of R_S stay at roughly the same value after the transformation ($R_S(1 + Q_S^2) \sim R_S$) and the equivalent shunt inductance of the leakage inductance ($L(1 - k^2)(1 + Q_S^2)$) to be much larger than the already existing shunt magnetization inductance (k^2L) and be neglected in the parallel configuration. As the load network ($R_{sh} \| C_{sh}$) is moved to the source side by getting multiplied by the factor k^2 (k being the coupling factor of the 1 : 1 transformer), the equivalent circuit looks like the one shown in figure 2.10.

Neglecting the transformed shunt equivalent of the leakage inductance with respect to the magnetization inductance reduces the parallel tank depicted in figure 2.10 to an inductance of k^2L in parallel with a capacitance of $\frac{C_{sh}}{k^2}$ with the resonance frequency of :

$$f_0 \sim \frac{1}{2\pi \sqrt{(k^2L) \cdot (C_{sh}/k^2)}} = \frac{1}{\sqrt{LC_{sh}}} \quad (2.15)$$

Where L is the self inductance of each loop of the transformer and C_{sh} was derived in equation 2.14. At the resonance frequency the apparent load impedance at the source side is k^2R_{sh}

where R_{sh} is described in equation 2.13. For an ideal case where coupling factor is unity and the transistor-only quality factor (Q_0) is infinity, the load impedance is transformed to the source side intact and the match is perfect. However for practical cases of $k \sim 0.7 - 0.9$, $Q_L \sim 2 - 3$ and $Q_0 \sim 5 - 7$ getting a -10dB of return loss (which is sufficient to meet the requirement for most communication systems) is fairly feasible.

2.3.3 Insertion Loss

The insertion loss due to the finite on-resistance of the switch (R_{ON}) will be the ratio of the power delivered to the effective on-thru load impedance and the power transferred to the total effective impedance, since we have these two impedances in series format (figure 2.9-b), the insertion loss is a resistance ratio and can be written as:

$$I.L. = \frac{R_L \frac{1+(Q_L \parallel Q_0)^2}{1+Q_L^2}}{R_{ON}(1 + (Q_0 \parallel Q_L)^2) + R_L \frac{1+(Q_L \parallel Q_0)^2}{1+Q_L^2}} \quad (2.16)$$

In practice Q_0 is few times higher than Q_L and as the design is moved to a more advanced technology in the future, Q_0 will be even higher with respect to Q_L and can be neglected in a parallel configuration ($Q_0 \parallel Q_L \sim Q_L$) and hence the insertion loss can be simplified as :

$$I.L. = \frac{RL}{RL + R_{ON}Q_L^2} \quad (2.17)$$

The transformer is assumed to be lossless in equations 2.16 and 2.17. Which means that the insertion loss of the transformer itself should be added to above equations to get the overall insertion loss of the switch. In practice, a typical transformer insertion loss is about 1-dB at mm-wave frequencies.

2.3.4 Leakage

To calculate the leakage signal we should consider that (1-I.L.) of the input power will be delivered to the network at the off-thru port. But since at that port, R_L is in parallel with R_{ON} , most of the power will be dissipated in R_{ON} and only a portion of it $\frac{R_{ON}}{R_{ON}+R_L}$ will reach the off-thru output port. Therefore the leakage can be formulated as :

$$Leakage = (1 - I.L.) \frac{R_{ON}}{R_{ON} + R_L} \sim \frac{R_{ON}}{R_L} \cdot \frac{R_{ON}Q_L^2}{R_L + R_{ON}Q_L^2} \quad (2.18)$$

Again the transformer was assumed to be lossless here. Non-ideal transformers help the leakage number as the leakage signal will be attenuated by the insertion loss of the transformer as well.

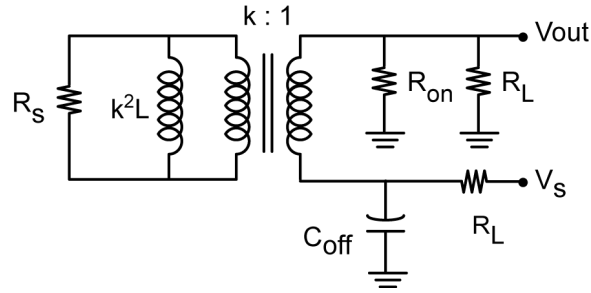


Figure 2.11: Equivalent circuit for calculating the isolation between two thru ports

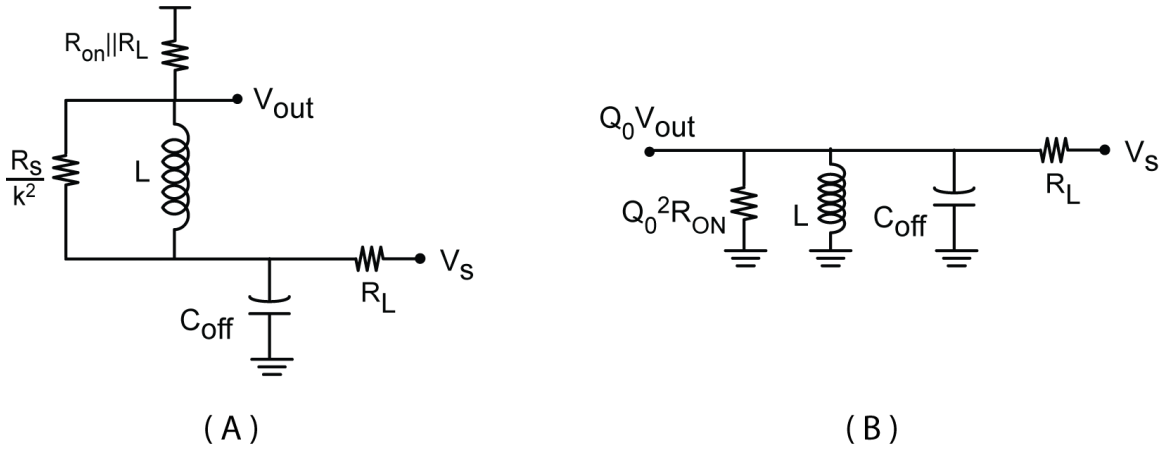


Figure 2.12: Source network is transferred to the load side (A) and then the structure is converted to a parallel tank configuration (B)

2.3.5 Isolation

To figure out the isolation, the input signal is connected to the off-thru port and the signal reached to the on-thru port is calculated (or vice versa). The situation is depicted in figure 2.11. The source impedance and the magnetization inductance can be moved to the secondary side as depicted in figure 2.12. Assuming the shunt resistance of $\frac{R_S}{k^2}$ is large enough to be neglected in a parallel configuration with the reactance of $L\omega$ (which is not a very accurate assumption but can be thought of as an overestimation for the quality factor of the inductor and hence as an underestimation of the isolation value) the equivalent network can be simplified further which results in the inductance L being in series with the resistance $R_{ON} || R_L$. Converting this series network to its shunt equivalent circuit (and take into account that the voltage gets multiplied by the quality factor which is in fact the Q_L) we arrive at the parallel tank network described in figure 2.12-b. At the center frequency

L and C_{off} will resonate out each other and the equation for the isolation can be written as :

$$Isolation = \frac{1}{Q_0} \cdot \frac{Q_0^2 \cdot R_{ON}}{Q_0^2 \cdot R_{ON} + R_L} = \frac{Q_0 R_{ON}}{R_L + Q_0^2 R_{ON}} \quad (2.19)$$

Impact of the transistor on-resistance R_{ON} is noticeable again in the performance of the switch and as more advanced technologies are employed, better isolation is also achievable in the transformer based T/R switches. These switches are reflective type of switches which means that the thru port is not matched at the off port and therefore the corresponding building block will be turned off in order not to cause any oscillation issue. Hence low isolations are somewhat tolerable in shunt switches.

2.4 Transformer based switch design example in 90nm CMOS

After deriving design equations of a transformer-based shunt T/R switch in section 2.3, a prototype is designed, fabricated and tested to verify the validity of proposed equations. In upcoming subsections, MOS transistor performance in switch mode is investigated and an optimum transistor layout to be used in the T/R switch design is figured out. Chosen sizes and design parameters for the transformer and shunt transistors are presented and measurement results are shown at the end.

2.4.1 MOS transistors performance in switch mode

Layout concerns of a MOS transistor when it is used as a high frequency switch is different from when it is expected to function as a high frequency amplifier. When a MOS transistor is laid out to be used in a mm-wave amplifier, smaller finger widths are used to decrease the series resistance of the gate terminal and as a result of that, lower losses and higher f_{max} is achieved. To decrease the loss through the back-gate effect, bulk resistance should be either close to zero (the case with a solid substrate ring surrounding the device) or approaching infinity (which means very few contacts are placed relatively distant from the transistor or in case of having the deep nwell option a bulk resonant network is employed). For high frequency amplifier design the former option of having a well defined substrate ring is preferred due to its compactness, more predictable modeling of the device and less vulnerability to substrate noise and coupling issues. However when a MOS transistors is laid out for switching purposes, especially transmit/receive switching, linearity is a major concern since any degradation to the output power will be extremely costly.

For MOS transistors acting as a shunt switch, the input signal is supposed to be directed to the output port that its corresponding shunt transistor is in the OFF mode ($V_{GS} = 0$). In case of having a small impedance at the gate, that terminal becomes an AC short which means that the gate voltage remains at the provided bias voltage throughout the operation. Therefore for positive swings of the output voltage, the terminal connected to the output acts as the drain terminal of the transistor while the AC fluctuation of the V_{GS} remains at zero and the transistor stays completely OFF at all times with no degradation to the output power linearity. On the other hand in negative

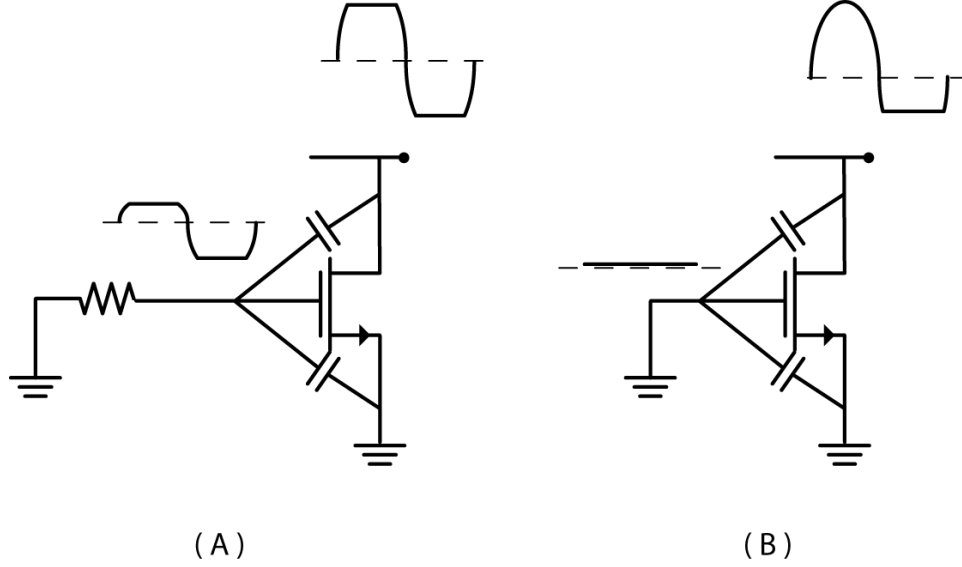


Figure 2.13: Higher biasing impedance at the gate not only improves the insertion loss but also results in less distortion to the output signal and a higher linearity number for the T/R switch

excursion of the swing the output node functions as the source of the transistor and if the output voltage goes below zero by a V_{TH} , the transistor will be turned on in the reverse direction which clamps the signal and produces distortion to the output voltage (figure 2.13).

On the other hand if there is a large biasing impedance placed at the gate node, the gate terminal will be an AC floating node and hence a feed through of output signal will appear at the gate terminal via parasitic paths provided by the transistor capacitances. As depicted in figure 2.13 for positive swings of output signal that the output node becomes the drain of the transistor, the feedthrough signal appearing across gate-source terminal through the $C_{GD} - C_{GS}$ network will be :

$$V_{GS} = V_G = \frac{C_{GD}}{C_{GS} + C_{GD}} V_{out} \quad (2.20)$$

For negative excursion of the output signal, the output node serves as the source of the transistor and therefore the voltage across gate-source terminal that modulates the channel will be :

$$V_{GS} = V_G - V_S = \frac{C_{GS}}{C_{GS} + C_{GD}} V_{out} - V_{out} = \frac{-C_{GD}}{C_{GS} + C_{GD}} V_{out} \quad (2.21)$$

So the gate-source voltage will be modulated by a fraction of the output voltage and the channel turns on when :

$$|V_{out}| \geq \frac{C_{GS} + C_{GD}}{C_{GD}} V_{TH} \quad (2.22)$$

Which is a factor of $1 + \frac{C_{GS}}{C_{GD}}$ larger than the turn-on voltage in the case of small biasing impedance being presented at the gate line ($|V_{out}| \geq V_{TH}$).

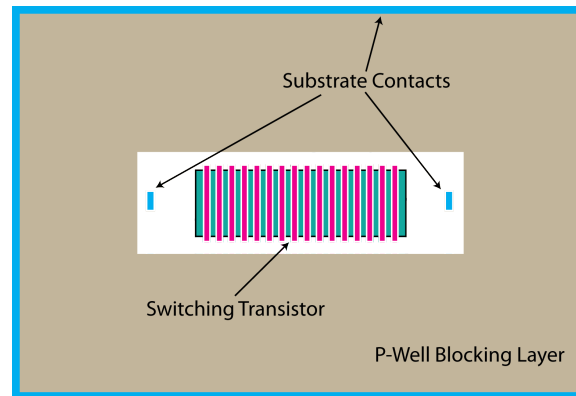


Figure 2.14: Layout example of a MOS transistor to be used as a switch

In conclusion, for the interest of linearity, large biasing resistors should be placed at the gate of MOS transistors that are intended for transmit/receive switching. Since that gate biasing resistance is in series with device's poly gate resistance and much larger than that, the value of intrinsic series gate resistance of the device loses significance and therefore larger finger widths are exploited in the layout to make the overall transistor structure more compact and the impact of parasitic inductances caused by long interconnections less effective.

The same argument applies to the bulk terminal through the back-gate effect which makes larger substrate resistances desirable. To do so one can use a deep-nwell option and bias the bulk of the transistor through an inductor which makes the bulk node open at the operating frequency. However having a triple-well transistor with a bulk resonant network is not appealing at mm-wave frequencies for following reasons:

1. Added deep nwell region increases parasitic capacitances of the MOS transistor and limits its operating frequency (for a maximum allowable capacitance, a smaller triple well transistor can be employed which results in a larger on-resistance of the switch and hence higher insertion loss)
2. Transformer based switches at mm-wave frequencies are quite compact and adding two bulk resonating inductors (one for each shunt transistor) roughly triples the size of the SPDT structure. Careful ground shielding should be placed around inductors in order to diminish the inductive coupling among the loops.

The other method to increase the substrate resistance is by having fewer number of bulk contacts (with respect to a MOS transistor specialized for high frequency amplifier design) and place them at a relatively farther distance. These few contacts will provide the correct bias voltage for the substrate. However for latch-up and substrate coupling purposes, it is still beneficial to have a well-defined ring at an outer area. The space between this ring and transistor's few bulk contacts can be filled with a pwell blocking layer in order to have a native substrate (as opposed to a highly doped surface) around the transistor to preserve the substrate resistance at a high desirable value.

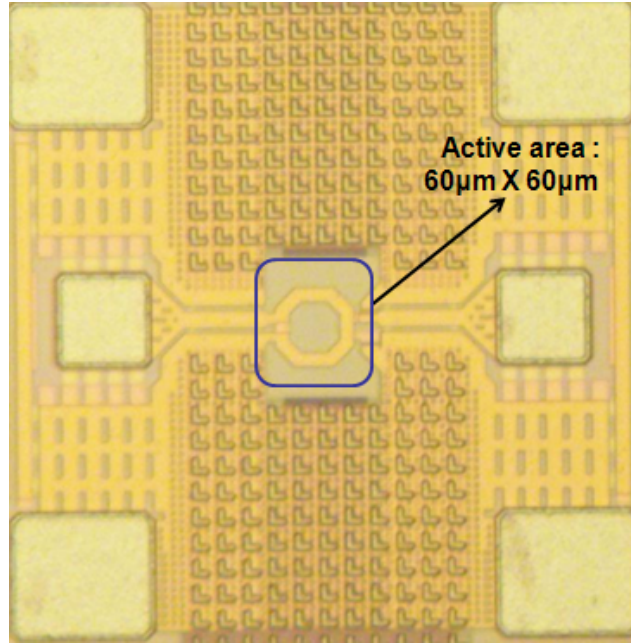


Figure 2.15: Die microphotograph of the miniaturized shunt switch employing a transformer

Since this is a customized layout configuration and not included in standard libraries provided by the foundry, extraction tools will not be accurate in capturing all the device parasitic parameters. To have an accurate model of the transistor, a test structure is required to be fabricated and characterized so that a measured based custom model of that transistor is available for maximum accuracy of the design.

2.4.2 Prototype design

To minimize the insertion loss of the switch, the on-resistance of NMOS transistors should be as small as possible. This in turn requires for an NMOS transistor as large as possible. Larger sizes increase the capacitive loading of the transistor and correspondingly a smaller inductance is required to tune out this capacitance at the operating frequency. However if the transformer gets too small, the primary and secondary self inductances get comparable to the trace inductances used for connecting to the switching transistor and the ground plane. Therefore too small of an inductance value makes the switching network more vulnerable to parasitics and is detrimental to the predictability of the design. In our design an overlay structure for a 1:1 transformer was used. Two equally thick top metal layers were employed. The diameter of the octagon loop is $42\mu m$ and the width of the winding is $W = 8\mu m$. The two loops are identical in shape and have a self inductance of $L_1 = 90pH$. Since the thickness and conductivity of the two top metal layers are equal the quality factor numbers are similar for the loops ($Q \sim 12$). Although more parasitic

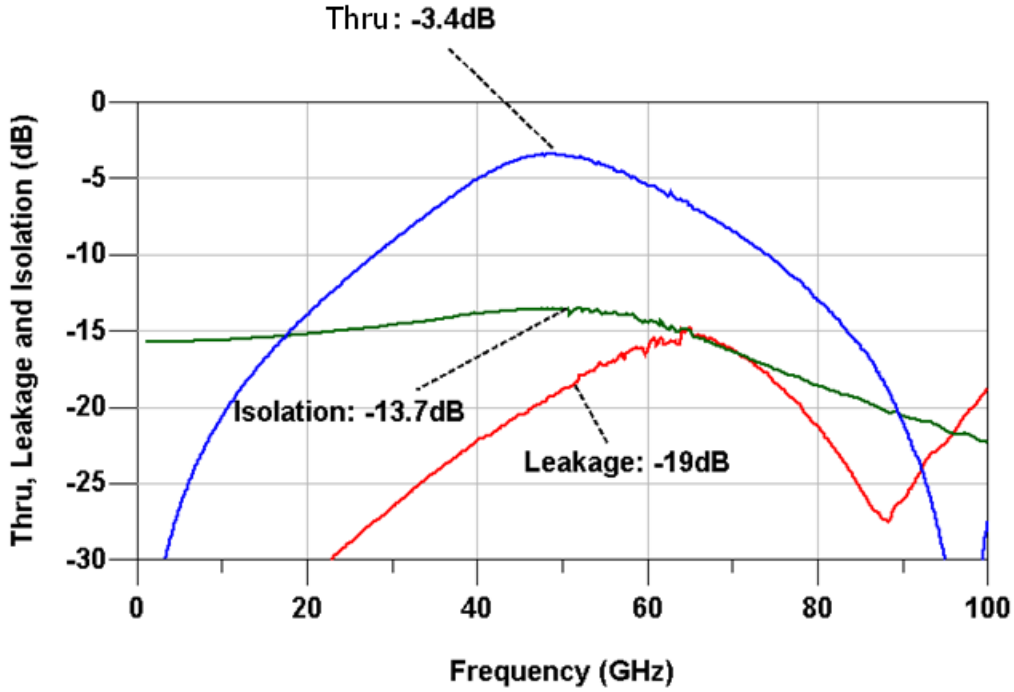


Figure 2.16: Measured insertion loss, leakage(input to the off-thru port) and isolation (between the on-thru and off-thru ports)

capacitances is being present at the bottom loop its effect on the Q is not significant at 60GHz . The coupling between the two loops is $k = 0.72$. These numbers for quality factors and mutual coupling results in a minimum insertion loss of 0.85dB at 60GHz for the selected transformer.

The prototype uses $40\mu\text{m}$ NMOS transistors in a 90nm standard CMOS process. Channel lengths are set to the minimum for smallest R_{ON} . However, since higher R_{Gate} is beneficial for the switch insertion loss (less power will be dissipated in the gate network) and improves the linearity (by making the gate terminal a floating node), unlike mm-wave amplifier design finger widths that result in maximum f_{max} were not used. To further increase the impedance at the gate a large biasing resistor of $R_G = 1\text{k}\Omega$ was inserted at the gate of each transistor in series. This large added R_G also alleviates the effect of any gate line inductance. Inductance at the gate line is detrimental to the linearity as it resonates out some part of parasitic gate capacitance to the ground (this is a capacitance that helps linearity as described in section 2.4.1). To eliminate the detrimental effect of the bulk network on linearity and insertion loss, a layout as described in figure 2.14 with few close by bulk contacts is devised.

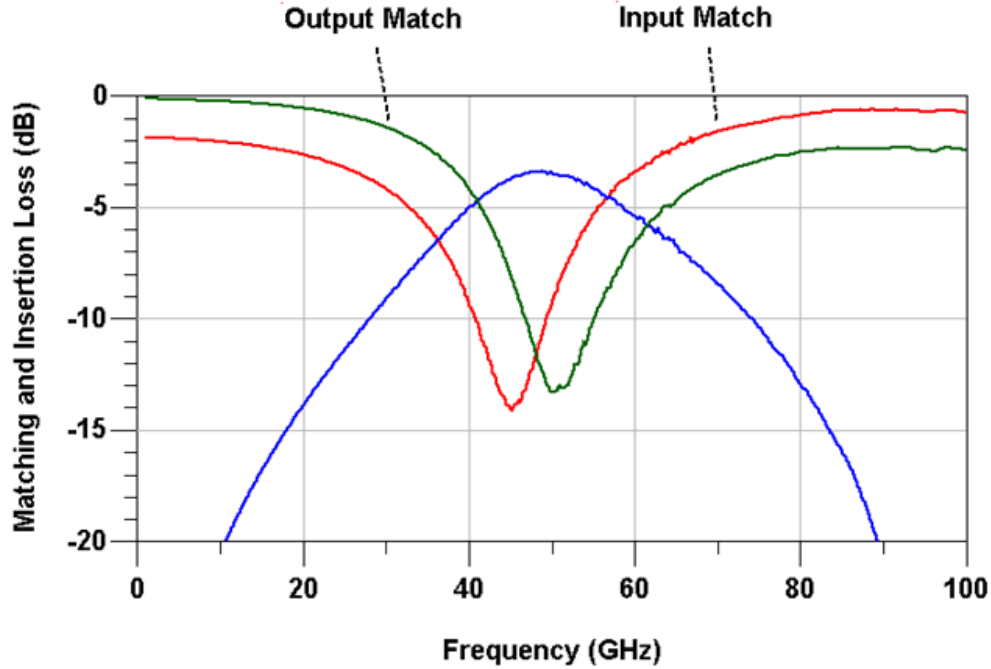


Figure 2.17: Measured gain, input and output return loss curves

2.5 Prototype Measurement Results

A prototype STDP TR switch has been fabricated in a 90nm CMOS process. The die photo is shown in figure 2.15. As can be seen in figure 2.15, employing the transformer and designing on a lumped component basis miniaturized the structure and the active area of the switch is only $60 \times 60 \mu\text{m}^2$. On-wafer measurement results are shown in figures 2.16 - 2.18. The input and output GSG pads have been de-embedded from the measurements using on-chip open and short de-embedding structures of the pad. As depicted in figure 2.16, the switch has its minimum insertion loss of 3.4dB at 50GHz and its 3dB bandwidth extends beyond the range of $40\text{GHz} - 60\text{GHz}$. The switch was initially designed for 60GHz , however due to the fact that the employed transistor layout as shown in figure 2.14 is a costume layout optimized for the switching performance, extraction tool estimation of its parasitic capacitances are not accurate. Taping out transistor test structures with the same exact layout and fitting the model to the measurement data or doing a measurement based design will capture the correct center frequency for future works. When the input is connected to an on-thru output, there is 19dB of leakage in the off-thru port. The isolation between two output ports is 13.7dB . Input and output return loss numbers are better than -10dB (figure 2.17). The

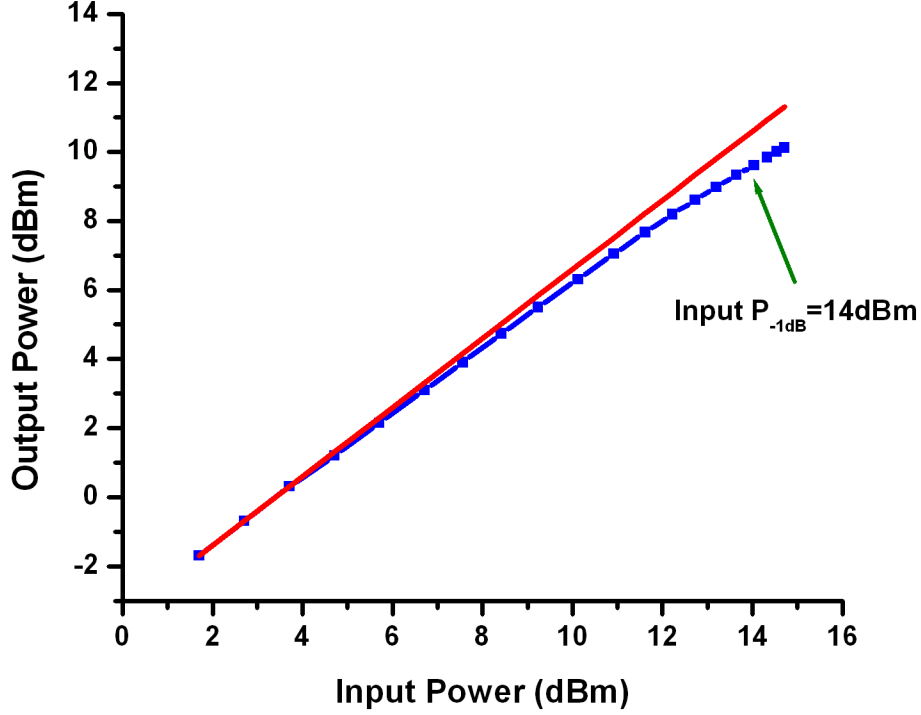


Figure 2.18: Large signal measurement ($P_{-1dB} = 14dBm$)

large-signal power measurements shown in figure 2.18 show an input referred 1dB compression point of +14dBm, which is adequate for most CMOS mm-wave applications. Due to this high input power required and power handling limitations of the VNA, an external power amplifier module was used to bring up the power provided at the probe tips to a range that P_{-1dB} curves can be captured. Table 1 summarizes the overall measured performance of the switch and compares to recently published results for mm-wave T/R switches. This work has the best performance in terms of insertion loss and die area. Beyond 50GHz it has a record linearity performance in a CMOS technology.

This transformer based shunt switch was implemented in a 90nm MOS process with an f_t of $\sim 100GHz$. According to equations derived in section 2.3, insertion loss, leakage and isolation numbers get better as the R_{ON} of transistor improves. More advanced CMOS technologies with improved f_t numbers directly enhances the performance of the switch by providing more g_m at a constant transistor size (and hence a constant capacitive loading). This makes the future deployment of transformer-based shunt switches in mm-wave systems promising. In addition to transmit/receive (T/R) switching, single pole double thru (SPDT) switches have many other useful

Ref.	This Work	[33]	[34]
Process	90nm CMOS	130nm CMOS	90nm CMOS
Frequency(GHz)	50	60	24
Insertion Loss (dB)	3.4	4.5	3.5
Tx-RX isolation (dB)	13.7	24	16
Supply voltage(V)	1	1.2	1.2
IP _{-1dB} (dBm)	14	4.1	28.7
Area (mm ²)	0.004	0.221	0.018

Table 2.1: Comparison to recently published mm-wave T/R switches

applications. Examples include modulators, stepped attenuators, and phase shifters.

Chapter 3

Phased Array Structures

3.1 Fundamentals of phased arrays

A phased antenna array structure is an array of single antenna elements spaced $\lambda/2$ apart to emulate one antenna with a higher effective aperture area (A_{eff}). Since directivity of an antenna is related to its effective aperture area through the equation:

$$D = \frac{A_{eff}}{\pi\lambda^2} \quad (3.1)$$

An antenna array shapes the radiation pattern to have a directional beam toward the desired angle of look and attenuates unwanted signals impinging on the array from other directions (spatial filtering).

To deviate the angle of look away from the broadside angle by θ degrees in a linear array (figure 3.1), the progressive delay of $T = \frac{\lambda \sin(\theta)}{c} [N - 1]$ should be inserted in each antenna path (1..N) to compensate for different times of arrival of signals at each antenna element. Having integrated delay elements to be able to electronically adjust each one provides the capability of steering the beam dynamically which is much beneficial for mobile and multi-user applications.

The phased antenna array technology has been employed in surveillance and weather radars for more than six decades now. However due to stringent requirements and the need for expensive technologies, deployment of phased arrays remained limited to high-end applications. Advances in silicon technologies made inexpensive implementation of relatively small phased antenna arrays (≤ 100 elements) promising. This makes many consumer electronic opportunities of radar, imaging and communication systems operating at mm-wave regime viable in terms of cost, power delivery/consumption and performance. In the next section we investigate the effect of a phased antenna array structure on a mm-wave system quantitatively and show that why at mm-wave regime phased array structures are so popular.

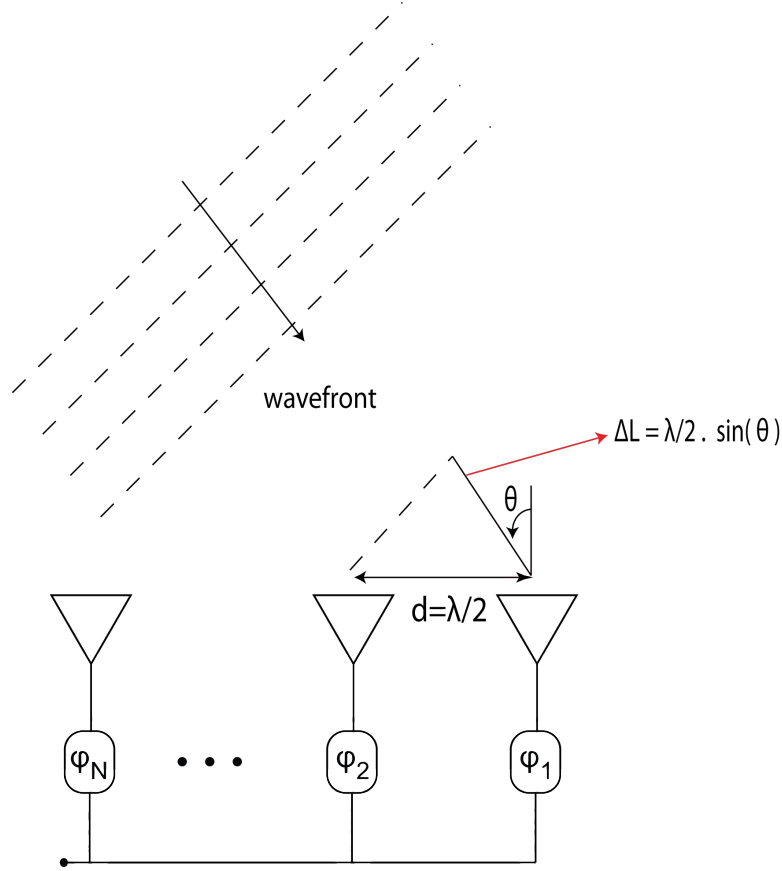


Figure 3.1: The incoming wavefront reaches each antenna element at a different time and corresponding delay element are needed in each path to compensate for that.

3.2 Advantages of phased arrays

Phased array structures help to exploit the power of numbers to create a high performance system out of many lower performance single elements. For example at the TX side, in modern silicon technologies, due to low break-down voltages and consequently lower allowable voltage swings, the power extractable from a device is limited. Decreasing the load impedance seen by the output of the power amplifier and pumping more current into it helps to obtain higher powers at constant voltage swings. However there is a limit in transforming down the antenna impedance to a low level load impedance to be seen by the PA, after which the high impedance transformation ratio implies a high quality factor number for the network ($Q_{network} = \sqrt{\frac{R_{high}}{R_{low}} - 1}$) which in turn results in a high insertion loss for the matching network ($I.L. = 1/(1 + \frac{Q_{Network}}{Q_{component}})$) due to the limited quality factor of on-chip passive components. Moreover having increasingly more figures in parallel to realize a bigger transistor to provide the high current required for that extremely low

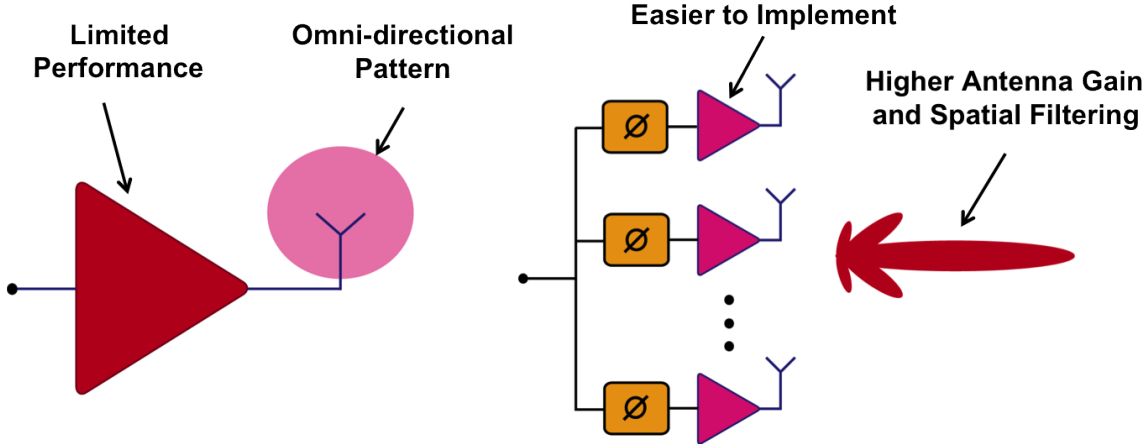


Figure 3.2: The passive gain achieved through array directivity increases the EIRP and relaxes PA design requirements.

load impedance limits the performance of the power transistor at mm-wave frequencies as interconnect inductances come into the picture. Passive power combining networks such as distributed active transformers ([62]), lumped transformer power combining ([57]), and synthesized quarter wavelength power combining([63]) help to add up powers provided by separate power amplifiers and deliver a higher power to the antenna; however they have the same limitation as high transformation ratio matching networks have and due to lossy passive components they consequently will have increasingly higher insertion losses for larger number of combiner ratios.

A phased array structure as depicted in figure 3.2 allows for spatial power combining which means that smaller PAs can work accordingly to add the power up in the space constructively for the desired direction of look and destructively for other directions. The effective isotropic radiated power will be:

$$EIRP = P_{PA} \times \text{Number of Antennas} \times \text{Array Gain} \quad (3.2)$$

In case of constant gain for each path and assuming isotropic antenna elements, the array gain will be equal to the number of antennas and therefore:

$$EIRP = P_{PA} + 20 \log(N) \quad (3.3)$$

For situations where spatial filtering is more important than the maximum achievable EIRP, different gain vectors can be applied to elements of the array to shape the radiation pattern and “main to side lobe ratio”. Figure 3.3 depicts beamforming gains achieved after uniform rectangular and triangular gain windowing vectors are applied to elements of the array. As can be seen the peak beamforming gain for a uniform rectangular gain windowing is $20 \log(N)$, whereas in case of triangular windowing the peak array factor is 6dB lower and a lower resolution (higher half power beam-width, HPBW) is obtained with respect to uniform windowing. But on the other hand side

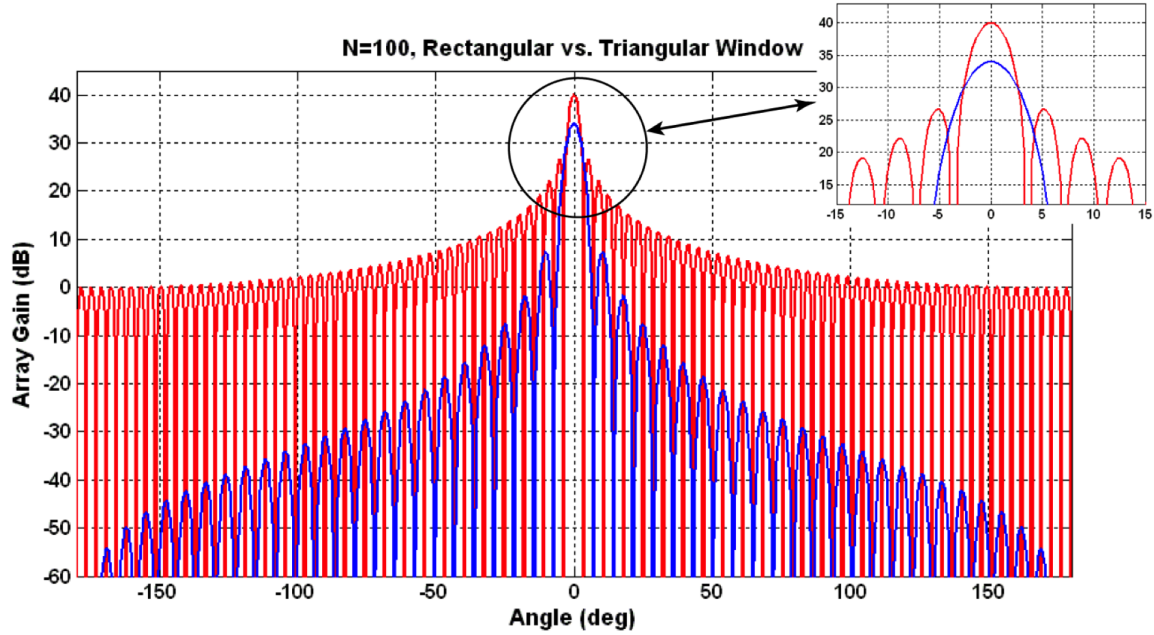


Figure 3.3: A simple triangular gain windowing significantly decreases side lobe levels at the price of less main lobe gain and lower main lobe resolution (wider half-power beam-width).

lobes are lowered significantly in a triangular windowing scheme to reduce blocker levels coming from unwanted directions. In an array, gain and phase vectors can be optimized for number and location of main lobes, main to side lobe ratios and placement of the nulls for more complex systems.

In the receive side, a phased array increases the sensitivity and distinguishes a weak signal buried in noise. This phenomenon is similar to SNR enhancement techniques through averaging that is widely applied in situations dealing with periodic signals accompanied with random noises. As the electromagnetic plane wave signal impinging the array is periodic in space domain with a periodicity of a wavelength ($\lambda = \frac{2\pi}{k}$, k being the wave number); having an N -element array provide N samples of the incoming signal that random noises are also added to them. With the aid of phase shifters or true time delay elements, received signals will be in-phase at the input of the combiner and hence will be added constructively at the combiner output and benefit from the array gain, while uncorrelated noises will not get that advantage.

Two types of combiners are demonstrated in figure 3.4. A Wilkinson power combiner which is realized through two quarter wavelength transmission lines and one isolation resistance is depicted in figure 3.4-a. For an $N : 1$ Wilkinson combiner, each individual input experiences a power gain of $G = \frac{1}{N}$ to reach to the output (while other inputs are terminated). This is due to the fact that the input signal sees the output port as well as all the terminated input ports. Therefore for N uncorrelated signals at inputs of the $N : 1$ combiner, the resultant power at the output can be

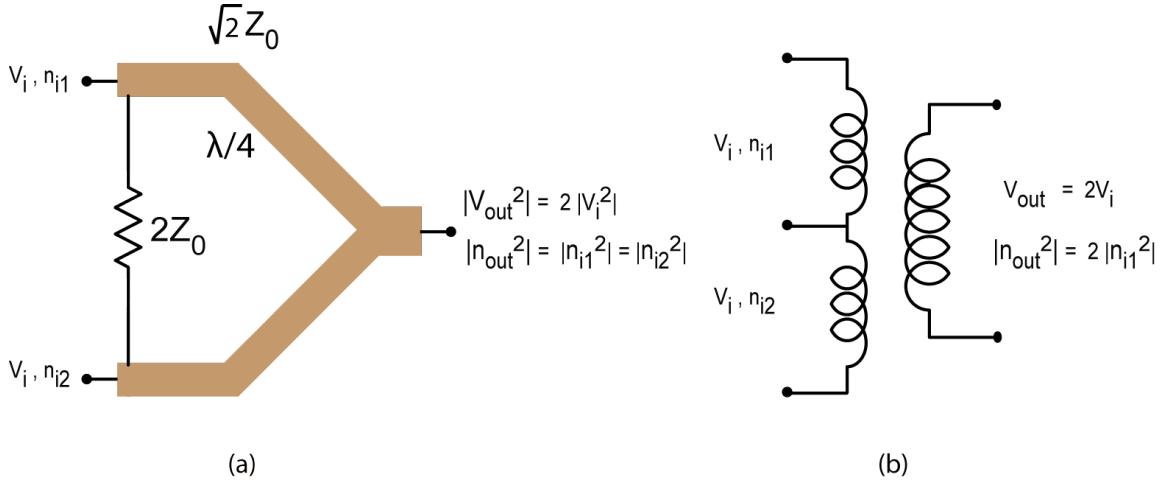


Figure 3.4: In a phased array receiver, signals add up in voltage domain whereas uncorrelated noises add up in power domain. This increases the overall sensitivity of the receiver array.

written as : $|n_{out}^2| = N \cdot \frac{1}{N} |n_i^2| = |n_i^2|$. For fully correlated signals (in-phase signals in the case of a phased array receiver), the Wilkinson combiner is excited in an even-mode and one input signal will not see other input ports and all powers will be directed to the output port yielding at an output power of : $|V_{out}^2| = N \cdot |V_{in}^2|$. Therefore the signal power will be gained up N times more than the noise power and the SNR directly benefits from that.

The magnetic field combining technique where lumped component transformers are exploited to boost up the voltage level is shown in figure 3.4-b. In this structure the secondary loop of the transformer is connected across the output load and N input signals drive each one of the N loops which are placed in a series configuration to form the primary winding of the transformer. This transformer combining technique adds up correlated signals in the voltage domain $V_{out} = N \cdot V_i$ and uncorrelated noises in the power domain ($|n_{out}^2| = N \cdot |n_i^2|$) and improves the SNR by a factor of N .

Regardless of the combining methods described above and their associated gains for signal and noise power, the signal to noise ratio enhances by a factor of N . This improvement in sensitivity can be either thought of as a directivity (passive antenna gain) in the receiver phased array or as an enhancement of the noise factor of the receiver ($F = \frac{SNR_{out}}{SNR_{in}}$). Therefore a phase antenna array at the receive side relaxes the link budget requirement by $10 \log(N)$ where N are the number of antenna elements in the array.

3.3 Different phase shifting schemes

Phase shifting and signal combining can be accomplished at various parts in the transceiver section. It can be done in baseband and digital section or at the LO domain or it can be moved up to the RF part of the chain.

3.3.1 Digital phased arrays

Figure 3.5 depicts the digital domain phase shifting scheme where signals are phase shifted and added up in the digital signal processor(DSP) after they are digitized by N parallel analog to digital converters. This scheme is the most flexible and programmable phase shifting scheme where N digitized samples (sampled in space) are available in the digital domain and different high resolution algorithms (such as MUSIC and ESPRIT [27]) can be applied to have a very high resolution beam for the direction of arrival. On the other hand the down-side of this structure which makes it impractical for mm-wave applications is the fact that no component sharing is done and the full chain from RF to ADCs should be replicated N times. This makes it costly in terms of area and also makes it prohibitively power consuming. Moreover since the spatial filtering gets effective after the signal combining is done, blockers from unwanted directions pass through the system all the way down to the digital section. Hence all building blocks in the chain should have enough linearity and dynamic range to cope with high blocker levels. This in turn makes the design of each building block challenging and more power hungry.

3.3.2 LO domain phase shifting scheme

LO phase shifting technique is demonstrated in figure 3.6 and an example is presented in [70]. In this method, all building blocks after the down converter are shared which helps conserving power consumption as well as lowering the component count with respect to the digital phase shifting scheme. The signal combining is done at the IF frequency band which makes the realization of the combiner easier than the RF combining technique due to much lower operating frequency. Linearity requirements of the IF and baseband building blocks will be relaxed as a result of the attenuation of unwanted signals through spatial filtering of the phased array. The other advantage is that since phase shifters are inserted in LO paths of mixers, their noise and non-linearity will have minimal effects on the overall performance of the chain. But there are still N mixers that should be designed for a good dynamic range to cope with large blockers. The distribution of the LO signal is another challenge because a symmetric LO distribution network is necessary to provide identical phases for all mixers. Due to the loss of transmission lines in the LO distribution network, additional amplifiers might be needed to restore the signal level which adds to the power dissipation of the entire system.

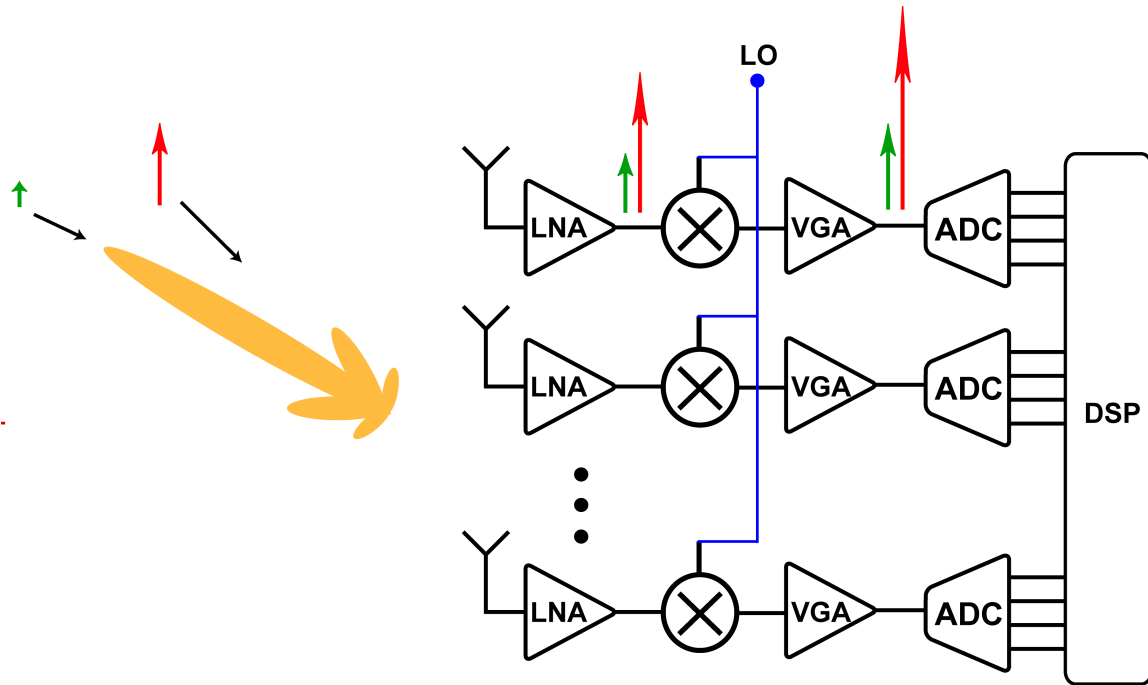


Figure 3.5: Accomplishing the phase shifting and signal combining in the digital domain has the advantage of ultimate flexibility and programability. It results in the highest component count and area consumption as well as higher required dynamic range for the entire chain

3.3.3 RF domain phased arrays

RF phase shifting/combining method is described in figure 3.7. The highest amount of component sharing is achieved through the RF phase shifting scheme and as a result of that, lowest component count and power consumption is achieved. Design requirements of the mixer and proceeding building blocks are relaxed due to the fact that spatial filtering is effective before the down-conversion. However since phase shifters are directly added to the RF signal path, their noise and nonlinearity will affect the overall performance of the chain and make the phase shifter design in RF domain phase shifting more challenging than the other two previously mentioned phase shifting schemes. Since the signal combining happens at mm-wave frequencies, designing a power combining network in order to be implemented in a reasonable footprint and not to degrade the overall bandwidth and gain of the chain is another challenge in this scheme.

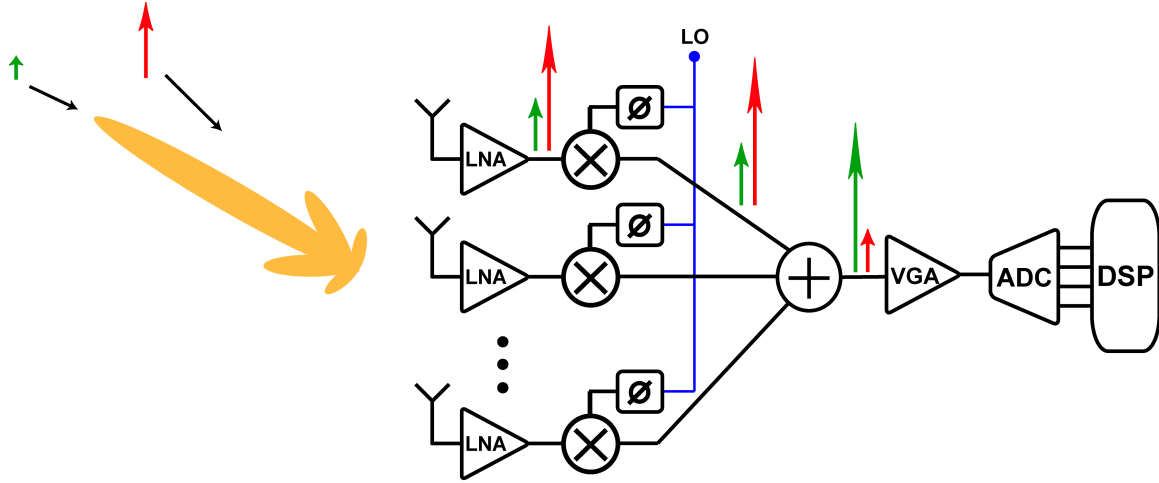


Figure 3.6: A LO phase shifting scheme has less component count with respect to digital phase shifting. Phase shifter non-idealities are not directly in the RF path. A fully symmetric LO distribution network is necessary

3.4 Automotive radar link budget and the required size of the array

To have an estimate on how important of a role phased arrays are playing in a silicon-based mm-wave system, the link budget requirement of a mm-wave (76-77GHz) FMCW automotive long-range radar (LRR) is investigated. Implementing a silicon-based long range automotive radar is challenging due to its relatively long detection range requirement (150m) which requires high effective isotropic radiated power (EIRP) at the transmitter and high sensitivity in the receiver. Also due to demanding requirements for range resolution, good phase noise and chirp linearity performance should be maintained in the signal synthesis part.

The proposed solution of an integrated automotive radar system in silicon technology employs the FMCW structure. The continuous-wave nature of a FMCW radar system eliminates the need for very high voltage circuits that are needed in pulse systems to meet that range requirement. This makes the complete solution realizable in an advanced silicon technology with low break down voltages.

The block diagram of an FMCW radar is depicted in figure 3.8. A continuous wave signal is modulated in frequency to produce a linear chirp which is radiated toward a target through an antenna after power amplification. The echo signal scattered back from the target and received T_{tof} seconds later at the receiver is multiplied with a portion of the transmitter signal to produce a beat signal at a frequency of $f_b = \frac{T_{tof}}{T_{chirp}} B$ where T_{tof} is the time of flight for the signal to make the round-trip from the radar to the target, T_{chirp} is the chirp duration and B is the total available

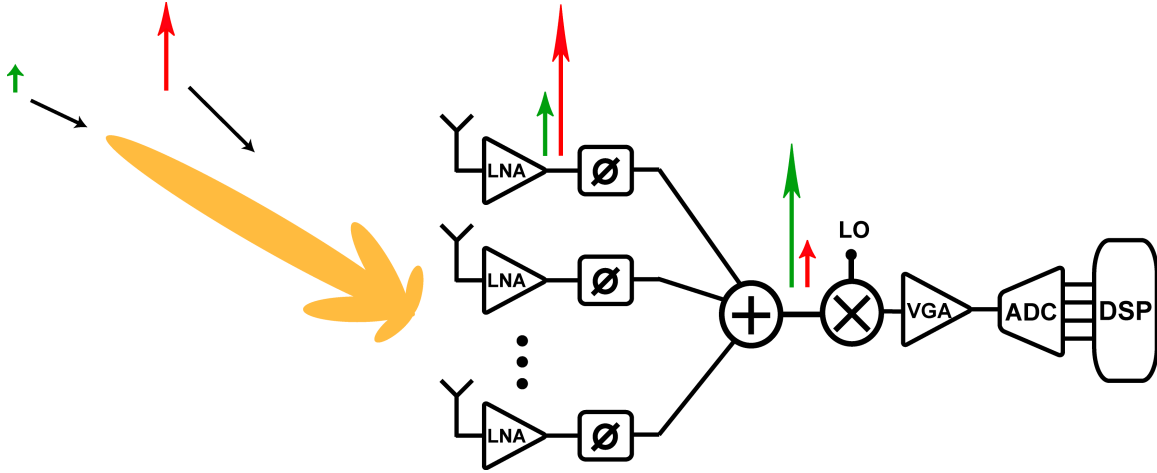


Figure 3.7: The RF phase shifting scheme has the least component count and area/power consumption. Phase shifter noise and non-linearity are directly in the signal path.

bandwidth for the chirp signal. The range resolution for such a radar system is $\delta R = \frac{c}{2B}$ [24]. Signal processing after the mixer is performed at a relatively low frequency band.

Having a single channel for transmit and receive sections will only resolve the range of the target and is not capable of determining the angular position of the object. To fully detect the position of the object at least a two-element receiver is needed. Having more number of antennas on the receive side increases the directivity of the antenna array pattern and hence the angular resolution. It also increases the sensitivity of the receiver and weaker incoming signals will become detectable. Moreover due to the antenna pattern shape a spatial filtering is performed and interference signals bouncing back from clutters and road surface in directions other than the target direction will be attenuated. Having multiple antennas on the transmit side is also beneficial due to the fact that the burden of providing the total large output power is divided on number of blocks and much higher output power is generated from numerous smaller power amplifiers operating on low break-down voltage conditions. Due to the directivity obtained by employing an antenna array at the transmitter, only certain angles will be illuminated at each scan, which mitigates the problem of signals bouncing back from clutters. To have a quantitative analysis and an estimate of the required number of antennas in the array we start with the radar equation.

3.4.1 Radar Equation

The radar equation for calculating maximum detection range is given by [24] and can be written as :

$$R_{max}^4 = P_{TX} \frac{G_{TX} A_{RX} \sigma}{(4\pi)^2 S_{min}} \quad (3.4)$$

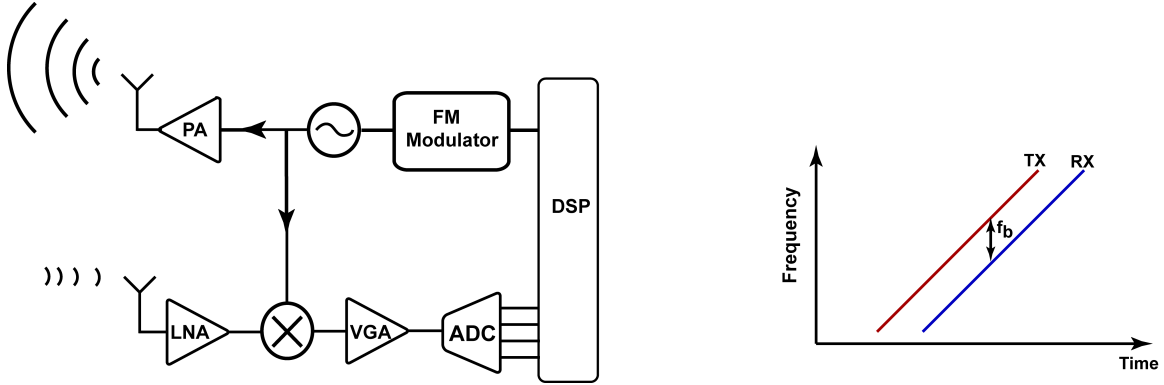


Figure 3.8: Bulding block diagram of an FMCW radar

Where P_{TX} is the total output power, G_{TX} is the gain of the transmit antenna and A_{RX} is the effective aperture area of the receive antenna, σ is the radar cross section(RCS) of the target and S_{min} is the minimum detectable signal at the receiver. From antenna theory, it is known that :

$$G = \frac{4\pi A_e}{\lambda^2} \Rightarrow A_{RX} = G_{RX} \frac{\lambda^2}{4\pi} \quad (3.5)$$

Antenna gain includes both directivity (D) and efficiency(η)of the antenna or the antenna array. Having an array of N_{TX} antenna for transmitting and an array of N_{RX} antenna for receiving the signal results in directivities of N_{TX} and N_{RX} for transmit and receive arrays respectively. Having both arrays implemented on a same substrate results in equal efficiencies for each antenna element in an array. Neglecting array non-idealities such as mutual coupling between antennas and array impedance variation as a function of the scan angle, we can assume that transmit/receive array efficiencies are equal to each other($\eta_{TX} = \eta_{RX} = \eta$). The radar equation can be written in the following form

$$P_{TX} = N_{TX} \cdot P_{out} \Rightarrow R_{max}^4 = P_{out} \frac{N_{TX}^2 N_{RX} \lambda^2 \eta^2 \sigma}{(4\pi)^3 S_{min}} \quad (3.6)$$

Where P_{out} is the output power of each power amplifier in the transmit array. This equation assumes line of sight signal propagation between the radar and targets in the free space. Although the line of sight assumption is a reasonable one for a radar system, absorption of the mm-wave signal traveling through the air (which at these frequencies and at these distances will be few tenths of a dB) and scattering losses (which effectively adds to the path loss) under rain, fog and other weather conditions should be considered and compensated for by having enough margin in the link budget.

Radar Cross Section (σ):

The radar cross section data for different objects of interest is given below. It should be noted that these numbers are approximate and vary greatly based on the angle of scattering, the frequency

(the relative size of the wavelength to the object dimension), and other such factors. At best many of these quantities should be treated as statistical averages:

- Car: $10m^2$
- Van: $30m^2$
- Road surface: $4-10m^2$
- Human: $1m^2$

The goal of an LRR automotive radar system is to detect pedestrians up to 50m and automobiles up to 150m. Due to the 4th power dependance on the range, the worst case scenario is detecting an automobile ($\sigma = 10m^2$) at $R_{max}=150m$.

Minimum Detectable Signal (S_{min}):

The minimum detectable signal can be written based on the total-noise power captured by the antenna, the noise figure of the receiver and the required SNR for distinguishing the signal

$$S_{min} = P_{noise} + NF + SNR \quad (3.7)$$

The available bandwidth for the automotive radar application is 1GHz (76-77GHz). Assuming that all of this bandwidth is exploited for maximizing the range resolution, the total noise power in a 50Ω environment will be

$$P_{noise} = -174dBm/Hz + 10\log(1GHz) = -84dBm \quad (3.8)$$

An N-element receive array enhances the sensitivity by $10\log(N)$. This can be assumed as a $10dB$ improvement in the NF or a $10dB$ enhancement in the array gain. Since the receive array gain is already taken into account in the radar equation, the single receiver noise figure should be considered in the equation 3.7. In state of the art silicon technologies at these frequencies the NF can be as good as $8dB$. Assuming minimum SNR of $6dB$ for detecting the signal, the minimum detectable signal comes out to be $-70dBm$. The averaging that is usually done in radar systems through integration helps to detect weaker signals.

In a FMCW radar system that integration factor can thought of as the RF to IF bandwidth ratio. There are multiple factors in determining the IF bandwidth such as maximum unambiguous rate, refresh rate, number of angular scans and the slope of the chirp. An RF to IF ratio of 100 (10MHz IF bandwidth) results in a minimum detectable signal of $-90dBm$.

3.4.2 Array size

Since in an FMCW radar the modulation is in the frequency of the signal and not in the amplitude domain, a nonlinear PA operating at the saturated output power suffices. State of the art silicon technology is capable of delivering a saturated output power as high as $+20dBm$. Assuming perfect antennas with 100% efficiencies and taking into account that the wavelength at $77GHz$ is $\lambda = 3.9mm$, the radar equation reduces to :

$$R_{max}^4 = P_{out} \frac{N_{TX}^2 N_{RX} \lambda^2 \eta^2 \sigma}{(4\pi)^3 S_{min}} \Rightarrow N_{TX}^2 N_{RX} = 30000 \quad (3.9)$$

If the same array is used for both receive and transmit sections via a T/R switch to lower the package size, then the number of antenna elements in such an array would be :

$$N_{TX} = N_{RX} = N = \sqrt[3]{30000} \sim 31 \quad (3.10)$$

It should be noted that perfect on-package antennas with 100% efficiencies were assumed to obtain the above mentioned number. Antenna non-idealities due to ohmic and dielectric losses of the on-package metallization will require more antenna elements to compensate for that. Significantly lower efficiency numbers of on-chip antennas (10 – 20%) make the system on chip (SOC) solution of a long range automotive radar inferior to the on-package (SOP) solution in terms of power consumption, die area and its associated cost.

In following sections it will be shown that for a large phased array structure, if the application deals with wideband signals, the array fails to track the target accurately as the angle of incidence gets closer to the end-fire angle. However since the FMCW radar system is a narrowband structure and moreover the field of view for a long range radar is only $20^\circ (\pm 10^\circ)$, that will not be an issue for it. On the other hand larger arrays for FMCW radar structures can benefit from advantages of higher number of antenna elements in phased array structures such as their insensitivity to the phase shift accuracy. This provides the possibility of using digital phase shifters with less number of bits and more relaxed quantization errors as the size of the array increases[28].

3.5 Phase shifters or delay elements

As described previously in section 3.1, beam steering capability is obtained as variable delay elements are inserted in individual antenna paths to compensate for each element's received signal path length difference in the space. However implementing integrated variable delay elements with both high delay resolution and high delay variability (maximum to minimum delay ratio) is challenging. In narrowband systems, a true time delay element (linear phase shift response) can be approximated with a constant phase shift (figure 3.9). This approximation fails if the instantaneous bandwidth of the signal gets large, and a timed array should be implemented in this case. It will be shown here that the delay-phase approximation degrades if the array size gets larger or the angle of incidence gets closer to the end-fire angle.

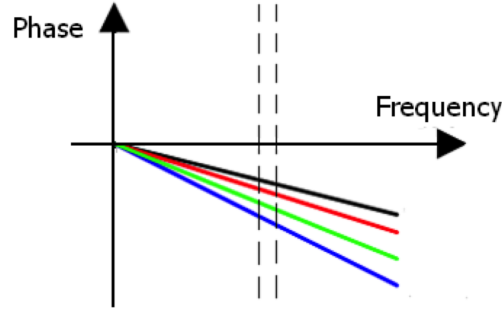


Figure 3.9: In a narrowband system, true time delay elements (linear phase response vs. frequency) can be approximated with phase shifters (constant phase response vs. frequency).

To have a quantitative comparison, we calculate the array gain for both timed and phased arrays as a function of the incident angle. Assuming a sinusoidal signal at an angular frequency of ω , in a linear array configuration (figure 3.1) each antenna ($k=0..N-1$) receives a delayed sample of the signal as :

$$S_k(t, \theta_{in}) = \cos(\omega(t - k \frac{d}{c} \sin \theta_{in} t)) \quad (3.11)$$

To compensate for the path difference in the space a corresponding delay should be applied to each sampled signal, which results in delayed samples that can be written as:

$$S_k(t, \theta_{in}, \Delta\tau) = \cos(\omega(t - k \frac{d}{c} \sin \theta_{in} t - (N - k)\Delta\tau)) \quad (3.12)$$

The array gain is the ratio of the power of the summed signal to the power of each individual signal:

$$Array\ Gain = (\frac{\sum S_k}{S_k})^2 = (\frac{\sin \frac{N(\omega\Delta\tau - \frac{\omega d}{c} \sin \theta_{in})}{2}}{\sin \frac{\omega\Delta\tau - \frac{\omega d}{c} \sin \theta_{in}}{2}})^2 \quad (3.13)$$

For timed arrays the progressive delay element can be set to $\Delta\tau = \frac{d \sin \theta}{c}$ so that the array gain will be equal to N^2 which is independent of the operating frequency and incident angle. However for the case of a phased array, the required time delay will be approximated as the equivalent phase shift at the center frequency:

$$\omega\Delta\tau \sim \omega_0\Delta\tau = \Delta\phi_0 \quad (3.14)$$

This results in an array gain which is a function of both operation frequency and incident angle:

$$Phased\ Array\ Gain(\omega, \theta_{in}) = (\frac{\sin \frac{N(\Delta\phi_0 - \frac{\omega d}{c} \sin \theta_{in})}{2}}{\sin \frac{\Delta\phi_0 - \frac{\omega d}{c} \sin \theta_{in}}{2}})^2 \quad (3.15)$$

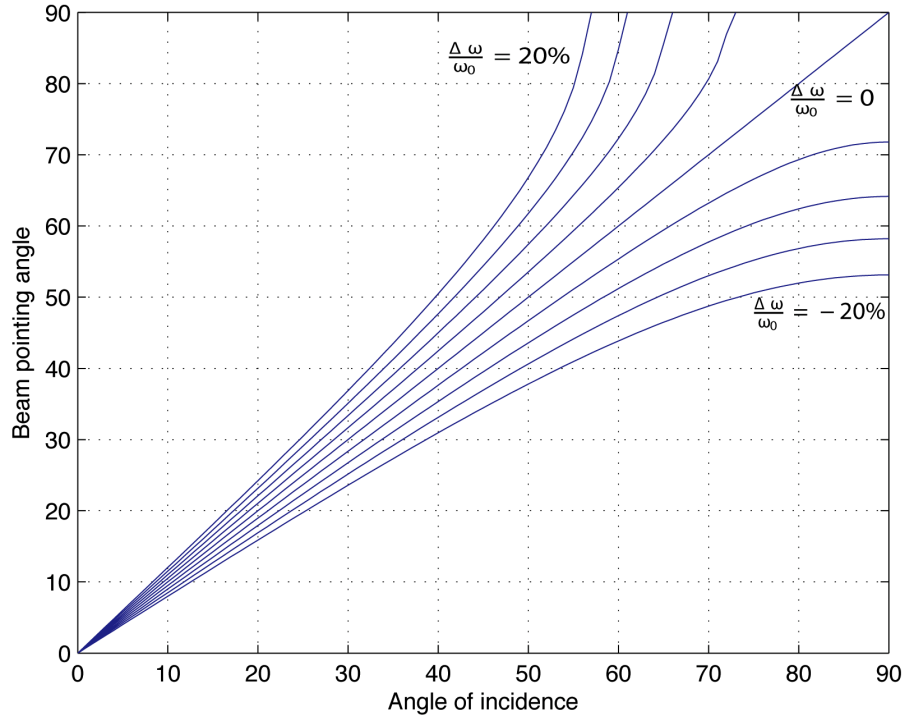


Figure 3.10: The nonlinear relation of equation 3.16 leads to erroneous estimation of direction of arrival for end-fire angles (frequency deviation is swept up to 20% in 5% steps)

Therefore due to approximation of a linear phase response with a constant phase at the center frequency, as the operating frequency deviates from the center frequency, the angle of look which is the angle that the peak array gain is pointing to, deviates from the incident angle by :

$$\theta_{look}(\omega) = \sin^{-1}\left(\frac{\omega}{\omega_0} \sin \theta_{in}\right) \quad (3.16)$$

Since equation 3.16 is a nonlinear equation, the deviation of angle of look from incident angle becomes larger as the incident angle approaches the end-fire angle. The situation is demonstrated in figure 3.10 for different signal bandwidths.

The situation exacerbates for larger arrays, as higher angular resolution is achieved for larger arrays, deviation of angle of look from the incident angle at band edges can be detrimental to the operation of the array (figure 3.11). From antenna theory, half power beam-width (HPBW) of an array is calculated as $\frac{\lambda}{D}$ where D is the aperture width. For an N element linear array with a nominal inter-element spacing of $\frac{\lambda}{2}$ the half power beam-width will be almost inversely proportional to the

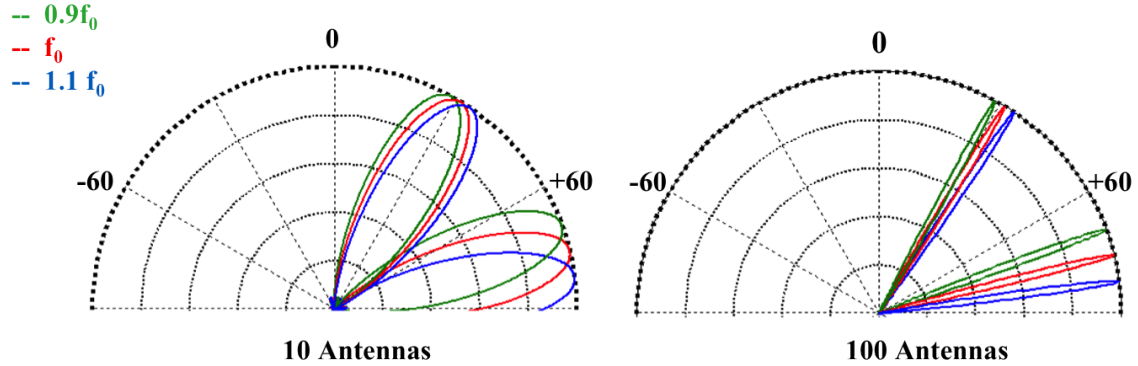


Figure 3.11: As the array gets larger, the HPBW shrinks and the phase shift approximation of delay elements fails for incident angles closer to the end-fire angle even for narrowband signals

number of antenna elements in the array:

$$H.P.B.W. = \frac{\lambda}{D} = \frac{\lambda}{(N-1)\frac{\lambda}{2}} = \frac{2}{N-1} \quad (3.17)$$

In conclusion, the choice between a timed or phased array depends on the signal bandwidth and number of antennas (the beam-width in each dimension is inversely proportional to the number of antennas in that dimension). For applications such as 60GHz connectivity and automotive radars, with maximum of 100 antenna elements in the array (10 elements in each dimension of a square array) and bandwidths up to 10% of the carrier frequency ($\pm 5\%$), phased array solutions are sufficient whereas for ultra wideband applications such as pulse based imaging systems, a timed array system should be investigated

Chapter 4

True Time Delay Elements

When implementing a phased array structure for ultra wideband applications such as impulse based systems for high speed communication or mm-wave imaging applications, true time delay elements instead of their substitute “phase shifters” are preferred. As described in chapter 3, section 3.5, for large scale phased antenna arrays with a sizable field of view where incident angles can be all the way up to end-fire angles, approximating true time delay elements with phase shifters degrades the accuracy of array pointing beam in calculating the direction of arrival and hence should be avoided. In this chapter various techniques for implementing integrated tunable delay elements are investigated and a new “inductance multiplication technique” is introduced which allows for miniature implementation of synthesized delay lines with a high delay tunability ($\frac{t_{D,high}}{t_{D,low}}$ ratio). Two implementations of the inductance multiplication technique, one in the passive mode where series pass transistor switches are employed, and the other one in the active mode where CML like current-mode switching is exploited will be demonstrated.

4.1 Tunable delay structures

The traditional microwave method of implementing a variable delay element is through a switched transmission line approach depicted in figure 4.4-a. Depending on the switching state, the input signal will take the shorter or the longer path to reach to the output and therefore an adjustable delay can be inserted in the signal path. On-chip transmission lines are highly accurate, completely linear and very broadband. The bandwidth of a switched transmission line structure is limited by series switches and their well known loss / operation-frequency trade-off described in chapter 2. Despite the advantages of this structure, it has a major issue of having a large footprint as a result of using transmission lines which are still too bulky to be implemented on-chip at mm-wave frequencies, if a maximum delay as large as half the period of the signal ($\frac{T}{2} = \frac{\lambda}{2c} \sim 16ps @ 60GHz$) is going to be realized.

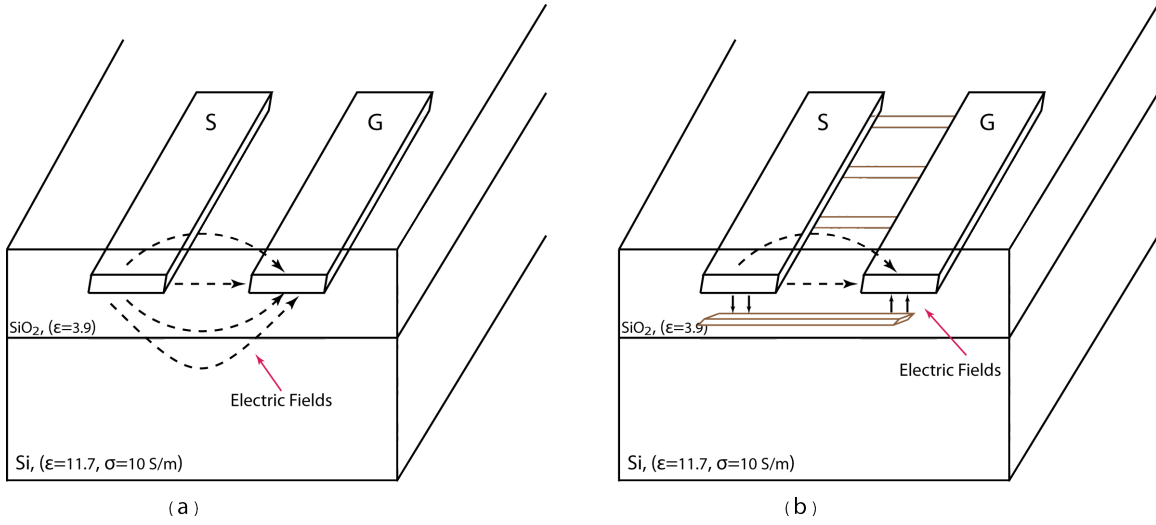


Figure 4.1: Electric field pattern in a normal (a) and a slow wave (b) transmission line.

4.1.1 Slow wave transmission lines

Using slow wave structures is one technique to shorten the length of a transmission line. In slow wave transmission lines, thin metal filaments are placed underneath signal and ground lines as depicted in figure 4.1. Compared to a typical transmission line, employing filaments results in more confined electric fields and makes them traverse a shorter path in the dielectric medium underneath the line. Since capacitance is calculated as $C' = \frac{\epsilon\epsilon_0 A}{d}$, a smaller d leads to a higher per unit length capacitance of the line which in turn results in a lower phase velocity of the line ($v_p = \frac{c}{\omega\sqrt{LC}}$, $c = 3 \times 10^8 \text{ m/s}$). Having lower phase velocity gives the name of slow-wave to these structures. A larger phase shift ($\phi = \omega\sqrt{LC}$) or equivalently a higher delay ($t_D = \sqrt{LC}$) is achieved via slow-wave lines with respect to the same length of a normal transmission line which is beneficial. As depicted in figure 4.1, filaments prevent electric fields from penetrating into the lossy substrate as well and as a result of that they decrease the attenuation constant ($\alpha = \frac{\sqrt{LC}}{2}(\frac{R}{L} + \frac{G}{C})$) of the line and enhance the overall resonance quality factor ($Q = \frac{\beta}{2\alpha}$) as plotted in figure 4.2-d.

Increasing the width of filaments underneath the transmission line increases eddy current loops induced in them which have losses associated with them and decrease the quality factor. At the limiting case where a solid ground plane is placed underneath ground and signal lines, substrate will be shielded completely, however there will be a path for the return current right underneath the signal line which decreases the inductance of the line considerably. This decrement in the line inductance offsets the effect of capacitance increment in obtaining more delay/phase-shift per unit length. Furthermore since lower metal layers have more conductive loss than the top metal layer, having a close by return path in a more lossy metal layer decreases the quality factor of the transmission line. As depicted in figure 4.2, slow wave transmission lines enhance the

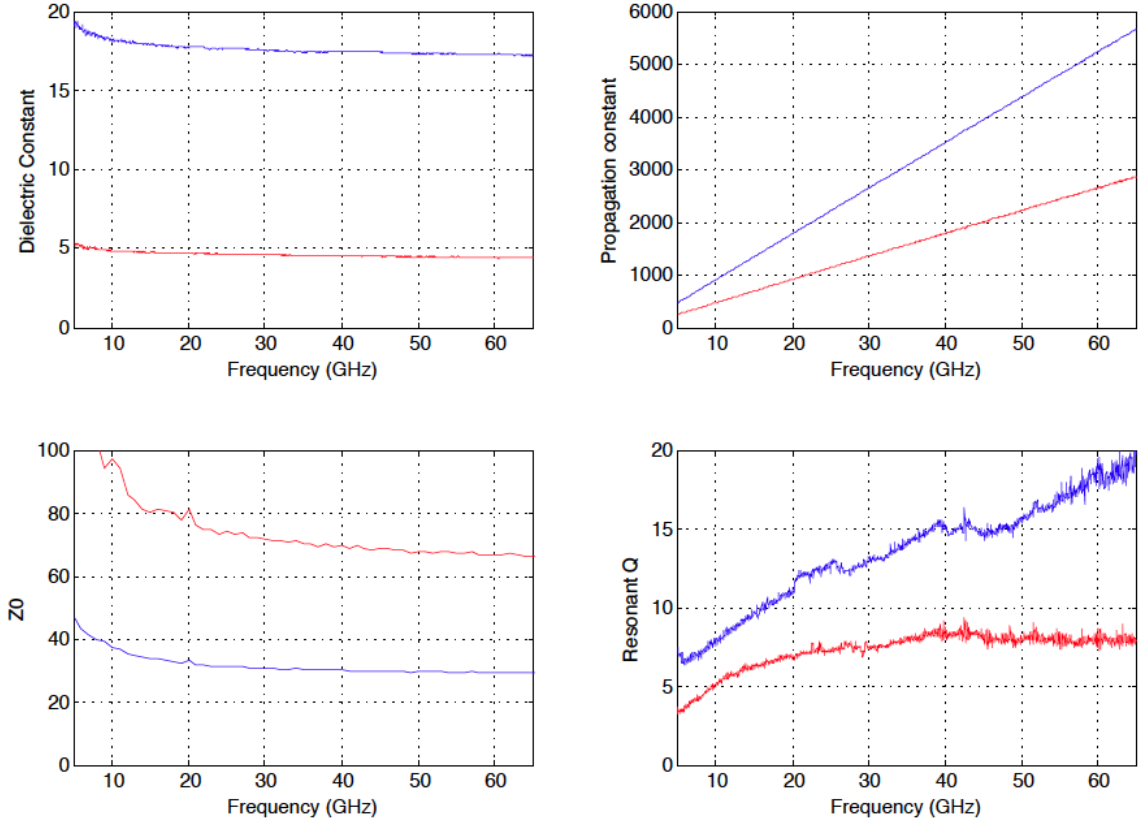


Figure 4.2: Measurement comparison of a slow wave structure (blue) with a normal transmission (red) line in terms of the dielectric constant(ϵ), propagation constant(β), characteristic impedance(Z_0), and resonant quality factor($Q_{res} = \frac{\beta}{2\alpha}$).

propagation constant (β) by increasing the effective permittivity of the dielectric (ϵ) and maintaining the same permeability (μ). This comes at the price of lower characteristic impedance ($Z_0 = \sqrt{\frac{L}{C}}$) which demands for more power being dissipated in preceding building blocks to drive this lower impedance delay stage.

A prototype slow wave transmission line was fabricated in a 90nm CMOS process. Signal and ground lines were implemented on the thick top metal layer and filaments were placed from M_1 up to one metal layer lower than the top metal. Measurement results of this slow wave transmission line and their comparison with the similar transmission line without filaments underneath are depicted in figure 4.2. This slow wave structure enhances the effective permittivity of the dielectric by a factor of ~ 4 which doubles the propagation constant(β) and the delay per unit length of such a structure. On the other hand filaments decrease the characteristic impedance (Z_0) by $\sim 50\%$. The resonant quality factor is also increased considerably(more than twofold at 60GHz) since dense

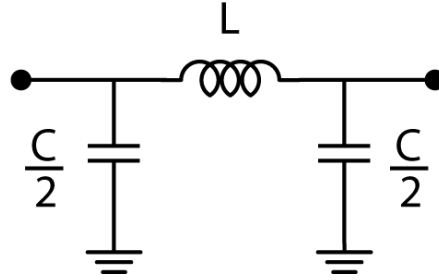


Figure 4.3: A π -section of an artificial transmission line synthesized out of lumped component inductors and capacitors instead of infinitesimal distributed inductance and capacitance of a classic transmission line

filaments made the lossy substrate completely shielded. Half the length of a delay line will be saved by applying this method. However it still occupies a noticeable area of the die and other techniques to decrease the size even further will be investigated next.

4.1.2 Synthesized transmission lines

Artificial transmission lines in which per unit length inductances and capacitances of the line are replaced with lumped component inductors and capacitors are synthesized structures that shrink the size of a transmission line to a great extent. Significant size reduction is attained at the price of limited bandwidth. A π section of an artificial transmission line (figure 4.3) is a second order Butterworth filter with pole frequencies at $\omega = \frac{2}{\sqrt{LC}}$. Each π section of the synthesized line provides a delay of $t_D = \sqrt{LC}$. If the required total delay and the impedance of the line $Z_0 = \sqrt{\frac{L}{C}}$ are known, total needed inductance and capacitance of the line can be calculated and the operating frequency determines how many π sections are needed to be cascaded so that the pole frequency ($\omega = \frac{2}{\sqrt{(L/N)(C/N)}} = \frac{2N}{\sqrt{LC}}$) is sufficiently higher than the frequency at which the circuit is desired to function. Higher required bandwidths necessitate for more sections, and smaller lumped inductors and capacitors in each section, which at limiting cases of very high bandwidths the structure approaches to a classic transmission line.

Delay or phase tunability in such structures come from the fact that variable capacitors (MOS varactors or a bank of switched capacitors) can replace shunt capacitances of the synthesized line. Varying the capacitance of the line changes the delay through the line at the price of altering its characteristic impedance as well. Variable characteristic impedance of the line can violate the return loss requirement as the delay/phase-shift changes from its nominal value, and this is the reason that in [71] the phase shift is only varied up to $\Delta\phi = \frac{\pi}{4}$ of its nominal value. To overcome this issue and mitigate the effect of Z_0 variation, in [38] a variable capacitor is added in series with the inductance of the line as depicted in figure 4.4-c.

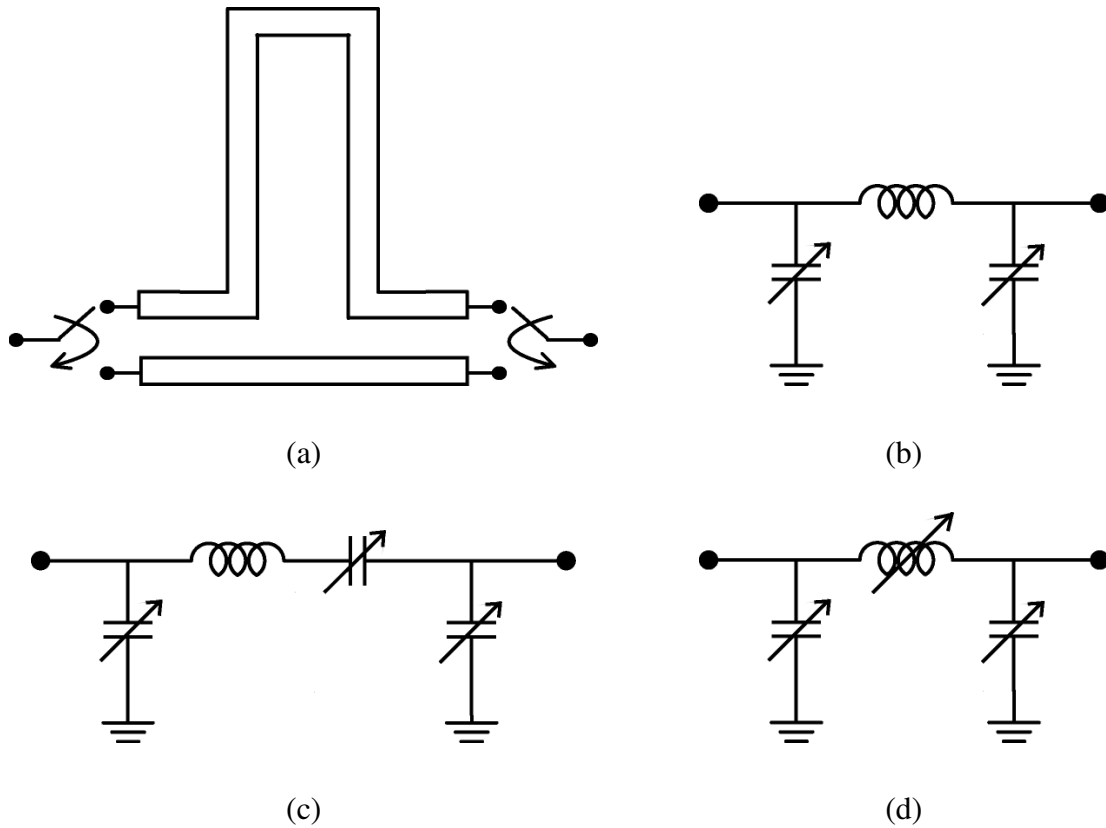


Figure 4.4: (a) Switched transmission lines (large footprint). (b) Varactor loaded artificial transmission line (Z_0 variation). (c) Modifying the effective series reactance by adding a varactor in series with the inductor (narrowband network). (d) Broadband solution for tunable synthesized transmission lines.

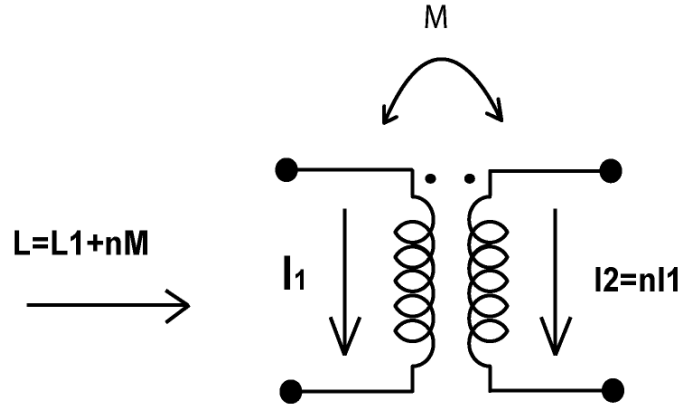


Figure 4.5: Inductance tunability, net magnetic flux crossing a loop is altered via the flux generated by another loop.

As shunt capacitances are increased in the figure 4.4-c to achieve more delay through the structure, the series capacitance value will be decreased, so that the reactance of the series network of L and C will increase and as a result of that there will be a larger effective series inductance in the line. This keeps the inductance to capacitance ratio constant and alleviates the Z_0 variation problem. However this solution is a narrowband technique in which a variable inductance is approximated by a series tank comprised of a fixed inductance and a variable capacitance. Lowering the quality factor of the series tank makes the approximation to be valid over a wider range of frequencies at the price of more insertion loss.

To make an artificial varactor loaded transmission line work with wideband signals, the actual inductance value should be adjusted instead of the effective series reactance. As depicted in figure 4.4-d, a true variable inductor is required in conjunction with variable capacitors to form a synthesized tunable transmission line. In the next section a technique to obtain a tunable inductance is demonstrated.

4.2 Inductance tuning technique

Inductance is determined by the geometry of a closed path of current and the permeability (μ) of the surrounding material. In traditional IC processing, magnetic materials of high permeability are not available and therefore the only way to change the self inductance of a single loop is to change the geometry, which implies that the path of the current or return current flow should be reconfigured. It is not trivial to change the geometry of an inductor electronically and efficiently since switches incur substantial loss.

The other way to change the effective inductance is to change the net magnetic flux passing through a loop via the flux generated by another loop. This could be done via a transformer if the

current passing through the secondary loop is controlled. As depicted in figure 4.5, if the current passing through the secondary is a multiplicative copy of the primary current ($I_2 = n \cdot I_1$), then the effective inductance seen through the primary is $L_{eff} = L_1 + n \cdot M$. Therefore either the secondary current or the mutual inductance should be adjustable in order to have a variable inductance at the primary.

In figure 4.6 several methods to change the magnetic flux of the primary loop are depicted. In figure 4.6(a-c), signals are single-ended, and therefore an input balun is used to convert the signal from single-ended mode to the differential version and to produce a copy of the current that is flowing in the primary. A multiplication of this current will be routed into the secondary to change the net magnetic flux crossing the primary loop.

The structure depicted in figure 4.6-a uses a current amplifier to adjust the current at the secondary. In this structure, the secondary current could be varied over a wide range but it comes at the price of DC power consumption and limited bandwidth due to the fact that the operation frequency of the current multiplier should be reasonably below $f_t/(N + 1)$. Furthermore, if the current amplifier is implemented with reasonable power consumption, its input impedance ($1/g_{m1}$) will be high and it needs a matching network to lower it down to the characteristic impedance of the top path. Otherwise if the impedance looking into the current amplifier is much higher than the impedance of the delay cell (Z_0), the voltage across the secondary loop will appear mostly on the current amplifier side rather than at the input of the delay cell. This results in a significant attenuation for the signal that appears at the output of the delay cell.

Figure 4.6-b demonstrates another structure that modifies the magnetic field induced in the primary by changing the number of turns of the secondary winding. Since $M = k\sqrt{L_1 \cdot L_2}$, changing the inductance of the secondary side modifies the mutual inductance and as a result the effective inductance of the primary loop. Unlike the previous configuration, this structure does not consume any DC power, however a large inductance in the bottom path lowers the bandwidth and the operation frequency of such a structure will be limited by the auxiliary bottom path which is not the main signal path. Furthermore, changing the inductance at the bottom path alters its impedance, which loads the input balun. Variable loading results in variable voltage levels that appears at the input of the delay cell, and this causes the signal to experience variable gain from input to the output depending on the inductance needed for each delay setting.

In the structure shown in figure 4.6-c, the magnitude of the current and the inductance of the secondary loop stay unchanged. However with the aid of a switching network, current will change direction in the secondary and in one switching condition it totally bypasses the secondary and directly goes into the termination resistor. Therefore current gain numbers of $n = (-1, 0, +1)$ are obtained and as a result of that the effective inductance seen at the primary will take on values of $L_1 - M$, L_1 , and $L_1 + M$. If the two loops are identical ($L_1 = L_2 = L$) and mutual inductance is designed to be half the value of the self inductance ($M = L/2$), then for effective inductance values looking into the primary loop a ratio of $L_{max}/L_{min} = (L + M)/(L - M) = 3$ will be achieved. To have a constant Z_0 , the same tuning range should apply to the variable capacitors: $C_{max}/C_{min} = 3$, yielding a delay variation factor of three while maintaining a constant Z_0 , which is considerably higher than the achievable delay variation of a varactor loaded synthesized transmission line.

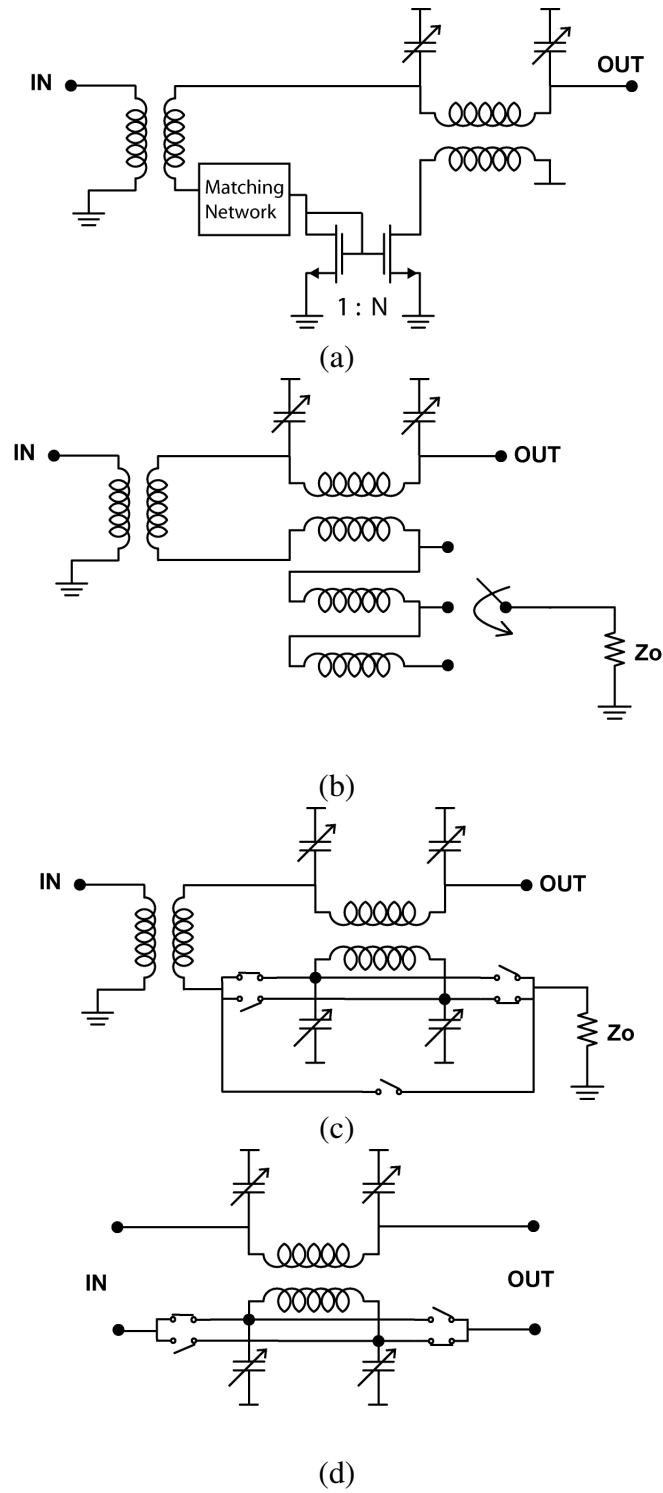


Figure 4.6: Four different ways to realize inductance tuning: (a) Using a current amplifier, (b) varying the mutual inductance, or by rerouting the current in the secondary in a (c) single-ended or (d) differential manner.

The circuit demonstrated in figure 4.6-c has no DC power consumption, and the bottom path does not impose bandwidth limitations on the top (delay) path, and there are no matching requirement in the bottom path. But still an input balun is devised to convert the signal from the single-ended mode to a differential signal to be applied to the core delay cell. This means that under the desired matched condition, half the signal power will be lost. On the other hand, if a differential delay line is adopted, with a configuration illustrated in figure 4.6-d, a fully symmetric structure is realized with the same L_{max}/L_{min} ratio as the previous single-ended version. The effective inductance can take on two values ($L - M$ and $L + M$) in the differential version rather than three values ($L - M, L, L + M$) obtained in the single-ended configuration. Despite the fact that only two discrete delay settings are available via employing the differential configuration (figure 4.6-d), being fully differential and not having the 3-dB inherent loss make this structure the preferred one. In next sections two versions of active and passive mode implementations of this inductance multiplication technique in wideband true time delay elements will be described.

4.3 Passive mode implementation of delay elements employing inductance multiplication technique

Looking at the differential variable delay element depicted in figure 4.6d, the first implementation idea that comes into mind is realizing the switching network via pass transistor switches. In chapter 2 it was described that why series MOS switches are not suitable for mm-wave applications and instead shunt switches along with a transformer were exploited to form a transformer-based shunt TR switch at mm-wave regime. Here the trade-offs for employing series MOS switches in tunable synthesized transmission lines are stated. Since each π -section of the artificial line is a Butterworth filter, for a known Z_0 , it can be written:

$$\omega_c = \frac{2}{\sqrt{L_{max} \cdot C_{max}}}, \quad Z_0 = \sqrt{\frac{L_{max}}{C_{max}}} \implies Z_0 = \frac{L_{max} \cdot \omega_c}{2} \quad (4.1)$$

In inductance multiplication technique, capacitance ratio is the same as inductance ratio and can be written as:

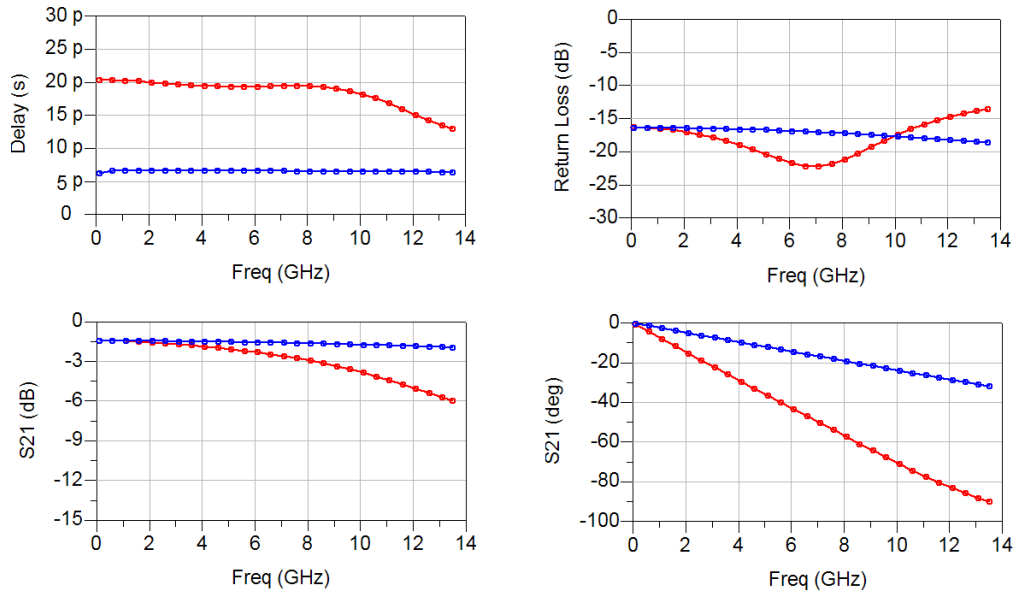
$$C_{max} = \frac{1+k}{1-k} \cdot C_{min}, \quad C_{min} \geq C_{sw} \quad (4.2)$$

Where k is the coupling factor for the delay cell transformer and C_{sw} is parasitic capacitance of the switching transistor. For a given technology, the unity current gain frequency (f_t) is fixed; which is:

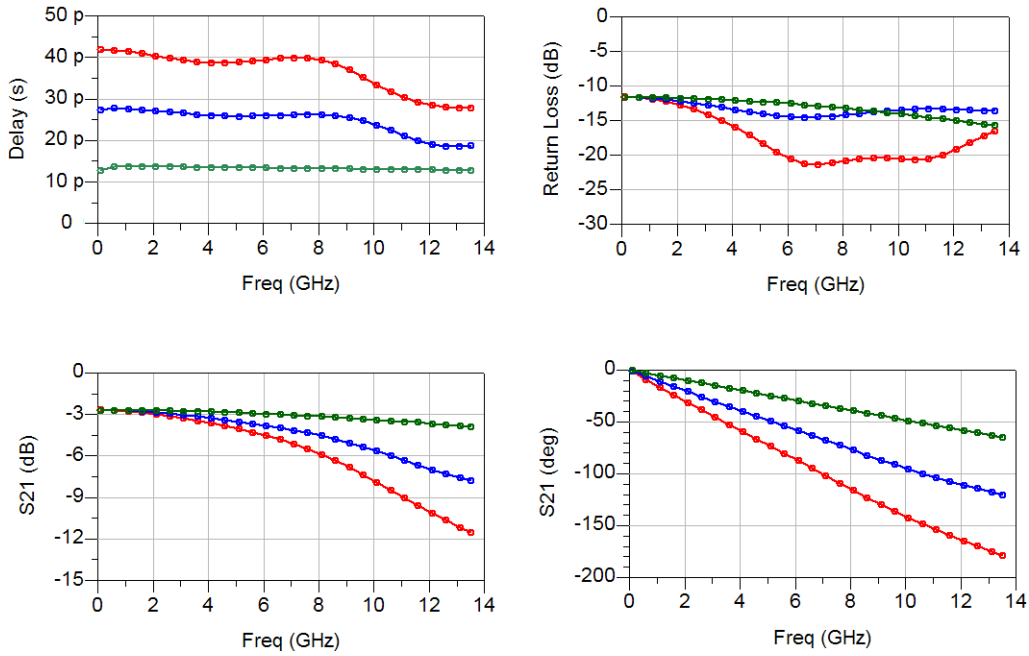
$$f_t = \frac{g_m}{C} \implies C = \frac{g_m}{f_t} = \frac{1}{f_t \cdot R_{ON}} \quad (4.3)$$

Where R_{ON} is the on-resistance of the switching transistor and it is desired to be as low as possible due to the facts that the insertion loss of each π -section can approximately be written as:

$$I.L._{stage} \sim \frac{Z_0}{Z_0 + R_{ON}} \quad (4.4)$$



(a)



(b)

Figure 4.7: Group delay, return loss (input/output) and phase response of a single delay cell and a cascade of two delay cells are demonstrated in (a) and (b) respectively.

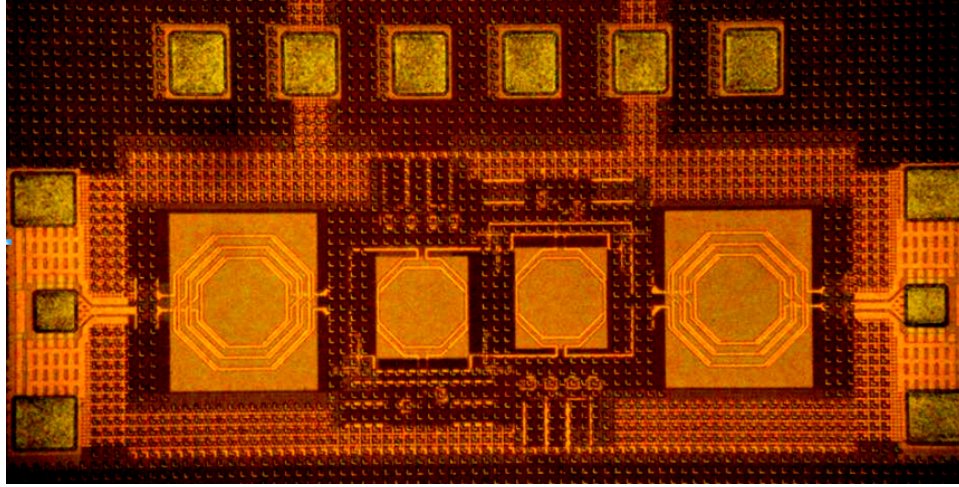


Figure 4.8: Die Microphotograph

Therefore for a low insertion loss, a small R_{ON} is required which demands for a larger transistors with higher parasitic capacitances. This in turn sets the minimum capacitance value of the varactor (equation 4.2) and eventually the cut-off frequency will be limited through equation 4.1.

A prototype was implemented in 90nm digital CMOS process. Based on above mentioned equations, calculations show that with this technology with an $f_t \sim 100GHz$, a cut-off frequency close to $\omega_c = 20GHz$ is achievable. Transformers were implemented as overlaid structures using two top metal layers and they are designed to have a coupling factor of $k = 0.5$. Varactors were realized by a bank of switched capacitors in which interdigitated MOM finger capacitors are switched in and out of the signal path depending on the required delay settings. $W = 50\mu m$ wide NMOS transistors with 7Ω of on-resistance was used as a switch. This transistor's parasitic capacitance ($100fF$) was absorbed in the inductance of the line. The self-inductance of each loop of the balun is $L = 300pH$ and capacitors are chosen to have differential characteristic impedance of 50Ω . The worst case cut-off frequency ($\omega_c = 2/\sqrt{LC}$) is around $18GHz$ that is associated with the high delay mode cut-off.

For demonstration purposes, a unit-cell and two cascaded unit cells were fabricated in 90nm CMOS process (figure 4.8). For the cascaded version, the first switching network was placed in the bottom path and the second one in the top path to maintain the maximum symmetry. To be able to do single-ended measurements, two baluns were added at the input and output of the structure and their effects were de-embedded after the measurement.

Measurement results of fabricated delay lines are shown in figure 4.7. In a single delay cell structure, by moving from a low delay mode to a high delay mode, T_D increases by a factor of 3, or from $t_{low} = 7ps$ to $t_{high} = 20ps$. For the cascaded version, delay values of $14ps$, $27ps$ and $40ps$ are obtained for low-low, low-high and high-high delay modes. Since both the inductance and the capacitance of the synthesized line change correspondingly, Z_0 stays constant and the structure is

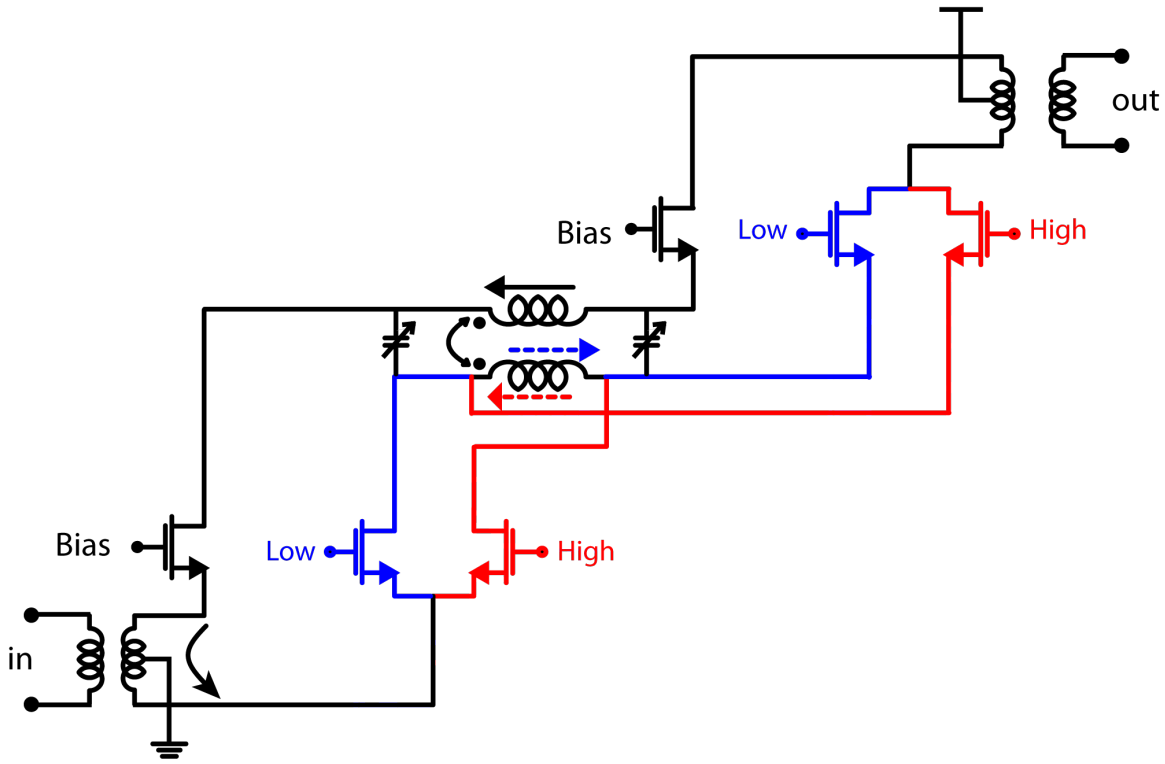


Figure 4.9: A variable delay amplifier implemented in CMOS technology

well-matched while the delay varies significantly. Delay variation for this implementation is less than $\pm 5\%$ up to 8GHz for the highest delay mode. As can be seen from the die microphotograph (figure 4.8), the core die area for two cascaded delay cells is roughly $200\mu\text{m} \times 400\mu\text{m}$ and a noticeable area saving was accomplished as a normal on-chip transmission line should be 6mm long to provide the 40ps delay.

4.4 Active mode implementation of delay elements employing inductance multiplication technique (variable delay amplifiers)

As investigated in section 4.3, an inductance multiplication technique employed in passive delay elements is not suitable for mm-wave operation. Voltage switching (MOS pass transistors) in series with inductors limit the frequency of operation while adding to the loss. Instead a CML-like current switching technique can be used which at the cost of dissipating DC power can operate at a higher frequency than a voltage switching structure. When fully implemented as demonstrated in figure

4.9, it resembles a cascode amplifier with an adjustable transmission line between top and bottom transistors. We call it a variable delay amplifier (VDA) where the series tunable transmission line between top and bottom transistors of the cascode amplifier not only absorb the transistor parasitic capacitances into the transmission line but also it inserts a variable delay in the signal path. Cascaded variable delay amplifiers not only provide the variable delay required for operation of a timed array system, but also provides gain for the RF chain and relaxes the burden on the LNA gain requirement.

There are two ways to implement a variable delay amplifier. One of them is depicted in figure 4.9 and is suitable for low voltage applications where there is not enough headroom to stack more than two transistors on top of each other (the case for highly scaled CMOS technologies). The structure shown in figure 4.9 is a differential cascode amplifier with one side having an additional branch. Depending on the V_{high} and V_{low} voltages, only one of these two branches is operating on the right side of the differential amplifier. Since these two branches pass the current through the bottom loop of the delay transformer in opposite directions, different net effective inductances of $L(1 + k)$ or $L(1 - k)$ is achieved depending on which one of these two branches are selected to be in the signal path. Differential input current is provided by a transformer at the input of the structure. MOS varactors are employed to provide the capacitance required for each inductance state in order to realize an artificial transmission line segment with a characteristic impedance of $Z_0 = \sqrt{\frac{L}{C}}$ between top and bottom transistors. Since bottom transistors of the VDA are excited in a common gate fashion, the input impedance will be $Z_{in,diff} = \frac{2}{g_m}$. 2:1 interstage transformers boost up this relatively low impedance level to a higher value being loaded at drain terminals of the previous stage. The voltage gain associated with each stage would be :

$$\frac{V_{out}}{2} = \frac{V_{in}}{2} \cdot g_{m,stage1} \frac{n^2}{g_{m,stage2}} \cdot \frac{1}{n} \implies A_V = n \cdot \frac{g_{m,stage1}}{g_{m,stage2}} \quad (4.5)$$

To have a higher gain out of each stage we should either have a n:1 transformer with a large value for n, or transistors should be sized and biased in such a way to have a decreasing g_m profile through the stages.

In a CMOS VDA chain design, transistor sizes are determined by their capacitive loading to the core delay cell and they are biased at a point to achieve the maximum f_t . Therefore transconductance values of VDA stages are comparable to each other and not much gain is extractable out of g_m ratios. As a result of that, voltage gain of each stage of VDA would be set by the transformation ratio (n). Having “n” larger than 2, even for very small loops results in a poor self resonance frequency for the transformer which is not suitable for a 60GHz design. Therefore 2:1 transformers were employed in the VDA chain to couple the stages. These interstage transformers were implemented in a lateral fashion on the top metal layer.

Although a voltage gain of 2 (6dB) out of each VDA stage sounds more than enough, imperfect coupling ($K \sim 0.75$) directly hurts the impedance transformation ratio, and the insertion loss of the transformer due to finite quality factor of the loops ($I.L. \sim 1dB$), and extra poles associated with different nodes of the VDA, make the overall obtainable gain out of a VDA implemented in this fashion small and it can easily turn out to a lossy structure.

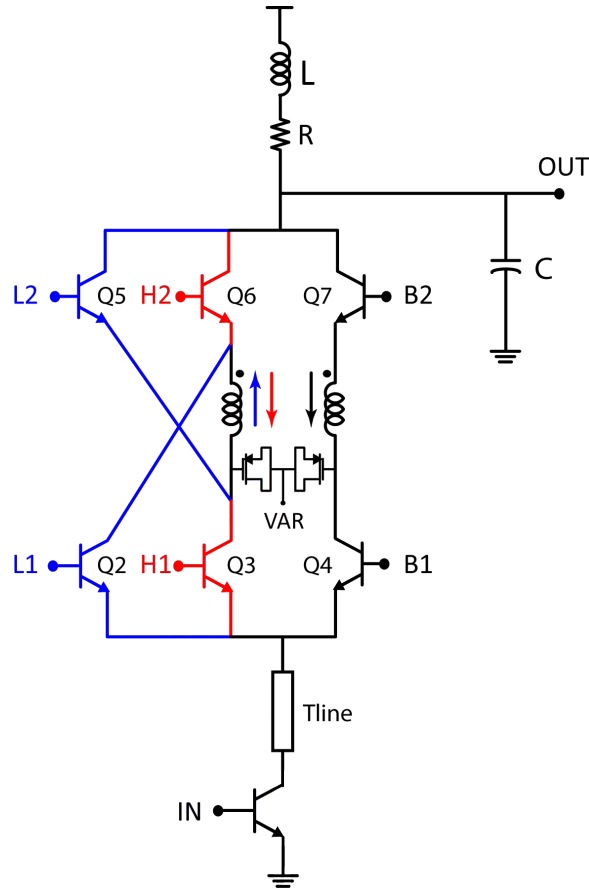


Figure 4.10: A variable delay amplifier in BiCMOS SiGe technology with $V_{DD} = 3.3V$ that can be used to stack more devices

Since the main limitation of the described structure comes from the fact that each stage is loaded with a low impedance loading of the next common gate stage which is proportional to $\frac{1}{g_m}$, a structure as demonstrated in figure 4.10 can help to alleviate this issue. In this structure one more transistor is added at the bottom which increases the total stack count to three transistors and poses a challenge for low voltage design. Since SiGe technologies have higher supply voltages compared to their CMOS counterpart ($V_{SiGe} = 2.2V, V_{CMOS} = 1.2V$), a SiGe technology is chosen to implement this structure. The bottom common emitter transistor acts as a g_m -cell as well as providing a higher loading impedance for the previous stage (buffering the variable delay cell). The current generated by the g_m transistor (Q_1) is divided in half via transistors $Q_2 - Q_4$ (depending on the switching state only one of Q_2 or Q_3 will be on at the same time). Then the two currents that are equal in magnitude and phase are passed through the delay element in such a way that the effective inductance of each loop will be either $L + M$ (if Q_3 is on) or $L - M$ (if

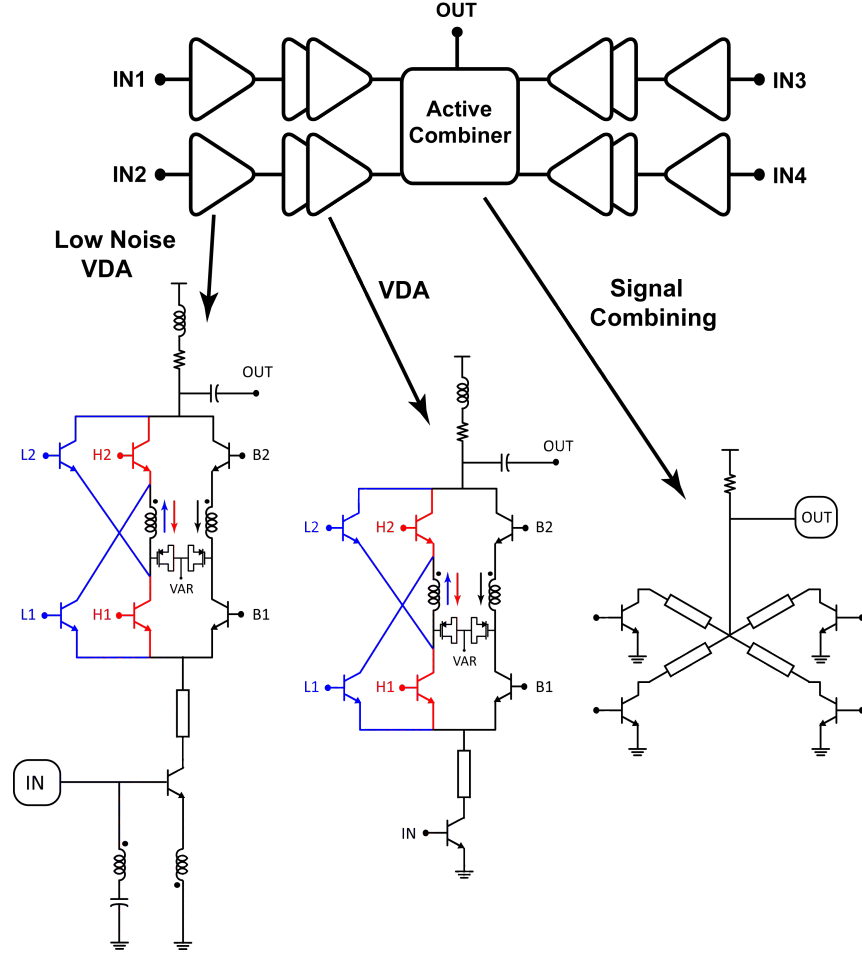
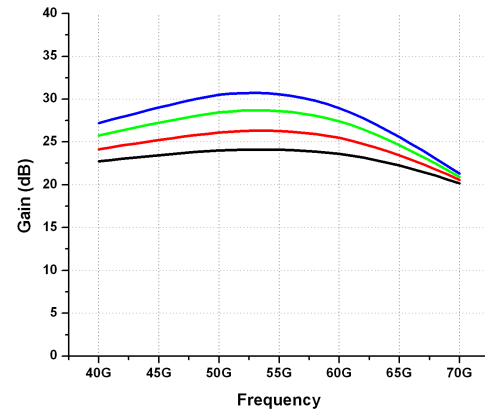


Figure 4.11: A 4-way phased array receiver comprising of 4 low noise VDAs, 8 VDAs and a 4 to 1 combining network implemented in 130nm SiGe technology

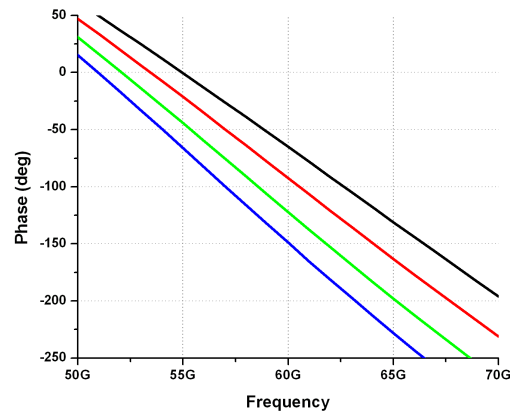
Q_2 is on). The two delayed versions of the input current will be collected by transistors $Q_5 - Q_7$ and then passed through the shunt peaking load. The shunt peaking load was chosen for its higher bandwidth characteristic suitable for wideband timed array systems. Cascaded VDA stages are coupled through MIM capacitors and the gain through each stage of VDA will be [17]:

$$|A_V| = g_m \cdot R \cdot \sqrt{\frac{(\omega L/R)^2 + 1}{(1 - \omega^2 LC)^2 + (\omega RC)^2}} \quad (4.6)$$

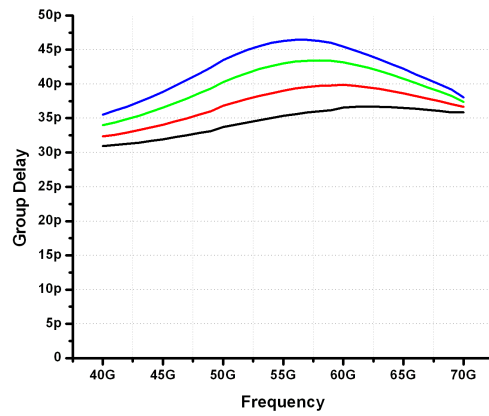
Where L and R are inductance and resistance of the shunt peaking load while C being the capacitive loading of the proceeding stage seen by the shunt peaking network. A shunt peaking factor of $m = \frac{RC}{L/R} = 2$ was chosen for the design which guarantees that at the frequency $\omega = \frac{1}{RC}$ the



(a)



(b)



(c)

Figure 4.12: (a) Gain response of each channel for different delay settings, (b) phase response of each channel for different delay settings (c) group delay $\tau = -\frac{d\phi}{d\omega}$

magnitude of the load seen by transistors is $|Z| = R$. A series transmission line connects the output of the g_m -transistor to the variable delay cell. This transmission line is designed so that its effective inductance absorbs the rather high parasitic capacitance being present at the input of the delay cell.

A 4-way timed array receiver was implemented in 130nm SiGe BiCMOS technology (figure 4.11). Each channel is a cascade of three VDAs. Since the first VDA interfaces the RF input (Antenna), it should be designed for a good matching and noise figure. To provide the match a transformer feedback was implemented at the input as shown in figure 4.11. This transformer feedback tunes out parasitic capacitances associated with the pad and input transistor and provides a match through a series-shunt feedback network.

For combining the 4 in-phase signals at the output of each channel, an active combiner as shown in figure 4.11 is exploited. Four identical L-match networks (each network being comprised of the shunt output capacitance of the transistor and a series inductance realized through a short section of a microstrip transmission line) transform down the output impedance of the transistors in the active combiner and when parallelized with each other a good match to the Z_0 loaded at the output will be provided.

Simulation results are demonstrated in figure 4.12. For different settings gain and phase response are plotted in figure 4.12-a,b respectively. Group delay ($\tau = -\frac{d\phi}{d\omega}$) is plotted in figure 4.12-c.

4.5 Conclusion

An inductance multiplication technique for making broadband and variable passive delay elements is described. By varying both the inductance and capacitance of a synthesized transmission line, the delay value can be altered significantly while keeping the Z_0 constant. Two different versions of this inductance tuning technique were implemented. In one, employing series pass transistor switching networks, a passive structure was realized in 90nm digital CMOS process. Delay values ranging from 14ps – 40ps were obtained from DC to 8GHz while maintaining matched condition over the bandwidth. Due to limited performance of MOS transistors acting as mm-wave series switches in the 90nm technology, the performance of such structures were limited and operating at mm-wave frequencies was out of reach. In the second active version, CML-like current switches were employed to extend the frequency of operation at the cost of burning DC power. When delay cells were fully embedded in CML switches and one g_m transistor was added, an amplifier structure with adjustable group delay was constructed. Nonetheless including variable delay structures in an antenna array is still costly in terms of die area, power dissipation, and functionality risks. As in the case of most mm-wave consumer electronic applications where phase shifters can do the task of beamforming, they should be utilized instead of delay cells as variable delay structures are an too costly. The next chapter discusses different phase shifting structures and an active RF-path phase shifter will be designed and implemented.

Chapter 5

Phase Shifting Structures

For phased antenna array systems that the phase shifting capability (instead of true time delay elements) is sufficient to steer the array pointing beam, there are multiple techniques to implement the phase shifting structure. Phase shifters can be passive building blocks in which the variability comes from varactors or MOS transistors functioning as switches. Phase shifters can also be implemented as active structures with some sort of amplification involved. In next sections we investigate both of these two types of phase shifting structures and demonstrate a mm-wave prototype phase shifter which is designed in a standard 40nm CMOS technology.

5.1 Passive phase shifters

5.1.1 Varactor-based passive phase shifting structures

Although the reactance of a tunable LC-tank (where fixed capacitors are replaced with varactors) can change from purely inductive ($+90^\circ$) to purely capacitive (-90°) and provide a total of 180° phase shift, but the magnitude of the variable tank impedance varies as the phase shift is changing. Therefore it is not acting as an all-pass filter with constant gain over the frequency band of interest and hence would not be considered as a phase shifter on its own (figure 5.1). However due to the fact that for a high-Q tank the impedance (regardless of its magnitude) is imaginary (either a capacitor or an inductor), the reflection coefficient when such an impedance is loaded on a transmission line has a magnitude of unity and a phase which depends on the value of loaded variable reactance.

Structures such as quadrature hybrids when loaded with two variable tanks at thru and coupled ports can exploit this fact and provide an output signal at the isolation port which has a constant gain through the structure and a variable phase shift depending on the value of loaded reactance (or on the other hand the offset of operating frequency and the resonant frequency of loading tunable tanks). These structures are called reflective type phase shifters (RTPS) and an example is depicted in figure 5.2.

To have a numerical understanding of the gain of RTPS structures and minimum/maximum

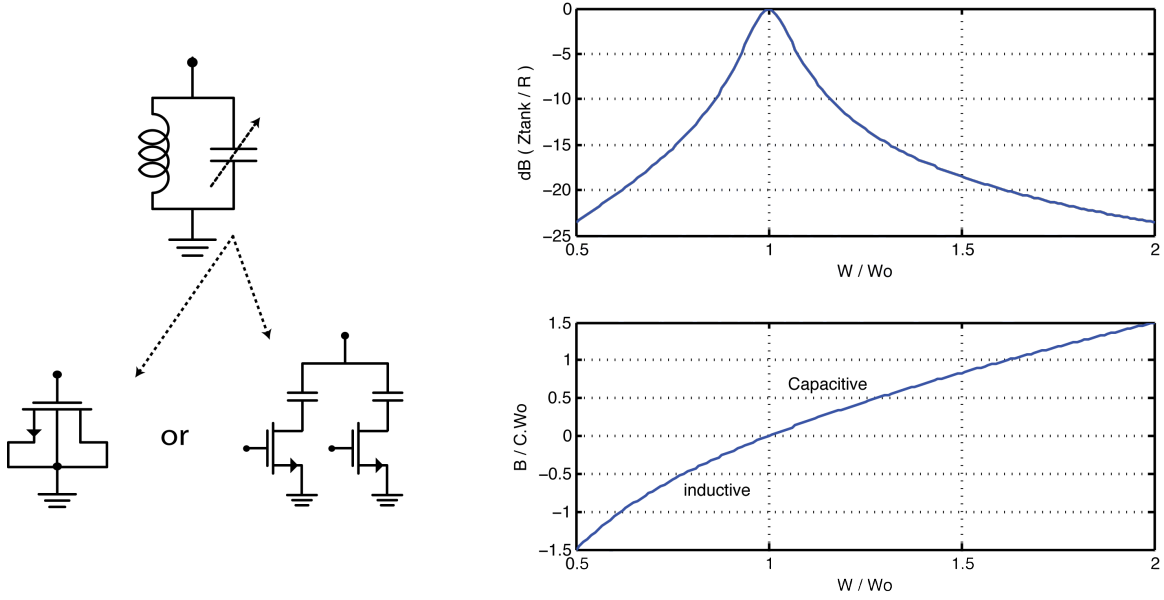


Figure 5.1: The reactance of a variable tank provides 180° phase shift, but due to the variation of its normalized impedance ($\frac{Z}{R_P}$) it can not act as an all-pass filter on its own.

phase shifts extractable through each stage of them, it should be considered that a parallel resonant tank has an admittance of :

$$Y_{tank} = j\omega C + \frac{1}{j\omega L} + \frac{1}{R_P} \quad (5.1)$$

For high-Q tanks at an offset $\Delta\omega$ of the resonant frequency ω_0 , the admittance can be written as:

$$@ \omega = \omega_0 + \Delta\omega \implies Y_{tank} = \frac{1}{R_P} \left(1 + 2jQ_0 \frac{\Delta\omega}{\omega_0} \right) \quad (5.2)$$

Where Q_0 is the quality factor at the center frequency ($Q_0 = \frac{R_P}{L\omega_0} = R_P C \omega_0$). Therefore the impedance of the parallel tank at $\omega = \omega_0 + \Delta\omega$ can be written as:

$$Z_{tank} = \frac{R_P}{1 + 2jQ_0 \frac{\Delta\omega}{\omega_0}} \quad (5.3)$$

The reflection coefficient at quadrature hybrid ports interfacing variable tanks will be :

$$\rho = \frac{Z_{tank} - Z_0}{Z_{tank} + Z_0} = \frac{R_P - Z_0 - 2jQ_0 Z_0 \frac{\Delta\omega}{\omega_0}}{R_P + Z_0 + 2jQ_0 Z_0 \frac{\Delta\omega}{\omega_0}} \quad (5.4)$$

Where Z_0 is the characteristic impedance that hybrid ports are matched to. The magnitude and

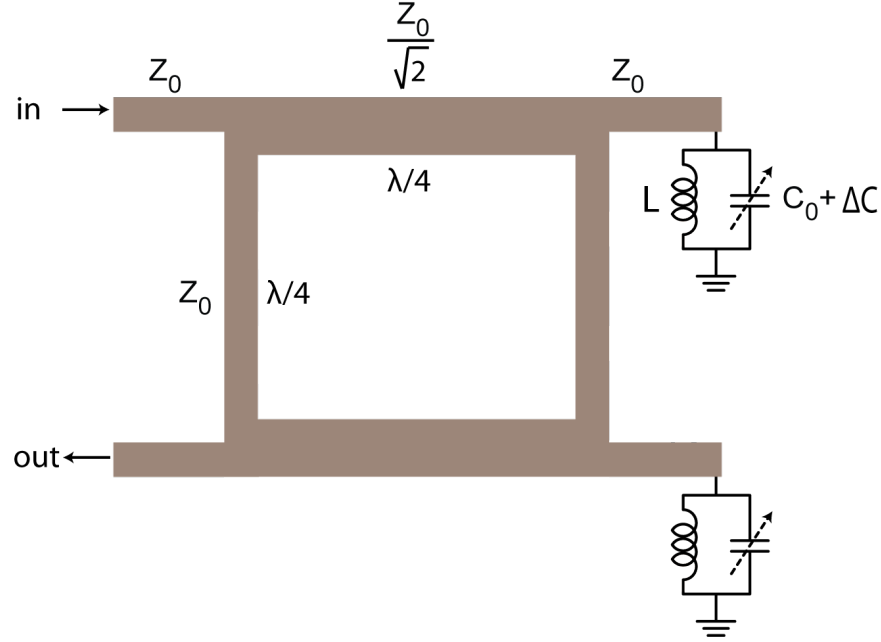


Figure 5.2: A quadrature hybrid structure in conjunction with two variable tanks form a reflective type phase shifting structure (RTPS).

phase of the reflection coefficient can be calculated as following:

$$|\rho| = \sqrt{\frac{(R_P - Z_0)^2 + (2jQ_0Z_0\frac{\Delta\omega}{\omega_0})^2}{(R_P + Z_0)^2 + (2jQ_0Z_0\frac{\Delta\omega}{\omega_0})^2}} \quad (5.5)$$

$$\angle\rho = -\tan^{-1}\left(\frac{2Q_0Z_0\frac{\Delta\omega}{\omega_0}}{R_P - Z_0}\right) - \tan^{-1}\left(\frac{2Q_0Z_0\frac{\Delta\omega}{\omega_0}}{R_P + Z_0}\right) \quad (5.6)$$

For typical realizable characteristic impedances of integrated transmission lines ($Z_0 \leq 100\Omega$) and high-Q tanks that present a parallel impedance (R_P) of on the order of a $K\Omega$, equations 5.5 and 5.6 will reduce to:

$$R_P \gg Z_0 \implies |\rho| = 1, \quad \angle\rho = -2\tan^{-1}\left(\frac{2Q_0Z_0\frac{\Delta\omega}{\omega_0}}{R_P}\right) = -2\tan^{-1}\left(2\frac{Z_0}{Z_{0,tank}}\frac{\Delta\omega}{\omega_0}\right) \quad (5.7)$$

Where the tank characteristic impedance ($Z_{0,tank}$) is defined as:

$$Z_{0,tank} = \frac{R_P}{Q} = L\omega_0 = \sqrt{\frac{L}{C_0}} \quad (5.8)$$

With adjusting the varactor and deviating the tank resonant frequency from its nominal value (ω_0) by the amount of $\Delta\omega$, the signal passing through the structure will be phase shifted (and the amount

of phase shift is $\phi = -2 \tan^{-1}(2 \frac{Z_0}{Z_{L0}} \frac{\Delta\omega}{\omega_0})$ with respect to its nominal value). In theory this phase shift can be as high as $+180^\circ$ and as low as -180° to cover the full 360° of total required phase shift in a one stage RTPS structure. However, in practice $\Delta\omega$ can not be set arbitrarily high or low (those limits are calculated next) and consequently more stages of RTPS structures are needed to cover the full 360° of phase. With two stages of RTPS cascaded, while each one is providing 180° of phase shift, the required tuning capability in the resonant frequency of the tank will be :

$$|2 \frac{Z_0}{Z_{L0}} \cdot \frac{\Delta\omega}{\omega_0}| = 1 \Rightarrow |\Delta\omega| = \frac{\omega_0}{2} \cdot \frac{Z_{0,tank}}{Z_0} \Rightarrow \frac{\omega_{max} - \omega_{min}}{\omega_0} = \frac{Z_{0,tank}}{Z_0} \quad (5.9)$$

Therefore to provide the 180° of phase shift with less frequency adjustability, a smaller inductor and larger capacitances should be used in the tank. Frequency adjustability is limited by the capacitance variation range of varactors ($\frac{C_{max}}{C_{min}}$). Assuming that the varactor is optimally designed ($C_0 = \sqrt{C_{min}C_{max}}$), the maximum achievable frequency tunability can be written as :

$$\frac{\omega_{max} - \omega_{min}}{\omega_0} \leq \frac{1/\sqrt{L \cdot C_{min}} - 1/\sqrt{L \cdot C_{max}}}{1/\sqrt{L \cdot C_0}} = (\frac{C_{max}}{C_{min}})^{\frac{1}{4}} - (\frac{C_{min}}{C_{max}})^{\frac{1}{4}} \quad (5.10)$$

Higher frequency tunability demands for higher C_{max}/C_{min} ration and consequently larger channel lengths in a MOS varactor. This in turn degrades the quality factor of the varactor by increasing the channel resistance in series with the capacitance. Figure 5.3 demonstrates the tradeoff between frequency tunability and varactor quality factor as the channel length of a MOS varactor is increased.

In order to provide the 360° of phase shift in two cascaded stages of RTPS structures, the following requirement should be met in terms of the characteristic impedance of the quadrature hybrid, tank inductance and capacitance ration and capacitance variability of the varactor:

$$\frac{Z_{0,tank}}{Z_0} \leq (\frac{C_{max}}{C_{min}})^{\frac{1}{4}} - (\frac{C_{min}}{C_{max}})^{\frac{1}{4}} \quad (5.11)$$

According to equation 5.11, for situations where the phase shift is not enough and to extend the phase span of the structure, the following solutions can be considered:

1. Larger characteristic impedance (Z_0) for the hybrid. Higher Z_0 values result in thinner signal lines and wider gap spacings between signal and ground lines which exacerbates the series conductive loss and shunt dielectric/substrate loss of the hybrid structure respectively. This makes a practical limit of $Z_0 \sim 100\Omega$ for characteristic impedances of transmission lines designed in modern silicon technologies. Since the signal travels through the hybrid twice, once to reach the reflective load and once to reach to the output port, the hybrid loss will hit the performance twice and degrades the overall gain of the RTPS structure as Z_0 increases.
2. Smaller inductances in the tank. Lowering the inductance of the tank and operating at the same frequency requires a higher value for the varactor in the tank which makes the overall

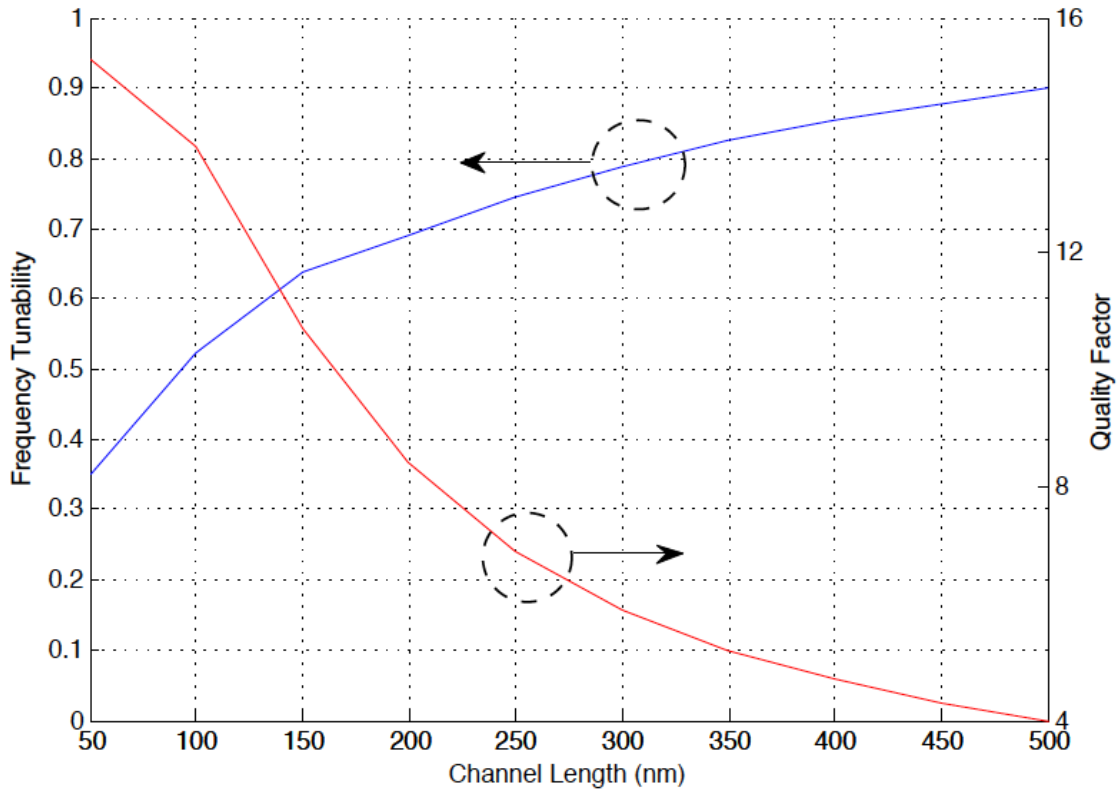


Figure 5.3: Increasing the channel length of a MOS varactor increases $\frac{C_{max}}{C_{min}}$ and hence the frequency tunability $((\frac{C_{max}}{C_{min}})^{\frac{1}{4}} - (\frac{C_{min}}{C_{max}})^{\frac{1}{4}})$ at the price of lowering the varactor quality factor.

quality factor of the tank limited by the Q of varactor rather than the inductor Q. since at mm-wave frequencies varactors have poorer quality factors than inductors, this makes the overall Q of the tank smaller which in turn increases the total loss according to equation 5.5.

3. Larger tunability in the varactor. At mm-wave regime, a bank of switch capacitors acting as digitized varactors are not suitable since very large transistors should be employed as switches in order not to degrade the overall Q. This limits the C_{min} of the structure, moreover due to the fact that at mm-wave frequencies tiny capacitors are needed; the layout of such a structure becomes very sensitive to parasitics. Therefore MOS transistors remain the choice for realizing varactors at mm-wave frequencies. In highly-scaled CMOS technologies, the overlap fringe capacitances of gate to source and drain terminals limit the minimum achievable capacitance and hence the $\frac{C_{max}}{C_{min}}$ ratio. In 40nm CMOS technology, the maximum to minimum capacitance ratio for minimum channel length device is about 1.7 while maintaining a Q of 14 for the varactor. The $\frac{C_{max}}{C_{min}}$ ratio can be enhanced to 4.8 by increasing the

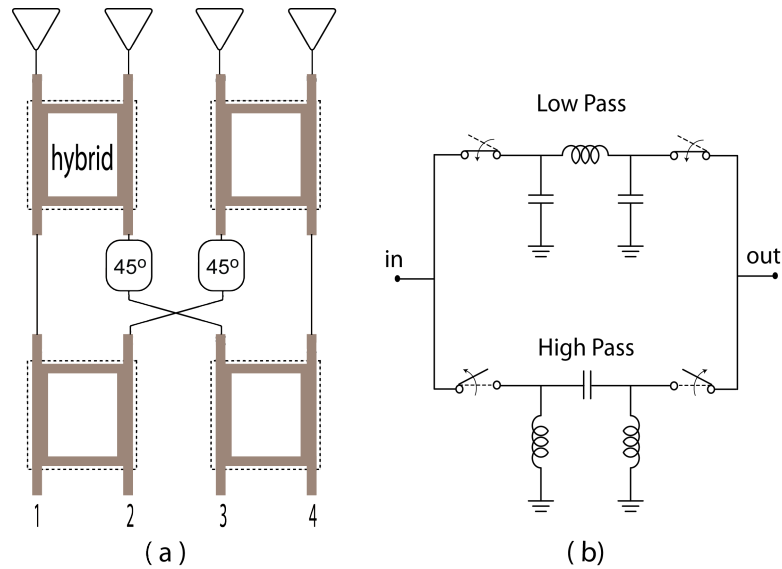


Figure 5.4: Two types of passive phase shifters that do not use varactors to achieve phase variation (Butler mixer and high-pass/low-pass networks).

channel length ten times the minimum channel length at the cost of reducing the Q to about 4 which is rather poor. Therefore similar to the previous two solutions, Expanding the phase shift span comes at the price of increasingly higher losses for the RTPS (figure 5.3).

As mentioned above, increasing the phase span of RTPS structures directly increases the loss of such a structure. The other issue is that, as the phase shift is varied by adjusting the capacitance value, varactor Q varies (figure 5.3). This results in a variable loss of RTPS structures as a function of the phase shift (or the beam pointing angle of the array). To overcome this, a variable gain amplifier (VGA) is required to be included in the RF path to compensate for the gain variation of the RTPS. Although integrated solutions of RTPS structures are quite lossy at mm-wave frequencies, they have good linearities due to their passive nature. The linearity advantage is an attribute which might favor these types of structures for applications that require large dynamic ranges.

5.1.2 Butler mixer and high/low pass filters

There are other passive phase shifting structures that do not use varactors (and hence not affected by their relatively low Q at mm-wave regime). Figure 5.4-a shows a Butler mixer structure that provides different beam pointing directions for each input signal through a network of quadrature hybrids and 45° constant phase shift elements. This structure is suitable for systems that multiple beams are required simultaneously. However using all these passive components increases the footprint of the Butler mixer to be prohibitively large even at mm-wave frequencies and even if

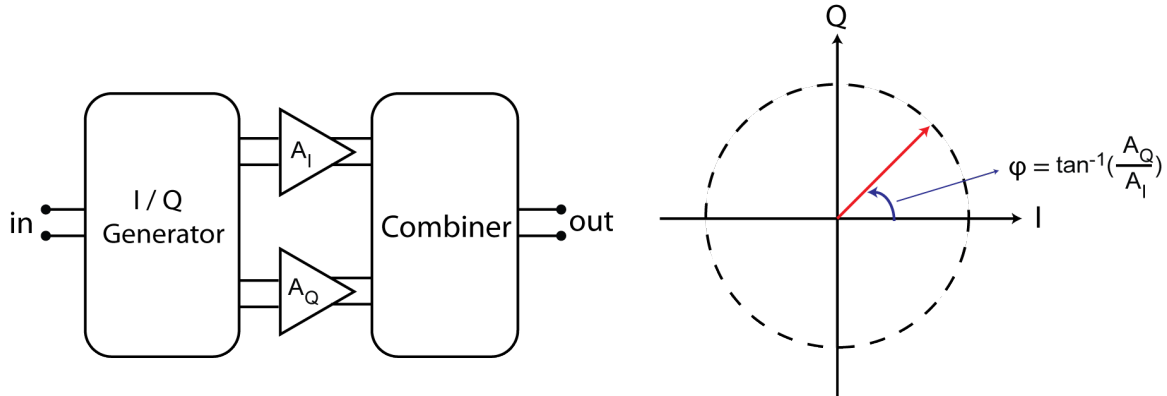


Figure 5.5: Schematic diagram of an I-Q interpolating active phase shifter.

synthesized transmission lines with lumped inductances and capacitances are used to make the structure more compact.

Figure 5.4-b shows a network of high-pass/low-pass filters that exploit MOS transistor switches for obtaining the phase variability in discrete phase steps. Since the gain through the high-pass and low-pass filters are designed to be equal, only the phase shift changes at the output terminal depending whether the high-pass or the low-pass filter is inserted in the signal path. Each high-pass/low-pass section requires three inductors and three capacitors to function properly and the overall structure can get fairly large for even moderate phase resolutions. MOS switch loss at mm-wave frequencies and its loss vs. operating frequency trade-off is another issue for high-pass/low-pass networks.

Among above-mentioned concerns, the main issue with these passive types of phase shifters are their large footprint that make them unsuitable for integrated solutions and hence we seek for other solutions for an RF-path phase shifting structure.

5.2 Active phase shifter : I-Q interpolation

Having in-phase and quadrature-phase components of the signal and adding up a scaled versions of them (figure 5.5) can cover the full 360° of phase span in a two-dimensional I-Q plane. Advantages of such a structure would be :

1. Providing the full 360° of phase shift in one stage.
2. Capability of calibrating gain and phase mismatches (between parallel RF chains in the array) through two I and Q variable gain amplifiers (VGA).
3. Possibility of acting as a VGA in the RF-chain as a side capability.

Depending on the operating frequency, the f_t of the technology and the way signals are divided and then combined in the two I and Q paths, the overall structure might have gain or loss. Operating at frequency limits of the transistor can make the overall structure prohibitively lossy and that is the reason why there has not been such an active phase shifter operating at 60GHz regime in the literature yet. Design equations will appear in later sections and a prototype design in a 40nm CMOS technology ($f_t \sim 200GHz$) that successfully passes the requirements are demonstrated. Due to the active nature of this structure, its down sides with respect to a passive structure can be commented as following:

1. Non-zero power dissipation
2. Excess noise and non-linearity of the phase shifter affects and should be considered in the system design of the RF path.

As depicted in figure 5.5, an I-Q phase interpolator in the RF domain requires:

1. Generating I and Q components of the signal.
2. Variable gain amplification with a capability of switching the polarity of the signal.
3. Dividing and combining the signal at mm-wave frequencies

In upcoming sections, design choices and their corresponding trade-offs for each of these three building blocks are investigated to reach to the optimum solution for a mm-wave RF-path active phase shifter.

5.2.1 I-Q Generation

RC poly phase filters

RC poly phase filters that are exclusively used in integrated solutions at lower frequency applications are compact solutions for I-Q generation. In this section we investigate the feasibility of implementing them at mm-wave frequencies.

A low-pass RC filter has a frequency response of $\frac{1}{1+j\omega RC}$ and its high-pass counterpart has a frequency response of $\frac{j\omega RC}{1+j\omega RC}$. So a combination of high-pass and low-pass filters realizes two output signals that are 90° apart in phase and at the pole frequency of $\omega = \frac{1}{RC}$ their transfer functions have equal magnitudes of $|H_{HP}(\omega = \frac{1}{RC})| = |H_{LP}(\omega = \frac{1}{RC})| = \frac{1}{\sqrt{2}} = -3dB$. 3dB loss of voltage does not seem to be a road blocker and at the first glance it looks feasible to compensate for that loss in the VGA design and hence use the RC poly phase filter for the I-Q generation and benefit its extremely compact nature. However that 3dB loss is not all the price to be paid to employ an RC poly phase filter.

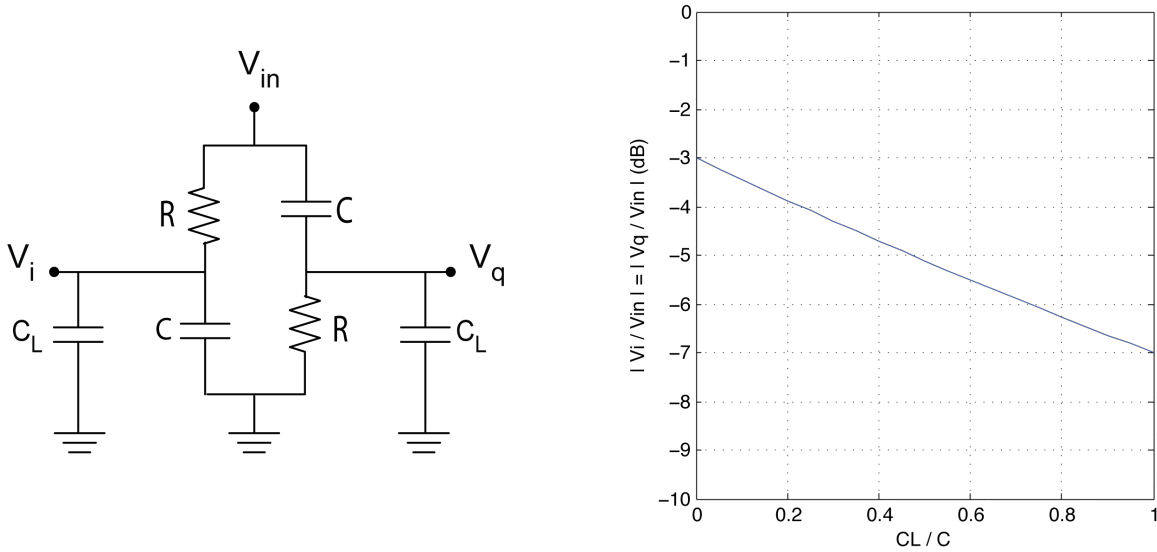


Figure 5.6: A voltage domain poly phase filter (left) and its gain degradation as loading capacitances being presented output nodes.

Since the poly phase filter needs to interface a next stage of amplification, it will be loaded by some capacitances as depicted in figure 5.6. Calculating the new transfer function for the poly phase filter including loading capacitances will result in the new transfer function as:

$$H_{LP}(\omega) = \frac{1}{1 + j\omega RC(1 + \frac{C_L}{C})}, \quad H_{HP}(\omega) = \frac{j\omega RC}{1 + j\omega RC(1 + \frac{C_L}{C})} \quad (5.12)$$

To have the two magnitudes equal, $\omega RC = 1$. Multiplying the pole frequency in the new transfer function by a factor of $(1 + \frac{C_L}{C})$ increases the loss of the poly phase filter. As numerical examples, to keep the extra loss below 1dB, C_L should be 25% of the capacitance in the poly phase filter and if $C_L = C$ then the additional loss will go up to 4dB which makes a total loss of 7dB. For operation at mm-wave frequencies small R and C values are needed for the poly phase filter. For typical transistor sizes, as the loading of the following amplification stage and the requirement that the poly phase capacitance (C) should be roughly four times the loading capacitance of the transistor, the resistance of the poly phase filter needs to be relatively small (less than 100Ω). This low value of resistance poses another issue for the RC poly phase filter demonstrated in figure 5.7.

In CMOS amplifiers the magnitude of the output voltage will be :

$$|V_{out}| = g_m \cdot R_L \cdot |V_{GS}| \quad (5.13)$$

Because there is a matching network at the input of the amplifier to match the parallel R_G of the transistor to the characteristic impedance of the input transmission line (Z_0), the V_{GS} can be related

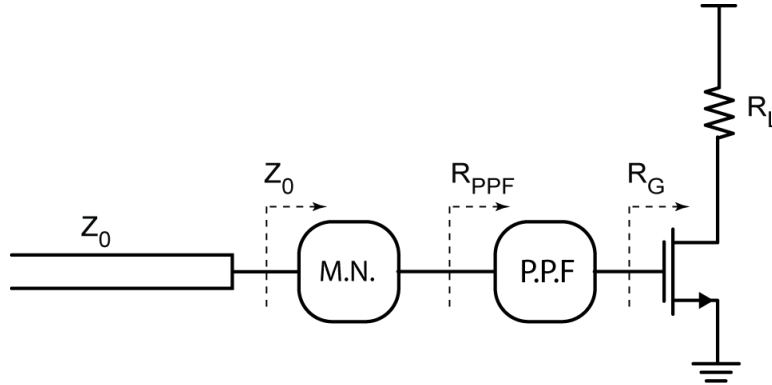


Figure 5.7: Adding the poly phase filter lowers the input impedance of a MOS amplifier which results in lower matching network Q and consequently lower gain through the chain

to V_{in} as :

$$|V_{GS}| = \sqrt{\frac{R_G}{Z_0}} \cdot |V_{in}| \Rightarrow \left| \frac{V_{out}}{V_{in}} \right| = g_m \cdot R_L \cdot \sqrt{\frac{R_G}{Z_0}} \quad (5.14)$$

In modern CMOS technologies, after optimal choice of device size and number of fingers, the shunt gate resistance (R_G) can be kept as high as $1K\Omega$ even at mm-wave frequencies. Therefore a considerable portion of the gain comes from the rather high-Q matching network at the input of the MOS transistor ($Q = \sqrt{\frac{R_G}{Z_0} - 1}$). In case of having an RC poly phase filter at the input of the amplifier, the voltage gain will be:

$$\left| \frac{V_{out}}{V_{in}} \right| = g_m \cdot R_L \cdot \frac{1}{\sqrt{2}} \cdot \sqrt{\frac{R_{ppf}}{Z_0}} \quad (5.15)$$

Where R_{ppf} is the poly phase filter resistance. So not only is there a minimum of 3dB loss for including the poly phase filter in the path, but also due to the fact that the matching network Q decreases when there is an RC poly phase filter, the voltage gain gets an additional hit of $\sqrt{\frac{R_{ppf}}{R_G}}$. As mentioned above R_G can be an order of magnitude higher than R_{ppf} and this additional 10dB of loss is too detrimental to include an RC poly phase filter in the signal path.

Current Mode RC Poly Phase Filters

Since the major drawback of the voltage mode RC poly phase filter was decreasing the input matching network Q, one might want to say : an RC poly phase filter in current domain as demonstrated in figure 5.8 solves the problem. In this setup the RC poly phase filter is inserted between top and bottom transistors of a cascode amplifier and bottom g_m transistors buffer the poly phase filter from the input and alleviate the need for a matching network. This removes the hit in the gain as a result of lowering the matching network Q by inserting the poly phase filter.

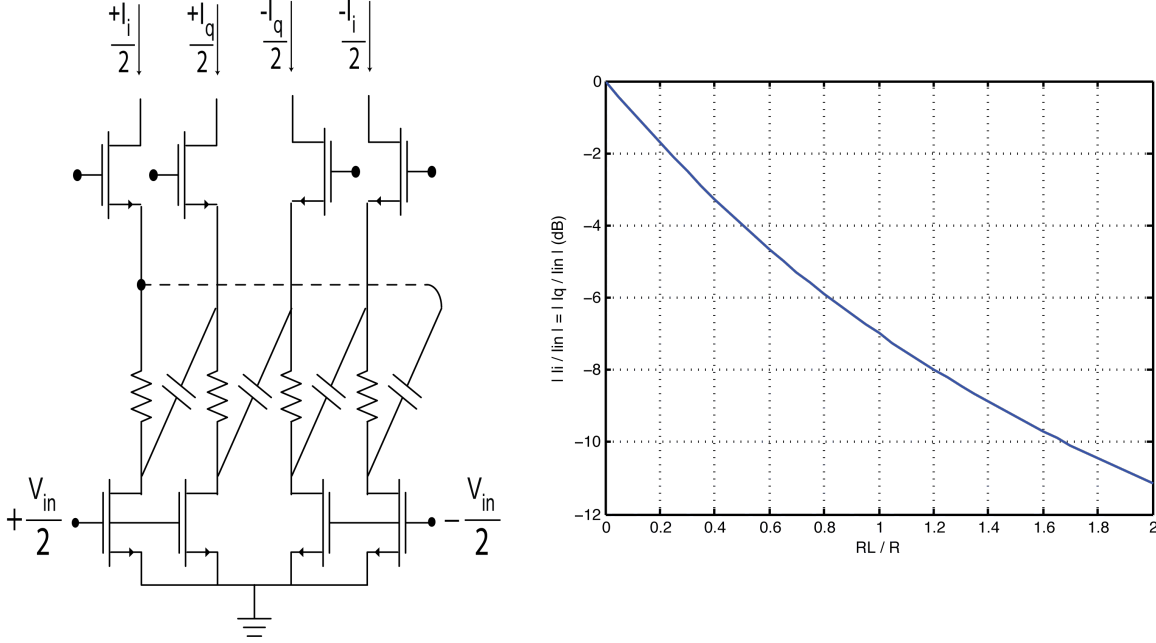


Figure 5.8: A current mode poly phase filter (left) and its gain degradation as a nonzero load impedance being presented at the output.

The current domain RC poly phase filter at the pole frequency $\omega = \frac{1}{RC}$, if loaded with a very low output impedance, has a transfer function of $|\frac{i_o}{i_i}| = 1$. However for a non-zero loading impedance of R_L , the transfer function will be multiplied by a factor of:

$$\frac{1 + j\omega RC}{1 + j\omega RC(1 + \frac{2R_L}{R})} = \frac{1 + j}{1 + j(1 + \frac{2R_L}{R})} @ \omega = \frac{1}{RC} \quad (5.16)$$

The loading resistance R_L will be determined by upper transistors transconductance $R_L = \frac{1}{g_{m2}}$. However due to supply voltage limitations of highly scaled CMOS technologies ($V_{DD} \leq 1.2V$), the $\frac{R_L}{R}$ ratio can not be made very small. There is a maximum allowable voltage drop on the resistance of the poly phase filter since $V_R = V_{DD} - V_{GS2} - V_{DS1}$ and to have a good f_t for each device they need to be biased with a sufficient V_{GS} and V_{DS} . For a fixed DC current, there is a maximum transconductance extractable from a transistor at 60GHz after the proper sizing and voltage biasing is done. In 40nm CMOS technology this maximum transconductance is roughly $g_m \sim 2.5 \frac{mS}{mA} \cdot I_D$. Therefore for the resistance ratio, it can be conferred that:

$$\frac{R_L}{R} = \frac{1}{g_m \cdot R} = \frac{1V}{2.5 \cdot V_R} \quad (5.17)$$

For 200mV of voltage drop across the resistance of the poly phase filter (which is still a large value and starts to degrade the f_t of the bottom transistor due to a non-sufficient bias of V_{DS}), the

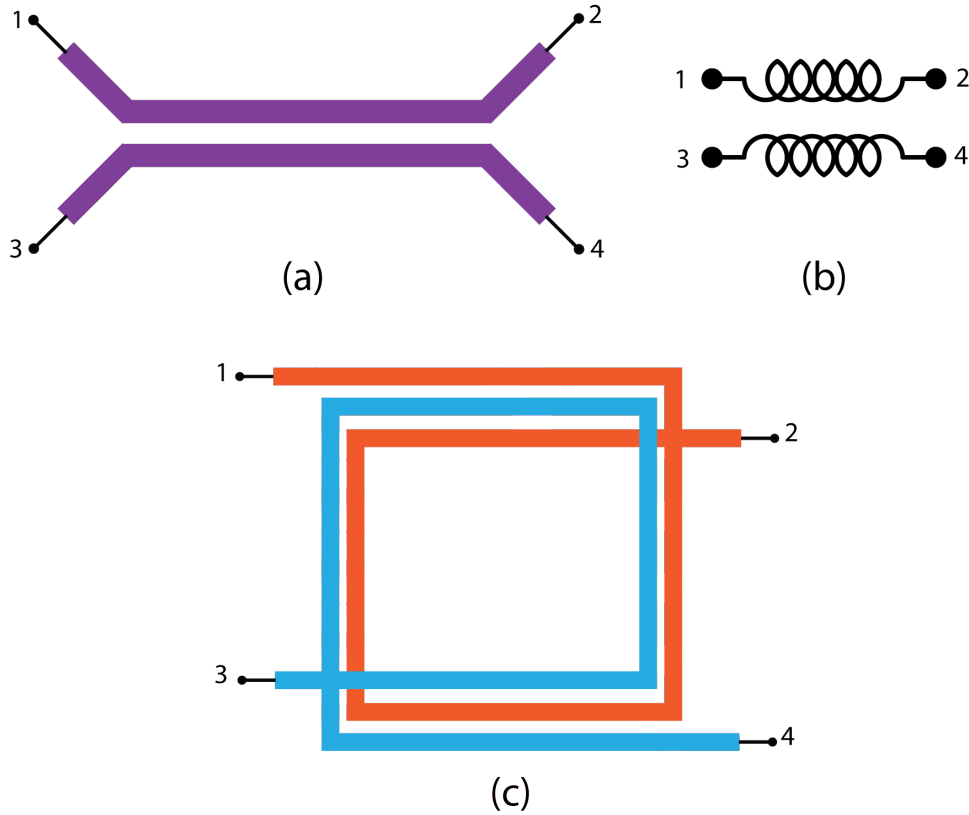


Figure 5.9: a) A transmission line coupler, b) A transformer can be exploited for the lumped version of a transmission line coupler, c) 1.25 turn loops comprising a transformer coupler resemble a quadrature hybrid.

resistance ratio will be $\frac{R_L}{R} \sim 2$. If this resistance ratio is plugged into the equation (5.16), it yields to an extra loss of higher than 10dB for the current mode poly phase filter which again makes RC poly phase filters too lossy to be implemented in the signal path of a mm-wave system.

Synthesized Quadrature Hybrids

Electromagnetic based design of quadrature hybrids can be quite bulky. However there are techniques to make them more compact. A quadrature hybrid as depicted in figure 5.2 can be realized in a smaller footprint if slow wave transmission lines are used in the arms. Capacitive loadings of preceding and following stages can be absorbed in the hybrid design to further reduce the size.

One other electromagnetic hybrid structure is the “coupled transmission lines” hybrid demon-

strated in figure 5.9-a. For this structure, a coupling factor is defined as [12]:

$$C = \frac{Z_{0e} - Z_{0o}}{Z_{0e} + Z_{0o}} \quad (5.18)$$

After working out the even-odd mode analysis technique and assuming that the input excitation is applied to port 1, following equations will be achieved for the signal level at other ports[12]:

$$V_2 = V_1 \cdot \frac{\sqrt{1 - C^2}}{\sqrt{1 - C^2} \cos \theta + j \sin \theta} \quad (5.19)$$

$$V_3 = V_1 \cdot \frac{jC \tan \theta}{\sqrt{1 - C^2} + j \tan \theta} \quad (5.20)$$

$$V_4 = 0 \quad (5.21)$$

Where θ is the electrical length of transmission lines. For $\theta = \frac{\pi}{2}$ (quarter wavelength transmission lines) the two output terms reduce to:

$$\theta = \frac{\pi}{2} \rightarrow \frac{V_2}{V_1} = -j\sqrt{1 - C^2}, \frac{V_3}{V_1} = C \quad (5.22)$$

Synthesizing an artificial transmission line out of lumped component inductors and capacitors help to reduce the size of the coupled line hybrid and make it size appropriate for being integrated in a modern CMOS IC design. A transformer as shown in figure 5.9-b with corresponding capacitive loadings can help to realize a miniature synthesized quarter wavelength coupled line structure that provides 90° out of phase signals at the output. As demonstrated in figure 5.9-c, two 1.5 turn loops comprising a transformer to be used as the lumped version of a transmission line coupler, resemble a quadrature hybrid as well.

Although electromagnetic hybrids take up more area than RC poly phase filters, their superior performances and the fact that the size can be miniaturized through employment of lumped inductors and capacitors result in net benefits. this is the technique that will be used as the I-Q signal generator for the RF-path active phase shifter.

5.2.2 Variable Gain Amplifiers

To change the gain through one stage of a transistor amplifier, bias conditions can be changed as depicted in figure 5.10. Bias current can be adjusted through varying V_{GS} , V_{DS} or switching parallel fingers of the transistor in and out of the signal path. At mm-wave frequencies devices are already pushed to their limits and are operating close to their functionality boundaries. Hence adding the flexibility of gain variation through simple methods of changing the biasing point and not employing more transistors in the signal path can be beneficial. The programability can come from tuning the V_{GS} as shown in figure 5.10 -a, or V_{DS} of the bottom transistor in a cascode amplifier as depicted in figure 5.10 -b. However due to nonlinear characteristics these methods

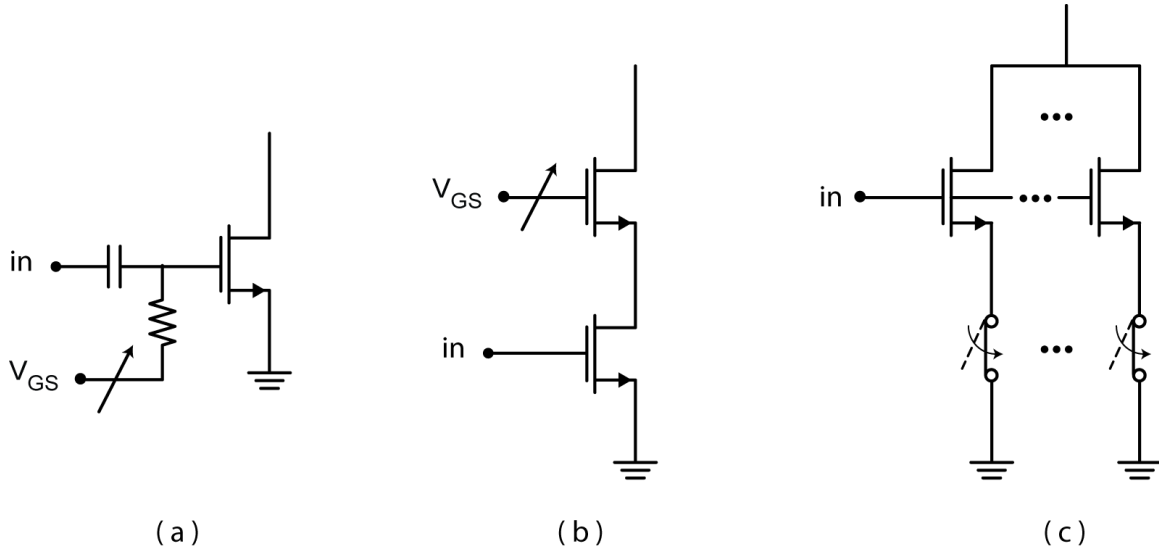


Figure 5.10: Gain variation capability can be achieved through changing bias conditions of an NMOS transistor.

become very sensitive to V_{GS} and V_{DS} accuracy. Moreover through these methods, gain variability comes at the price of the transistor current density deviating from the optimum current density for achieving the maximum f_t . To maintain the same f_T and consequently the same current density, a structure as depicted in figure 5.10 -c can be used. However to have a very low R_{ON} for switches, in order not to degrade the transconductance through source degeneration, large transistors should be used as switches which in turn makes the layout very cumbersome.

One shortcoming of all three structures shown in figure 5.10, is the fact that they can not invert the polarity of the signal. Without inverting the polarity of the signal, only one out of four quadrants in the I-Q plane is covered which is not sufficient. Having an extra phase inverter block as presented in figure 5.11 not only takes up more area but also simulations show that the optimum design has more than 3dB of loss at 60GHz. Therefore using alternative VGA structures that use more transistors in the signal path to provide the opportunity of phase inversion, despite having lower extractable gain, it is favorable as long as the gain difference is less than or comparable with the loss of the phase inverter. That is due to its compactness with respect to above mentioned solutions of VGA structures cascaded with a required phase inverter.

A current commuting type VGA as shown in figure 5.12 functions as the difference between the two cascode gate voltages ($V_{up} - V_{down}$) determines how much of the differential current generated by bottom transistors will cancel out each other and how much will flow to the output load and provide gain. If V_{up} and V_{down} values are flipped then the output phase will be inverted and hence such a VGA structure removes the need for phase inverters to cover all four quadrants in the I-Q plane. The maximum gain of the current commuting VGA structure happens when two of top

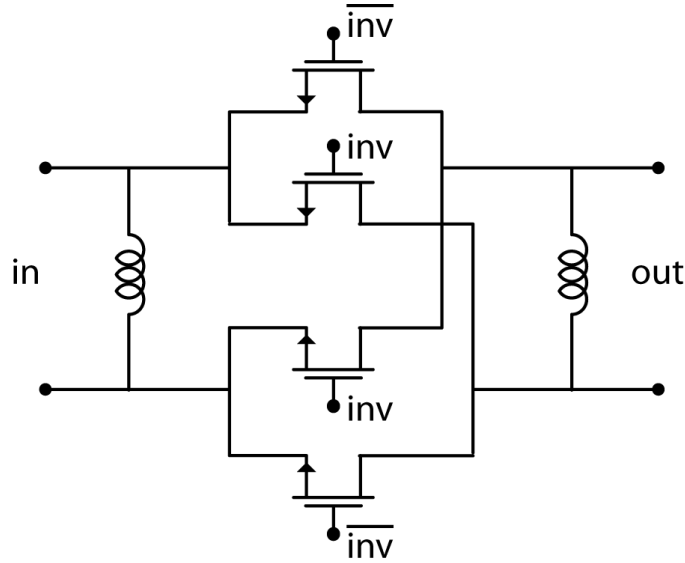


Figure 5.11: A phase inverting building block with input and output tuning inductances.

cascode transistors are completely on and the other two are completely off. In this case it has less gain compared to a fixed gain cascode amplifier because of the extra loading at the middle node of the cascode amplifier (node X). A shared junction cascode structure can not be realized for current commuting VGA due to the need for access to that middle node for connection to the second top transistor. Not only that, but also due to the fact that the second cascode transistor –even when it is off– adds to the capacitive load of the middle node, the pole associated to that middle node comes down in frequency and decreases the gain. Rough calculations estimate that middle node pole to be $\frac{g_m}{3C} \sim \frac{f_t}{3}$ for the current commuting VGA when it is providing the maximum gain. When compared to a fixed gain cascode amplifier with the middle node pole of $\frac{g_m}{2C} \sim \frac{f_t}{2}$, current commuting VGA at its maximum gain setting gives about $20 \log(1.5) \sim 3.5dB$ less gain. On the other hand a current commuting VGA alleviates the need for a phase inverter which adds to the area and has about the same loss as the difference between the maximum gain of a current commuting VGA and a fixed gain cascode amplifier. The core layout of a current commuting type VGA can be accomplished in a very compact fashion which reduces the loss attributed to interconnect parasitics as well and justifies the use of this structure as the VGA for RF path I-Q interpolator.

5.2.3 Signal dividers / combiners

There are multiple ways for dividing and then combining the signal after they are passed through VGAs. In traditional microwave design a Wilkinson divider and a 90° hybrid with according matching networks as depicted in figure 5.13 do the task. The total voltage gain for the worst case

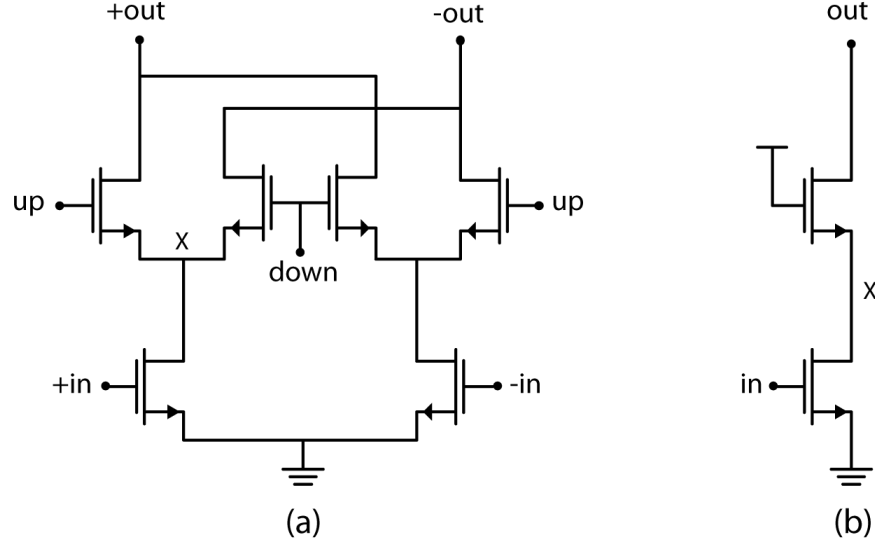


Figure 5.12: Schematic of core transistors for a current commuting type VGA (a) and a fixed gain cascode amplifier (b).

scenario where only one of the two I or Q channels is active will be :

$$A_V = -6dB + 20 \log\left(\sqrt{\frac{R_G}{Z_0}} g_m r_0 \sqrt{\frac{Z_0}{r_0}}\right) - 4I.L. = -12dB - 4I.L. + 20 \log(g_m \sqrt{R_G r_0}) \quad (5.23)$$

Where -6dB comes from the theoretical signal division/combining loss of 3dB for hybrid and Wilkinson divider and I.L. is the insertion loss of each stage of passive components (hybrid, Wilkinson divider and matching networks with an assumption that they are comparable and can be approximated to be equal to each other) due to the finite Q of integrated passive components. This is the gain for when only one channel is active; for other cases where both I and Q channels are amplifying the signal, the voltage gain would be higher. But for the sake of comparison we look into worst case scenarios for the gain with one active channel.

But since integrated circuit design in highly scaled CMOS technologies (even at mm-wave frequencies) can be assumed a lumped component design as a result of tiny transistors and small footprints of active circuitry; analog techniques for dividing/combining signals in the voltage domain (connecting input gates of transistors) as demonstrated in figure 5.14-a or in the current domain (connecting output drain terminals) as shown in figure 5.14-b can be utilized. For the first scenario (figure 5.14-a) the voltage gain for the path will be :

$$A_V = -3dB + 20 \log\left(\sqrt{\frac{R_G}{2Z_0}} g_m \frac{r_0}{2} \sqrt{\frac{Z_0}{r_0}}\right) - 3I.L. = -12dB - 3I.L. + 20 \log(g_m \sqrt{R_G r_0}) \quad (5.24)$$

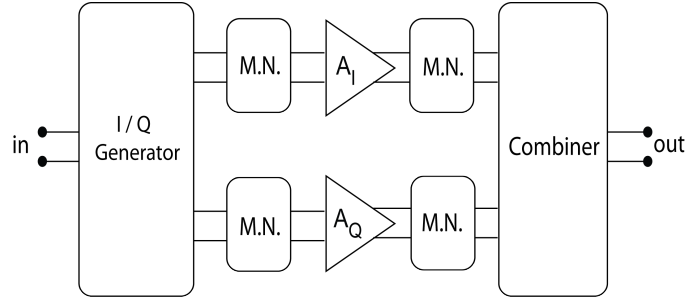


Figure 5.13: Traditional microwave method of signal dividing and combining.

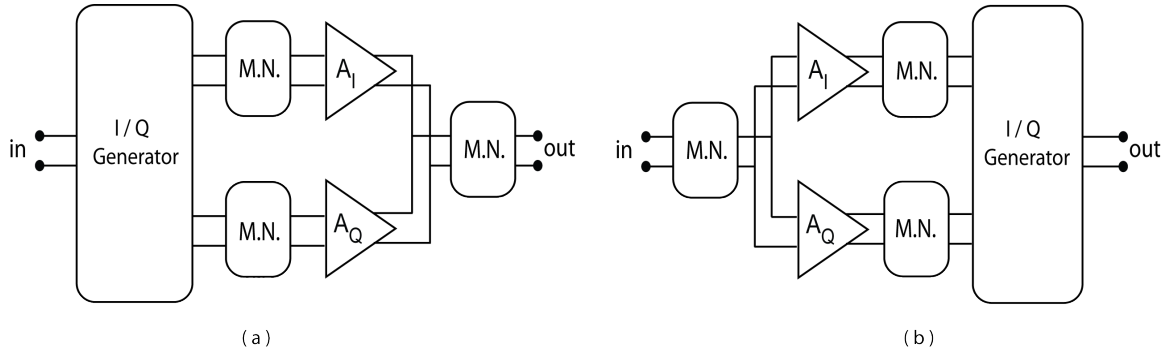


Figure 5.14: Two lumped component version for dividing and combining signals in current (a) or voltage (b) mode

And for the second scenario (figure 5.14-b) the voltage gain can similarly be calculated as :

$$A_V = -3dB + 20 \log\left(\sqrt{\frac{R_G}{Z_0}} g_m \frac{r_0}{4} \sqrt{\frac{2Z_0}{r_0}}\right) - 3I.L. = -12dB - 3I.L. + 20 \log(g_m \sqrt{R_G r_0}) \quad (5.25)$$

As suggested by equations 5.24 and 5.25, the voltage and current combining methods give equal voltage gains in first order approximations. Due to the fact that they have one less passive component, voltage or current combining methods have smaller footprints and less loss compared to a traditional microwave combining technique and hence they are preferred. At the first glance current and voltage combining have the same performance, and voltage gain is identical for them, however since the VGA output resistance changes as the gain is varied, the I-path gain will be strongly dependent on the Q-path gain if the outputs are connected directly. On the other hand if outputs are connected through a network of matching networks and a 90° hybrid (which provides some isolations between its ports) voltage gains of I and Q paths will be independent of each other (the input impedance of VGAs will not change as the gain is altered) which makes it easier for predictability and gain tuning programmability. As a result the structure shown in figure 5.14-a is

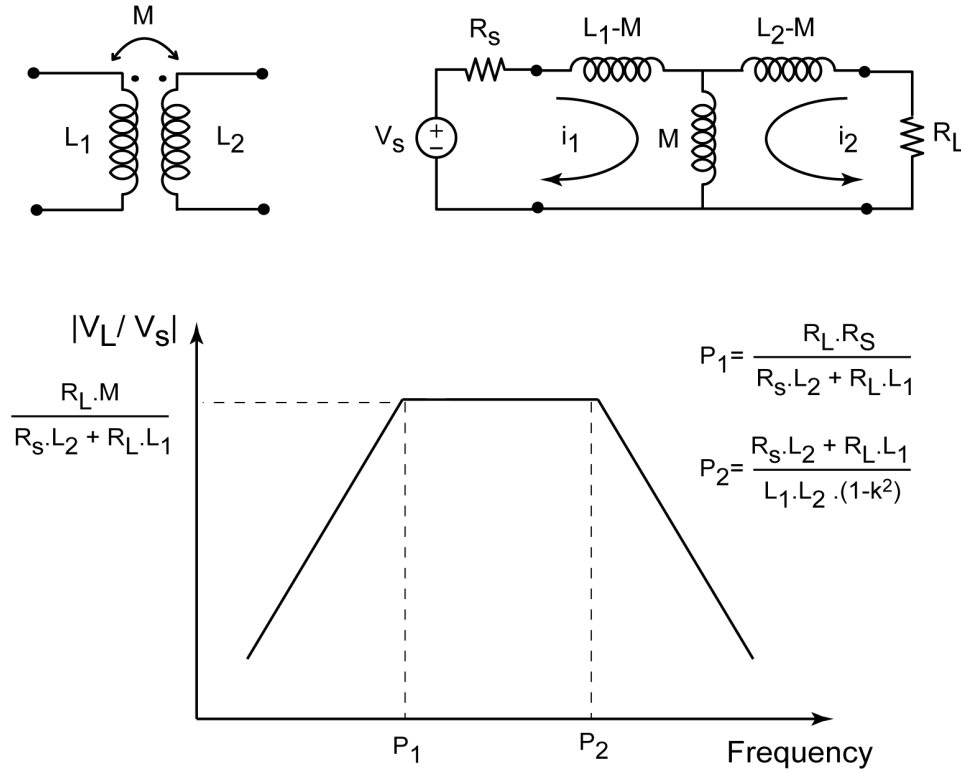


Figure 5.15: A transformer acts as a wideband matching network

chosen for signal dividing /combining between the two I and Q paths.

5.3 mm-Wave implementation of the active phase shifter

5.3.1 Transformer matching networks

Since the design is fully differential and also a relatively wide bandwidth (7GHz around 60GHz) is to be addressed, transformers are used for matching networks. Transformer matching networks have higher bandwidths compared to L-shape matching networks. An ideal transformer (perfect coupling between the loops and large self inductances associated with each loop) has the capability of multiplying an impedance by n^2 (n being the turn ratio) over all frequencies. A practical transformer still has more bandwidth than an L-shape matching network. Neglecting parasitic capacitances and assuming two loops with self inductances of L_1 and L_2 and a mutual inductance of M , the network demonstrated in figure 5.15 can be solved through mesh analysis and the resultant

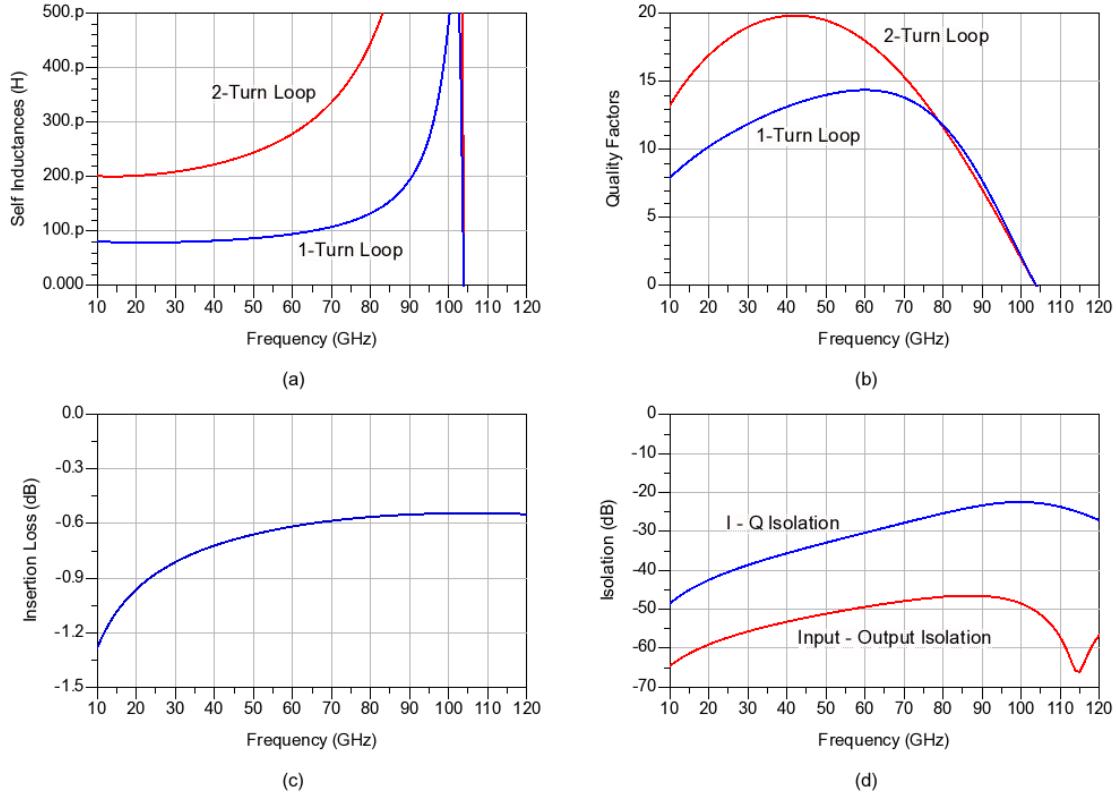


Figure 5.16: (a)-Simulated Self inductances of a 2 : 1 transformer, (b)- quality factors of primary and secondary windings, (c)-transformer insertion loss, (d)- coupling between transformers

transfer function can be written as:

$$\frac{V_L}{V_S} = \frac{R_L \cdot M \cdot s}{s^2(L_1 L_2 - M^2) + s(R_S L_2 + R_L L_1) + R_S R_L} \quad (5.26)$$

At midband the voltage ration can be written as $A_V = \frac{V_L}{V_S} = \frac{R_L M}{R_S L_2 + R_L L_1}$ and the load impedance of R_L will be multiplied by A_V^{-2} and appear across the source winding. From the above equation for the transfer function, relation between the two low frequency and high frequency poles can be written as :

$$P_1 + P_2 = \frac{R_S L_2 + R_L L_1}{L_1 L_2 - M^2} \quad (5.27)$$

$$P_1 P_2 = \frac{R_S R_L}{L_1 L_2 - M^2} \quad (5.28)$$

Assuming that the two poles are well separated from each other:

$$P_2 \gg P_1 \implies P_1 = \frac{R_S R_L}{R_S L_2 + R_L L_1}, \quad P_2 = \frac{R_S L_2 + R_L L_1}{L_1 L_2 (1 - k^2)} \quad (5.29)$$

As can be seen from equation 5.29, higher tight couplings ($k \rightarrow 1$) and larger self inductances of the loop makes the two poles diverge from each other and result in higher bandwidths for a transformer matching network. In practice self inductances can not be made arbitrarily large as parasitic capacitance of each loop to the substrate increase as well as the capacitance between the two loops which eventually limits the operation frequency of the matching network.

Matching networks are realized as 2 : 1 lateral transformers on the top metal layer. The primary and secondary inductances are $L_1 = 96pH$ and $L_2 = 278pH$ respectively and the self resonance frequency (SRF) of the structure is $f_{res} = 105GHz$ (figure 5.16-b). The coupling factor between is $k = 0.84$ at $60GHz$. Two loops of a 2 : 1 lateral transformer are tightly coupled as the 1-turn primary loop is in the middle and two turns of the larger secondary winding are placed on the two sides of the primary loop (compared to the $k = 0.72$ obtained in the 1 : 1 transformer that was employed in the T/R switch in chapter 2). Windings are made wide enough not to degrade the shunt resistance that they need to match the Z_0 to, but not too wide to lower the SRF. Primary and secondary windings have quality factors of $Q_1 = 14.4$ and $Q_2 = 17.7$ at $60GHz$ respectively (figure 5.16-b) and as a result of those the minimum insertion loss of the transformer is $I.L. = 0.62dB$ at $60GHz$ (figure 5.16-c).

There are three transformers for matching the input and the two outputs of VGAs to Z_0 . Hence there will be three closely laid out transformers as shown in figure 5.18, and coupling between them might be an issue. These three transformers were simulated altogether in a full-wave electromagnetic simulator (HFSS) to capture all the parasitic and coupling effects. As plotted in figure 5.16-d, at $60GHz$, $30dB$ of isolation is achieved between I and Q transformers while a $50dB$ of isolation is obtained for the coupling between the input and either I or Q output transformers. The extra $20dB$ of isolation came from the fact that there are low impedance ground and supply lines separating the input transformer from I/Q transformers at the outputs of the VGA. These low impedance lines act as a shield for electric fields and the current induced in them due to the Lenz law cancels out the residual magnetic field leaked out of the windings of the transformer. $30dB$ of isolation between I and Q transformers was enough for our case but for situations where more isolation is required, a similar ground line between I and Q transformers can be inserted. A slight resizing for the transformer compensates for the extra parasitic capacitances introduced by that ground line. The complete schematic of the phase shifter is depicted in figure 5.17 and its corresponding layout is shown in figure 5.18.

5.3.2 Measurement results

The prototype was fabricated in a 40nm CMOS process and the chip microphotograph is shown in figure 5.18. On-wafer measurements were performed using mm-wave differential probes that convert the single ended signal of the "Vector Network Analyzer (VNA)" to the differential mode at the probe tip via embedded Merchant baluns. Having differential open and short structures for the input / output pads allows for de-embedding pad parasitics from the measurement data. De-embedded data for the gain, phase and return losses at input and output along with the simulation prediction is plotted in figure 5.19.

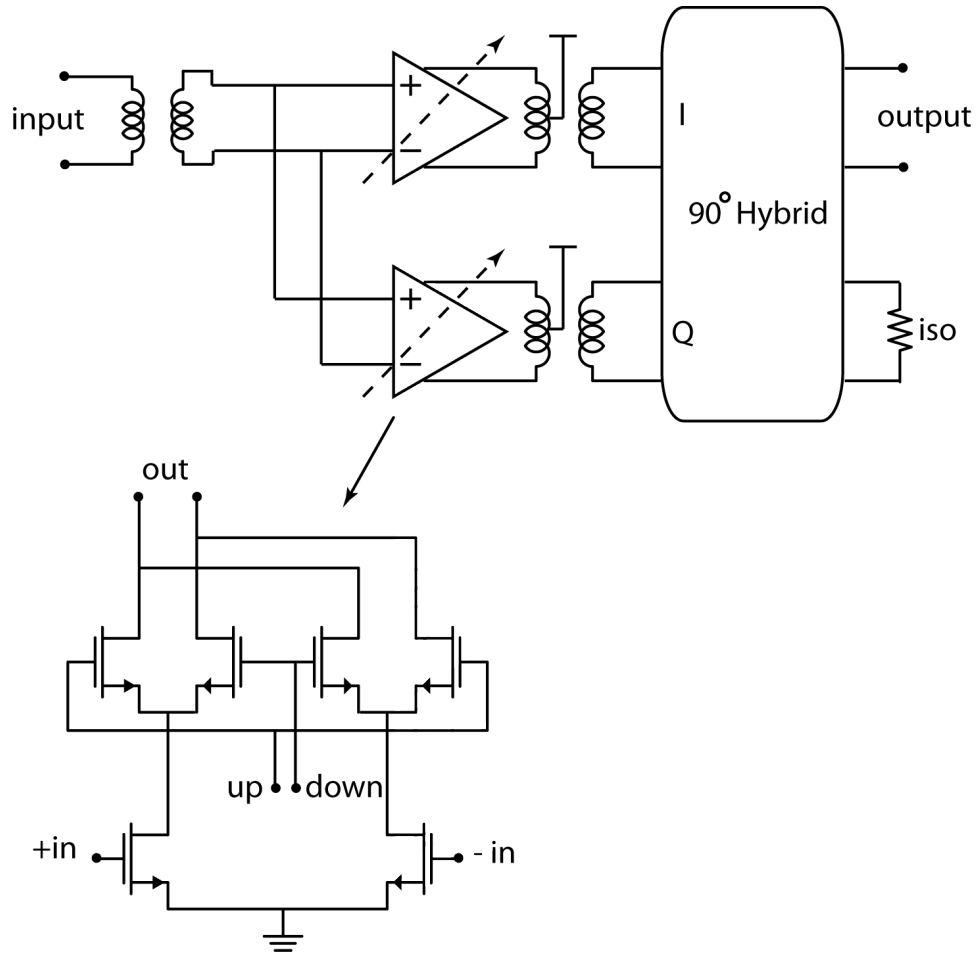


Figure 5.17: Schematic of the complete active I-Q interpolating phase shifter

As can be seen in figure 5.19, measurement and simulation data are in a good agreement with each other. The reason that the center frequency is leaned toward the higher side of the 60GHz band is that design kit models at the time of the design were premature and had a pessimistic estimation for parasitics. However when the chip came back, measurement results matched the simulation data based on updated design kit models.

Measured phase response for different phase shift settings are demonstrated in figure 5.20. Phase plots for eight different modes of $a_i \times I + a_q \times Q$ where $a_i, a_q \in (-1, 0, 1)$ are demonstrated. As can be seen from the graphs, full 360° of phase shift is achievable through an active I-Q interpolating phase shifter. Advantages of using such an structure for phase shifting is that, there is a VGA in each I and Q paths that can adjust the gain for I and Q and hence calibrate initial mismatches in gain and phase responses. These two VGAs can operate in conjunction with each other and function as a variable gain amplifier in the signal path as well. A measurement versus

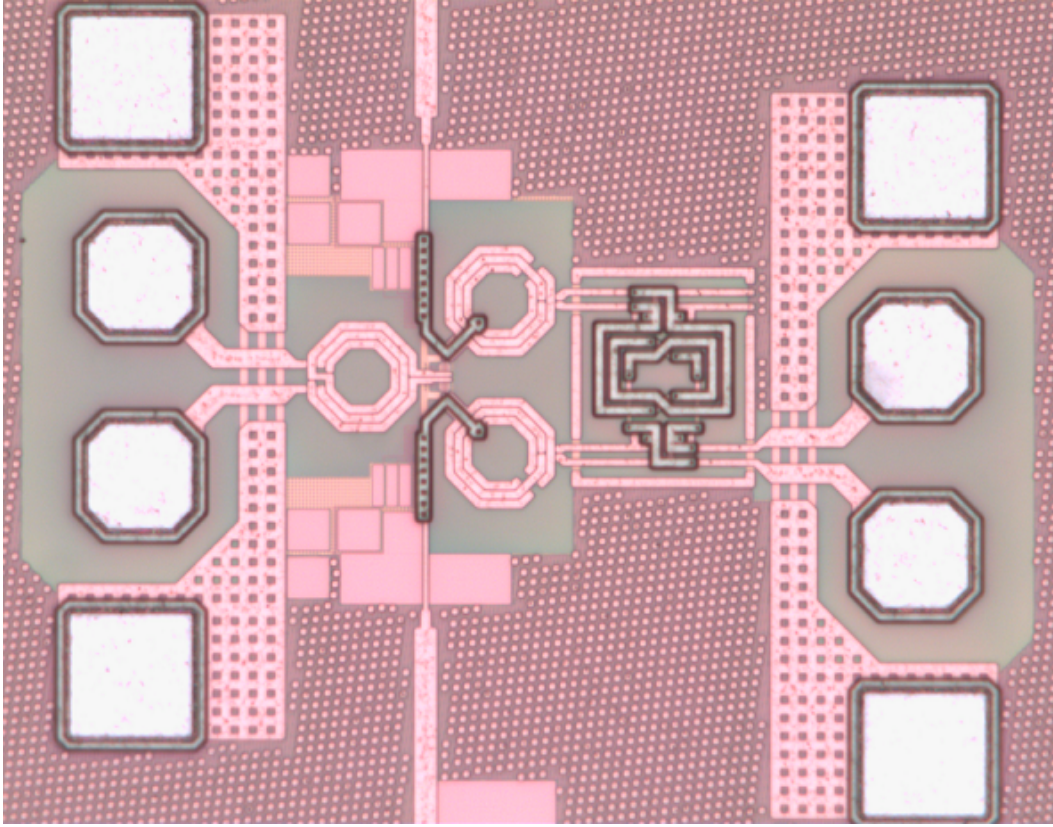


Figure 5.18: Die microphotograph of the I/Q interpolating active phase shifter

simulation of this gain variation capability is depicted in figure 5.21.

In a Gilbert type VGA (figure 5.12), the V_{up} voltage is set to V_{DD} and the V_{down} voltage can vary from V_{DD} to zero. However since the bias current will be fully switched to one branch when the differential voltage ($V_{up} - v_{down}$) reaches the V^* (Where V^* is the overdrive voltage of the ON transistor), the minimum voltage for V_{down} can be set to $V_{DD}/2$. For inverting the phase of the signal in each I and Q paths, voltage settings for V_{up} and V_{down} can be swapped.

Large signal measurements were done at 65GHz where the non de-embedded s-parameter data showed a reasonable input return loss of $-8dB$. A signal generator and a power sensor were employed to perform the measurement and cable+probe losses were de-embedded from the data. Graphs of the large signal measurement data and their comparison with simulation prediction are plotted in figure 5.22. As can be seen from the two different cases, one when only the I-path is active and in the other case both I and Q paths are in the RF path and amplifying the signal, measurement data match the simulation prediction pretty well. Despite the gain difference of these two extreme cases, the input referred $1dB$ compression points are similar and that is because the overall structure has a gain of less than $0dB$ and as a result of that the linearity is input limited

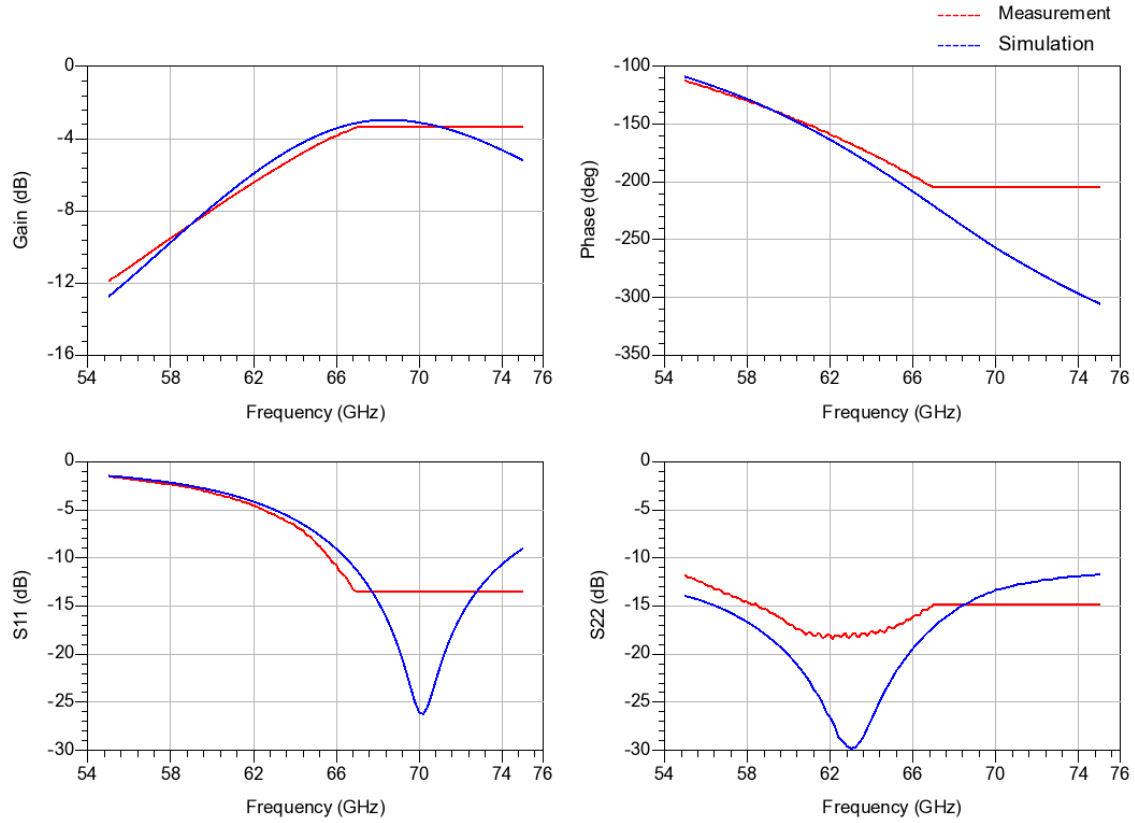


Figure 5.19: Measurement data for gain, phase and return losses at input and output and their comparison with simulation prediction

determined by the maximum voltage swing across gate-source terminals where the gain is still within 1dB of its nominal value.

In conclusion, the first RF-path active I-Q interpolating CMOS phase shifter at 60GHz was designed and implemented. To be as area efficient as possible, transformers were used exclusively in matching networks, I-Q generation and signal combining. A differential design made the circuit more robust against uncertainties such as ground and supply line inductances which are not negligible at 60GHz. Having a Gilbert-cell type of VGA shows the possibility of applying analog and RF techniques at higher frequencies of mm-wave regime in 40nm standard CMOS process. As technologies get more advanced in the future, more and more transistors will be present in the mm-wave signal path to conduct functions such as signal amplification, switching, current mirroring, feedback, and so on.

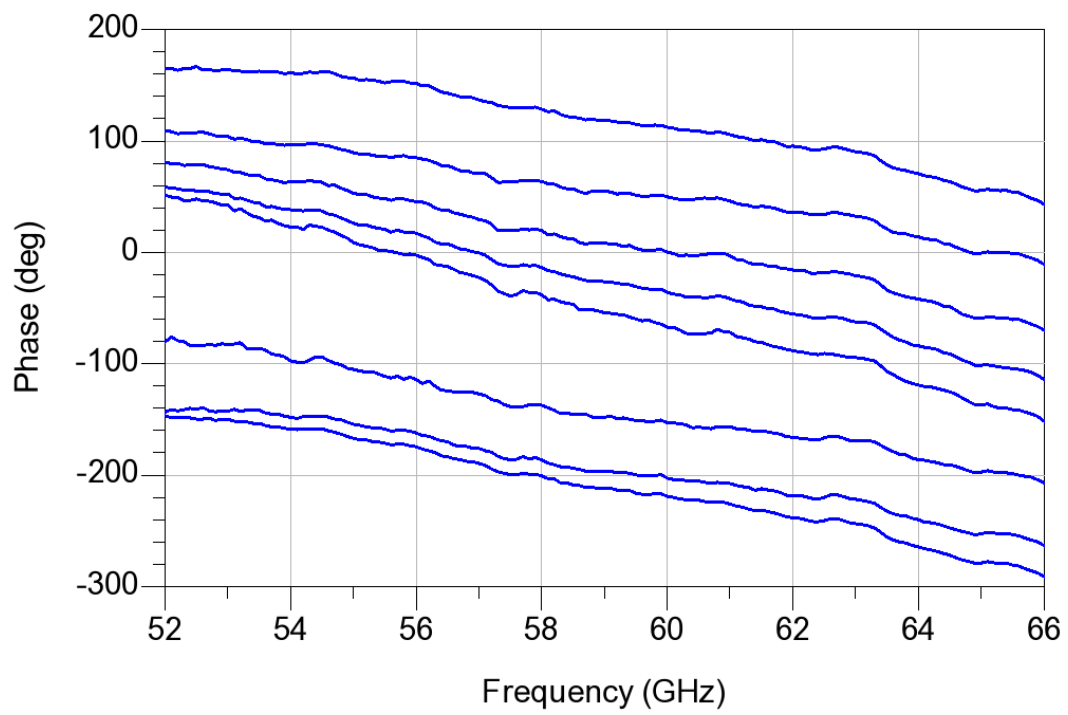


Figure 5.20: Measured phase responses for different phase shift settings

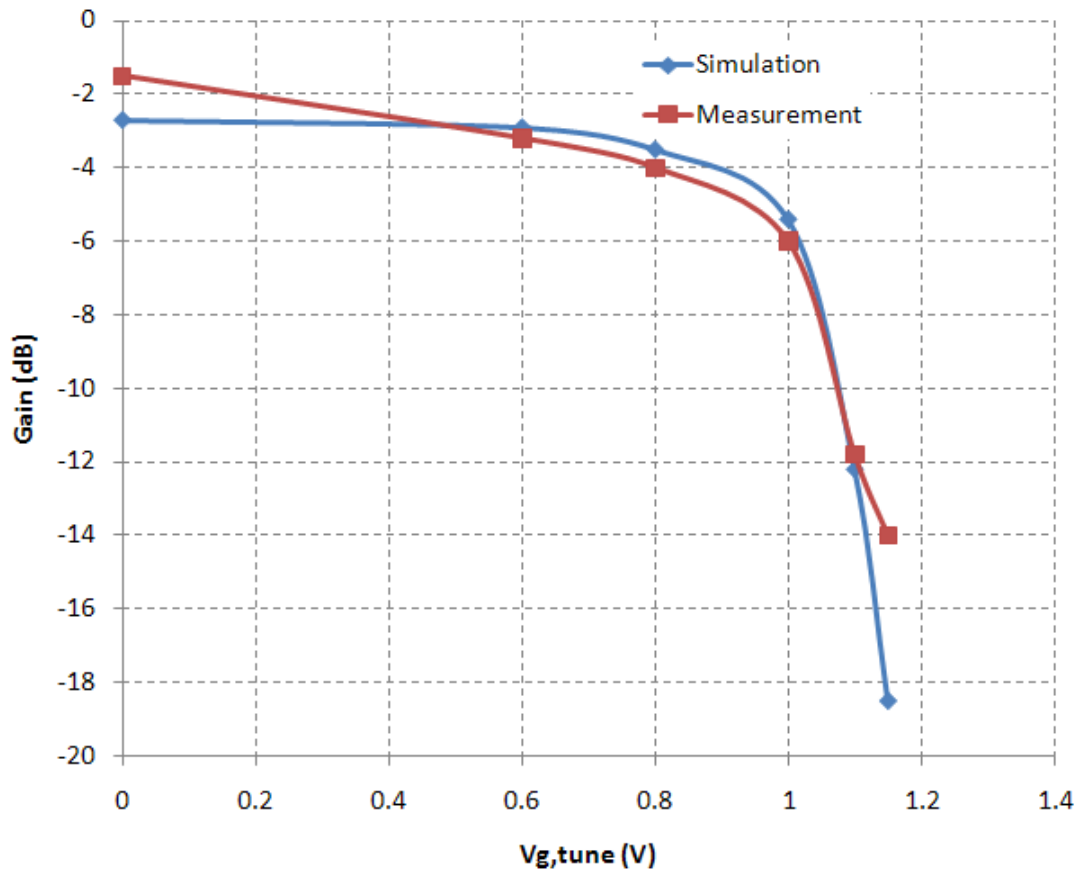


Figure 5.21: $V_{g,tune}$ can be varied from V_{DD} down to $V_{DD}/2$ in order to adjust the gain

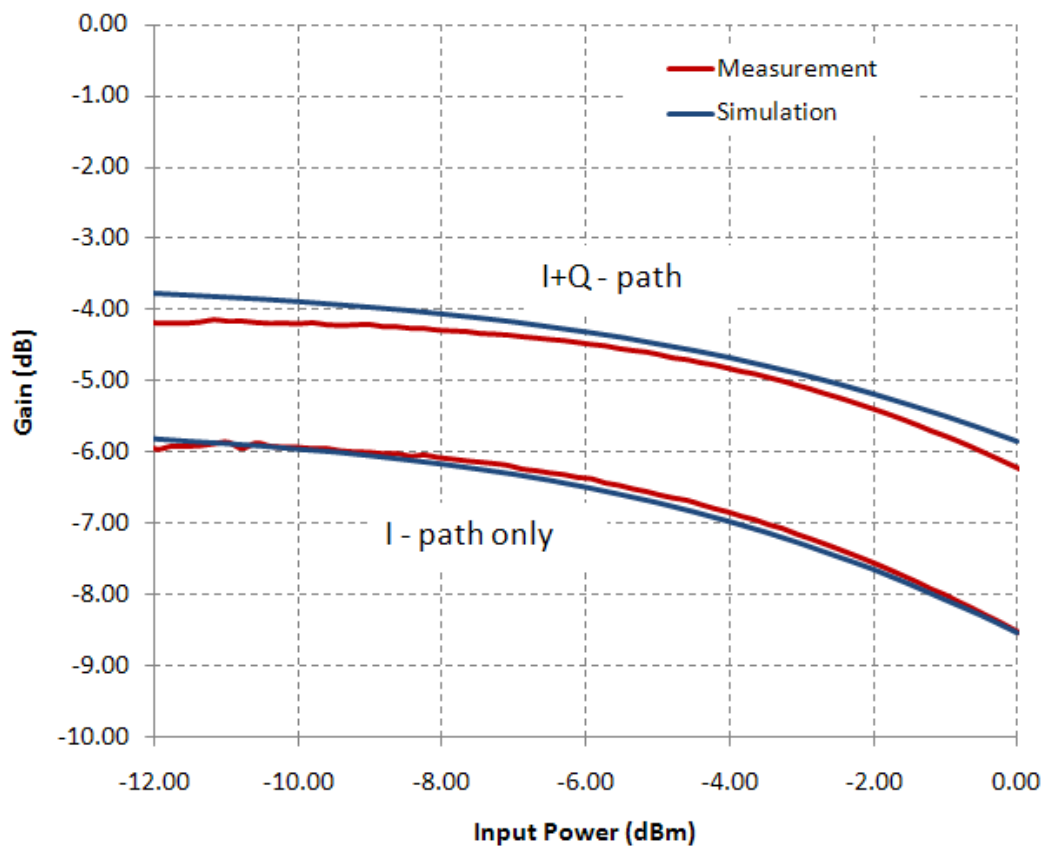


Figure 5.22: Large signal measurements for two extreme cases of I+Q and only-I settings and their comparison with simulation data

Chapter 6

Conclusion

A yet to bloom mm-wave market of consumer electronic products will open up various opportunities for extremely high data rate applications, ranging from high-volume WPAN applications to low-cost automotive radar modules to be embedded in every single vehicle, and to more performance stringent applications such as point to point mm-wave links to replace the fiber link in sparse areas. For high volume applications, being low-cost, power efficient and easily deployable are key metrics in determining the success of a product. Therefore endeavors to migrate a solution to a more mainstream technology with less RF specific options, and bringing down the chip area and power consumption numbers are much beneficial. Pioneer mm-wave research [44][48][53] and development [77] demonstrated the feasibility and advantages of a system which enables the 60GHz connectivity. However to de-risk these initial solutions, traditional microwave methods were adopted in their design and implementation. Transmission line based matching networks were utilized extensively and handful of carefully designed and characterized transistors were responsible for signal synthesis, low noise/high power amplification, and frequency translation. As the mm-wave field becomes more mature and gets ready to enter the commercial market of consumer electronics, more die-area savings and power efficiency enhancements should be investigated. As saving the die area or decreasing the power consumption adds to the functionality risk of such designs, intensive research should be carried out to guarantee the robustness of these mm-wave circuit and system design.

A theme that was pursued in this work was to replace bulky transmission lines with transformers. Miniature on-chip transformers were widely employed in matching networks, signal combining/dividing, small footprint quadrature hybrids, and providing variable inductances to be used in tunable synthesized delay lines. Transformer networks sustain more bandwidths which make the system less vulnerable to frequency mistuning as a result of added parasitics and their variations. They also obviate the need for interstage coupling capacitors. And in differential designs, transformers provide an AC ground center-tap node for supply and ground connections which completely removes the uncertainty associated with inductances in supply and ground lines.

As described in chapter 2, for large arrays or moderate size array being implemented on a mainstream (and hence lossy) package, employing transmit/receive switches results in a larger range for

the communication link. A highly compact transformer-based shunt switching structure was introduced and its design equations were derived. A prototype was fabricated and characterized for the validity check.

Due to significantly higher free space path loss numbers for mm-wave signals compared to lower frequency RF signals and because of less performance being available by transistors at these high frequencies close to their activity region, passive antenna gains provided by phased antenna array structures must be exploited to meet the link budget requirement of a mm-wave system. Fundamentals of phased arrays were described in chapter 3 and the next two chapters after that were devoted to implementation of RF beamformers. Chapter 4 employs true time delay elements for the beamforming function and RF-path phase shifters are exploited in chapter 5 to provide the beam steering capability of the array.

For ultra wideband applications and for large arrays with a sizable desired field of view, true time delay elements should not be replaced with phase shifters as the beam steering capability of the array degrades by doing so. Chapter 4 explores different tunable delay structures and propose an inductance tuning technique, in which with the aid of transformers, variable inductors are realized in differential synthesized transmission lines. These variable inductances along with variable capacitances enhance the delay tunability of the line while maintaining the characteristic impedance of the line constant. Two different versions of this inductance tuning technique were implemented. In one, employing series pass transistor switching networks, a passive structure was realized. Due to limited performance of MOS transistors acting as mm-wave series switches in the 90nm technology, the performance of such structures were limited and operating at mm-wave frequencies was out of reach. In the active version, CML-like current switches were employed to extend the frequency of operation at the cost of burning DC power. When delay cells were fully embedded in CML switches and one g_m transistor was added, an amplifier structure with adjustable group delay was constructed. Nonetheless including variable delay structures in an antenna array is still costly in terms of die area, power dissipation, and functionality risks. For applications in which phase shifting structures are sufficient to do the task of beamforming (most consumer electronic mm-wave applications), they should be utilized instead of delay cells as variable delay structures are an overkill.

Chapter 5 describes the first RF-path active I-Q interpolating phase shifter at 60GHz. To be as area efficient as possible, transformers were used exclusively in matching networks, I-Q generation and signal combining. A differential design made the circuit more robust against uncertainties such as ground and supply line inductances which are not negligible at 60GHz. Having a Gilbert-cell type of VGA shows the possibility of applying analog and RF techniques at higher frequencies of mm-wave regime in 40nm standard CMOS process.

As it was stated earlier, to remain competitive, current mm-wave solutions employing traditional microwave methods should adopt more lumped component techniques. As was shown in this thesis, transformer-based differential design are promising solutions. Despite their robustness, differential designs have a down-side of consuming more power. To remain appealing in the years to come, single ended designs should be also investigated for applications that power dissipation is the primary concern. As demonstrated in this work, unlike microwave approaches, transistors

were not only employed as mm-wave amplifiers but also they were utilized as mm-wave switches and were used to provide gain variability as well. As technologies get more advanced, more and more transistors will be present in the mm-wave signal path to conduct functions such as signal amplification, switching, current mirroring, feedback, etc.

Bibliography

- [1] "International Technology Roadmap For Semiconductors," 2007 Edition, Radio Frequency and Analog/Mixed-Signal Technologies for Wireless Communications
- [2] J. Powell, and D. Bannister, "Business Prospects for Commercial mm-Wave MMICs," *Microwave Magazine, IEEE*, vol. 6, Issue: 4, pp. 34-43, 2005
- [3] A. H. Pawlikiewics, and S. E. Rai, "RF CMOS or SiGe BiCMOS in RF and Mixed Signal Circuit Design", *Mixed Design of Integrated Circuits and Systems Digest of Papers*, pp. 333-338, Jun. 2007
- [4] A. J. Joseph, D. L. Hamee, B. Jagannathan, D. Coolbaugh, D. Ahlgren, J. Magerlein, L. Lanzerotti, N. Feilchenfeld, S. Oge, J. Dunn, and E. Nowak, "Status and Direction of Communication Technologies SiGe BiCMOS and RFCMOS," *Proceedings of the IEEE* , vol. 93, Issue: 0, pp. 1539-1558, 2005
- [5] Ali M. Niknejad, S. Emami, B. Heydari, M. Bohsali, E. Adabi "Nanoscale CMOS for mm-Wave Applications", *2007 IEEE Compound Semiconductor Integrated Circuit Symposium, IEEE*, 1-4, October, 2007
- [6] P. D. Munday, J. Powell, D. Bannister, P. J. Rice, "The Development of Affordable Front-End Hardware for mm-Wave Imaging Using Multi-Layer Softboard Technology," *Proceedings of the SPIE*, vol. 6548, pp. 65480G, 2007.
- [7] S. P. Voinigescu, T. O. Dickson, R. Beerkens, I. Khalid, and P. Westergard, "A Comparison of Si CMOS, SiGe BiCMOS, and InP HBT Technologies for High-Speed and Millimeter-Wave ICs," *Silicon Monolithic Integrated Circuits in RF Systems, Digest of Papers*, pp. 111-114, Sept. 2004
- [8] R. D. Isaac, "The Future of CMOS Technology," *IBM J. Res. Develop.*, vol. 44, no. 30, pp. 369-378, May 2000.
- [9] S. Lee, L. Wagner, B. Jagannathan, S. Csutak, J. Pekarik, N. Zamdmer, M. Breitwisch, R. Ramachandran, and Greg Freeman, "Record RF Performance of Sub-46 nm Lgate NFETs in Microprocessor SOI CMOS Technologies," *IEDM proceeding*, pp. 241-244, Dec. 2005.

- [10] B. Gaucher, T. Beukema, S. Reynolds, B. Floyd, T. Zwick, U. Pfeiffer, D. Liu, J. Cressler, "MM-Wave Transceivers Using SiGe HBT Technology," *Silicon Monolithic Integrated Circuits in RF Systems Digest of Papers.*, pp. 81-84, Sept. 2004
- [11] G. Gonzalez, "Microwave Transistor Amplifiers", 2nd Edition , Prentice-Hall, 1996
- [12] D.M. Pozar, "Microwave Engineering", 3rd Edition, Wiley ,2004
- [13] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, "Analysis and Design of Analog Integrated Circuits", John Wiley and Sons, 2001.
- [14] J. M. Rabaey, A. P. Chandrakasan, B. Nikolic, "Digital Integrated Circuits", second edition, Prentice Hall, 2002
- [15] A.M.Niknejad, "Electromagnetics for High-Speed Analog and Digital Communication Circuits",1st Edition. Cambridge press, 2007
- [16] B. Razavi, "RF Microelectronics", First Edition, Prentice Hall PTR, Nov. 1997
- [17] T. H. Lee, "The Design of CMOS Radio-Frequency Integrated Circuits", second edition, Cambridge University Press, 2003
- [18] G. D. Vandelin, A. M. Pavio, and U. L. Rohde, "Microwave Circuit Design, using linear and non-linear techniques",second edition, Wiley, 2005
- [19] K. K. Clarke, and D. T. Hess, "Communication circuits : analysis and design", third edition, Krieger, 2002
- [20] B. L. Anderson, and R. L. Anderson, "Fundamentals of Semiconductor Devices", first edition, McGraw-Hill, 2005
- [21] S. Ramo, J. R. Whinnery, and T. Van Duzer, "Fields and Waves in Communication electronics", third edition, Wiley 1993
- [22] B. Razavi, "Design of Analog CMOS Integrated Circuits", first edition, McGraw-Hill, 2001
- [23] A. M. Niknejad, and H. Hashemi, "mm-Wave Silicon Technology: 60 GHz and Beyond", Springer, 2008
- [24] M. Skolnik, "Introduction to Radar Systems", third edition, McGraw-Hill, 2003
- [25] I. V. Komarov, and S. M. Smolskiy, "Fundamentals of Short-Range Fm Radar", Artech House Publishers, 2003
- [26] J. D. Jackson, "Classical electrodynamics", third edition, Wiley 1998

- [27] D. H. Johnson, and D. E. Dudgeon, "Array Signal Processing", Prentice Hall, 1993
- [28] R. C. Hansen, "Phased Array Antennas", second edition, Wiley, 2009
- [29] C. H. Doan, S. Emami, A. M. Niknejad, R. W. Brodersen, "Millimeter-wave CMOS Design," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 144-155, Jan. 2005.
- [30] A. Babakhani, X. Guan, A. Komijani, A. Natarajan and A. Hajimiri "A 77-GHz Phased-Array Transceiver With On-Chip Antennas in Silicon: Receiver and Antennas," *IEEE J. Solid-State Circuits*, vol. 41, NO. 12, Dec. 2006.
- [31] N. A. Talkwalker, C. P. Yue, H. Gan, and S. S. Wong, "Integrated CMOS transmit-receive switch using LC-tuned substrate bias for 2.4GHz and 5.2GHz applications ,"*IEEE J. Solid-State Circuits*, vol. 39, pp. 863-870, June. 2004.
- [32] F. Huang and K. K. O., "A 0.5 μ m CMOS T/R switch for 900MHz wireless applications ,"*IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 486-492, March. 2001.
- [33] C. M. Ta and R. J. Evans, "A 60 GHz CMOS transmit/receive switch," *IEEE RFIC Symp. Dig.*, pp. 725-728, June 2007.
- [34] P. Park, D. H. Shin, J. J. Pekarik, M. Rodwell and C. P. Yue, "A high-linearity, LC-tuned, 24GHz T/R Switch in 90nm CMOS," *IEEE RFIC Symp. Dig.*, pp. 369-372, June 2008.
- [35] M. Y. Chia, T. Lim, J. Yin, P. Chee, S. Leong, and C. Sim, "Electronic Beam-Steering Design for UWB Phased Array," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 6, pp. 2431-2438, Jun. 2006.
- [36] B. Analui, and A. Hajimiri, "Statistical Analysis of Integrated Passive Delay Lines," *Proceedings of CICC.*, pp. 107-110, Sept. 2003.
- [37] T. Chu, J. Roderick, and H. Hashemi, "An Integrated Ultra-Wideband Timed Array Receiver in 0.13 μ m CMOS Using a Path-Sharing True Time Delay Architecture," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 201-204, Dec. 2007.
- [38] T. M. Hancock, and G. M. Rebeiz, "A 12-GHz SiGe Phase Shifter With Integrated LNA," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 3, pp. 977-983, Mar. 2005.
- [39] F. Ellinger, H. Jackel, and W. Bachtold, "Varactor-Loaded Transmission-Line Phase Shifter at C-Band Using Lumped Elements," *IEEE Trans. Microw. Theory Tech.*, vol. 51, no. 4, April 2003.
- [40] F. Ellinger, "26-42GHz SOI CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 522-528, Mar. 2004.

- [41] J.-H. Tsai, W.-C. Chen, T.-P. Wang, T.-W. Huang, and H. Wang, "A miniature Q-band low noise amplifier using 0.13- μ m CMOS technology," *IEEE Microwave and Wireless Component Letters*, vol. 16, no. 6, June 2006.
- [42] M. A. Masud, H. Zirath, M. Fendahl, and H.-O. Viques, "90nm CMOS MMIC amplifier," *IEEE RFIC Symp. Dig.*, pp. 201-204, June 2004.
- [43] H. Shigematsu, T. Hirose, F. Brewer, and M. Rodwell, "Millimeter CMOS circuit design," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 2, pp. 472-477, Feb. 2005.
- [44] C. H. Doan, S. Emami, A. M. Niknejad, R. W. Brodersen, "Millimeter-wave CMOS Design," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 144-155, Jan. 2005.
- [45] M.A.T.Sanduleanu, G. Zhang, and J. R. Long, "31-34GHz Low Noise Amplifier with On-chip microstrip Lines and Inter-stage Matching in 90-nm Baseline CMOS ," *IEEE RFIC symposium* , June 2006.
- [46] Y. SU, and Kenneth K. O., "An 800- μ W 26-GHz CMOS Tuned Amplifier," *IEEE RFIC symposium*, June 2006.
- [47] T. Yao, M. Gordon, K. Yau, M. T. Yang and S. P. Voinigescu "60GHz PA and LNA in 90nm-CMOS ," *IEEE RFIC symposium*, June 2006.
- [48] B. Heydari, M. Bohsali, E. Adabi, and A. Niknejad "Low-Power mm-Wave components up to 104GHz in 90nm CMOS ," *to appear in ISSCC*, Feb. 2006.
- [49] D. K. Shaeffer and T. H. Lee, "A 1.5-V 1.5-GHz CMOS Low Noise Amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745-759, May. 1997.
- [50] B. Afshar, A. M. Niknejad, "X/Ku Band CMOS LNA Design Techniques," *Proceedings of CICC*, 2006, pp. 389-392.
- [51] T. Nguyen, C. Kim, G. Ihm, M. Yang, and S. Lee, "CMOS Low-Noise Amplifier Design Optimization Techniques," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, pp. 1433-1442, May. 2004.
- [52] E. Adabi, A. M. Niknejad, "A mm-Wave Transformer Based Transmit/Receive Switch in 90nm CMOS Technology, EuMW, September, 2009.
- [53] C. Marcu, D. Chowdhury, C. Thakkar, L.-K. Kong, M. Tabesh, J.-D. Park, Y. Wang, B. Afshar, A. Gupta, A. Arbabian, S. Gambini, R. Zamani, A. M. Niknejad, E. Alon, "A 90nm CMOS low-power 60GHz transceiver with integrated baseband circuitry," *ISSCC Dig. Tech. Papers*, pp. 314-315, Feb. 2009.

- [54] B. Heydari, P. Reynaert, E. Adabi, M. Bohsali, B. Afshar, M. A. Arbabian, A. M. Niknejad, "A 60-GHz 90-nm CMOS cascode amplifier with interstage matching," *Microwave Integrated Circuit Conference (EuMIC)*, Oct. 2007, pp. 88-91.
- [55] B. Heydari, M. Bohsali, E. Adabi, A. M. Niknejad, "A 60 GHz Power Amplifier in 90nm CMOS Technology," *Custom Integrated Circuits Conference (CICC)*, 2007, Sept. 2007, pp. 769-772.
- [56] E. Adabi, A. M. Niknejad, "CMOS Low Noise Amplifier with Capacitive Feedback Matching," *Custom Integrated Circuits Conference (CICC)*, 2007, Sept. 2007, pp. 643-646.
- [57] D. Chowdhury, P. Reynaert, A. M. Niknejad, "A 60GHz 1V +12.3dBm Transformer-Coupled Wideband PA in 90nm CMOS", *ISSCC Dig. Tech. Papers*, pp. 560-561, Feb. 2008.
- [58] B. Heydari, E. Adabi, M. Bohsali, B. Afshar, M. A. Arbabian, A. M. Niknejad, "Internal Unilateralization Technique for CMOS mm-Wave Amplifiers, *RFIC Digest of Papers*, June 2007, pp. 463-466.
- [59] E. Adabi, B. Heydari, M. Bohsali and A. M. Niknejad, "30 GHz CMOS Low Noise Amplifier, *RFIC Digest of Papers*, June 2007, pp. 625-628
- [60] A. Arbabian, B. Afshar, J.-C. Chien, S. Kang, S. Callender, E. Adabi, S. Dal Toso, R. Pilard, D. Gloria, A. M. Niknejad, "A 90GHz-Carrier 30GHz-Bandwidth Hybrid Switching Transmitter with Integrated Antenna, *ISSCC Dig. Tech. Papers* pp. 420-421, Feb. 2008 .
- [61] A. M. Niknejad, M. Bohsali, E. Adabi, B. Heydari, "Integrated circuit transmission-line transformer power combiner for millimetre-wave applications," *Electronic Letters*, 43, 5, 290-291, March, 2007
- [62] I. Aoki, S. D. Kee, D. B. Rutledge, and A. Hajimiri, "Distributed active transformer-a new power-combining and impedance-transformation technique," *Microwave Theory and Techniques*, vol. 50, issue 1, pp 316-331, Jan. 2002
- [63] A. Shirvani, D. K. Su, and B. A. Wooley, "A CMOS RF power amplifier with parallel amplification for efficient power control," *IEEE J. Solid-State Circuits*, vol. 37, pp. 684-693, Jun. 2002
- [64] B. Cetinoneri, Y. A. Atesal, and G. M. Rebeiz, "A miniature DC-70 GHz SP4T switch in 0.13- μ m CMOS," *Microwave Symposium Digest*, pp 1093-1096, Jul. 2009
- [65] K. J. Koh, and G. M. Rebeiz, "An X- and Ku-Band 8-Element Phased-Array Receiver in 0.18- μ m SiGe BiCMOS Technology," *IEEE J. Solid-State Circuits*, vol. 43, pp. 1360-1371, Jun. 2008

- [66] J. R. Long, "Monolithic transformers for silicon RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1368-1382, Aug. 2002
- [67] T. S. D. Cheung, J. R. Long, "Shielded passive devices for silicon-based monolithic microwave and millimeter-wave integrated circuits," *IEEE J. Solid-State Circuits*, vol. 41, pp. 1183-1200, Apr. 2006
- [68] W. L. Chan, J. R. Long, "A 56–65 GHz Injection-Locked Frequency Tripler With Quadrature Outputs in 90-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, pp. 2739-2746, Dec. 2008
- [69] J. Paramesh, R. Bishop, K. Soumyanath, and D. J. Allstot, "A four-antenna receiver in 90-nm CMOS for beamforming and spatial diversity," *IEEE J. Solid-State Circuits*, vol. 40, pp. 2515-2524, Dec. 2005
- [70] X. Guan, H. Hashemi, A. Hajimiri, "A fully integrated 24-GHz eight-element phased-array receiver in silicon," *IEEE J. Solid-State Circuits*, vol. 39, pp. 2311-2320, Nov. 2004
- [71] A. Natarajan, B. Floyd, and A. Hajimiri, "A Bidirectional RF-Combining 60GHz Phased-Array Front-End," *ISSCC Dig. Tech. Papers* pp. 202-204, Feb. 2007 .
- [72] J. Roderick, H. Krishnaswamy, K. Newton, and H. Hashemi, "Silicon-Based Ultra-Wideband Beam-Forming," *IEEE J. Solid-State Circuits*, vol. 41, pp. 1726-1739, Jul. 2006
- [73] T. S. Chu, J. Roderick, and H. Hashemi, "An Integrated Ultra-Wideband Timed Array Receiver in 0.13 μ m CMOS Using a Path-Sharing True Time Delay Architecture," *IEEE J. Solid-State Circuits*, vol. 42, pp. 2834-2850, Dec. 2007
- [74] A. Parsa, and B. Razavi, "A New Transceiver Architecture for the 60-GHz Band," *IEEE J. Solid-State Circuits*, vol. 44, pp. 751-762, March 2009
- [75] B. Razavi, "A 60-GHz CMOS receiver front-end," *IEEE J. Solid-State Circuits*, vol. 41, pp. 17-22, Jan. 2006
- [76] S. Alalusi, and R. W. Brodersen, "A 60GHz Phased Array in CMOS," *CICC Digest of Papers*, Sepy. 2006, pp. 393-396
- [77] J. M. Gilbert, C. H. Doan, S. Emami, C. B. Shung, "A 4-Gbps Uncompressed Wireless HD A/V Transceiver Chipset," *Micro, IEEE*, vol. 28, pp. 56-64, March 2008