Self-healing Techniques for RF and mm-Wave Transmitters and Receivers

Thesis by

Kaushik Dasgupta

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended October 27, 2014)

To Ma and Baba....

Acknowledgements

Pursuing a doctoral degree is a long but rewarding endeavor and is often not possible without the help and support of numerous individuals. To begin with, I would like to thank my thesis supervisor, Professor Ali Hajimiri, who has always been a source of inspiration for me. He has encouraged me to pursue new and challenging ideas while simultaneously infusing confidence in me to push my own boundaries. While being the best technical adviser one could ever ask for, he has also helped me a great deal on a personal level with invaluable advice, discussions and debates on a wide variety of topics including ethics, philosophy, travel as well as technology, culture, books, and movies. He is one of the very few advisers who places equal importance on both professional excellence as well as personal development and I am thankful for that. As anyone from the CHIC laboratory would know, Ali has a unique way of operating his lab which promotes a sense of togetherness as well as mutual respect among all the lab members, a feeling often absent in most of the other groups I have interacted with. It has been my privilege to be part of such a rich academic heritage, and I am sure that the things I have learnt from him professionally and personally will last a lifetime.

I would like to extend my gratitude towards my thesis committee members: Prof. Azita Emami, Prof. Sander Weinreb, Prof. David Rutledge, and Prof. Hyuck Choo. Time and again they have provided me with invaluable advice and feedback on my work here at Caltech. I am indebted to Prof. Emami and Prof. Weinreb for allowing me to use their laboratory facilities as well as equipment.

I am thankful to my labmates Steve and Amir for their friendship as well as for the countless technical discussions from which I have learnt so much over the past few years. We have had so much fun over board games, sailing, and wine-tasting, as well as some not

so fun sleepless nights during the several tapeouts. I would also like to thank Kaushik (K_1) and Behrooz who have been great labmates and from both of whom I have learnt a great deal. Special thanks to Behrooz for keeping me motivated to go to the gym. I would also like to acknowledge all past and present members of the CHIC laboratory whom I have had the pleasure of working with, including Alex, Costis, Florian, Hua, Joe, Ed, Firooz, Tomoyuki, Aroutin, and Brian. I am proud to be in the company of such talented and motivated individuals during my stay here. Past and present members of the MICS lab have also helped me a lot over the years; this includes Mayank, Manuel, Meisam, Saman, Juhwan, Matt, Krishna, and Abhinav, and I thank them for their support. Of course none of this would have been possible without the continued help and support of Michelle, our group administrator, as well as Tanya and Carol. Outside of the lab, I have had the pleasure of knowing two of my closest friends, Bose and Hemanth. Several 'adventures' come to mind including the impromptu road-trips, photographic 'expeditions', and last but not the least trying out unknown food spots all over the city. Thanks to all of my other friends both at Caltech as well as outside of it.

Finally I would like to extend my gratitude towards my parents for their unconditional love and support in whatever I have pursued in my life. Without their motivation and patience none of this would have been possible. I would also like to thank my sister for her constant encouragement.

Abstract

With continuing advances in CMOS technology, feature sizes of modern Silicon chip-sets have gone down drastically over the past decade. In addition to desktops and laptop processors, a vast majority of these chips are also being deployed in mobile communication devices like smart-phones and tablets, where multiple radio-frequency integrated circuits (RFICs) must be integrated into one device to cater to a wide variety of applications such as Wi-Fi, Bluetooth, NFC, wireless charging, etc. While a small feature size enables higher integration levels leading to billions of transistors co-existing on a single chip, it also makes these Silicon ICs more susceptible to variations. A part of these variations can be attributed to the manufacturing process itself, particularly due to the stringent dimensional tolerances associated with the lithographic steps in modern processes. Additionally, RF or millimeter-wave communication chip-sets are subject to another type of variation caused by dynamic changes in the operating environment. Another bottleneck in the development of high performance RF/mm-wave Silicon ICs is the lack of accurate analog/high-frequency models in nanometer CMOS processes. This can be primarily attributed to the fact that most cutting edge processes are geared towards digital system implementation and as such there is little model-to-hardware correlation at RF frequencies.

All these issues have significantly degraded yield of high performance mm-wave and RF CMOS systems which often require multiple trial-and-error based Silicon validations, thereby incurring additional production costs. This dissertation proposes a low overhead technique which attempts to counter the detrimental effects of these variations, thereby improving both performance and yield of chips post fabrication in a systematic way. The key idea behind this approach is to dynamically "*sense*" the performance of the system, identify when a

problem has occurred, and then "*actuate*" it back to its desired performance level through an intelligent on-chip optimization algorithm. We term this technique as **self-healing** drawing inspiration from nature's own way of healing the body against adverse environmental effects. To effectively demonstrate the efficacy of self-healing in CMOS systems, several representative examples are designed, fabricated, and measured against a variety of operating conditions.

We demonstrate a high-power mm-wave segmented power mixer array based transmitter architecture that is capable of generating high-speed and non-constant envelope modulations at higher efficiencies compared to existing conventional designs. We then incorporate several sensors and actuators into the design and demonstrate closed-loop healing against a wide variety of non-ideal operating conditions. We also demonstrate fully-integrated self-healing in the context of another mm-wave power amplifier, where measurements were performed across several chips, showing significant improvements in performance as well as reduced variability in the presence of process variations and load impedance mismatch, as well as catastrophic transistor failure. Finally, on the receiver side, a closed-loop self-healing phase synthesis scheme is demonstrated in conjunction with a wide-band voltage controlled oscillator to generate phase shifter local oscillator (LO) signals for a phased array receiver. The system is shown to heal against non-idealities in the LO signal generation and distribution, significantly reducing phase errors across a wide range of frequencies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 CMOS RF and millimeter-Wave Systems

The past decade has witnessed an unprecedented growth in the mobile wireless market with an estimated 1.9 billion mobile devices shipped in 2014 alone. In fact, by 2018, mobile data traffic is expected to increase 11 fold based on a 61% annual growth rate. This ever increasing demand in high data rate communication has stimulated development not only in the short range Wi-Fi space but also in emerging long range LTE/cellular networks. The most modern cellular handsets today have a plethora of sophisticated chips dedicated to multiple applications — cellular, Wi-Fi, Bluetooth, NFC, GPS, etc. More advanced and spectrally efficient modulation schemes have seen rapid deployment particularly in Wi-Fi applications to improve data rates in an indoor setting. Additionally, millimeter-wave based 5-G mobile chipsets are also being developed with predicted widespread usage by 2020. These technological advancements have primarily been enabled by the continuing development of Complementary Metal Oxide Semiconductor (CMOS) technology providing progressively smaller and faster transistors over the years, leading to low cost chipsets with billions of transistors per chip.

1.2 CMOS Scaling

In the past 40 years, CMOS transistors have continued to scale aggressively almost exactly in accordance with Moore's law [1]. Figure 1.1 shows how processors (CPU/GPUs) have continued to scale over the past few decades roughly in accordance with Moore's law, with present generation processors having well above a billion transistors. Both the footprint of these

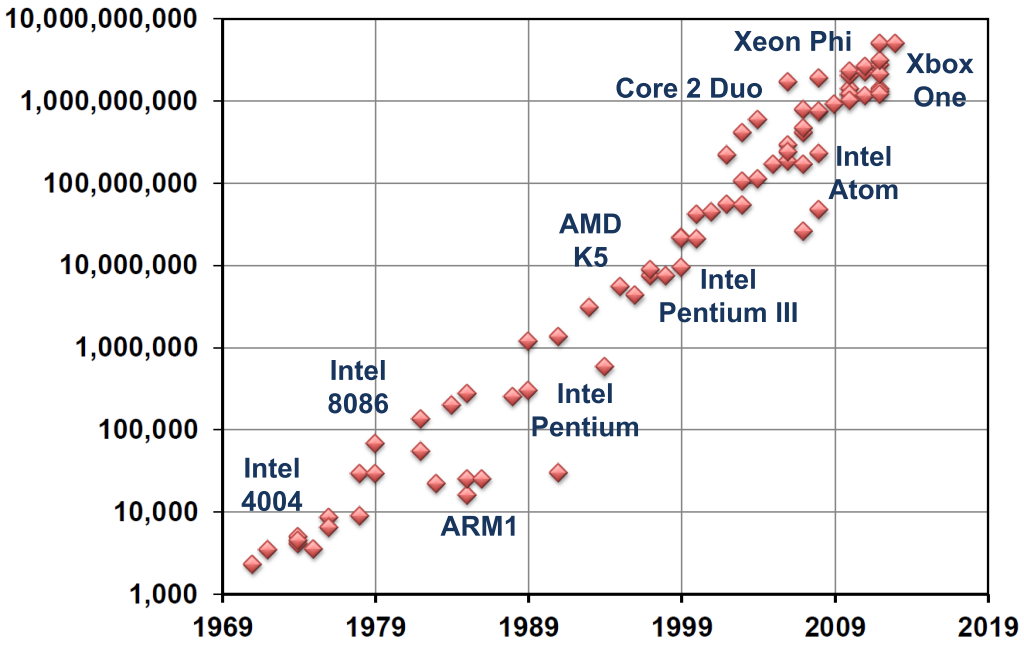


Figure 1.1: Transistor count progression over the past four decades.

CMOS chips as well as the cost/transistor of these processes have gone down dramatically over the past few technology generations (Figure 1.2).

However, more recently this method of scaling has become less effective as features are moving closer to the fundamental dimensions and performance enhancing techniques like strained Si [3] [4], high-k metal-gate [5], as well as tri-gate [6] devices are widely being adopted in the fabrication industry.

The origins of Moore's law lie in a scaling law proposed by Dennard in [7], where, in order to make an equivalent device for a smaller technology node (channel length), three variables need to be transformed accordingly: dimension, voltage, and doping. All linear dimensions,

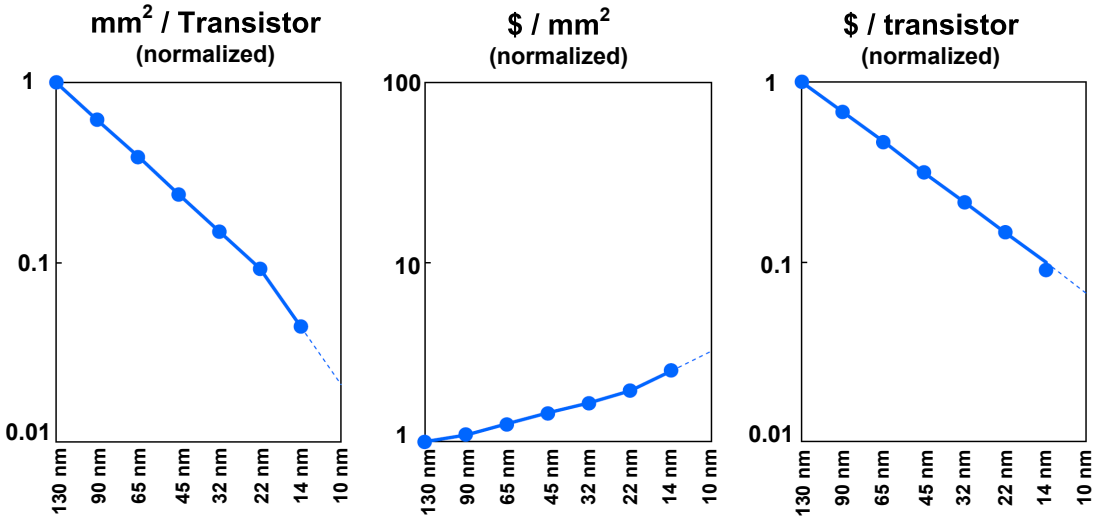


Figure 1.2: Cost and area scaling per transistor over the past few technology generations [2].

including vertical dimensions such as junction depth as well as horizontal dimensions like channel width, must be scaled down by a factor κ, for example, $t'_{ox} = t_{ox}/\kappa$. The device operating voltage must also be scaled down by the same factor as $V'_{DS} = V_{DS}/\kappa$. Finally, the substrate doping concentration must also be increased by the same factor: $N'_a = N_a.\kappa$.

We can immediately conclude that the gate-oxide capacitance scales linearly with κ. As shown in Equation 1.1, this type of scaling also reduces propagation delay and hence improves speed.

$$\tau'_{ox} \propto \frac{L^2_{min} V_{DD}}{(V_{DD} - V_{th})^2} = \tau/\kappa \tag{1.1}$$

If we look at the dynamic power dissipation, it scales *down* by a factor of κ^2 (Equation 1.2). This also implies that the power density stays unchanged as we scale dimensions.

$$P'_{dynamic} \propto W.L'_{min}.C'_{ox}.V^2_{DD}.f = P_{dynamic}/\kappa^2 \tag{1.2}$$

However, as scaling continues, velocity saturation effects come into play. Using expressions for delay and power dissipation for velocity saturated devices (Equation 1.3, 1.4), it

is evident that there is no difference in the scaling behavior over the one for long-channel devices.

$$\tau'_{ox} \propto \frac{L_{min} V_{DD}}{(V_{DD} - V_{th})} = \tau/\kappa \tag{1.3}$$

$$P'_{dynamic} \propto W.L'_{min}.C'_{ox}.V^2_{DD}.f = P_{dynamic}/\kappa^2 \tag{1.4}$$

It is clear that to leverage the performance benefits of constant field scaling, the transistor threshold voltage also needs to be scaled in the same way. This assumption is not strictly true, since the threshold voltage is eventually governed by the sub-threshold slope of the transistor, which is in turn limited by the thermal voltage kT/q. In addition, the threshold voltage choice is also limited by the static power dissipation, or in other words the off-state current of the device.

1.3 Variations in CMOS

1.3.1 Process and Mismatch

In addition to the performance benefits, CMOS scaling also comes with trade-offs in terms of variations both between chips as well as between transistors on the same chip. Such variations have been extensively discussed in [8–11]. One of the two major sources of these variations is random dopant fluctuations (RDF), which deals with variations in doping number in the channel of a MOSFET. In nm-scale CMOS devices, the placement of dopant atoms in the channel becomes critical to the performance of the device, especially because only a handful of such atoms can be accomodated in such small linear dimensions, as shown in the trends in Figure 1.3 (a). Figure 1.3 (b) shows simulated potential distribution in a 35nm MOSFET subject to RDF. Besides the random positioning, the actual number of dopant atoms present in the channel region can also vary. It has been experimentally demonstrated that the V_{th} fluctuation is mainly caused by depletion layer charge fluctuation and can be approximated

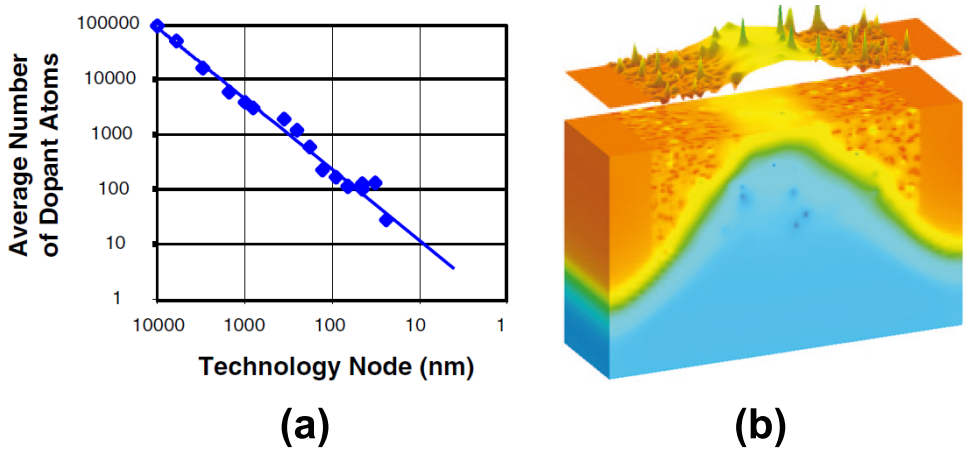


Figure 1.3: (a) Average dopant atoms as a function of technology node [12] and (b) Potential distribution in a MOSFET subject to RDF [13].

by a Gaussian function. In addition, it can be shown that the standard deviation of the threshold voltage variation depends inversely on the square root of the channel length, as shown in Equation 1.5.

$$\sigma_{V_{th}} = \frac{q}{C_{inv}} \cdot \sqrt{\frac{N_{sub} W_{dep}}{3LW}} \tag{1.5}$$

As an example, Figure 1.4 shows threshold voltage histogram for 1000 Monte Carlo runs, including process variations and mismatch. The standard deviation of this threshold voltage variation is simulated to be 34.6mV, which is about 20% of the mean V_{th} of 177mV.

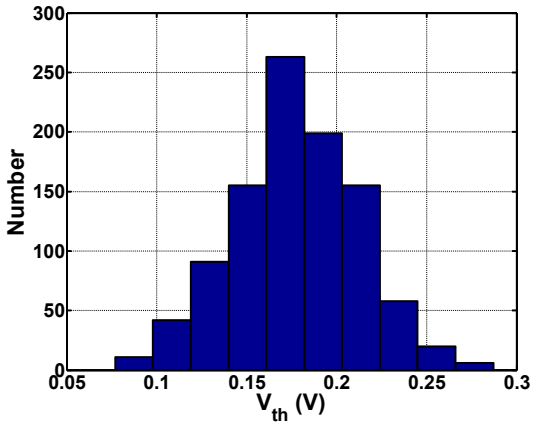


Figure 1.4: Threshold voltage variation for 1000 Monte Carlo runs in IBM's 32nm SOI CMOS process.

In addition to RDF, the other main factor contributing to CMOS variations is Line edge roughness (LER) caused by lithographic and etching steps. This effect is caused by the variations in the number of incident photons during lithographic exposure and is also due to the variations in the molecular composition of the photoresist itself, which in turn affects the overall reaction kinetics. Figure 1.5 shows SEM images of fabricated lines in a nm CMOS process and the width variation caused by LER.

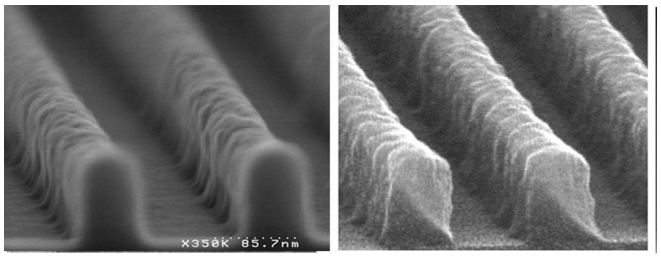


Figure 1.5: Line edge roughness shown in SEM images from an example process [14].

Figure 1.6 shows the threshold voltage variation versus process technology node. As is evident, this variation is much more manageable at larger nodes, and the variation is expected to continue to increase for smaller nodes.

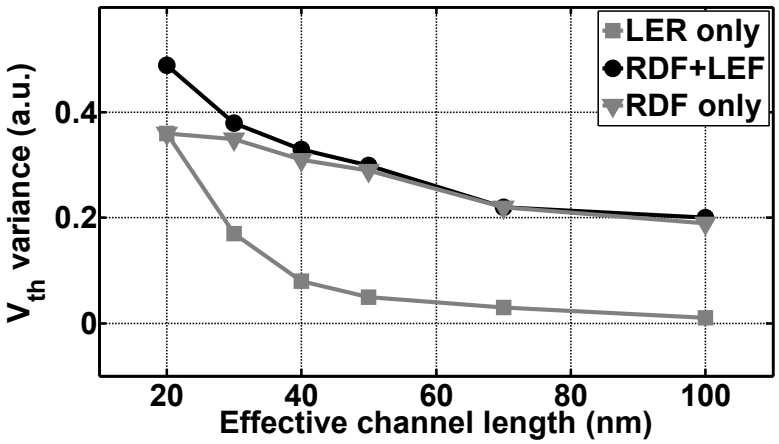


Figure 1.6: Combined effects of RDF and LER on threshold voltage variation.

The line-edge-roughness also causes variations in the separation between two adjacent metal lines, which leads to capacitance variation. A popular choice in CMOS RF design is a MOM capacitor where the capacitance value relies on the separation between two adjacent plates. As shown in Figure 1.7, LER can cause capacitance variation, which can cause severe design challenges in high-frequency/mm-wave design.

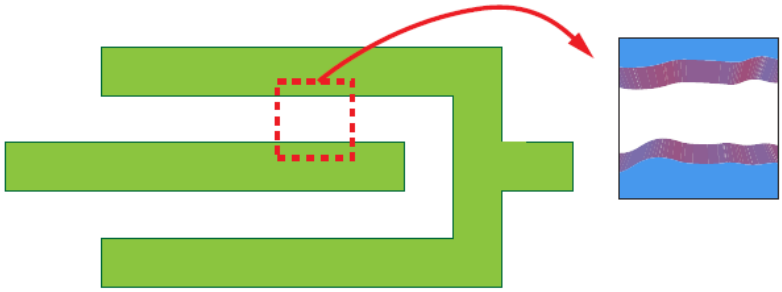


Figure 1.7: LER causing capacitance variation [15].

In addition to RDF and LER, some other sources of variations also dictate yield of modern CMOS processes. Oxide thickness variation (OTV) is caused by surface roughness at the Si/SiO_2 interface. Similar to LER, OTV can also cause capacitance variations in C_{gs}, C_{gd}, as well as other overlap capacitances and metal-to-metal capacitance changes in MOM capacitors, transmission lines, transformers, etc.

1.3.2 Dynamic Variations

Dynamic variations encompass all time varying changes in the operating conditions of the CMOS IC. These may include thermal variations — long or short term, transistor aging, electro-migration and in case of power amplifiers that are driving antennas, load impedance variation caused by voltage standing wave ration (VSWR) events.

The heat generated due to device activity and interconnect losses may heat up the substrate or the package. This raises the ambient temperature of operation of the circuit which can adversely affect the performance of the system. Depending on the activity of the circuit block, different parts of the same die may have vastly different thermal profiles. This

often leads to variation in delay and power consumption of circuits across the same chip. As a reference, Figure 1.8 shows variation in die temperature across an Intel Itanium processor in a 180nm CMOS process. The core area consumes maximum power, leading to an approximately 50^{o}C temperature difference between itself and other parts of the chip.

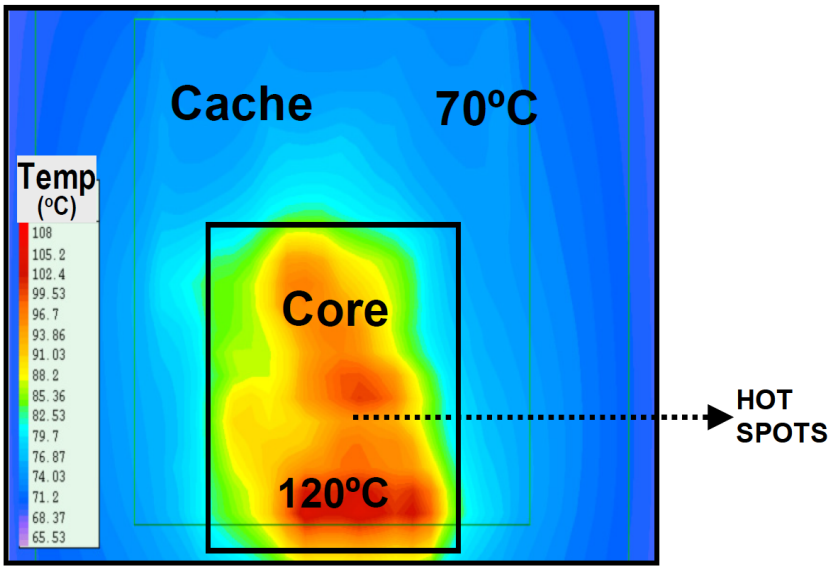


Figure 1.8: Thermal profile across an Intel Itanium processor.

Due to prolonged operation of the integrated circuit at high temperature or high voltages, the transistors themselves may degrade over time. These effects generally follow a bathtub curve, where the failures, after an initial fall, enter into a plateau region and then rise again as time progresses. There are several contributing mechanisms contributing to aging: hot carrier injection (HCI), time-dependent dielectric breakdown (TDDB), bias temperature instability (BTI), and electromigration.

Hot carrier effects occur when particles accelerated by a high electric field get injected into the gate oxide and stay trapped there, causing interface states. This leads to shifts in V_{th} and other characteristics of the transistor over time. It can be shown that the HCI based shift follows a power law versus the stress time. TDDB is an irreversible gate oxide reliability phenomenon where the conductance of the gate oxide increases suddenly and leakage current through the gate oxide increases over time. The time to this kind of breakdown can be

modeled using a Weibull distribution. Another effect contributing to long term failure of transistors is BTI. This is typically characterized by a V_{th} shift when a bias voltage is applied at the gate at higher temperatures. Eventually these shifts lead to degradation of carrier mobility and ultimately transistor failure. BTI is also known to be frequency as well as drain bias dependence.

Electromigration occuers in the interconnect wires and vias in any IC over time. It causes material transport due to momentum transfer between conducting electrons and metal atoms. This effect becomes particularly detrimental at high current densities, which is fast becoming a problem in modern CMOS processes where metal thicknesses are in the range of a few 100s of nm. The mean time to failure of a wire can be expressed in terms of the current density and the temperature, as shown in Equation 1.6 [16].

$$MTTF = \frac{A}{J^n} e^{\frac{\phi}{kT}} \tag{1.6}$$

where the constant A is dependent on the cross sectional area of the wire, J is the current density, ϕ is the activation energy for the metal, k is the Boltzmann constant, and T is the temperature. n is a scaling factor determined experimentally. As is evident, apart from the current density, the MTTF reduces dramatically at higher temperatures.

Today's integrated circuits, especially RF and mm-wave systems, may also be susceptible to short term environmental variations. This may come in the form of load impedance mismatch caused by VSWR events [17] [18] that occur when objects in the environment interact in the near field of the antenna. This may lead to lower signal-to-noise (SNR) in receivers as well as reduced power and efficiency in transmitting systems. It becomes particularly important to address this issue in phased array designs where the interaction between the array elements can significantly distort the overall array beamforming performance. Specifically for the case of power amplifiers, this mismatch may present a non-optimal load to the PA transistor leading to exceedingly high voltages, causing catastrophic transistor failure if care is not taken to "shield" the PA from extreme VSWR events. Figure 1.9 represents one such case where the variation in the antenna environment may cause the peak voltage to exceed

breakdown limits of the process.

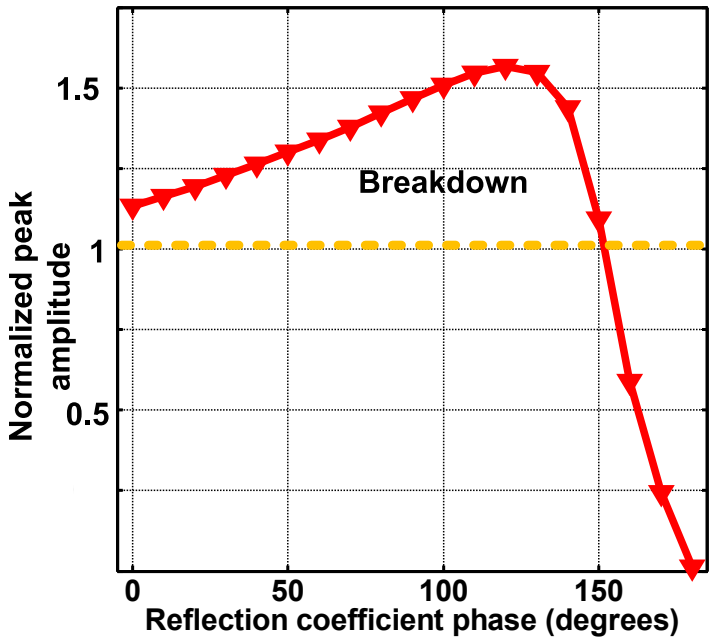


Figure 1.9: Peak voltage at the PA transistor when subject to load variations.

1.4 Countering Variations

1.4.1 Review of Existing Techniques

The commonly used approach to solve the issue of performance degradation due to variations is to design more variation tolerant systems. This entails adopting architectures and circuit topologies, which are less sensitive to process and mismatch variations. These techniques have been widely adopted in CMOS analog and digital designs over the years.

[19] discusses a new technique where there exists a gate bias voltage for which the variations in carrier mobility compensates the (V_{GS}-V_{th}) variation when temperature fluctuates. At this optimum voltage, the drain current can become less dependent on temperature. In fact, to reduce variations for some processes, the supply voltage needs to be lowered from

nominal for optimum operation. In [20] a low frequency CMOS oscillator has been reported where a bias generator circuit generates a process dependent voltage, which is then utilized to control the VCO through a temperature compensating block. In [21], a robust CMOS operational amplifier is designed where a constant g_m stage is implemented without the necessity of matching between n and p-channel transistors. A tuning curve linearization technique and switched capacitor current source were used in [22] to reduce sensitivity of a GPS VCO to PVT variations. Adaptive body biasing was demonstrated in [23] to reduce leakage power in nm domain CMOS circuits.

These techniques discussed above are difficult if not impossible to implement at RF/mm-wave frequencies. Efforts have been made over the past several years to incorporate techniques like tunable matching network and adaptive biasing among others to improve tunability of RF transmitting/receiving frontends. In [24], a CMOS VCO has been demonstrated to be less sensitive to process corners using an iterative design technique. [25] proposes a tunable output and input matching network based LNA at 2.4GHz which can be used to compensate for optimum performance across process corners. The technique leads to significant improvement in variation against process and mismatch, as shown in Figure 1.10. A PVT tolerant LC-VCO was reported in [26] where a process dependent reference voltage

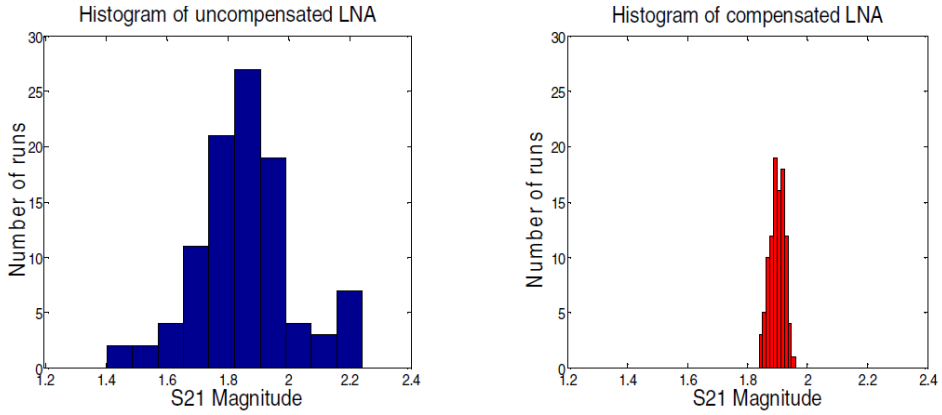


Figure 1.10: Reduction in standard deviation of LNA gain as reported in [25].

is used to control the supply regulator of the VCO, thereby reducing the sensitivity of the oscillation frequency on process variation. While these techniques have been demonstrated

to reduce PVT variations, their application is limited in high frequency RF/mm-wave CMOS design, where some of these techniques tend to severely affect the performance of the system. Moreover, the systems discussed above tend to be insensitive to only a few types of variations and often cannot counter dynamic or unexpected variations like aging and/or environmental variations. Thus, to truly build high-frequency systems that are tolerant against a wide variety of variations, both static and dynamic, we need a completely autonomous closed loop system. We discuss some of the salient aspects of this new approach in the following sub-section.

1.4.2 Self-healing to Counter Variations

A more scalable approach to deal with performance hits due to variations is to sense the degradation caused by variations and then change the system performance by using various knobs. The ability to dynamically sense and actuate critical high-frequency blocks of the system eliminates the additional design complexity of variation insensitive circuits/techniques. We term this closed-loop technique as self-healing. Figure 1.11 shows the basic block diagram of a representative RF/mm-wave self-healing system where data from sensors are processed using a digital processing block which then decides the optimum actuation settings.

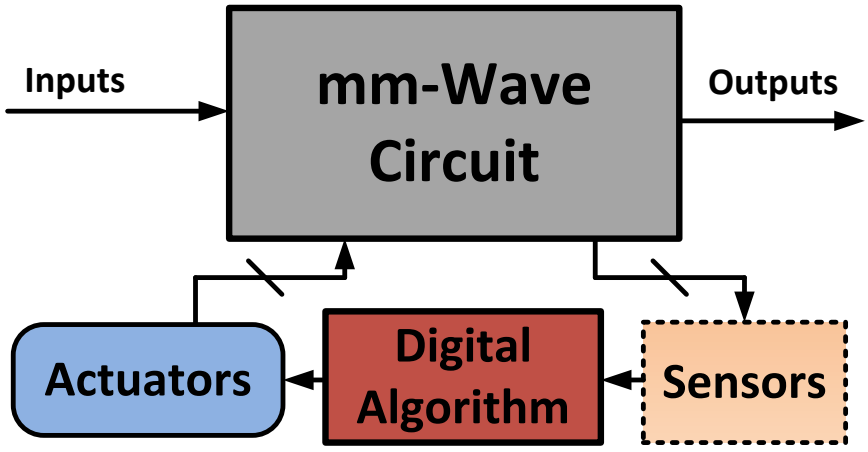


Figure 1.11: Conceptual self-healing mm-wave system.

Due to the digital driven aggressive scaling of CMOS in the past few years, most modern CMOS processes offer very limited high-frequency analog modeling of transistors. Iterative design based on measurements from such test transistors significantly slows down the design cycle and increases production costs. In addition, as discussed in the previous sections, the variations increase with further process scaling. An ideal closed-loop autonomous self-healing system should address both these problems, as shown in Figure 1.12, thereby improving overall yield.

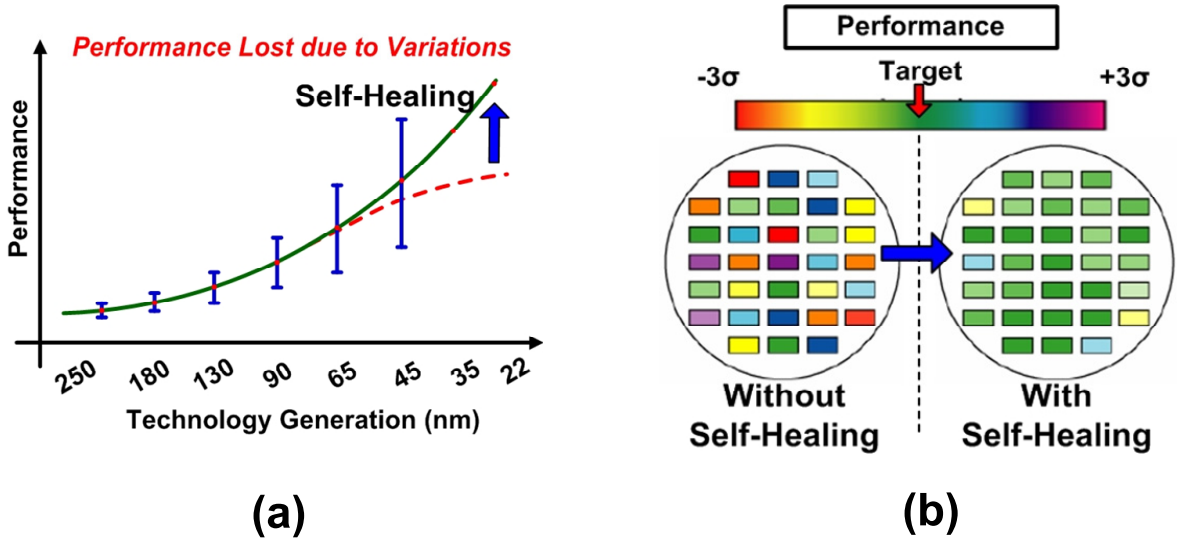


Figure 1.12: Self-healing (a) improving performance as well as reducing variations and (b) improving yield.

The self-healing loop involves integrated sensors that detect the performance of the mm-wave circuit. Several stringent requirements apply to these sensors, especially since they are subject to the same set of variations as the mm-wave circuit. Robustness of these sensors is especially critical, since their outputs should be true representations of the circuit performance and not dependent on variations within the sensor. To harness the vast digital processing power of modern CMOS processes, the self-healing algorithm must be implemented digitally. The analog output voltages of these sensors thus need to be converted to digital form using on-chip analog-to-digital converters (ADC) that send data to an integrated digital core. The

self-healing core then controls the RF system through digital to analog converters (DAC) that set the actuation points of the circuit. The actuators also need to cover a wide enough actuation space to be able to heal against unforeseen variations in the chip performance. Moreover, these actuators and sensors need to have minimum impact on the high-frequency performance of the system. This loop may be iterated until an optimum actuation state is reached.

Digital systems have deployed similar concepts primarily because sensors and actuators in the form of digital control can be easily integrated into the system without significant loss in performance. [27] describes a system which utilizes several adaptive processing units in addition to regular ones and periodically turns off the system and performs a functional verification. If necessary, the regular processing units are replaced with their adaptive counterparts to enable optimum system operation. The LNA in [28] incorporates a tunable matching network and a input return loss sensing mechanism in the form of a small valued resistor in the source of the input transistor to enable a closed loop system of input match auto-calibration. A reconfigurable test scheme is reported in [29], where the LNA performance is sensed by re-structuring it in a feedback configuration to produce an oscillation signal, which is then utilized as test signal to test the amplifier. Bias control and matching network tuning is used subsequently to heal the LNA back to its desired performance. As shown in Figure 1.13, the yield of a SiGe LNA was improved from 87% to 97% after self-healing using this technique.

A self-healing mm-wave power amplifier is reported in [30] where, based on output power levels, the PA bias is adjusted to provide constant gain. An extension of the work in [31] uses the adaptive biasing scheme to improve P-1dB of the 60GHz PA by 5.5dB from nominal. On-chip PVT compensation has been demonstrated in a 2.4GHz LNA [32], where the input of the LNA is switched between an off-chip and on-chip VCO during measurement and calibration phases, respectively.

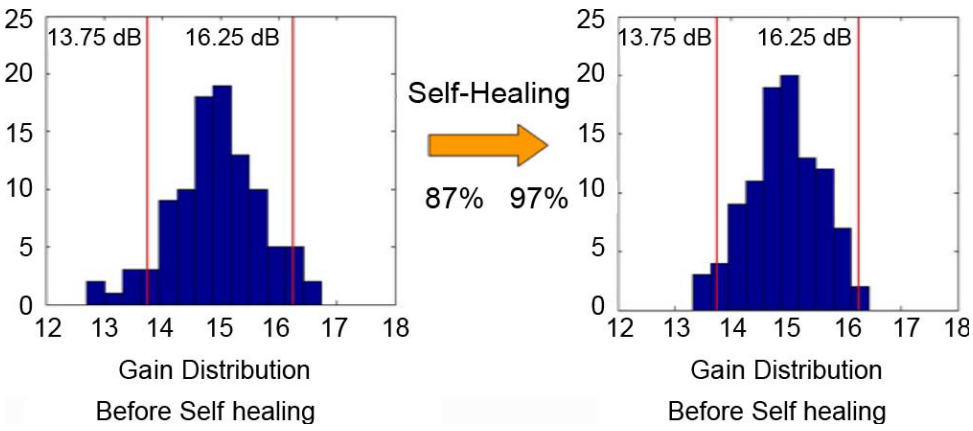


Figure 1.13: Self-healing to improve yield of a SiGe LNA [29].

1.5 Contributions

In this dissertation, we propose self-healing as a low-overhead technique to improve yield and simultaneously improve the performance of CMOS mm-wave systems in the nanometer regime. Using this technique, it is possible to overcome the limitations imposed by process variations, mismatch, and modeling inaccuracies, thereby improving the overall yield of an RF system. We demonstrate the usefulness of this design methodology using three representative examples.

Specifically we demonstrate:

- A fully integrated CMOS power mixer based transmitter in 32nm SOI CMOS capable of generating a high output power of 19.1 dBm at 51 GHz. The architecture enables the transmitter to produce several non-constant envelope modulation schemes at higher efficiencies compared to conventional architectures.

- Demonstration of self-healing in the above transmitter showing closed-loop healing against a variety of unforeseen conditions leading to overall performance improvement.

- Demonstration of direct digital modulation at mm-wave frequencies and at high data-rates for the above transmitter.

- Demonstration of self-healing in a 28 GHz mm-wave power amplifier in 45nm SOI

CMOS process, where fully integrated autonomous healing is shown to enhance overall performance significantly compared to default simulations, thereby improving yield.

- A similar design approach has been used to demonstrate healing in receivers as well, where closed-loop autonomous healing is shown to improve the phase accuracy for reliable LO generation in a wide-band tunable phased array receiver in 65nm CMOS.

1.6 Organization

The thesis is organized as follows. Chapter 2 introduces the building blocks of a mm-wave/RF self-healing system. Design considerations of the various blocks enabling a self-healing design, namely sensors, actuators, and data converters, have been discussed. Measurement results from several representative examples of the above mentioned blocks, implemented in nm CMOS processes, have also been presented.

Chapters 3, 4, and 5 describe three representative examples of self-healing systems in CMOS at RF/mm-wave frequencies. Chapter 3 first introduces a segmented power-mixer based mm-wave transmitter architecture capable of multi-Gbps wireless communication, implemented in a 32nm CMOS process. After a brief description of the implementation of the several blocks in the transmitter, several measurement results demonstrate the capability of the power-mixer array to handle non-constant envelope modulation at high data rates and at high output power without having to sacrifice efficiency to maintain sufficient linearity. We then discuss some of the building blocks designed to implement the self-healing capability before moving into the measurement results showing performance improvement with closed-loop healing.

Chapter 4 discusses the implementation details of another self-healing mm-wave PA, implemented in 45nm CMOS, where measurement results demonstrate healing not only across PVT variations, but also against catastrophic transistor failure and antenna impedance mismatch. Measurement results have been made from 20 chips and remarkable improvements in overall yield have been demonstrated.

Self-healing is also demonstrated in Chapter 5 in the form of closed-loop calibration of random and systematic phase offsets in a phase rotator for RF phased array applications in 65nm CMOS. A look-up table based approach is implemented, leading to significant improvements in phase accuracy over a wide range of frequencies. Some of the design aspects of a wide-band tunable voltage controlled oscillator (VCO) essential to such a system are also presented.

Chapter 6 discusses some of my other work which include high-power mm-wave PAs as well as polarization control and modulation in RF systems. An important aspect in any self-healing system is the on-chip digital ASIC controlling the sensors and actuators. For this purpose, the Appendix outlines some of the guidelines for designing a synthesizable digital library starting from standard cell design, layout, and characterization all the way to automated layout in a specific CMOS process. Such a digital library is critical to enabling fully integrated self-healing in CMOS transmitters/receivers.

Chapter 2

Enabling Self-Healing : Sensors, Actuators & Data Converters

2.1 Introduction

In this chapter we will discuss some of the design aspects as well as representative examples of the critical blocks required in any self-healing system: the sensors for dynamically sensing the performance of the system, actuators for adaptively reconfiguring the system to its optimum operating point, data converter units which convey information to and from the sensors/actuators, and the on-chip digital algorithm block.

2.2 Sensing Key Performance Metrics

The most vital aspect of sensor design for self-healing circuits is robustness. The mm-wave circuit is designed to be cutting edge, to push the envelope of possible performance given a particular process, and that means that variations can significantly degrade performance, especially when using minimum channel length transistors for the key designs. The sensors, on the other hand, can be designed more conservatively, with robust design topologies and other techniques such as using non-minimum length transistors. Also, because the self healing loop can be duty cycled by turning it ON only when required, the DC power requirements are far more relaxed than in the mm-wave circuit. This implies that robust sensor designs

that are more power hungry can still be acceptable in some circumstances. Various sensing metrics may be applicable based on the particular system. Conversion gain, noise figure, and input match may be important metrics to sense for a receiver, whereas output power and efficiency are critical performance parameters for a PA/transmitter. Depending on the metric, it can be categorized as bounded or un-bounded. For a PA, the user may want to continue to operate it at a specific output power which is insensitive to variations. In this case this is a bounded metric and the absolute value of the sensed metric is of utmost importance; so, care should be taken to minimize variations in the sensor itself. However, if it is desired the PA be operated at maximum efficiency, then only the relative value of the metric is important and variations in the sensor, as long as they stay monotonic, may be acceptable.

In the specific case of a power amplifier, the relevant parameters of interest include input and output power (and thus gain), DC power consumption, and linearity measurements such as P_{1dB}, OIP_3, IIP_3, spectral mask requirements, and error vector magnitude (EVM) of output modulated signal constellation. In this section, we will discuss the overview of sensor design, design considerations, and trade-offs, all of which are some design examples of low-over head sensor implementation for a mm-Wave PA amplifier. Specifically, we will discuss sensor design related to measurement of true RF and mm-Wave power (which takes into consideration load mismatches) and low-overhead DC current sensors. The application of such sensors is very broad in a transceiver design and could be applicable to various other circuit blocks such as LNA, mixers, oscillators, etc.

2.2.1 Sensor Requirements

There are several key requirements which a robust on-chip sensor must satisfy to be a viable candidate towards use in self-healing systems. **Sensor responsivity** is defined by the change in output of the sensor per unit change in the parameter which it senses. An ideal sensor must have the highest possible responsivity at the outset before any amplifying stages. **Sensitivity** of the sensor should also be very high in the presence of noise. This noise

may often include $1/f$ noise, since most sensor outputs are DC and special techniques like chopping and correlated double sampling may be essential to maximizing performance of such sensors. The sensors must also have sufficient **dynamic range** to enable self healing from a wide variety of unknown sources of variation. Sensors need to be **monotonic** across all possible variations so that the digital healing algorithm always converges. Depending on the metric being sensed, the **response time** of the sensors may need to be very small. In turn this speed will limit the overall speed of the self-healing loop itself. Lastly, as is evident, to keep the overall overhead due to self-healing low, the sensors need to have a **power and area overhead**.

2.2.2 RF Power Sensing

In a PA, sensing the input and output power provides information about the gain of the system in addition to the output power. Voltage sensors placed at the output can provide us with an estimate of the RF power as long as the output load does not change. However, in the presence of process and environmental variations, this is no longer a valid assumption. Thus we need sensing of both forward and reflected powers to get an idea of the true RF power delivered to the load.

A coupled transmission line (Figure 2.1) can give us the relevant information as long as we can sense powers in both the coupled and isolated ports.

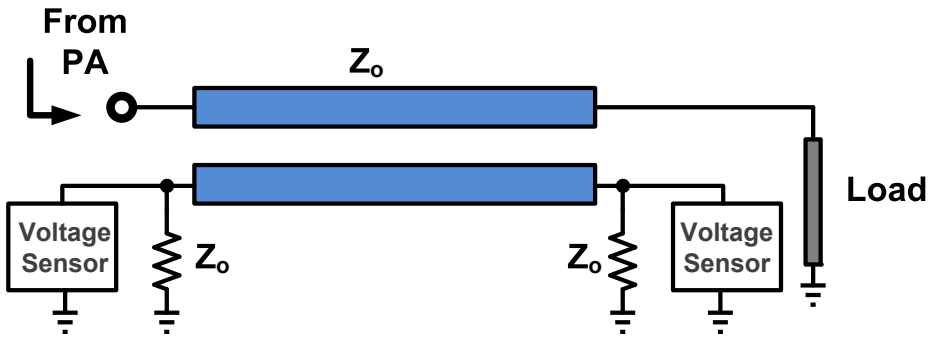


Figure 2.1: Coupled line sensor sensing both coupled and isolated port powers.

Such a sensor at the output and the input can provide us information about output match, output power delivered, input match/input power, as well as RF gain. In high power mm-wave systems, the couplers can be short and can also have low coupling ratios so as to have minimum impact on high frequency performance. As shown in Figure 2.2, for a 28 GHz PA with an output power of 16 dBm, the couplers can be about 220 μm having coupling coefficients of approximately 20 dB at the output.

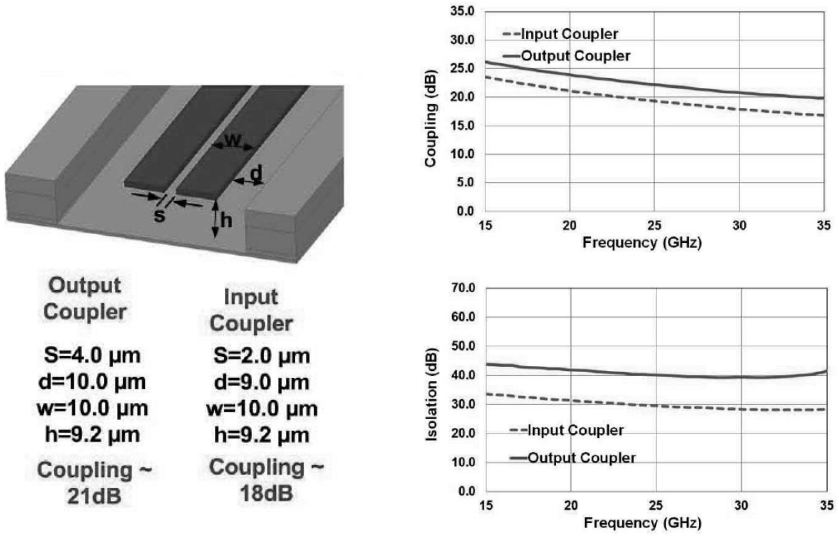


Figure 2.2: Coupler dimensions and measurement results for example 28 GHz PA with 16 dBm output power.

On the other hand for a higher power design, 27 dBm at 60 GHz, the coupler can be as short as 100 μm with a coupling ratio of 33 dB (Figure 2.3) so as to ensure the sensor circuitry sees the same voltage amplitude as in the previous case.

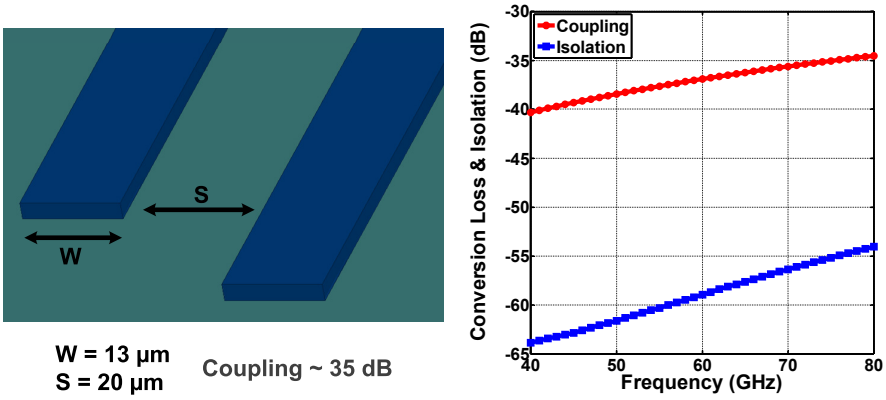


Figure 2.3: Coupler dimensions and simulation results for example 60 GHz PA with 27 dBm output power.

Because the coupling ratio in these two cases is on the low side, this type of power sensing has minimal impact on the main line match as well as the loss, as shown in Figure 2.4. While the loss is about 0.2 dB throughout the frequency range of the PA (55 GHz to 65 GHz), the return loss stays around 20 dB within the same bandwidth.

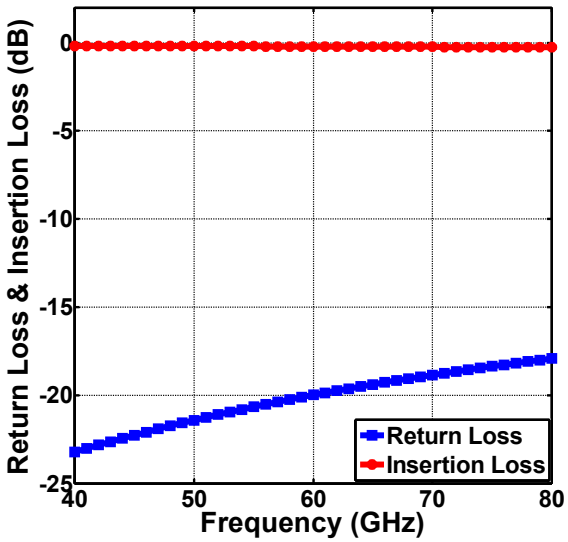


Figure 2.4: Simulated return loss and insertion loss for the 60 GHz coupler.

Once the RF power has been converted to a voltage swing at the coupled and isolated ports, a rectifier circuit then converts the voltage amplitude to a DC voltage, which can

subsequently be digitized and used by the digital algorithm. Kaushik Sengupta designed the rectifier circuitry, as shown in Figure 2.5, where transistor M_1 is biased at cut-off and a current is generated similar to that in a class-B stage. This current is then filtered and amplified in current domain and finally converted to a voltage across a resistor. The real power delivered to the load can now be expressed as the difference between the voltages sensed at the coupled and isolated port.

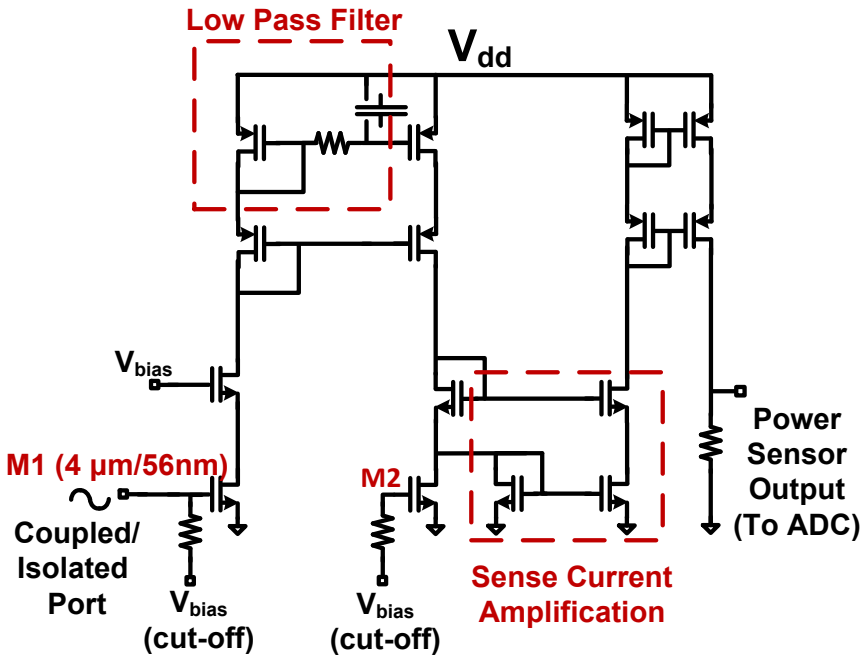


Figure 2.5: Schematic of the RF power sensor implemented in 45nm SOI CMOS.

Measurements were performed on RF sensors as part of the full PA, including the ADCs from 6 chips fabricated in 45nm SOI CMOS process, and are shown in Figure 2.6. Both responsivity and variation between sensors are shown. The 3-σ spread of the true RF power for a given measured sensor output is approximately 1dB for the output power sensor. Note that the variation between the sensors includes that due to the ADC as well.

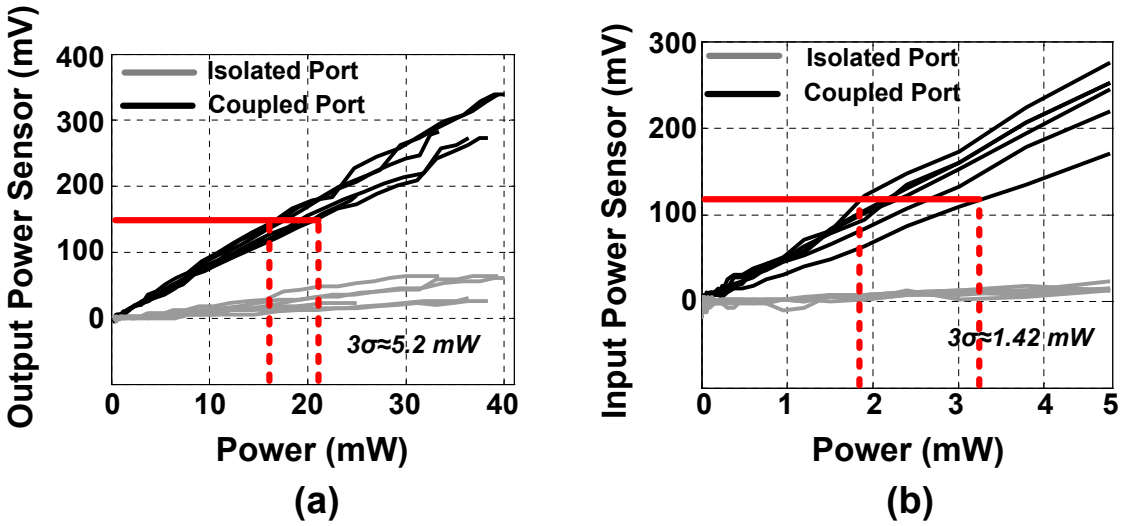


Figure 2.6: Measured RF power detector response of the output and input power sensor at 28 GHz over 6 chips (coupled and isolated ports).

2.2.3 DC Power Sensing

At first glance, sensing the DC power drawn by a PA may look trivial. In fact, a small series resistor will provide a DC voltage proportional to the current drawn by the PA. In [33], a resistor has been added in series with the supply line which monitors the current consumption of an LNA. However, for high power, performance critical PAs, this may lead to additional loss and thus a significant hit on the overall efficiency. An indirect way of sensing the current is to utilize the fact that most commercial PAs have a voltage regulator to maintain a stable V_{DD} in the presence of current variations. The technique relies on mirroring the PA current accurately without sacrificing efficiency. This in turn implies the voltage drop across the mirroring transistors should be very low (10-30mV). The circuit (designed by Kaushik Sengupta) is shown in Figure 2.7, where, to ensure accurate mirroring, the source nodes of M_1 and M_2 are held at the same potential through the opamp A_1 [34]. The sensed current is eventually converted to an analog DC voltage across a resistor.

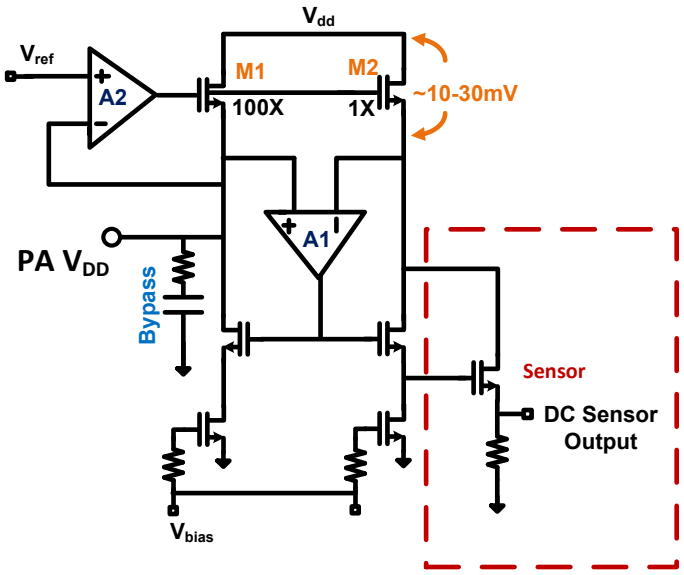


Figure 2.7: Measuring DC current drawn by PA through mirroring of regulator transistor.

Figure 2.8 shows measurement results of a DC sensor (measured as part of the entire PA) from 5 chips fabricated in the 45nm CMOS process. The 3-σ spread of the PA current for a given measured sense voltage is less than 14% over all the chips.

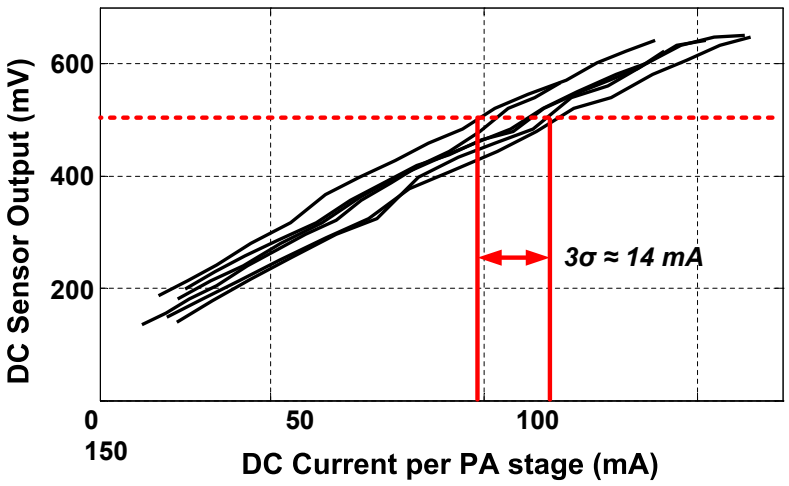


Figure 2.8: Measured sensor responses for 5 chips.

A test structure fabricated in a 32nm SOI CMOS technology for a different PA also shows a similar response.

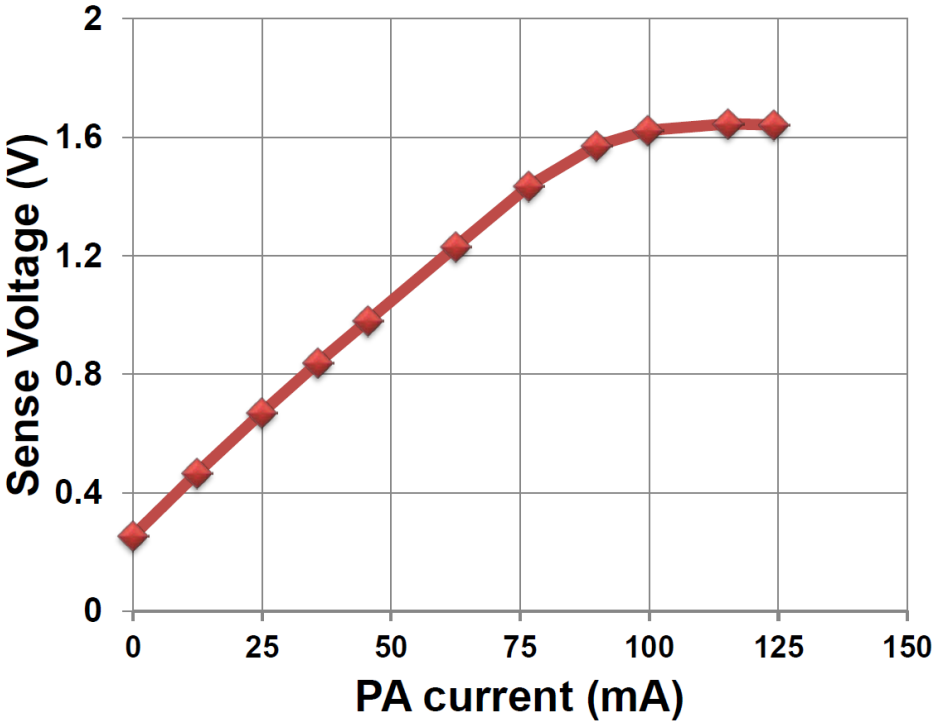


Figure 2.9: Measured DC sensor response (32nm SOI CMOS).

To verify the proper operation of the transistor M_1 in the test structure, DC probing was performed to obtain swept measurements of the regulated PA V_{DD} versus the op-amp reference voltage at a main V_{DD} of 1.8 V. Figure 2.10 verifies operation of the regulator from two separate DC sensors showing an excellent tracking with a $\sim$ 40 mV drop across the transistor.

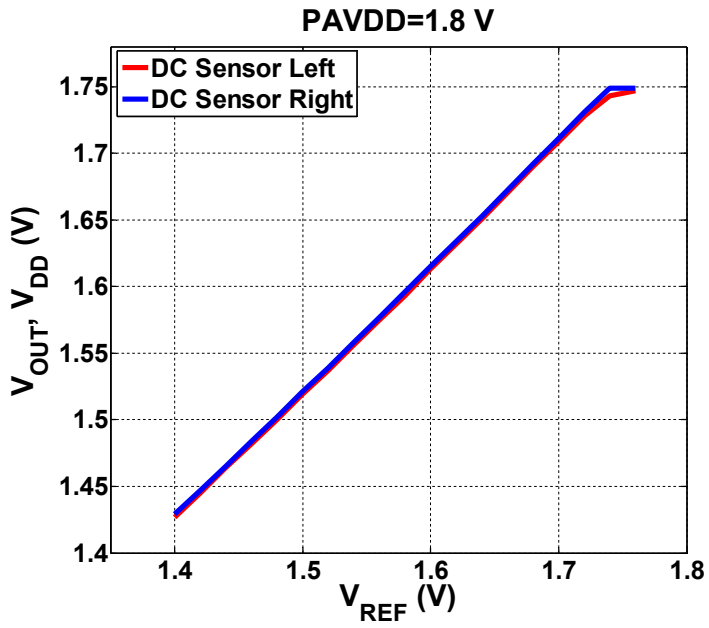


Figure 2.10: Measured response of regulated supply voltage versus reference voltage (32nm SOI CMOS).

Figure 2.11 shows the die photo of the test structure fabricated in 32nm SOI CMOS process showing the RF test pads as well as the 1:100 mirror transistor of the DC sensor.

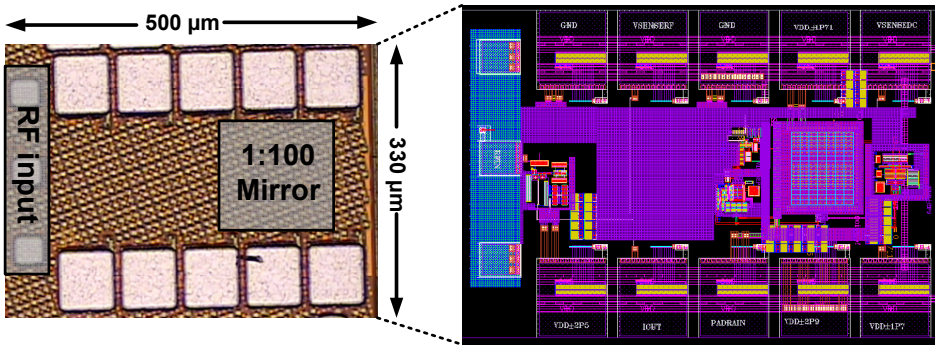


Figure 2.11: Die photograph of test structure.

2.2.4 Temperature Sensing

In any high power system, the local temperature around the core device can be shown to increase with power dissipation. This enables another indirect method for sensing DC power

by converting the temperature to the electrical domain. One low overhead way to achieve this is to use p-n junction sensor diodes placed close to the thermally active region and then using their output voltages to get a sense of the local temperature. To define the temperature relative to the other "cold" parts of the chip, identical diodes may be placed away from the "hot" devices. Simulations and theoretical predictions indicate that the temperature profile falls sharply beyond a 20-30 μm radius of a PA transistor. It has also been measured that the PA core can rise up to 10° while it conducts current. Figure 2.12 shows the placement of the sensor diodes for a representative cascode PA. The reference "cold" diode was placed about 40 μm away.

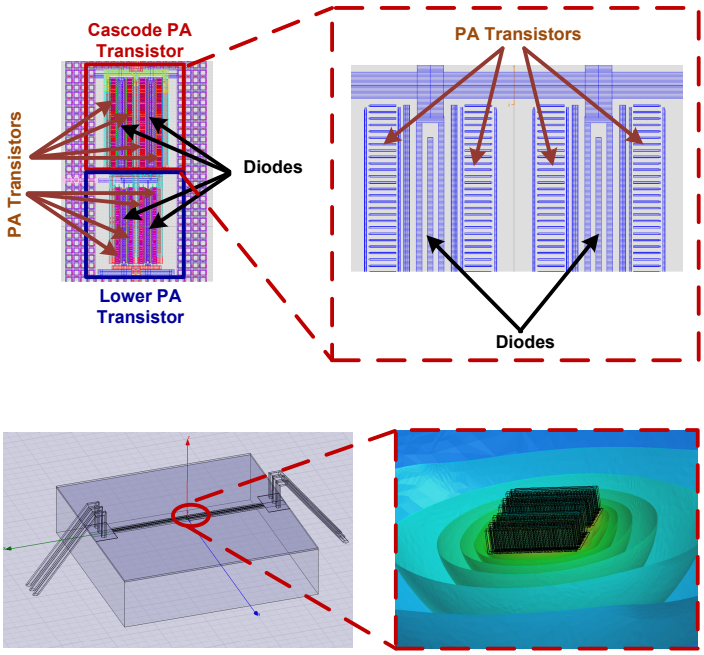


Figure 2.12: Layout of sensor diodes interspersed within the PA transistor and simulated thermal profile for 80 mW of power dissipation.

Figure 2.13 shows the schematic of the implemented thermal sensor where the difference voltage between the hot and cold diodes is amplified to produce a DC voltage within the dynamic range of the ADC. The thermal simulations and circuit design for this sensor were performed by Kaushik Sengupta.

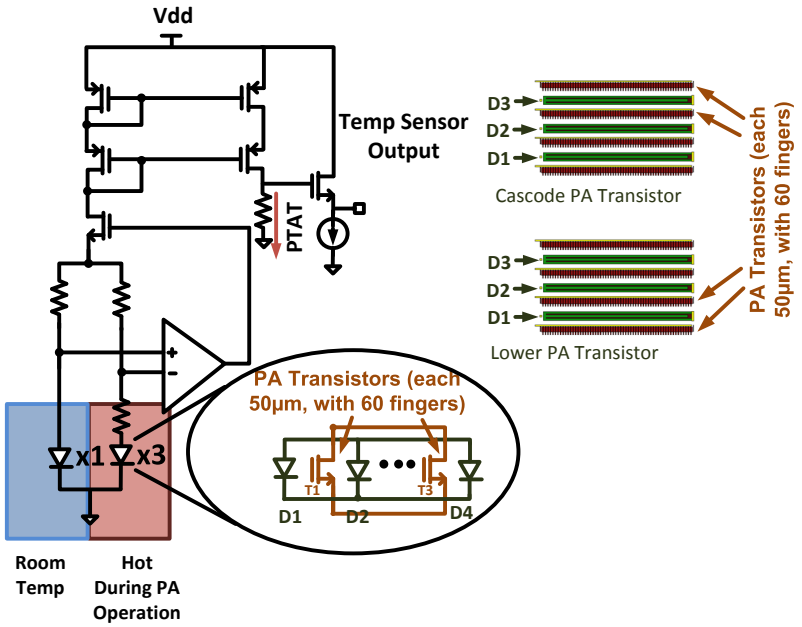


Figure 2.13: Thermal sensor schematic.

The thermal and the DC sensor show similar monotonic variation with DC power drawn by the example PA (Figure 2.14. The response of the thermal sensor is, however, limited by the thermal time constant. This restricts its usage in self-healing situations where a large actuation space and hence a large number of sensor reads have to be completed to minimize healing time.

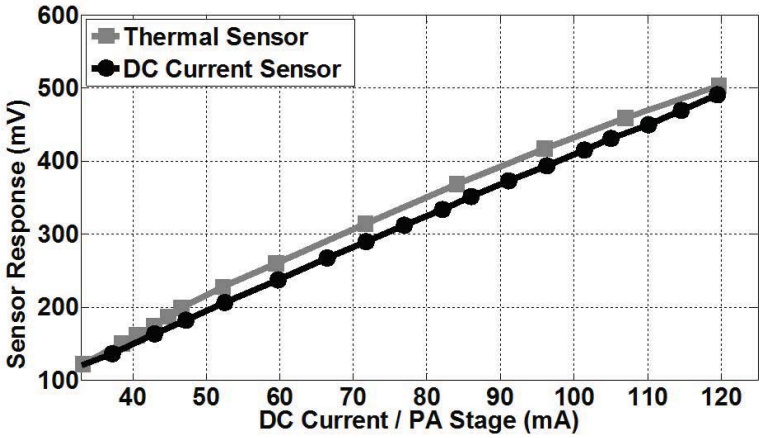


Figure 2.14: Thermal sensor schematic.

2.3 Actuation: Countering Performance Degradation

As we discussed in the previous chapters, both static (process/mismatch) as well as dynamic (temperature, aging, antenna mismatch) variations can cause performance degradation in RF/mm-wave circuits. With regards to a PA example, these variations will affect both the $G_{\max}$ and $f_{\max}$ of the transistors, leading to reduced output power. In addition, because of capacitance and V_{th} variations, the input and output impedances of the PA may vary, causing significant performance hits. Thermal variations will cause bias of the transistors leading to reduction in saturated output power from the amplifying transistors. Transistor aging will also have a similar effect of having lower saturated power as well as non-optimal output impedance.

How do we design appropriate actuators to counter the effects from all these variations? As mentioned briefly in the previous chapter, the actuators must cover a large enough actuation space while having minimal impact on RF performance. For a mm-wave PA, the actuators can be of three broad categories: gate bias voltage actuators, passive matching network tuning actuators, and transistor architecture actuators.

2.3.1 Gate Bias Actuators

A significant portion of the variations associated with a mm-wave power generation system is due to quiescent point fluctuations due to process and temperature changes. In fact, almost all the performance metrics of a mm-wave PA directly correlate to the bias current including efficiency, saturated output power, etc. Most PAs are designed to operate at or near their maximum saturated power or maximum efficiency point. However, in a typical communication system, the PA operates <10% of the time near its peak output power. A dynamic biasing scheme addresses both these issues; it optimizes the operation of the PA for maximum performance when required and it can also reduce DC power consumption of the system at back-off, leading to significant improvement in efficiency.

For a typical cascode PA, both the common-source (CS) as well as the cascode (CG) transistor biasing points have significant effects on the overall system performance. Thus,

bias control in the form of DACs can be implemented for both these transistors. The requirements for these DACs are the following: they need to be extremely low power blocks so as to reduce overall self-healing overhead, however they also need to drive the output stage transistors at relatively high speeds, which keeps healing time to a minimum. In the present design, current-source based DACs are implemented. These binary weighted current sources are laid out in a common-centroid fashion to ensure good matching. Figure 2.15 shows schematics of a 5-bit DAC where cascode current mirrors have been implemented to ensure accurate current matching. To minimize variations due to process, the transistors in

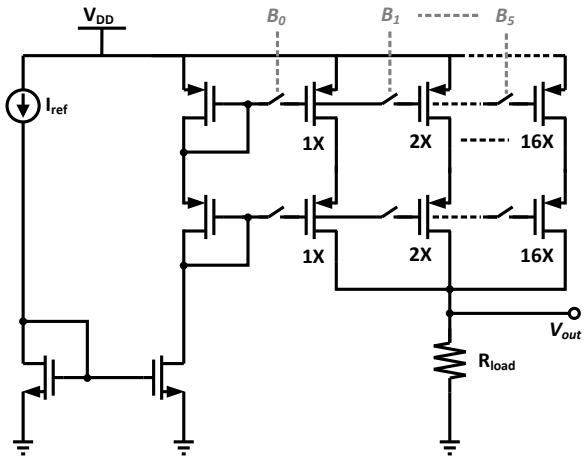


Figure 2.15: Gate bias actuator schematic.

the DAC are long-channel devices which are less susceptible to variations such as line-edge-roughness. Measurement results for two DACs implemented in 45nm SOI CMOS process are shown in Figure 2.16, where both CS and CG DACs are used. The DAC for the CS transistor provides output voltages in the range of 450 mV to 1.05 V, whereas the CG DAC provides 1.1 V to 1.95 V. The two DACs were verified to operate at 25 MHz, which was limited by the test setup.

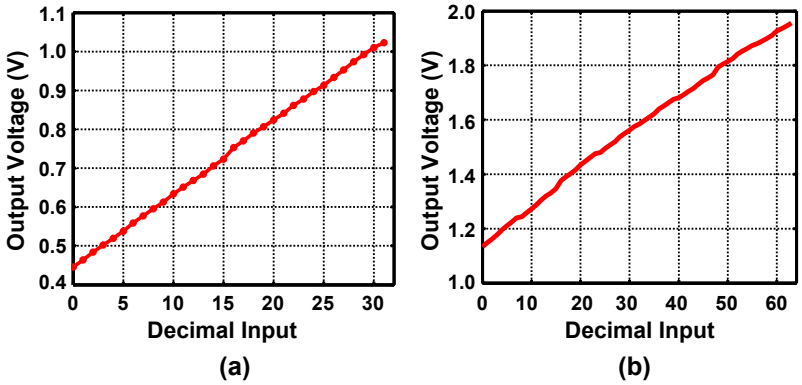


Figure 2.16: Measurement results from (a) CS and (b) CG DACs in 45nm CMOS.

A similar DAC was implemented in 32nm SOI CMOS and was measured as part of a test structure. Results are shown in Figure 2.17.

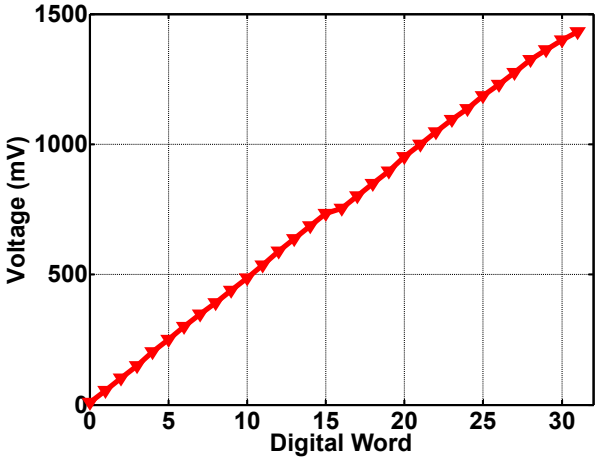


Figure 2.17: Measurement results from a full range DAC in 32nm CMOS.

Because of the extremely low currents used in the DAC mirrors, the overall power consumption of the DACs is minimal. For example, the DAC designed in 32nm CMOS consume a best and worst case power consumption of 250 μW & 1.03 mW, respectively.

2.3.2 Passive Matching Network Tuning Actuators

Passive matching network tuning actuators [35], [36] provide flexibility to a mm-wave/RF system by providing it the ability to dynamically tune the matching networks on the chip after fabrication to counter process, mismatch, as well as environmental variations. The commonest form of such actuators is the tunable capacitor, which may be implemented either as a varactor or discrete switched capacitor tuning. However, depending on the application, these actuators may be rendered unusable because of the typical high losses associated with them. For a power amplifier, keeping the loss in the output power combining network is critical to the overall performance, since any loss at the output has a direct impact on the output power as well as the efficiency. In such situations there exists a direct trade-off between the additional loss contributed by the tunable component and the tuning range. Moreover, for optimum performance, the amplifying transistors in the PA will be operating near their voltage breakdown and the matching network will transform the impedance to 50 Ω from a much smaller impedance, which means that the voltage swing within the output network and at the output will be much larger than the transistor breakdown. This implies that the transistors used in the switchable matching components or the varactors themselves may be susceptible to breakdown and should be placed at appropriate low-swing points along the matching network.

Here, we will discuss these trade-offs in the specific case of a tunable transmission line stub, which is a common element used in mm-wave matching networks. As shown in Figure 2.18, the effective length of the transmission line stub can be changed by shorting out the stub at various points along the length.

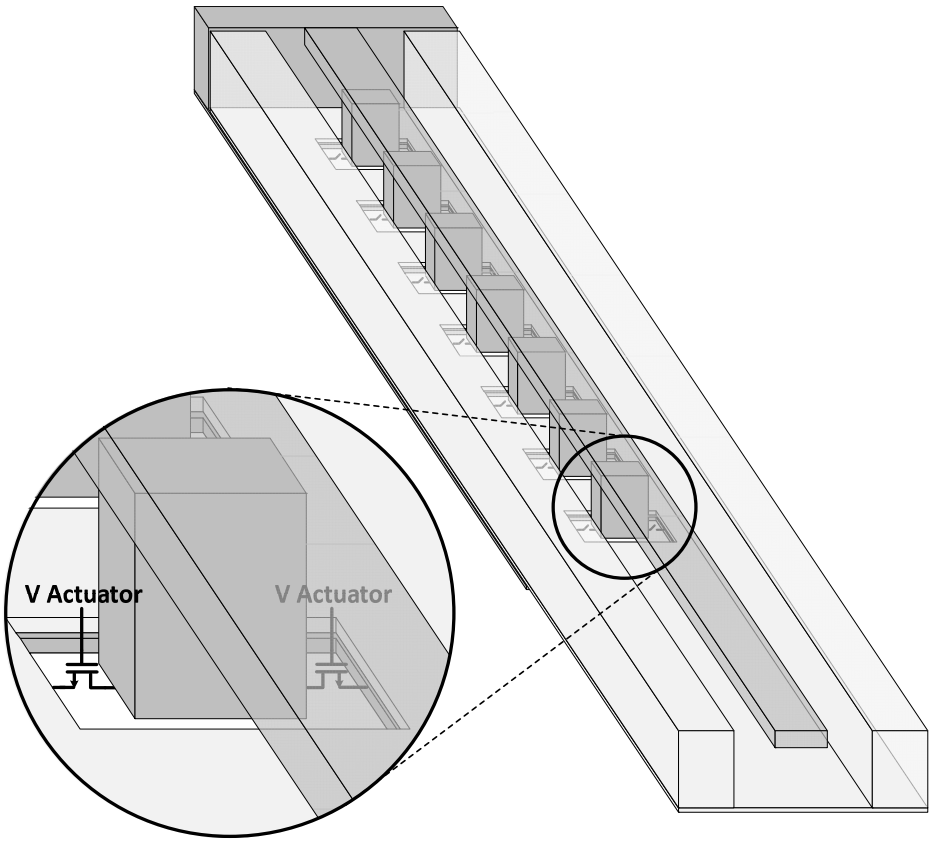


Figure 2.18: 3-D view of tunable transmission line stub actuator.

The switches are implemented as transistors, which leads to an interesting trade-off between loss and parasitic capacitance. As the size of the switch is increased, the loss reduces, but the off-state capacitance increases. If not taken into account, this additional capacitance, at mm-wave frequencies can cause significant de-tuning itself. An easy way to eliminate such de-tunings is to include the transistor capacitance as part of the transmission line itself. By placing these switches at regular intervals, the capacitance may be absorbed into the transmission line distributed model. This enables use of much wider transistors having low ON resistances while providing tunability.

Measurement results from one such tunable stub based on a grounded coplanar waveguide transmission line are shown in Figure 2.19. An actuation range of 25 to 71 pH was achieved with 8 switches, demonstrating the large actuation range. Moreover, because these switches

are digitally controlled, these actuators can directly be controlled by the on-chip self healing digital algorithm.

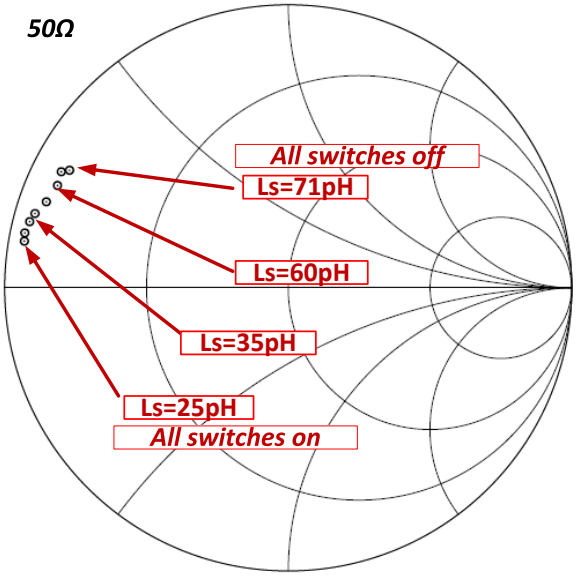


Figure 2.19: Measurement results from an example tunable transmission line stub for various switch settings.

2.3.3 Transistor Architecture Actuators

Changing the size or architecture of the transistors themselves is also a possibility for actuators for mm-wave PAs. One can imagine that an easy way to deal with saturated output RF power degradation is simply to increase the size of the amplifying transistors. Switching fingers of transistors in and out of the circuit, however, can cause significant losses in performance from two dominant sources. The first is that the switches used to switch the transistors in or out will either have higher ON resistance or OFF capacitance, and both will cause signal degradation. Secondly, even if perfect switches were available, changing the size of the transistors will significantly change the desired matching network, resulting either in more loss in the matching networks or in mismatch for some transistor sizes. For the example mm-wave systems in the subsequent chapters, it was determined that the best performance could be achieved by making the transistors large enough to cover the desired

saturated output power and to use the gate bias actuators to achieve higher efficiencies while in back-off, so transistor architecture actuators were not used.

2.4 Analog to Digital Converters

As with all other self-healing enabling blocks, the on-chip ADCs need to be extremely low power as well as area efficient designs. However, lower power generally implies lower speeds which will lead to slower digital healing algorithm, which in turn affects the total healing time. In addition to the overhead vs speed trade-off, resolution of the ADC also directly affects the accuracy of the healing algorithm.

The fastest ADCs are flash ADCs which operate in a parallel fashion, leading to simultaneous generation of output bits [37, 38]. However, because of this parallel computation, these converters usually are among the most power hungry ADCs. In addition, matching requirements for comparing elements become extremely stringent for higher resolutions. Both power as well as design complexity thus limit the use of flash ADCs in self-healing systems. A good compromise between resolution, design complexity, and speed is a pipelined ADC [39]. However, these ADCs usually occupy a larger area as well as consume significant amounts of power. The successive approximation register (SAR) based ADCs are serial ADCs, which offer a good compromise between power consumption, resolution, design complexity, and area [40]. In addition to being suitable for medium data rate applications, due to the inherent serial nature of the data output, it is an ideal candidate for self-healing systems where routing complexity for multiple high-speed digital signals across the chip also needs to be minimized.

An 8-bit SAR ADC was chosen for the current self-healing system. Data from various sensors (4 DC sensors, 4 thermal sensors, and 2 RF sensors) were multiplexed and fed to three ADCs placed throughout the chip. The sensor choice is governed by the digital ASIC. The block diagram of the implemented SAR ADC is shown in Figure 2.20a. Global clock and the digitization initialization signals are the only inputs to the ADC in addition to the sensor select bits. The ASIC waits for a serial data valid output bit from the ADC

indicating that the sensor data has been digitized. The 8-bits immediately following the serial data valid pulse correspond to the sensor data, most-significant-bit first. The SAR ADC utilizes an R-2R digital-to-analog converter to enable the successive approximation logic. Significant attention was paid to the layout of the R-2R DAC, since matching of the resistors will severely affect the linearity as well as monotonicity of the ADC. Two phases of the clock are generated at half the frequency which are in quadrature. One of them feeds the comparator and the other feeds the SAR register. This is to ensure that the comparator does not change its output over the approximation window itself. The initialization signal also gets re-timed using this divided clock.

A successive approximation register usually contains a sequencer and a code register. The code to be compared to is set by the sequencer and it is set or reset on the next clock cycle depending on the comparison results. In most designs, the sequencer is synchronous whereas the code register is asynchronous, which is prone to glitches, bit swallowing, etc. To alleviate these problems, a fully synchronous SAR design [41] was chosen based on modified mux-based flip-flops shown in Figure 2.20 (b).

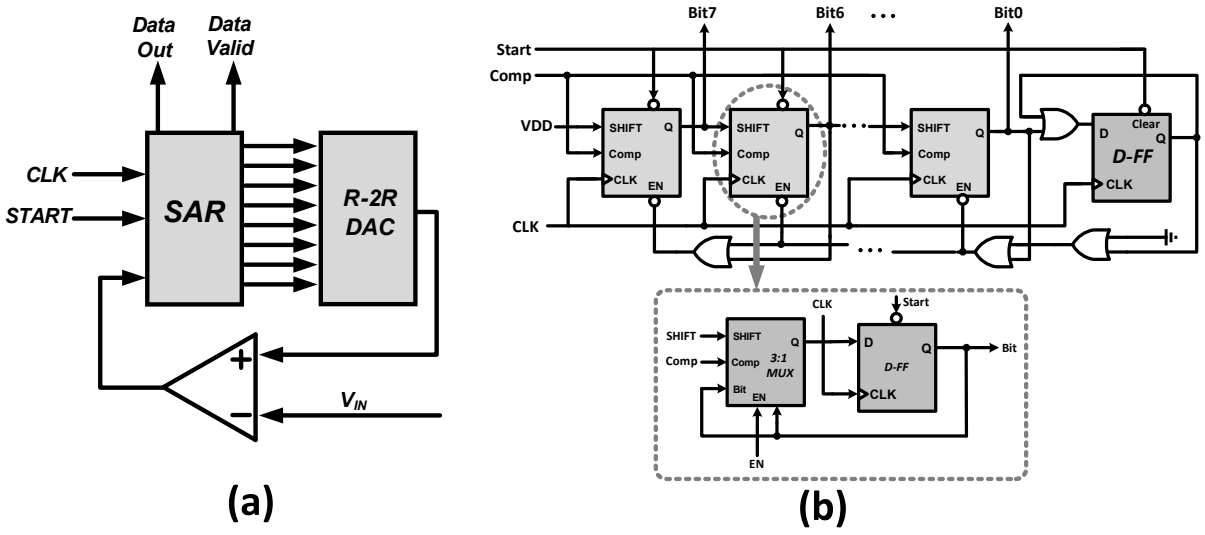


Figure 2.20: (a) ADC block diagram, (b) Fully-synchronous SAR.

Figure 2.21 shows simulated waveforms from the ADC showing a full digitization sequence

and the various timings associated with it.

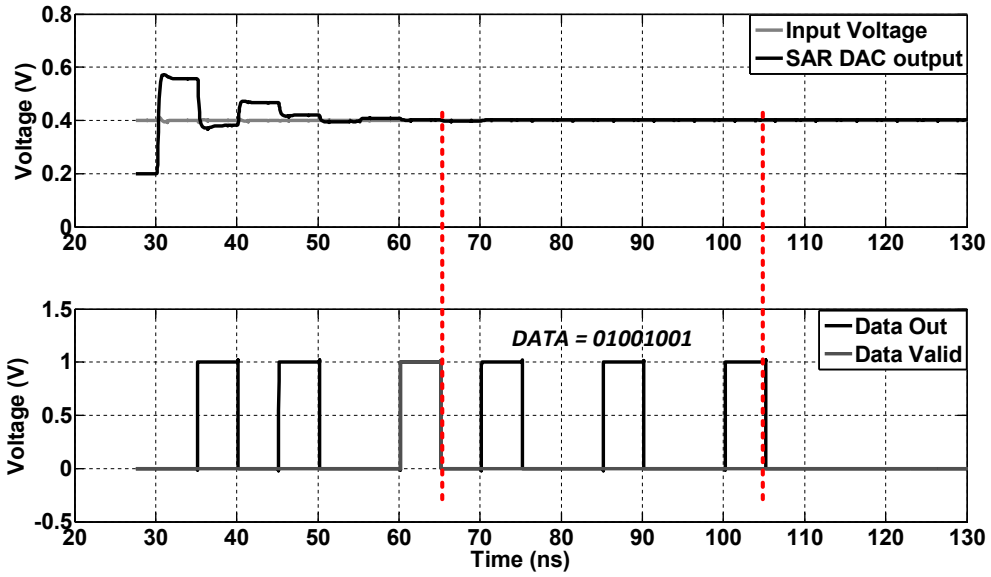


Figure 2.21: Simulated voltages for an input voltage of 400 mV. Data output is decimal 73 for an 8-bit output.

The ADC measurements were performed as part of the full self-healing system implemented in 45nm SOI CMOS. Sensor voltages were DC probed and the corresponding readouts were obtained through the digital ASIC. The ADC was verified to operate at 25 MHz clock frequency, which translates to 2.5 Msps for an 8-bit SAR with initialization and data ready bits. Figure 2.22 (a) shows measurement vs simulation results of the implemented ADC. The average DNL was -0.04 LSB and the worst case DNL was measured to be -0.605 LSB, as shown in Figure 2.22 (b). These ensure that the ADC is monotonic, which ensures proper operation of the digital ASIC. The SAR ADC draws 1.6 mW from a 1 V supply.

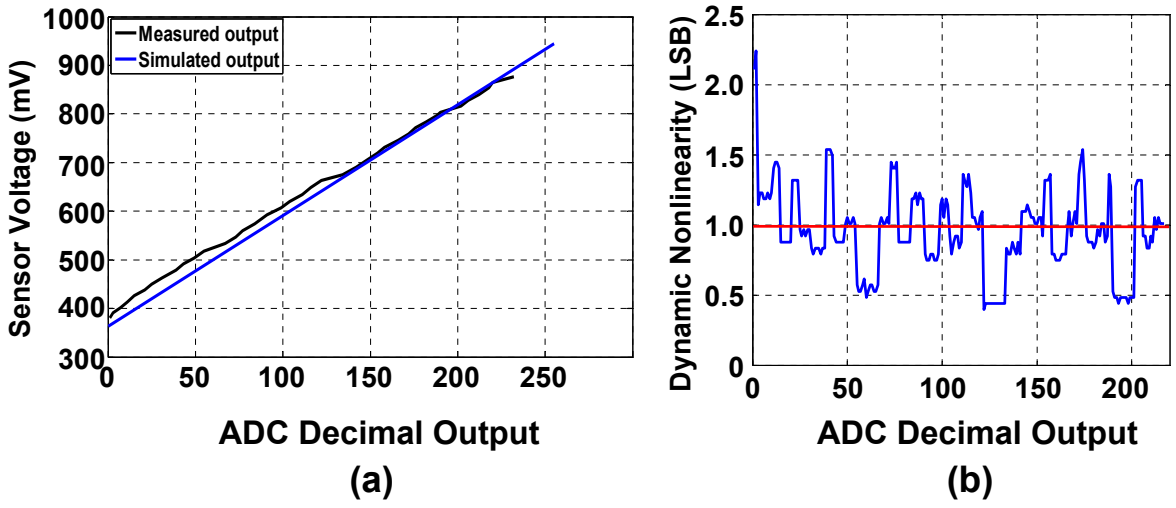


Figure 2.22: (a) ADC measured and simulated characteristics (V_{refn}=350 mV, and V_{refp}=950 mV), (b) measured dynamic non-linearity (DNL).

2.5 Digital Healing Algorithm

The self-healing digital algorithm block of the system reads in sensor data as inputs, and based on that it decides the next optimum actuation state and communicates the same to the actuators. For fastest operation and for fully integrated healing, the algorithm must be implemented on-chip, leveraging the vast digital processing power and high level of integration of modern CMOS designs. The digital core is also responsible for handshaking between the DACs, the optimization algorithm, and the ADCs. This is critically important, since the actuators should not change state while the optimization algorithm is reading in sensor data. Figure 2.23 shows the basic flowchart for a representative self-healing system, where convergence is achieved when the performance goal is met after an iterative algorithm.

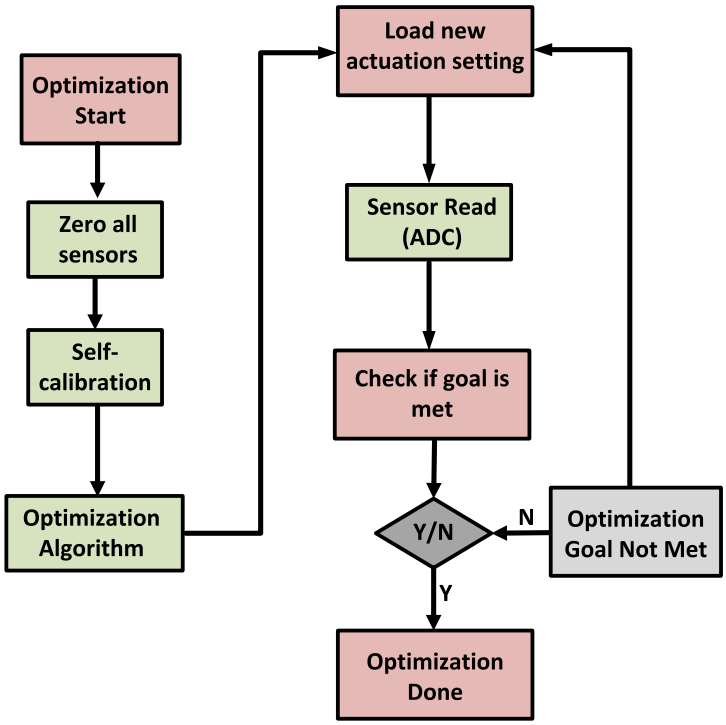


Figure 2.23: Example self-healing algorithm flowchart.

The choice of the algorithm for a particular application is governed by the behavior of the actuation space as well as the specific performance metric. For example, if the space is convex, various commonly available algorithms can guarantee convergence given an achievable performance metric. Another aspect of the self-healing algorithm is the choice between local and global optimization. In a mm-wave/RF transmitter or receivers, there may be separate algorithms to improve performance metrics pertaining to individual blocks — for example, improving linearity of the up/down-conversion mixers, increasing LNA gain and input match, improving power amplifier PAE, etc. In most cases block-level optimization will eventually lead to system level optimization. However, the possibility exists for a global optimization, where information from multiple blocks is utilized for overall system healing against unknown variations.

As mentioned before, in a fully integrated closed loop self-healing system, the digital core needs to be synthesized on-chip using available standard cells and a hardware description language. Depending on the system and the number of actuators as well as the nature of the

algorithm used, the speed of these standard cells will eventually limit the healing time. For static variations (process/mismatch), the healing time is not of much concern, since these variations need to be healed only once, or over large periods of time. However, for dynamic variations, such as example antenna impedance variation, this healing time is of utmost importance and the exact requirements are based on the actual wireless application. The other critical aspect of any self-healing algorithm is the self-calibration of sensors which any healing algorithm must perform to eliminate offsets from the sensor readings or the ADC. During such a calibration, the RF/mm-wave system is essentially unusable; depending on the application such a down time may not be acceptable every time the self-healing process occurs. In such situations, the auto-calibration needs to be performed during low traffic periods where system down time is acceptable.

2.6 Conclusion

In summary, we have discussed several building blocks of a self-healing system namely, actuators, sensors, data converters, and the controlling algorithm. Several representative sensors have been described in the context of mm-wave power generation along with associated simulation and measurement results. A directional coupler based sensing of RF power has been demonstrated to be an energy efficienct technique to sense output power in PAs and convert both the coupled and isolated port powers to a voltage across a 50-Ω resistor. Efficient DC current sensing is also demonstrated using a large mirroring ratio for the supply regulator transistor and subsequent current-to-voltage transformation. Apart from these two sensors, another low overhead method of sensing DC current is also discussed based on local temperature sensing by placing diodes close to the PA transistor. In practical applications however, due to the prohibitively large thermal time constant, only the RF and DC current sensors can be used for self-healing. Measurement results have been shown which verify the functionality as well as robustness of the designed sensors.

Two types of actuators are presented which enable bias/quiescent point adjustments as well as dynamic matching network tuning. Measurement results from both these actuators

show a large actuation space, enabling a typical mm-wave/RF circuit to heal from a wide variety of unforeseen variations. Analog to digital converters are discussed in the context of digitizing sensor data for use in the optimization algorithm. Measurement results for the implemented 8-bit SAR ADC are presented which demonstrate the utility of such low power and low area overhead ADCs in self-healing applications. Finally, critical aspects of design of the digital healing algorithm have been discussed with particular emphasis on auto calibration of the sensors.

Chapter 3

Self-healing mm-Wave Segmented Power Mixer

3.1 Introduction

Recent years have seen a surge in demand for high-speed, short-range, wireless applications, particularly those in handheld portable devices. This insatiable demand for such high-speed communication systems has led to the development of mm-wave (>30 GHz and beyond) transmitters and receivers, exploiting the higher bandwidth available in this portion of the electromagnetic spectrum compared to lower RF frequencies. This has been assisted by the continuing advances in scaling CMOS beyond the 32nm node with faster transistors capable of providing higher gain at these frequencies. While scaling up carrier frequencies is advantageous with respect to available bandwidth, use of complex modulation schemes is essential to optimally utilize the available spectrum. This bi-directional approach is necessary to truly scale up short-range wireless data transmission speed efficiently into the multi-Gbps range. Therefore, there has been a major focus toward non-constant envelope modulation schemes, such as QAM, which provide higher spectral efficiency. Many such mm-wave transmitters utilizing a variety of modulation schemes have been reported in recent literature [42–45].

Due to linearity requirements of a spectrally-efficient transmitter, power amplifiers generally have to be operated significantly at back-off, which leads to a direct trade-off between energy efficiency and spectral-efficiency. For example, efficiency of a class-A PA drops down

to a near quarter of its peak value when operated at 6 dB back-off [46] [47] [48]. This is a major problem since these communication chipsets need to be extremely energy efficient so as to have a minimum impact on the overall battery life of the portable device. Furthermore, these mm-Wave transmitters need to provide large output power levels to achieve necessary communication ranges in practice. Switching power amplifiers can provide high power at high efficiencies even at mm-wave frequencies, however they are only suited towards constant envelope/phase modulation schemes. Therefore high power, high-efficiency, and high linearity transmitter architectures at mm-Wave frequencies are critical toward enabling ultra-high-speed and energy-efficient wireless transceivers, but have been a major bottleneck so far.

To address this fundamental trade-off, several architectures have been proposed, including the polar modulation scheme, which relies on separating amplitude and phase paths to generate non-constant envelope modulations with saturated high-efficiency power amplifiers. These approaches have become very practical at lower RF frequencies and at lower data rates where supply modulators are generally utilized to generate amplitude modulated data [49], [50]. At mm-wave frequencies, however, their application has been limited due the design challenges of high-efficiency supply modulators operating at high speeds and the stringent requirements for delay matching between the amplitude and phase paths. In [51], a Cartesian I/Q based generation of modulated data at mm-wave frequencies has been reported using digitally controlled power DAC cells each operating at full power and maximum efficiency. The architecture relies on off-chip spatial power combining using patch antennas and multiple ICs, which requires larger chip area and/or multiple radiating elements. A similar architecture, also relying on spatial power combining, was reported in [52].

Efficient on-chip power combining at mm-wave frequencies overcomes many of the challenges associated with multi-antenna based spatial power combining transmitting systems. Transformer based power combining techniques have been extremely popular due to their high efficiency, relatively large bandwidth, and compact layout at these frequencies [53] [54]. Among the existing on-chip transformer based power combining techniques, the distributed active transformer (DAT) power combiner, first reported in [55], provides an efficient method-

ology for combining output powers from multiple differential power stages while at the same time providing convenient supply connectivity through the formation of virtual short circuits along the length of the transformer. However, its use at mm-wave frequencies has been mainly restricted to combining only two differential power stages [56–58] due to implementation challenges in the input distribution network. For example, the 4-to-1 DAT reported at 60 GHz in [59] uses an area consuming Wilkinson divider based input distribution network. This chapter presents a compact, energy-efficient, fully-integrated mm-wave transmitter architecture capable of generating complex modulation schemes at high output powers and at high data rates. The architecture utilizes a polar scheme where several segmented power mixers [60] are power combined using an on-chip, high-efficiency, dual primary DAT based power combiner. The segmented power generation technique eliminates the requirement for complex, high-speed, high-efficiency supply modulators. Similar to [51] and [52], each power mixer segment is operated near saturation, thereby providing higher efficiencies during non-constant envelope modulations compared to conventional linear power amplifiers.

In this chapter, first we are going to discuss design and implementation details of the core power mixer itself and present measurement results, both continuous wave and modulation for a power mixer chip without self-healing [61]. In the following section, we incorporate several sensors and actuators (discussed in Chapter 2) into the core architecture to enable self-healing. Measurement results will be shown to both increase in output power as well as reduced DC power consumption with closed-loop self-healing.

3.2 High-power mm-Wave Segmented Power Mixer

3.2.1 System Architecture

The block diagram overview of the power mixer transmitter architecture is shown in Figure A.1. An on-chip transformer converts a single-ended mm-wave LO signal to a differential one, which is then distributed to four differential power mixer stages using inverter based driver stages. Baseband signals are also applied to each power mixer either in pure analog or

digital form, as will be discussed in the subsequent sections. The power mixers up-convert the baseband signal at the mm-wave carrier frequency to generate modulated signals which are then power combined using a dual-primary distributed active transformer (DAT). Each power mixer thus performs both operations of modulation/up-conversion and power amplification within the same block. This eliminates the additional loss and area associated with matching networks between the up-conversion mixer and the power amplifying stage in conventional transmitting systems. Power control in such a system is achieved by dividing each power mixer into 8 smaller segments, each of which can be independently and digitally controlled by baseband data. In addition, by turning different segments on or off based on the required power levels, this technique also ensures reduced DC power consumption at back-off power levels.

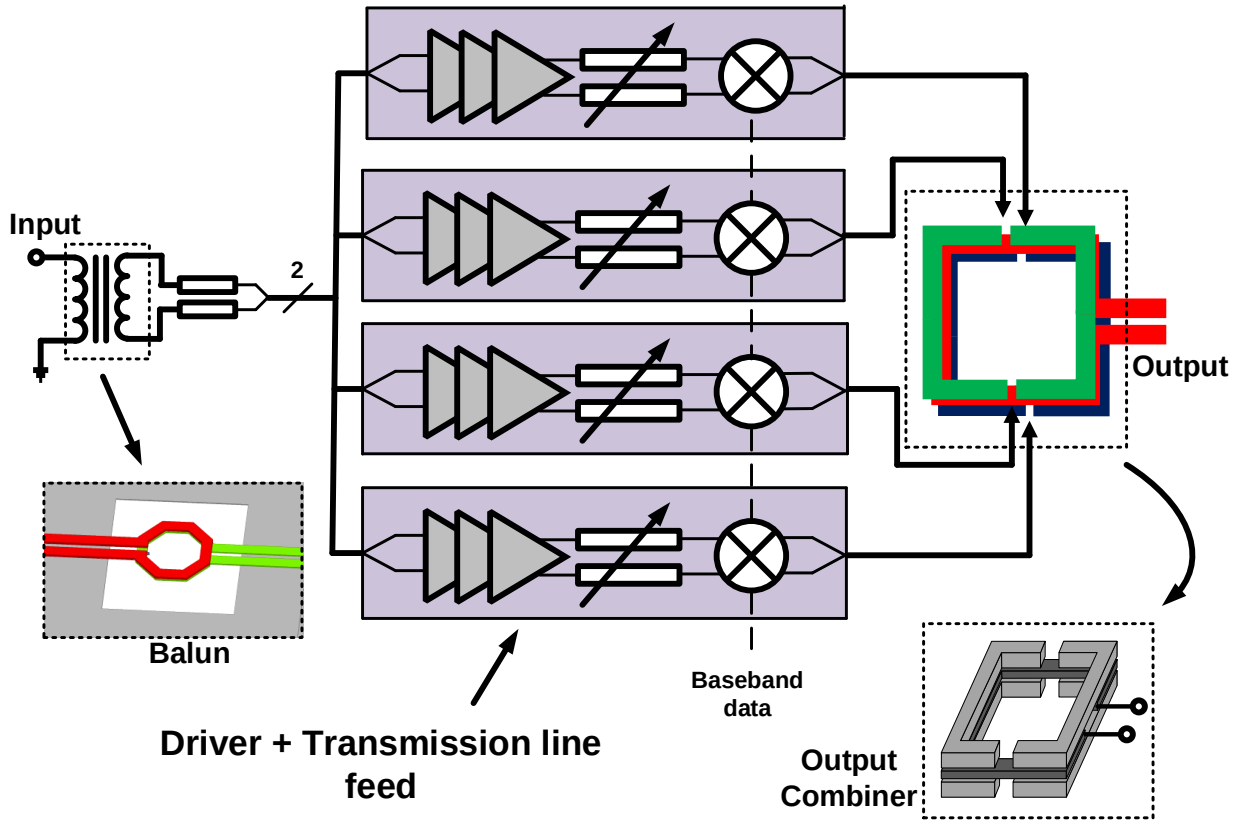


Figure 3.1: Overall architecture of the power mixer based transmitter.

While single-primary DAT based mm-wave power combiners have been reported in liter-

ature, the dual-primary DAT allows efficient power combining of four differential elements while significantly reducing the area, design, and layout complexity of the input LO drive network. Each primary is driven by two differential power mixer stages having identical segmentation control to maintain drive symmetry, a feature essential to the basic operation of the DAT. This implies that by driving each primary with independent segmentation controls, 64 segmented output power levels are possible, combining output powers from the two primaries. An important feature of this power mixer transmitter is the ability to generate non-constant envelope modulations while simultaneously not suffering from efficiency degradation issues when compared to switching mode transmitters/PAs. During such modulations, the system dynamically switches different segments on or off depending on the power requirement of the desired symbol. Figure 3.2 depicts generation of two representative symbols from a 16-QAM constellation where the amplitude and phase paths have been separated. In the power mixer, the mm-wave LO is phase modulated and the amplitude modulation is achieved by actuating the segmentation bits, thereby leading to generation of arbitrary mm-wave modulated QAM signals at the output. For example, 12-phases and 3 amplitude levels are required for 16-QAM, whereas 52 phases and 9 amplitude levels are essential for 64-QAM.

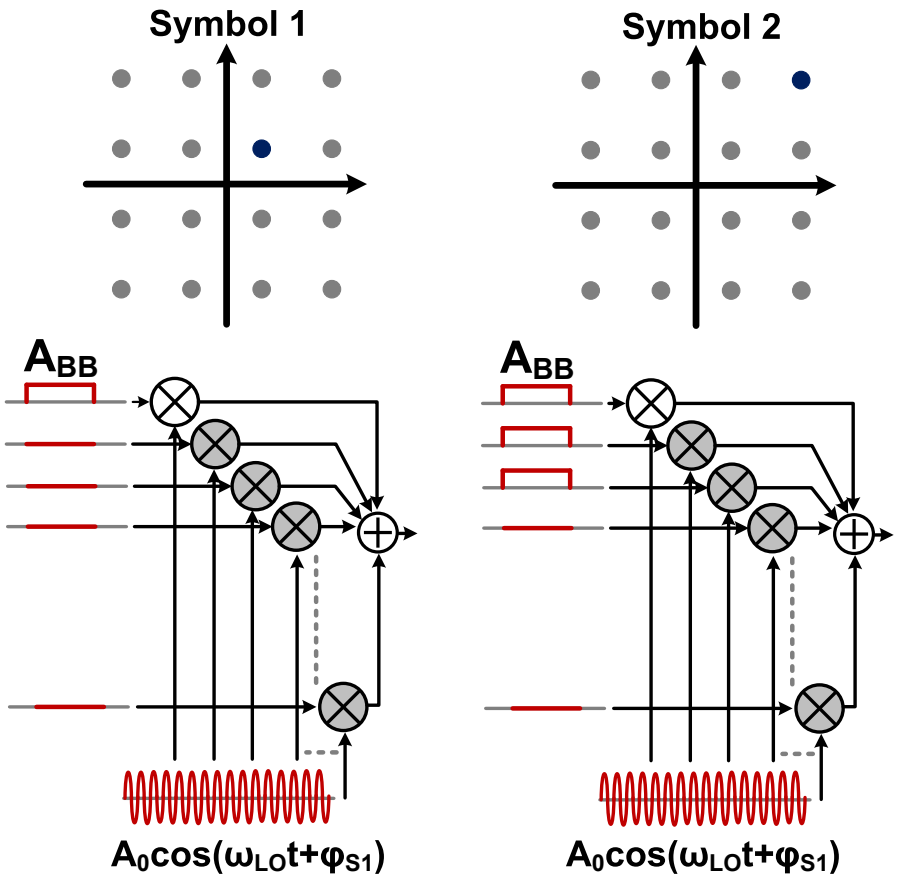


Figure 3.2: Example generation of two symbols in a16-QAM constellation with phase modulation through mm-wave LO and amplitude modulation through digital baseband paths.

Systematic delay offsets between the amplitude and phase paths in such a polar architecture can have a considerable effect on the overall EVM of a non-constant envelope modulated signal, as shown in Figure 3.3, showing simulated EVM degradation of a 16-QAM signal at a data rate of 4 Gb/s at a carrier frequency of 50 GHz as the phase delay between the two paths is swept. The programmable delay lines in the input distribution network of the present architecture allow us to calibrate for such systematic delays.

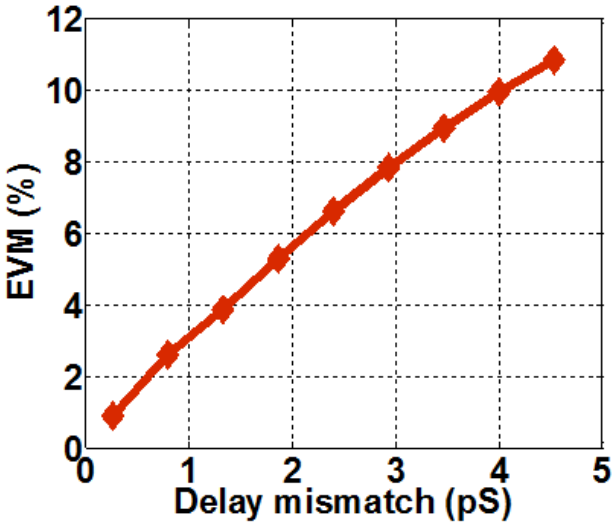


Figure 3.3: Effect of systematic delay between amplitude and phase paths for a 16-QAM signal.

3.2.2 Power Mixer Stage

The power mixer stage has been implemented as a Gilbert cell switched-conductor mixer [60] where the lower-tree common-source transistors (M_1, M_2) are driven by the mm-wave local oscillator (LO) signal, as shown in Figure 3.4. Thick-oxide transistors form the upper-tree quad transistors (M_{3-6}), thereby allowing higher voltage swings at the output. Baseband amplitude signals are applied to the gates of these transistors either in purely digital or differential analog form depending on the requirement. For efficient mixing operation it must be ensured that the differential signals are able to switch the mixing quads fully ON or OFF during modulation. Pass-transistor based analog multiplexers enable three different types of operation for these baseband mixing quad transistors. Figure 3.5 shows the three possible modes of operation (i) where the BB signals are applied purely digitally and differentially, which enables direct 0/180 BPSK modulation; (ii) where the BB signals are applied digitally but in a single-ended fashion, which allows power segmentation by directly turning different segments ON or OFF (Efficiency segmented mode – ES mode); and (iii) where the BB signals are purely analog and differential, thereby providing both phase modulation as well as analog power control (Baseline analog mode – BA mode). In the present design, all

high-speed amplitude modulations have been performed using the ES mode, which leads to significant DC power savings during modulation and thereby improves efficiency during non-constant envelope modulation.

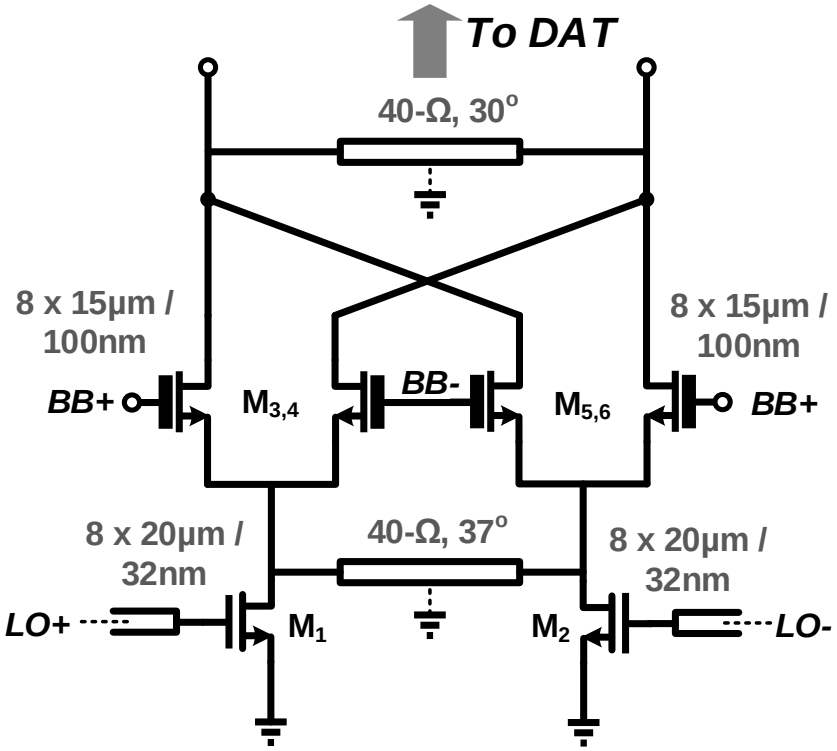


Figure 3.4: Power mixer stage schematic.

As discussed previously, direct digital power control is enabled by dividing each power mixer into eight smaller segments. In the ES mode, seven of these segments are controlled by digital signals, while the eighth one is controlled by purely analog signals to maintain power continuity. The digital BB signals are distributed throughout the chip using high-speed scaled inverter buffers. The thermometric BB segmentation controls are realized by on-chip 3-to-7 decoders to minimize off-chip high-speed signaling.

To accurately model delay and amplitude mismatches between the different segments at mm-wave frequencies, the interconnection between these stages was simulated using Ansofts HFSS and modeled as distributed transmission line structures. Due to layout constraints as well as to avoid additional signal routing loss, a binary-tree based distribution structure was not implemented in the present design, thereby creating minor mismatches between

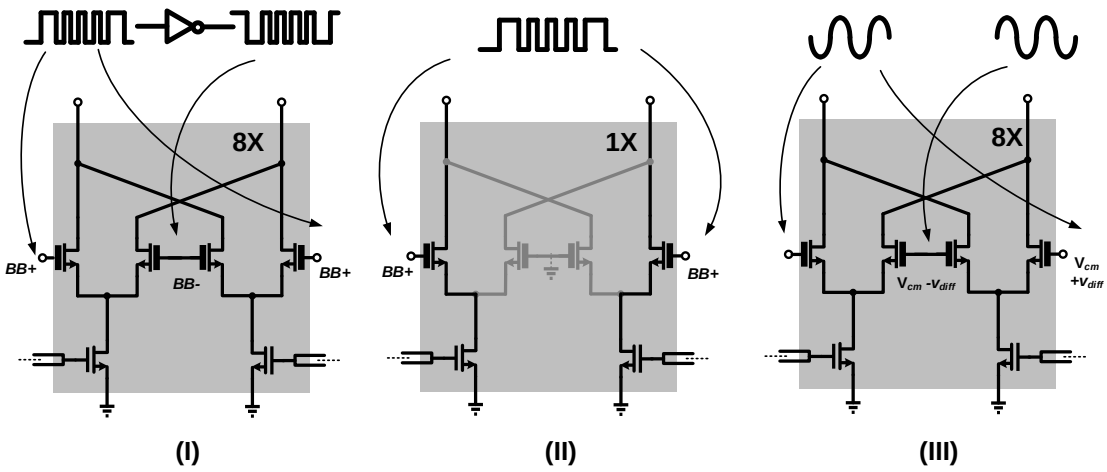


Figure 3.5: Three possible modes of operation of the power mixer.

segments. In simulation, the difference in output power contributed by the first segment (closer to combiner) and the eighth segment (farthest from combiner) was less than 0.9 dB. Figure 3.6 shows the interconnection between the LO transistors and the cross-coupling strategy used for the BB transistors is depicted in Figure 3.7. The cross-coupling structure consisting of transmission lines separated by a ground plane ensures isolation between the differential output nodes, leading to lower LO feed-through at the output.

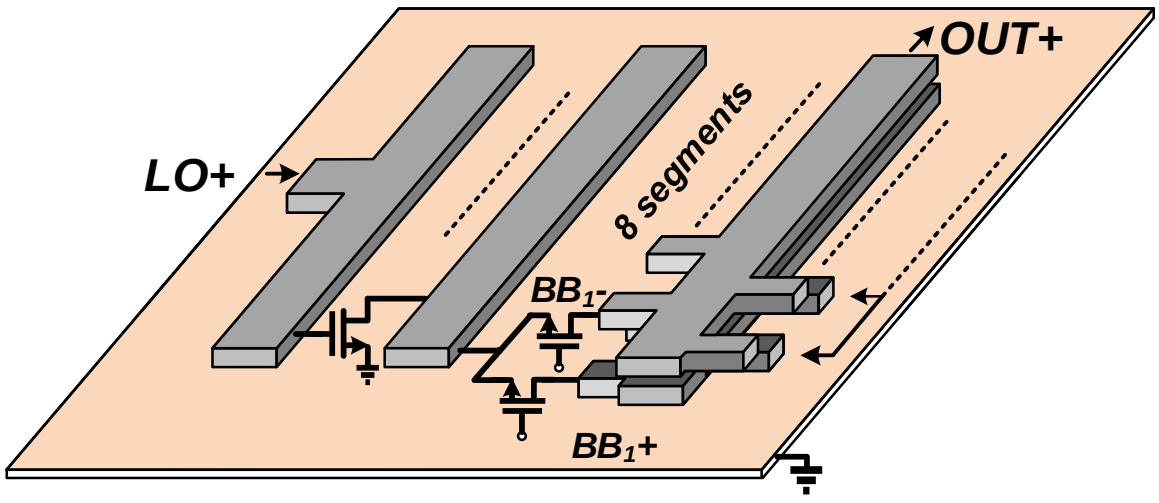


Figure 3.6: Interconnection between segments of the power mixer.

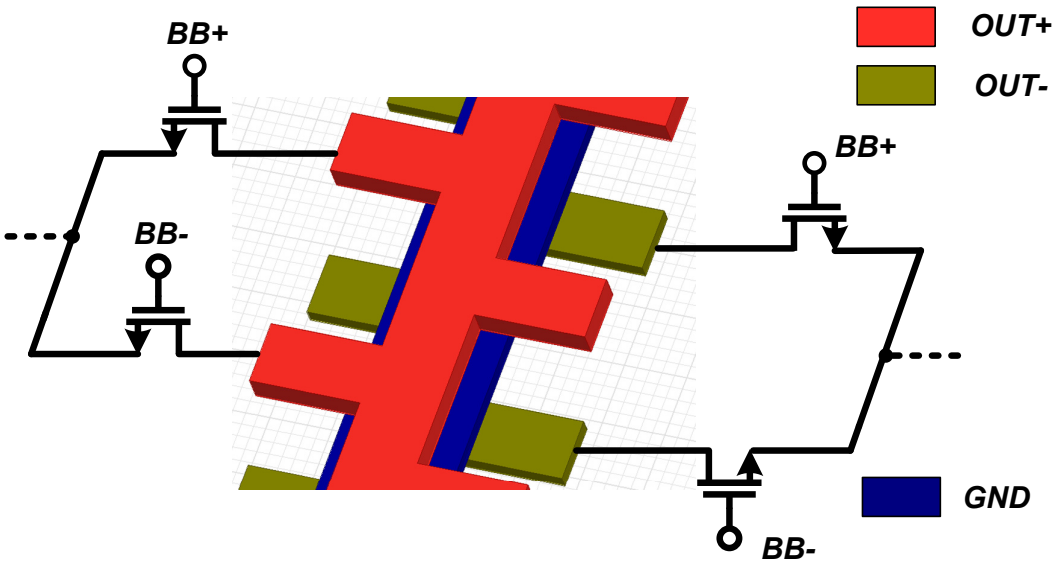


Figure 3.7: Cross-coupling structure for Gilbert cell quad transistors.

3.2.3 Power combining structure : Distributed Active Transformer

One of the major challenges associated with generating high output powers at mm-wave frequencies in deep submicron CMOS processes has been on-chip power combining. Due to breakdown voltage limitations of the transistors, output powers from several stages must be combined in an efficient way while at the same time enabling impedance transformation. The DAT presents an excellent choice for on-chip power combining as well as impedance transformation, while at the same time providing convenient supply connectivity through virtual ground nodes. Although the power combining itself is still efficient at mm-wave frequencies, implementation of the input distribution network feeding input power to more than two differential power stages can become challenging [59].

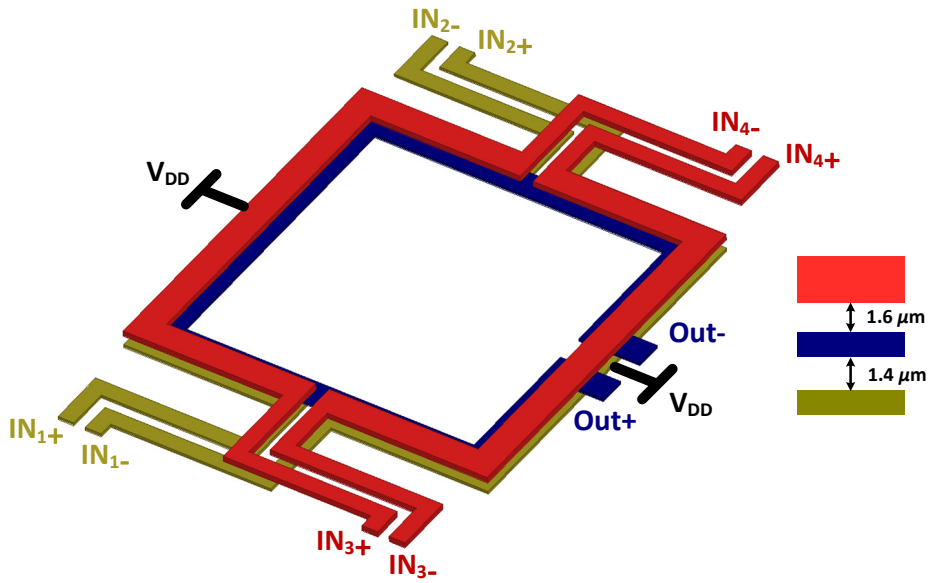


Figure 3.8: Structure of the dual primary DAT based power combiner and cross-section of the metal structure.

Use of such DATs has thus been mostly limited to power combining of two differential output stages. In this design, a dual-primary based DAT has been utilized to combine output power from four output stages in an efficient fashion while minimizing routing challenges of the input network. The secondary of the DAT is sandwiched between the two primaries where each primary is driven by two differential power mixer stages.

This ensures maximum coupling from each of the primaries into the secondary while keeping the input distribution network almost identical to an equivalent two stage power combiner. The three top metal layers 2.25μm Al (primary 1), 1.2μm Cu (secondary), and another 1.2μm Cu (primary 2) form the DAT combiner, as shown in Figure 3.8.

Figure 5 (b) shows the broadband nature of the DAT showing excellent power combining efficiency over frequency. Dimensions were optimized for maximum efficiency at 51 GHz; the DAT occupies a core area of 140 x 140μm^2.

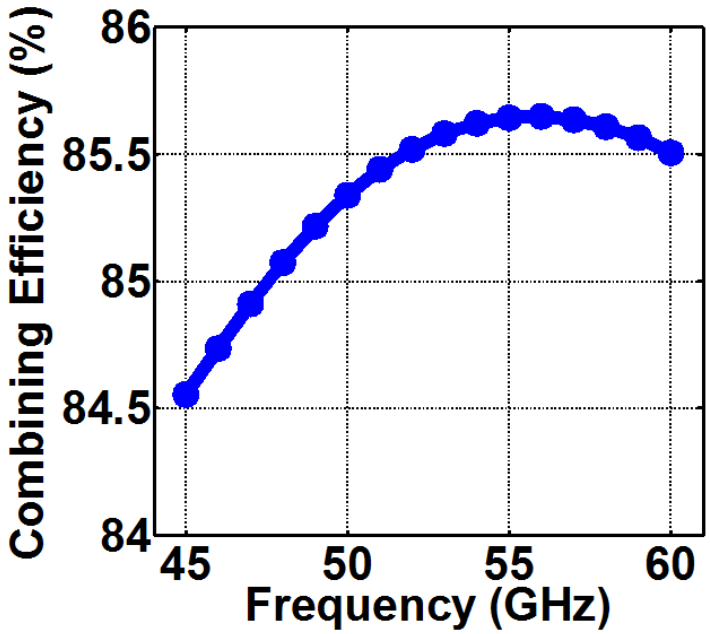


Figure 3.9: Simulated combining loss for the dual-primary DAT.

3.2.4 Input Distribution and Driver Stages

The input single-to-differential balun is implemented using the top two metal layers forming the primary and the next four Cu layers forming the secondary. As with the DAT, the primary and the secondary are placed on top of each other for minimum insertion loss. Figure 3.10 shows the structure of the transformer and the simulated input S_{11} versus frequency is depicted in Figure 3.11. The input return loss is better than 10 dB from 40 GHz to 60 GHz.

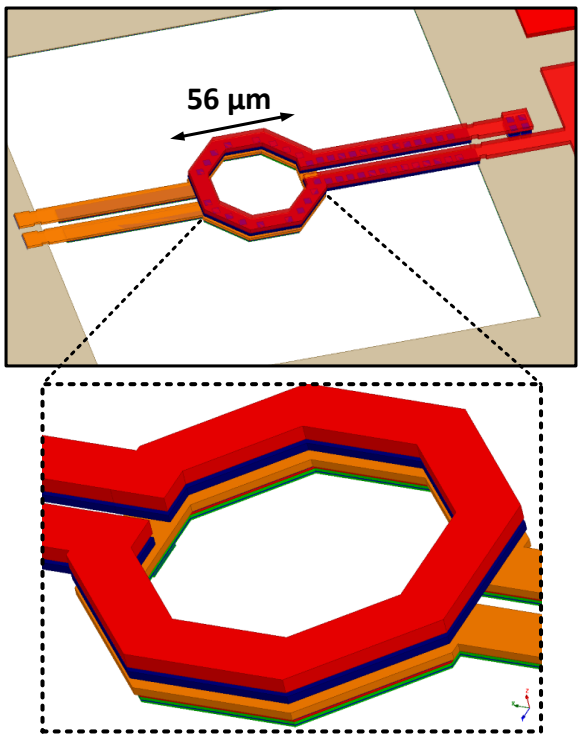


Figure 3.10: Input transformer for single to differential conversion.

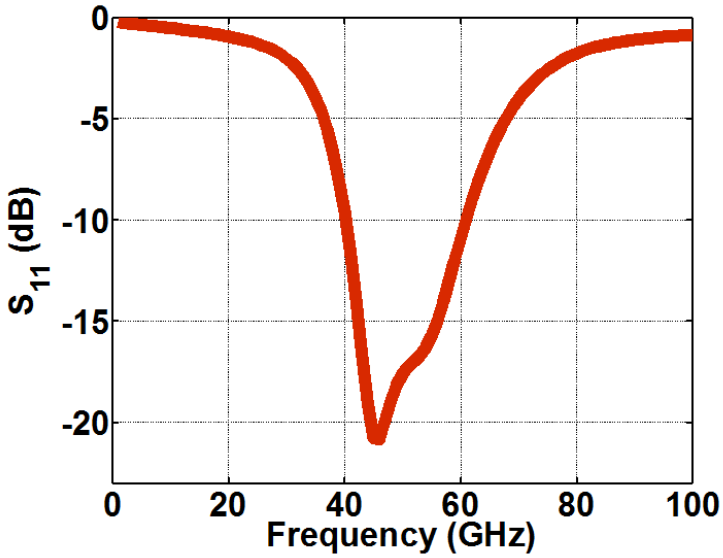


Figure 3.11: Simulated input return loss.

The center frequency of the input matching network together with the input balun is

sensitive to fill inside the inductor layer. Careful electromagnetic simulations including the metal fill, were performed to achieve optimum match at 51 GHz. Simulated excess insertion loss of the balun was 2.6 dB. Simulated differential phase and amplitude errors are depicted in Figure 3.12 over a wide range of frequencies, confirming the operation of the transformer as a single-to-differential converter. Since the driver amplifier stages are driven into saturation at the mm-wave LO frequency, this differential amplitude mismatch is not critical.

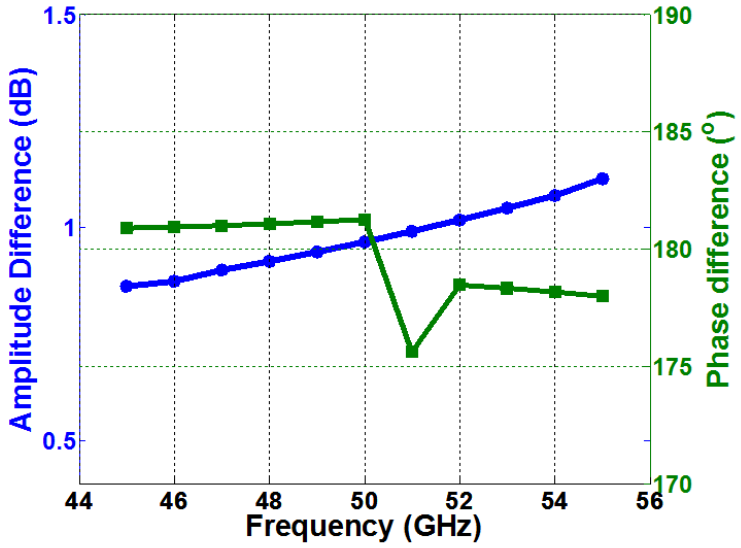


Figure 3.12: Simulated differential amplitude and phase mismatch of input balun over frequency.

Differential output signal from the balun is then amplified by four differential inverter based amplifying chains (Figure 3.13) which compensate for the loss in the input distribution network to provide maximum LO voltage swing to switch the power mixer stages. These inverting amplifiers are increasingly sized from the input transformer onwards. A maximum fan-out of 2 was selected for the chain, as shown in Figure 3.13.

In addition to systematic phase mismatches due to layout constraints, random mismatches due to process variations as well as due to temperature and modeling uncertainties can contribute to significant phase error between the four power mixers. These mismatches affect the drive symmetry of the DAT, thereby reducing the efficiency of the power combining

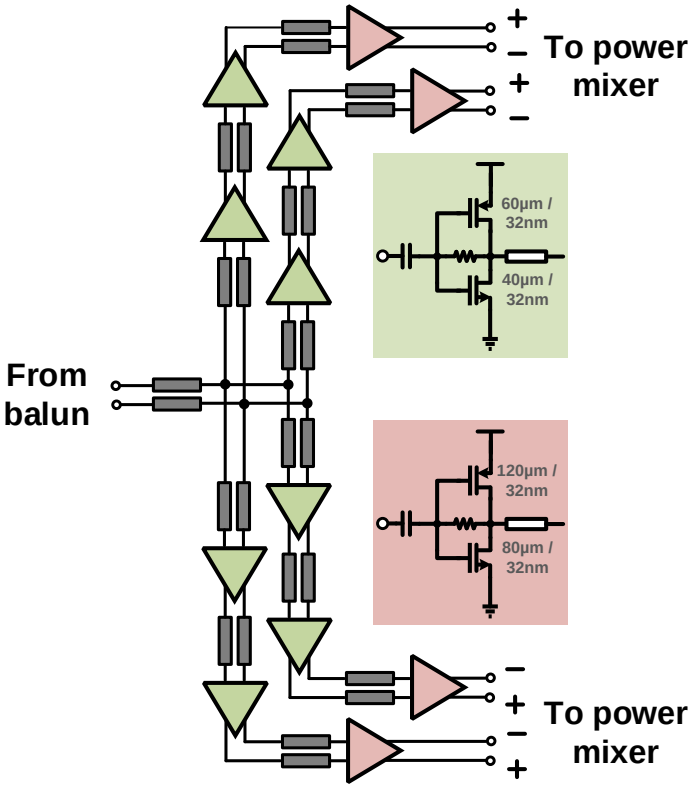


Figure 3.13: Driver stage and input distribution network.

network. Simulation results show output power degradation by more than 1 dB for phase mismatches up to 45 degrees.

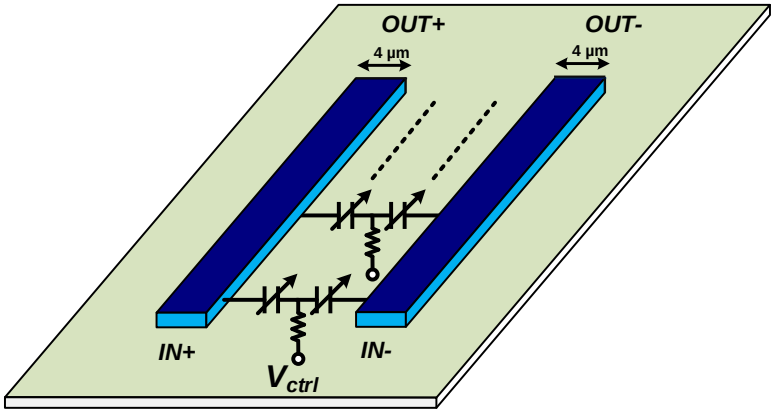


Figure 3.14: Distributed varactor based transmission line.

To compensate for such phase variations, varactor based phase shifters have been incorporated into part of the input distribution network in the form of a controllable delay line,

as shown in Figure 3.14. In simulation, these delay lines achieve more than 90^o of phase shift at 50 GHz by adjusting the varactor control voltages (Figure 3.15).

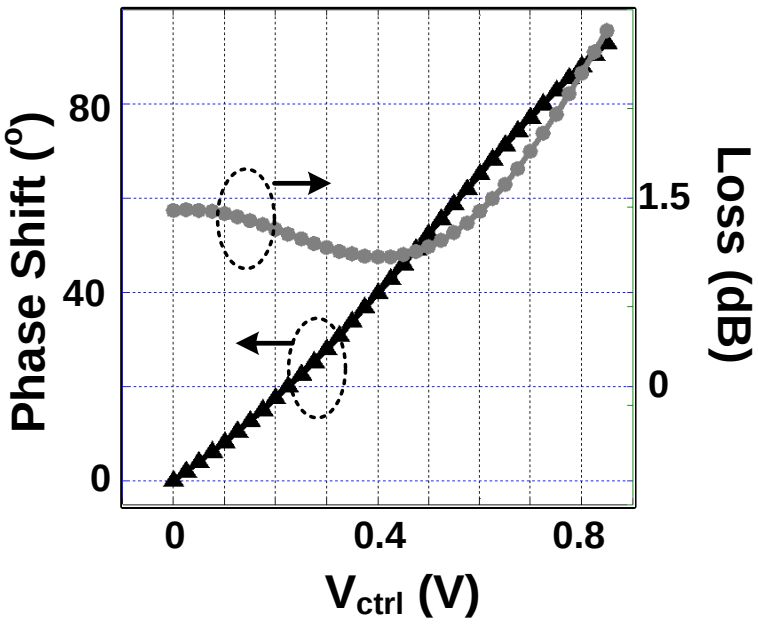


Figure 3.15: Simulated phase shift and loss of phase shifter over control voltage.

3.2.5 Technology and Device Layout

The power-mixer based transmitter and associated transistor test structures were fabricated in a commercial 32nm SOI CMOS process with 11 copper and 1 thick aluminum metal levels. Figure 3.16 shows the die photo of the pad-limited transmitter occupying a core area of 0.51mm^2 including input and output pads, the input balun, and the DAT power combiner.

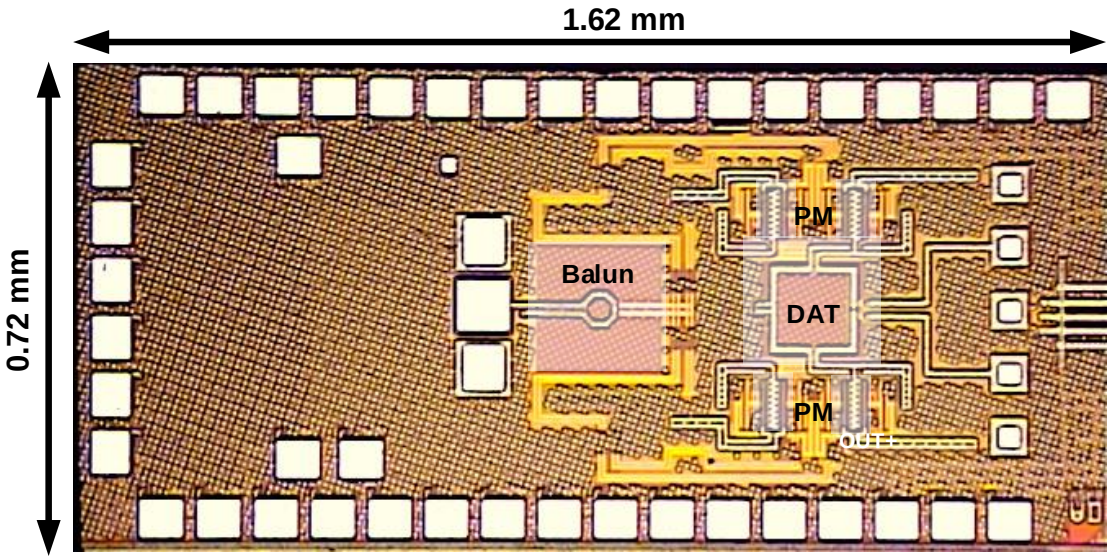


Figure 3.16: Chip micrograph of implemented power mixer based transmitter.

Two types of transistor layouts were fabricated to evaluate and validate the transistor and extraction models (Figure 3.17). For the same current density, a staggered metallization on the source and drain contacts [62] leads to higher transistor G_{max} than the simple stacked metallization.

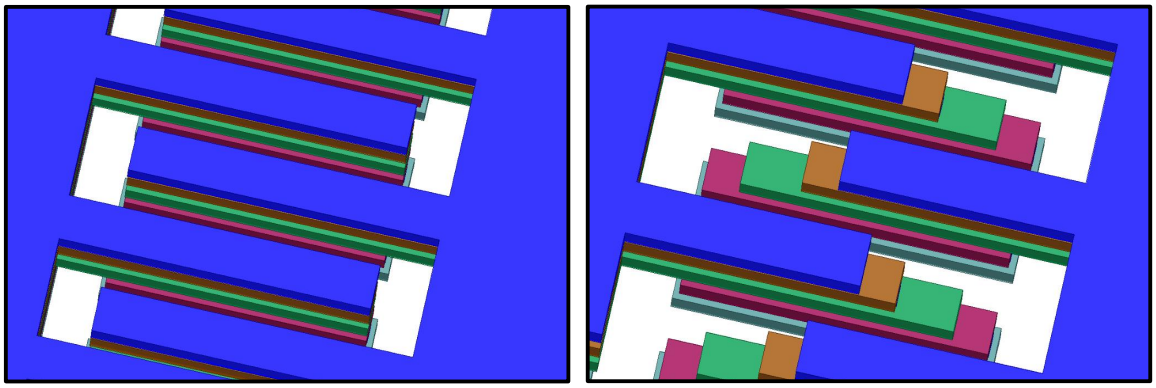


Figure 3.17: Stacked metal connections (Source-Drain) versus staggered for a power transistor layout showing less sidewall capacitance between source and drain.

Figure 3.18 depicts measurements of G_{max} from a minimum gate length, 25μm NFET with 1μm finger width after on-chip SOLT calibration showing good agreement with extracted

simulations. In the present design, only floating body FETs with staggered metal contacts have been used due to their higher performance.

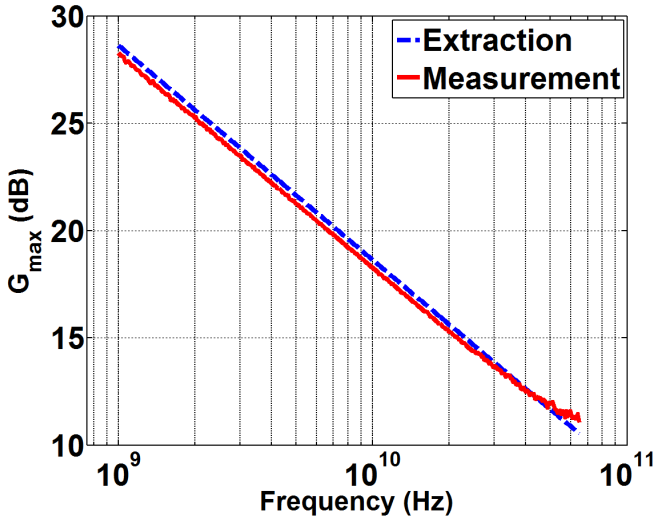


Figure 3.18: Measurement results against post-extracted simulations of a 25 x 1 μm / 32nm transistor up to 67 GHz.

3.2.6 Measurement Results

3.2.6.1 CW measurements

For continuous wave power measurements, the transmitter was driven by a single-ended mm-wave LO signal through a GSG input probe, while the differential output was probed by an SGS probe with one output terminated with a 50-Ω load. Figure 3.19 shows the measured differential output power versus input LO power as well as LO-to-RF gain at small-signal and large signal. A peak output power of 19.1 dBm was measured at 51 GHz with a drain efficiency of 14.2% and a PAE of 10.1%. The inverter based driver stages as well as the power mixer itself contribute to a small signal LO-to-RF gain of 16.2 dB, which reduces to 7 dB in saturation, including the additional loss of the input balun.

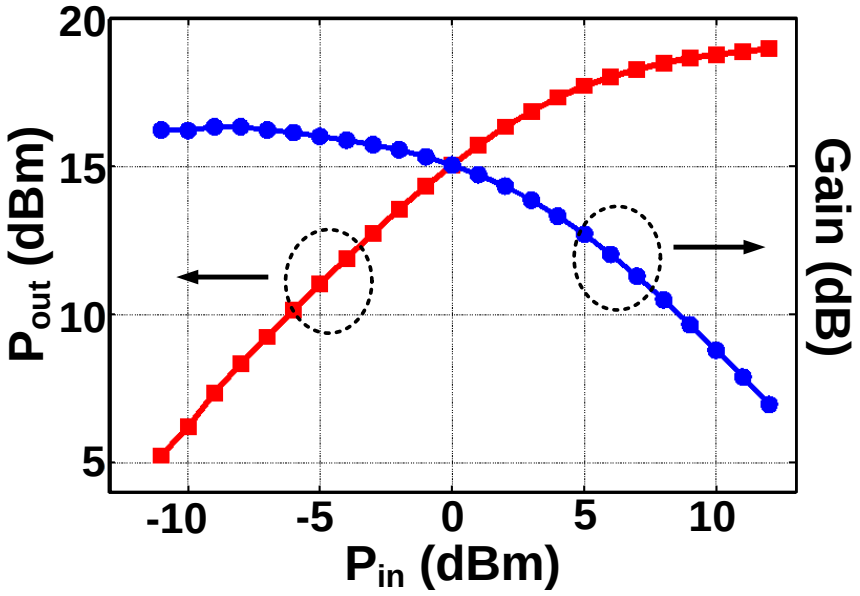


Figure 3.19: Differential output power versus input LO power and LO-to-RF gain.

The segmented power generation scheme of this transmitter allows the DC power to scale with the output power, thereby improving efficiency at back-off. Measurement results in Figure 3.20 show the DC current drawn by the power-mixer stages for two modes: (1) where the output power is varied by controlling the input power and (2) ES mode. As is evident from the measurements, the ES mode provides significant reduction in DC power consumption, especially at back-off power levels. As an example, a 40% reduction in DC power is observed in this mode at 12 dBm output power while improvements are observed across all power ranges. At 6-dB back-off power levels, the drain efficiency improves from 2.8% in mode (1) to about 4.8% in the ES mode, confirming the class-A operation of the power mixer stage. This effect can be further exploited during non-constant envelope modulation leading to higher efficiencies when compared to conventional linear transmitters, as will be shown in later sections.

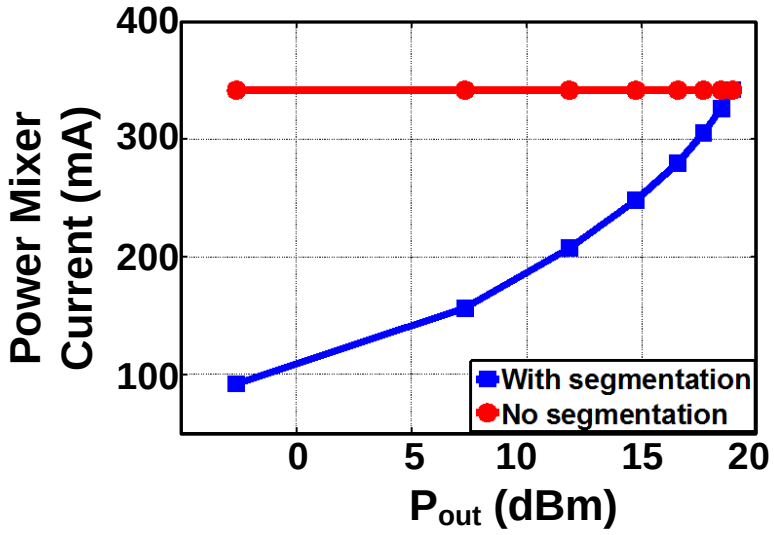


Figure 3.20: Measured DC power savings in segmentation mode.

Measurement results for the ES mode of operation are depicted in Figure 3.21 where the analog segment is controlled by an off-chip DAC. It must be noted that the occasional discontinuities in power at the segment transition points can be attributed to the fact that at mm-wave frequencies, depending on the layout location, all segments do not contribute identically to the output. Output power measurements in the BA mode are also shown in Figure 3.22.

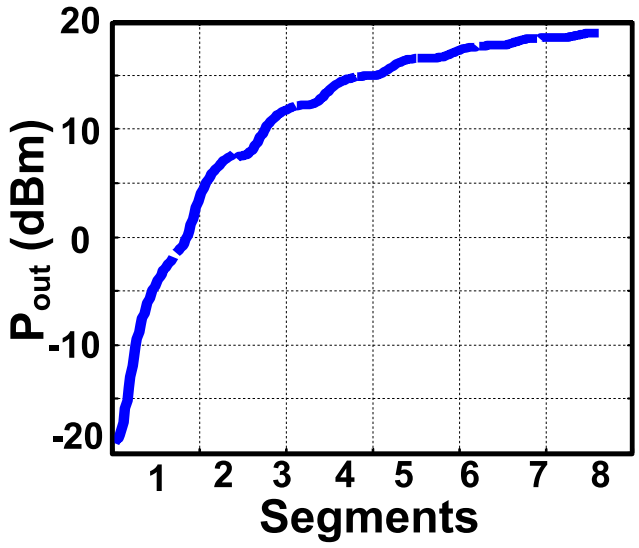


Figure 3.21: Measured output power versus segments in ES mode.

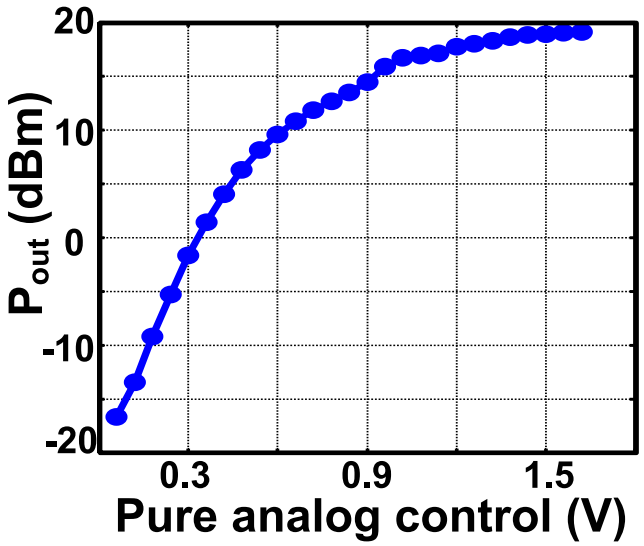


Figure 3.22: Measured output power versus segments in BA mode.

3.2.6.2 Modulation Measurements

The power-mixer transmitter was tested for various modulation schemes at varying rates for both constant and non-constant envelope modulation schemes. Figure 3.23 shows the measurement setup used for high-speed modulation measurements.

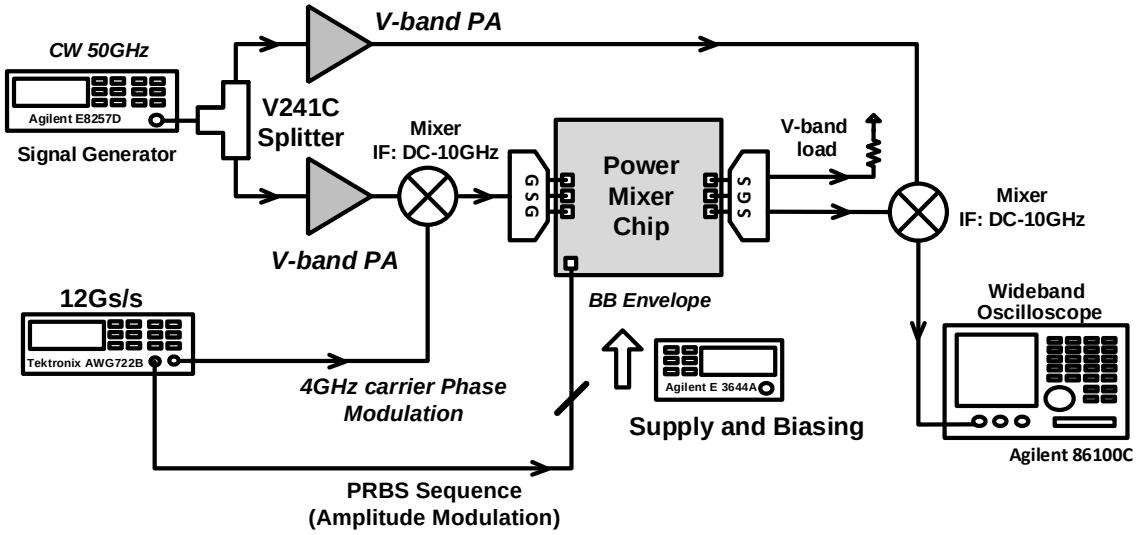
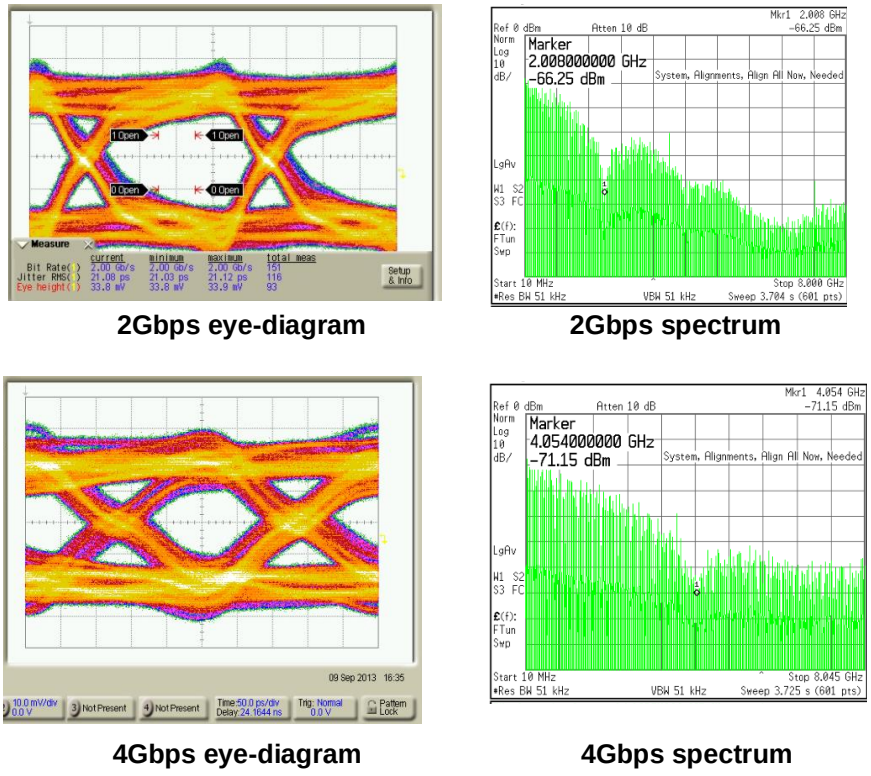


Figure 3.23: Measurement setup for modulation based measurements.

An arbitrary waveform generator (AWG) is utilized to generate constant envelope phase

modulations at a carrier frequency which is then up-converted using a wide-IF mixer. Down-conversion is performed at one of the differential outputs using an identical mixer and the same local oscillator signal while terminating the other differential output. This down-converted data is then captured in time domain using a 20 Gs/s oscilloscope and demodulated in MATLAB. Presence of an IF carrier is required for m-PSK measurements (m≠2), while for simple BPSK modulations, the digital data ($\in \{-1,1\}$) can be directly up-converted using the external mixer.

High-speed modulations are demonstrated at data rates up to 4 Gbps both for BPSK and QPSK with a 50 GHz LO carrier. Figure 3.24 (a) shows measured BPSK eye diagrams at 2 Gbps and 4 Gbps. Spectrum of the modulated output of the transmitter as well as the down-converted signal are also presented in 3.24 (b) at 2 Gbps.



2Gbps eye-diagram **2Gbps spectrum**

4Gbps eye-diagram **4Gbps spectrum**

Figure 3.24: High-speed BPSK measurements showing eye diagrams (a) as well as down-converted spectra in (b) at 2 Gbps and 4 Gbps.

For QPSK modulations, a 2 Gsymbol/sec data stream was generated at a 4 GHz IF

carrier through the AWG. This signal was then up-converted using the mm-wave LO and the transmitter output, then down-converted and demodulated. Figure 3.25 shows the recovered constellation for QPSK modulation at 4 Gbps without any equalization.

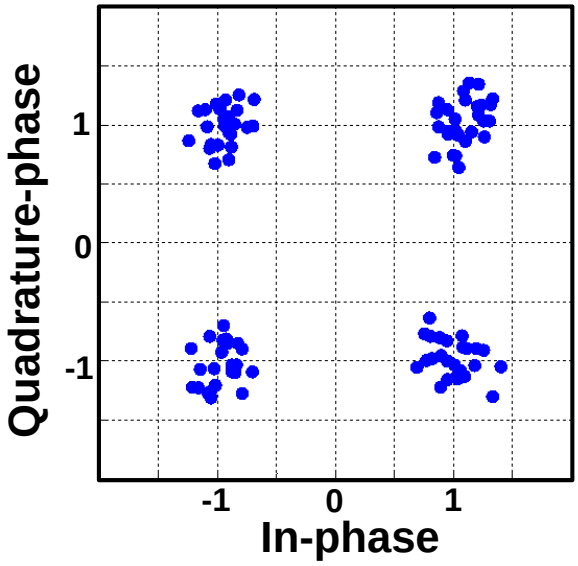


Figure 3.25: Demodulated constellation diagram for 4 Gbps QPSK.

The measured raw error vector magnitude (EVM) was -15.56 dB, of which 67% is attributed to the limited sampling rate of the AWG.

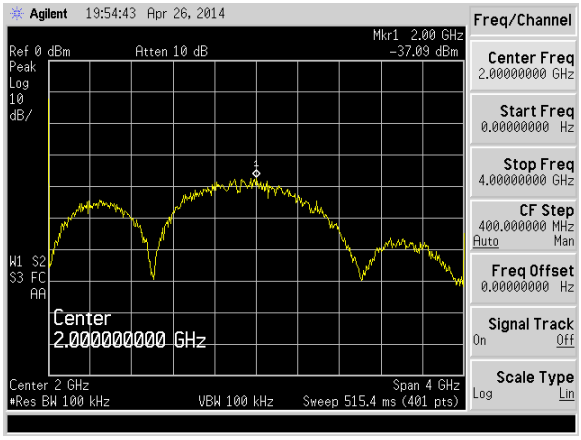


Figure 3.26: Down-converted spectrum for QPSK at 2 GHz carrier and 2 Gbps.

The wide-IF up and down-conversion mixers also contribute to this EVM degradation. Spectral measurements at 2 Gbps and 2 GHz carrier frequency are also shown in Figure 3.26.

Due to the segmented power generation scheme, direct digital control of these power mixer segments leads to high-speed ASK modulations. Additionally, because the DC power also scales with the output power level during segmentation, the overall efficiency during amplitude modulations is much higher than conventional transmitters. For example, the average DC current drops by 29% during binary ASK modulations at full power. The amplitude modulated output signal from the transmitter is down-converted using the external mixer and then viewed in a sampling oscilloscope. Eye diagram of binary ASK modulation at 1 Gbps (limited by test setup) is shown in Figure 3.27.

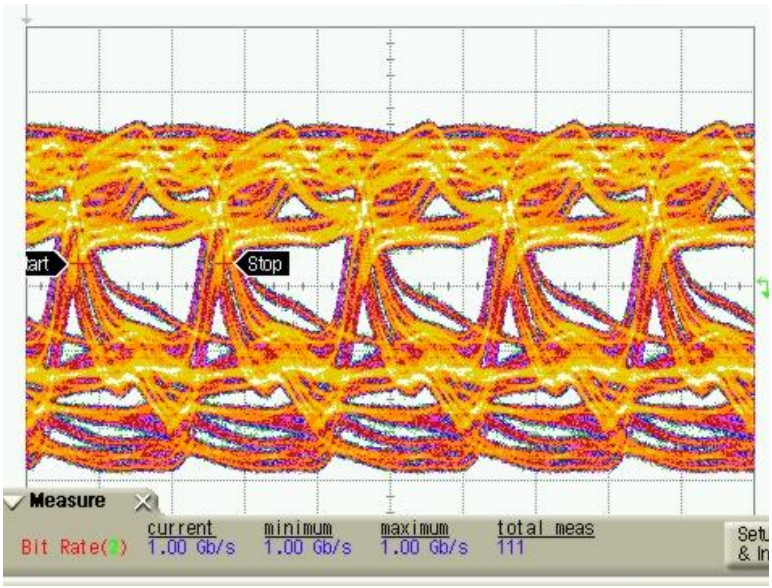


Figure 3.27: Binary ASK modulations using segmentation at 1 Gbps (limited by measurement setup).

By independently controlling the segmentation bits, m-ASK modulations are also possible in this architecture, as shown from the snapshot in Figure 3.28 of the recovered symbols in a 3-ASK modulation at 500 Mbps.

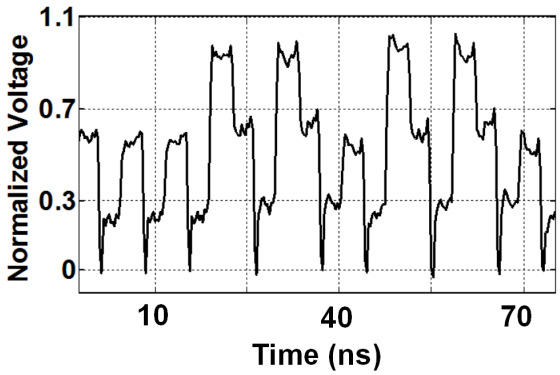


Figure 3.28: Demodulated symbols for duo-binary coding (3-ASK) at 500 Mbps.

By combining the constant envelope LO phase modulations and the segmented power generation discussed above, it is also possible to generate any non-constant envelope modulation schemes with this architecture. For example, to generate a 16-QAM signal, a 12-step phase modulation is generated through the mm-wave LO path and 3 amplitude levels are selected directly by the segmentation bits (Figure 3.29).

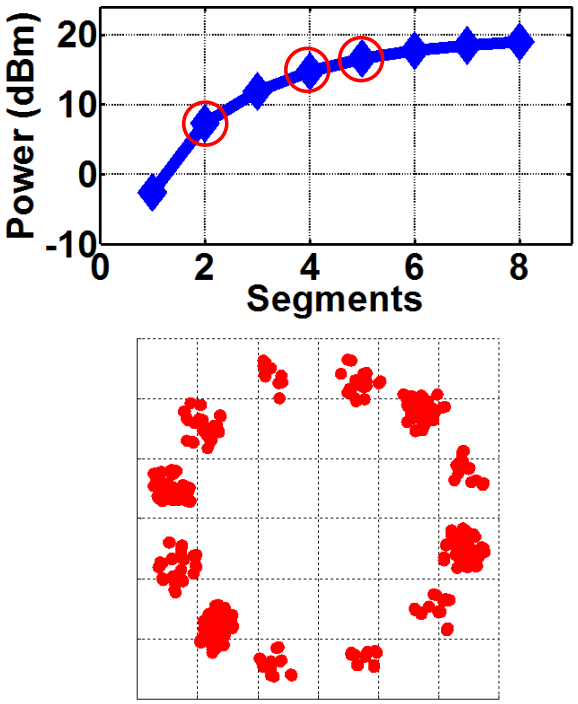


Figure 3.29: Generation of 16-QAM signal using 3 segmentation levels and 12 phases through LO.

By simultaneously modulating both paths, 16-QAM modulated signals can be generated, as shown in Figure 3.30, with a raw EVM of -18.45 dB.

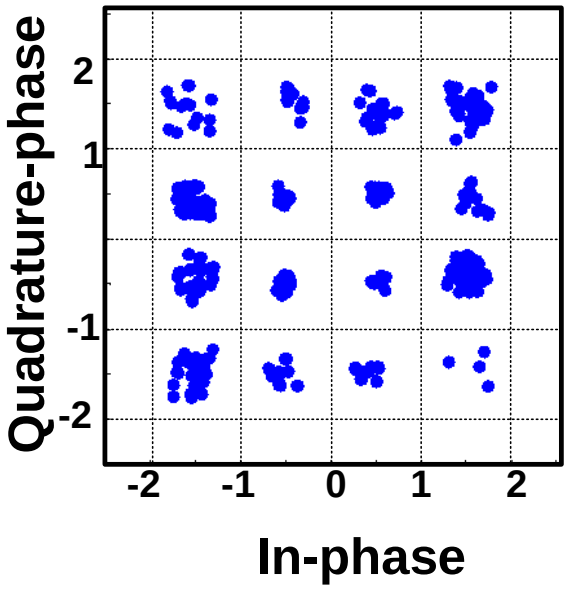


Figure 3.30: Generated constellation for 16-QAM.

3.2.6.3 Calibration Against EVM Due to Measurement Setup

Generation of arbitrary phase modulations (m-PSK) in polar architectures using an arbitrary waveform generator is a widely adopted technique at millimeter wave frequencies. The IF carrier frequency used for gigabit phase modulations is usually in the range of a few gigahertz. Sampling such high frequency waveforms at limited rates of the AWG (12 Gs/s or 24 Gs/s in the present case) introduces significant phase error at the source itself, thereby degrading the input constellation. For example, as shown in Figure 3.31, in an m-PSK modulation, a phase change due to the data can be delayed in the worst case by a full sampling period, causing severe degradation of the modulated constellation. In addition to this sampling error, other effects can be introduced by the non-idealities of any external up/down conversion mixers and/or amplifiers used in the setup. For a given pseudo-random-binary-sequence (PRBS) of symbols (generally used to characterize such systems), these systematic errors repeat for

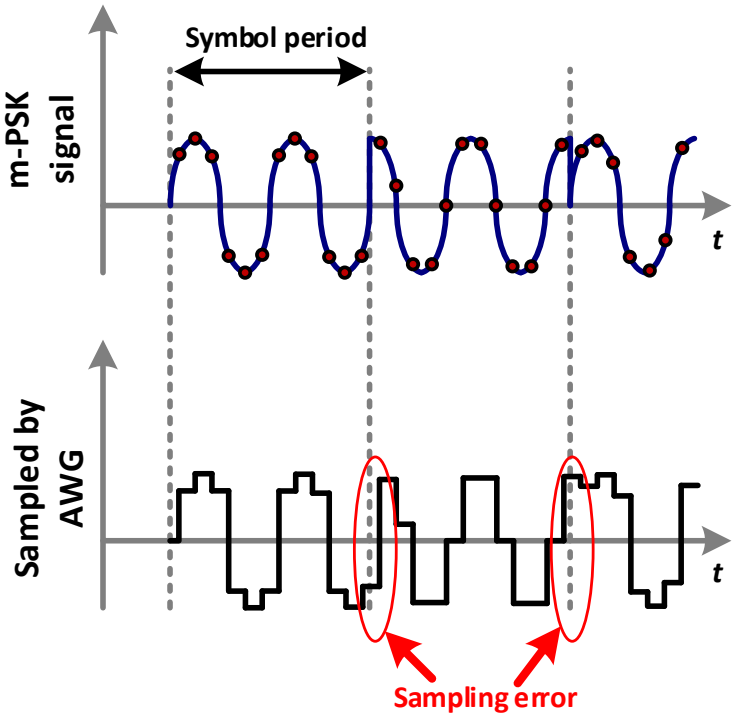


Figure 3.31: EVM introduced by finite sampling rate of AWG for an m-PSK signal.

each sequence and as such the total EVM can be completely characterized by computing error vectors for the first sequence only. The systematic EVM contributed by the setup and the AWG source can be characterized by making a through measurement of the entire setup without the transmitter chip and then calibrating the demodulated constellation at the output using the error vector obtained only from the first sequence. Figure 3.32 shows demodulated data obtained from the through measurement without the chip, for a 2 GHz carrier, 2 Gb/s QPSK signal sampled at 12 Gs/s using the AWG. The EVM was improved from -17.93 dB to -27.7 dB using the calibration method.

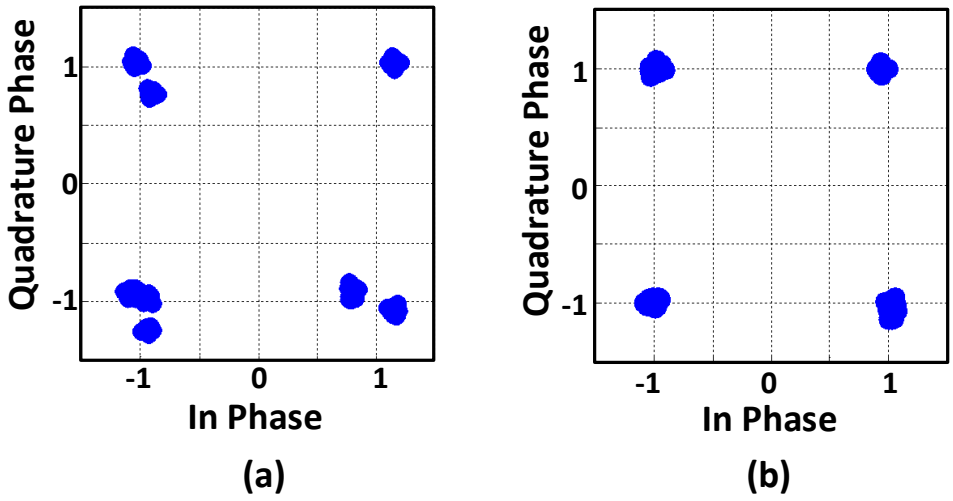


Figure 3.32: Demodulated symbols for QPSK modulation (2 GHz IF carrier, 2 Gb/s) from thru measurement without the chip (a) before and (b) after calibration.

Once this systematic EVM has been characterized, by deploying the same error vector on the recovered data from the transmitter chip, the degradation in EVM contributed by the test setup can be calibrated out. Figure 3.33 depicts the flow-chart of the calibration scheme.

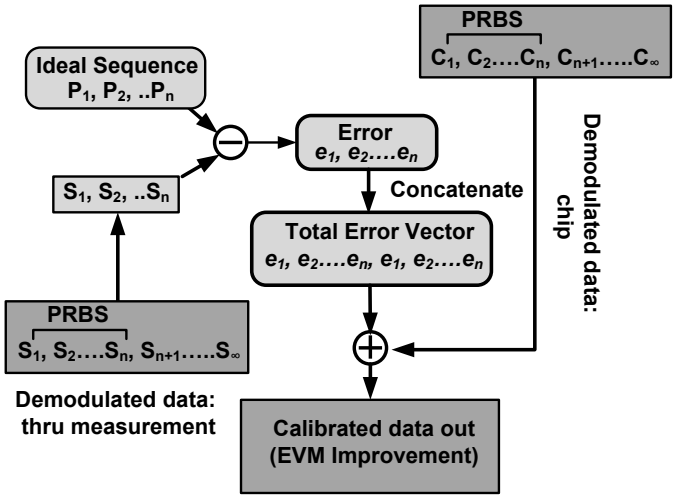


Figure 3.33: Flowchart showing EVM calibration to remove effects due to AWG sampling and systematic mixer non-idealities.

At the same carrier frequency and data rate, the EVM from the transmitter was improved

from -15.57 dB to -23.74 dB. It must be noted that in addition to this systematic EVM due to the setup, additional errors are also introduced by inter-symbol-interference caused by the external components which are not calibrated out in the present scheme. Figure 3.34 shows recovered constellations for a 2 Gb/s QPSK at an IF carrier frequency of 2 GHz from the transmitter before and after calibration. The technique is general and can be applied to the generation of modulations through any sampling based scheme as well as for non-constant envelope modulation schemes.

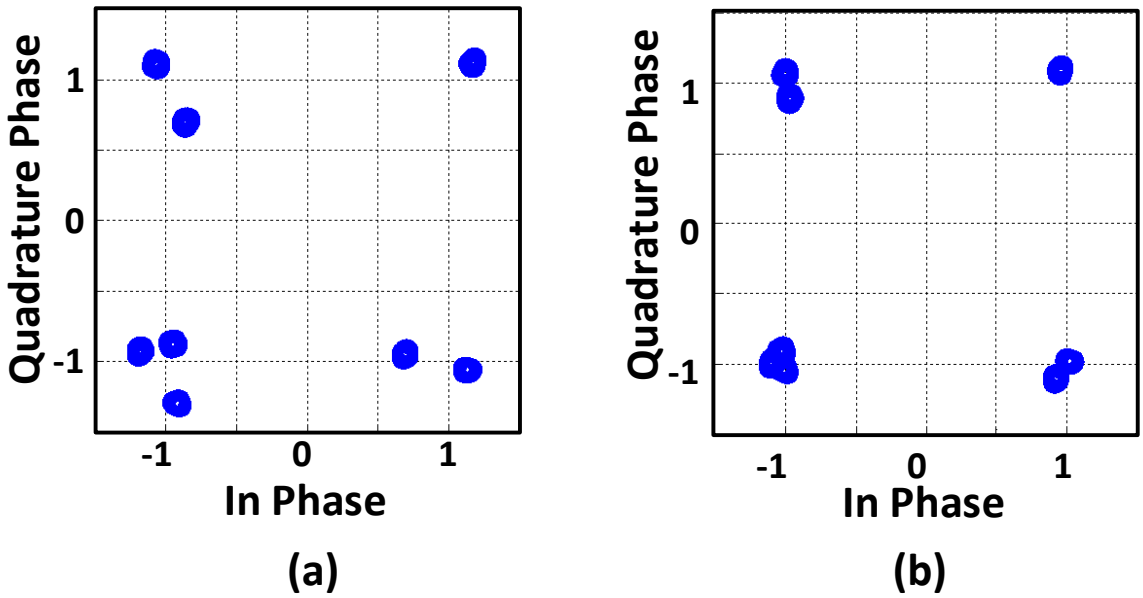


Figure 3.34: Demodulated symbols for QPSK modulation (2 GHz IF carrier, 2 Gb/s) from chip without and (b) with EVM calibration.

3.2.6.4 Reliability Measurements Under Stress

Due to the high voltage swings between the device terminals, high-power transmitters or power amplifiers in CMOS are susceptible to transistor aging and performance degradation when operated for sustained periods of time and/or at increased supply voltages and elevated temperatures. Another problem arises in polar PAs or segmentation based high-power transmitters when the amplifier is operating at close to its maximum output power with one or two segments turned off. In such a situation, the OFF segments experience a large voltage

swing across them since the overall output swing is still close to its peak value. Repeated operation in these conditions can lead to either long term performance degradation or in some cases, catastrophic transistor breakdown. Reliability of the power-mixer based transmitter in the present design has been verified against both these mechanisms of breakdown by operating it at elevated supply voltage over long periods of time and with worst case segmentation, as shown in Figure 3.35. Table 3.1 shows the present work compared with some of the recently published work.

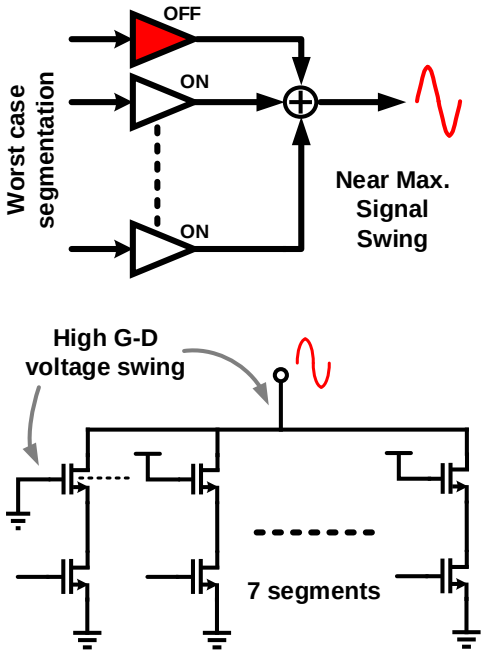


Figure 3.35: Stress on OFF segment near maximum output power/swing.

Only one of the 8 segments was kept off for progressively longer time periods and was turned on after each period to evaluate its performance degradation, if any. As shown in Figure 3.36, no degradation in output power was observed at an increased supply voltage of 2.2-V (30% higher than nominal) over an 8-hour measurement window.

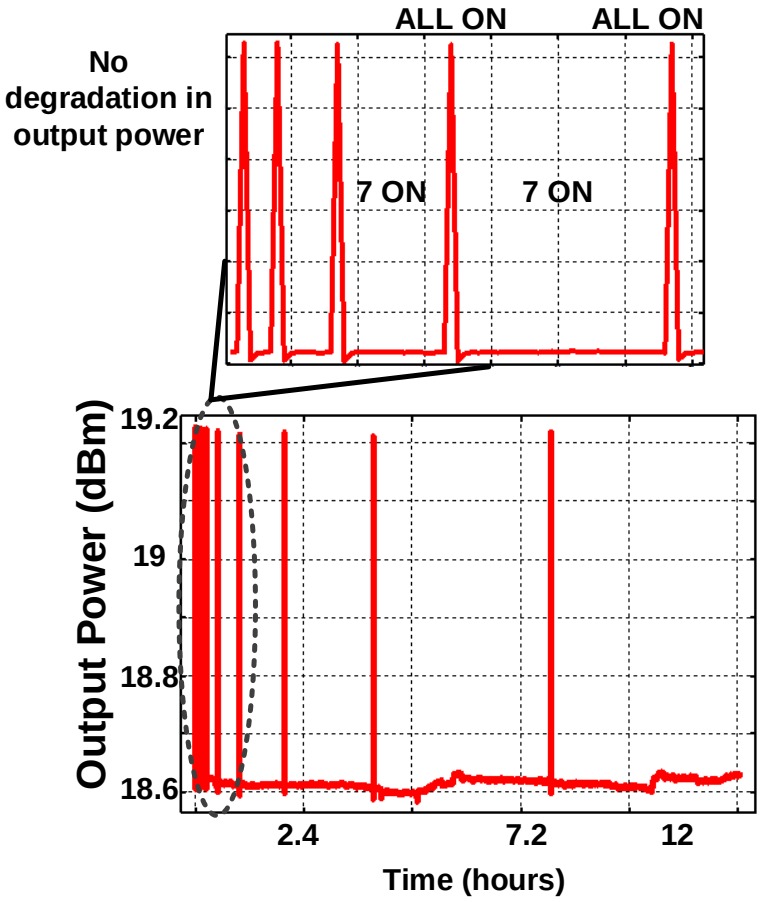


Figure 3.36: Measured output power over an 8 hour window showing no degradation in output power.

Table 3.1: Comparison with recently published work.

Metric	This Work	ISSCC'10 [63]	ISSCC'11 [56]	ISSCC'13 [52]	JSSC'14 [64]	JSSC'13 [51]
Frequency (GHz)	51	60	60	60	85-90	45
Process Technology	32nm SOI CMOS	65nm CMOS	65nm CMOS	65nm CMOS	45nm SOI CMOS	45nm SOI CMOS
P_{sat} (dBm)	19.1	17.9	18.6	9.6	19	24
Supply Volage (V)	1.7	1.0	1.0	1.0	6.8	5.1
Peak PAE (%)	10.2	11.7	15.1	17.4	8.9	14.6
Core Area (mm^2)	0.51	0.83	0.28	N/A	N/A	0.77
Modulations	ASK, m-ASK, BPSK, QPSK, 16-QAM	N/A	N/A	QPSK, 16-QAM	ASK, OOK	ASK, BPSK, QPSK

3.3 Digitally Modulated Self-healing mm-Wave Transmitter

3.3.1 Performance Variations

Process and environmental variations affect the performance of the power-mixer transmitter. Figure 3.37 shows output power and PAE variations versus gate bias voltage variation for a single differential cascode power stage. Significant reductions in P_{out} and PAE are observed as the gate bias voltage varies.

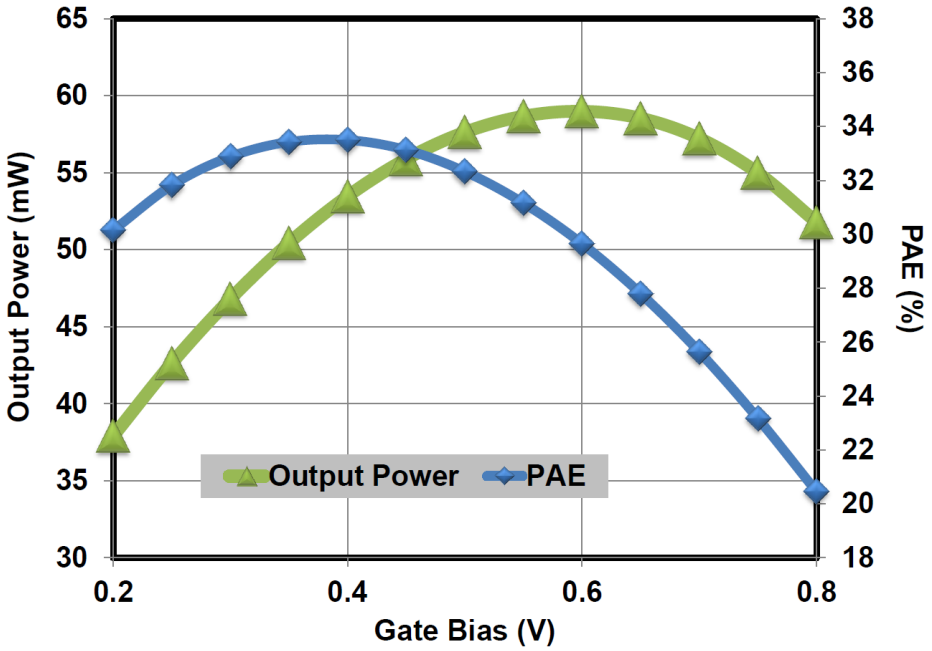


Figure 3.37: Simulated output power and PAE from a differential power stage versus V_{GS}.

Monte Carlo simulations were also performed to evaluate performance variations due to process variations and mismatch. Histograms of output power and PAE are shown in Figure 3.38 and Figure 3.39 for 100 Monte Carlo runs. A 3-σ variation of 10.78 mW and 5% were observed for P_{out} and PAE.

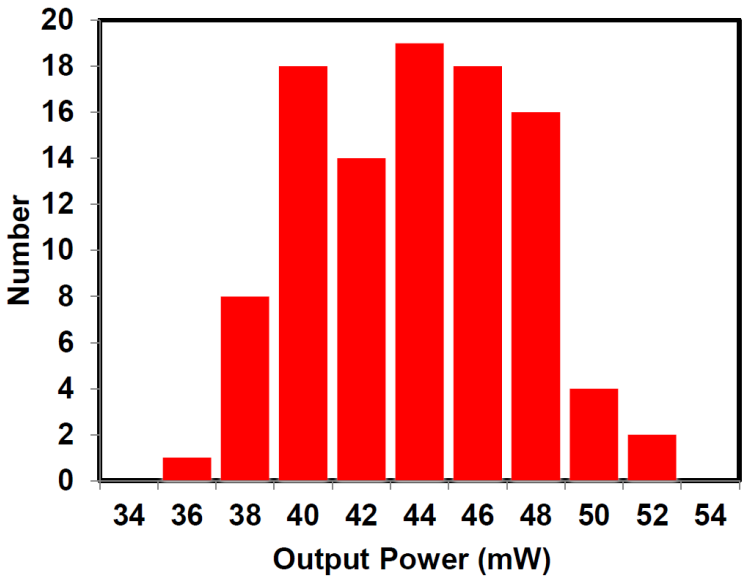


Figure 3.38: Monte Carlo simulation showing variations in P_{out} of one differential power stage.

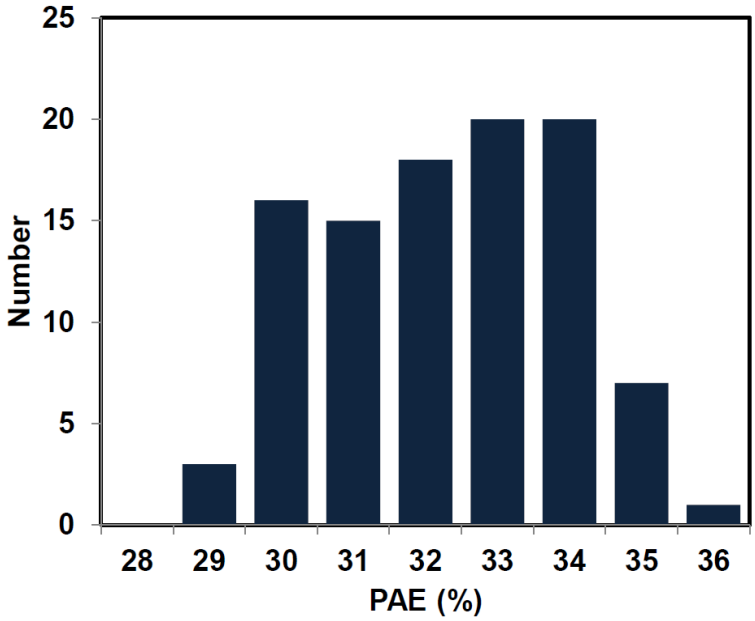


Figure 3.39: Monte Carlo simulation showing variations in PAE of one differential power stage.

To counter these variations, and utilizing some of the sensors and actuators discussed

in the previous chapter, self-healing was implemented for the power mixer transmitter as discussed in the following subsection.

3.3.2 Chip Architecture

Figure 3.40 shows the architecture of the self-healing power-mixer transmitter chip fabricated in the same 32nm SOI CMOS process.

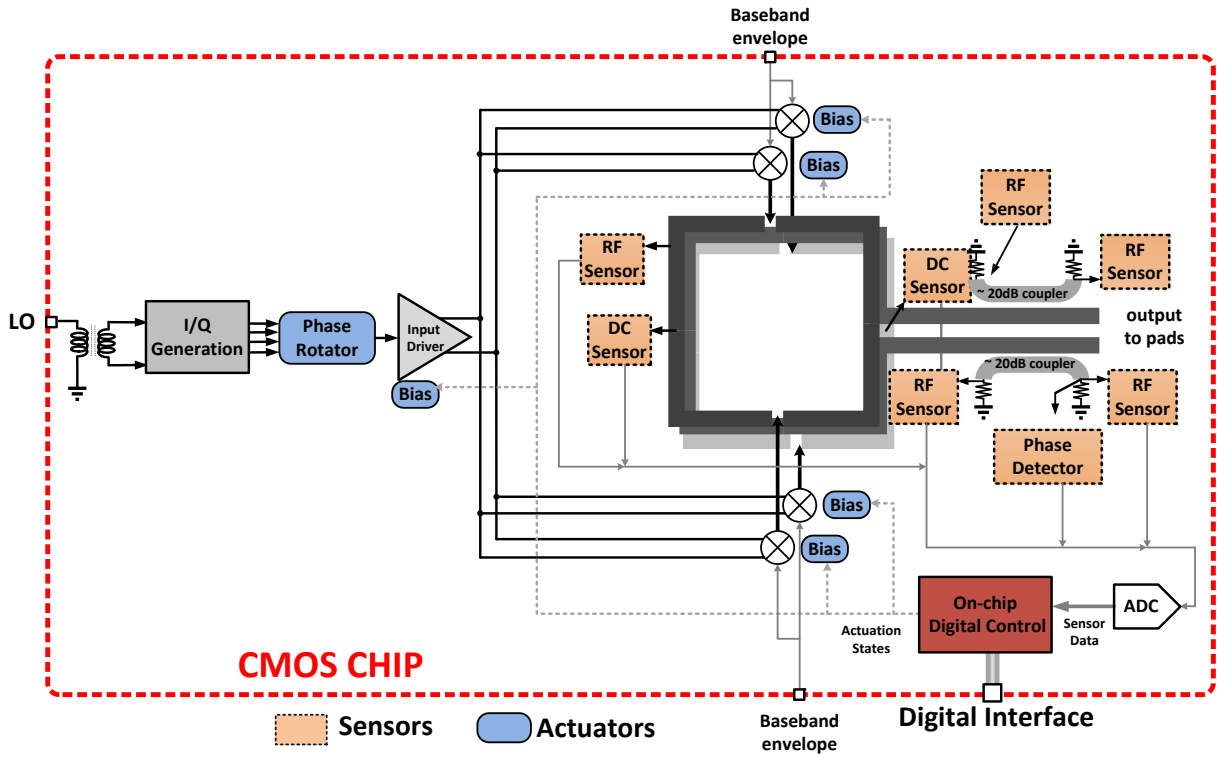


Figure 3.40: Self-healing power mixer based transmitter architecture.

Similar to the architecture in section 3.2.1, the chip utilizes a single-ended mm-wave LO signal as input, which is then converted to a differential form using the on-chip balun. After the conversion, in-phase and quadrature phase signals are generated using true time delays. Any arbitrary phase can now be generated using an interpolation based phase shifter [65]. The phase shifted mm-wave signal then goes through two driver stages before feeding the differential cascode power stages. The differential outputs of these stages are then combined

using an identical dual-primary DAT as discussed in section 3.2.3.

To enable self-healing, output power both in the forward and reverse direction are sensed using RF power sensors described in Chapter 2.2.2. Because of the differential nature of the output, there are four power sensors sensing both the differential imbalance as well as the real power delivered to the load. At virtual ground nodes along the length of the DAT primaries, two DC sensors similar to that described in section 2.2.3 are placed which provide regulated supply voltage as well as sense a fraction of the current drawn by the four power stages. DC output voltages from these sensors are digitized using on-chip 8-bit SAR ADCs (section 2.4) interfaced through a multiplexer to save area and power. Bias voltage actuators similar to those discussed in section 2.3.1 are used to control gate bias voltages of the power stages as well as the two driver stages.

The digital control is implemented in the form of a 8051 micro-controller synthesized using IBM's standard cells. The bias voltage actuators are directly controlled through a register widget which can be directly accessed by the micro-controller or written externally using a serial interface. Figure 3.41 shows the basic digital infrastructure.

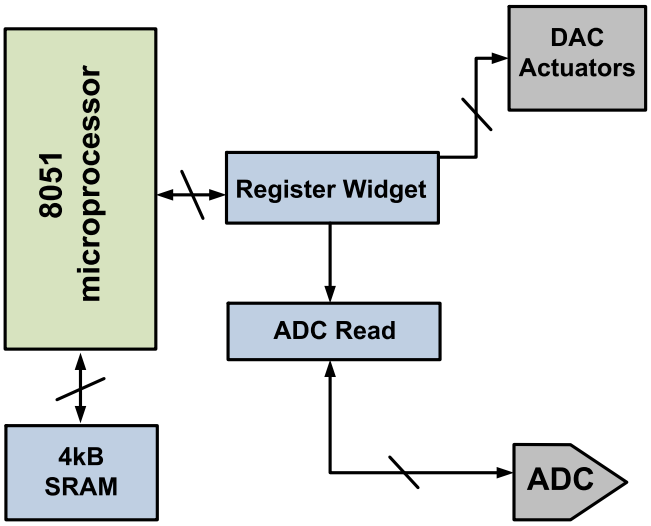


Figure 3.41: Digital infrastructure for the self-healing power mixer chip.

ADCs are read using a separate state machine, the flowchart of which is shown in Figure 3.42.

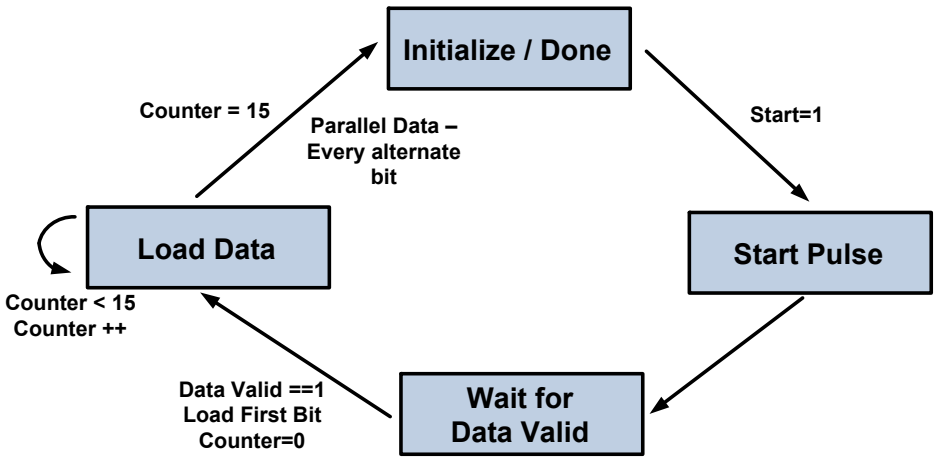


Figure 3.42: Flowchart showing ADC read operation.

Depending on the which sensor is being read, the digital control block sets the MUX selection bits and sends a start pulse to the ADC. After comparison, once the sensor output has been digitized, the ADC sends out a Data Valid. The state machine waits till the Data Valid arrives and then starts loading the serial data into a register. Note that, as discussed in section 2.4, the ADC generates two quadrature clock edges from one clock division to ensure that the comparator does not change states while its output is being clocked. This means that the output serial data arrives at half the clock rate as the state machine. To ensure data reliability, the state machine now samples the serial data at the full clock rate (16 bits for an 8-bit serial data) and compares each of the two samples corresponding to one serial data bit to check for errors. Once all the 8 serial bits have been captured, the state machine signals the global control that the parallel data is ready.

3.3.3 Power Stage Design

The power stages were implemented in the form of cascode stages with segmentation (Figure 3.43). Similar to section 3.2.2, the cascode transistors are thick-oxide devices to sustain high drain voltage swings. As shown in the figure, each power stage is segmented into four smaller stages. The bottom transistors in each such segment (M_1 and M_2) are 32nm transistors with

25 fingers of 1 μm width each. The top transistors similarly (M_3 and M_4) are sized as 1×25 μm/100nm. The power stage operates off a supply voltage of 1.8-V.

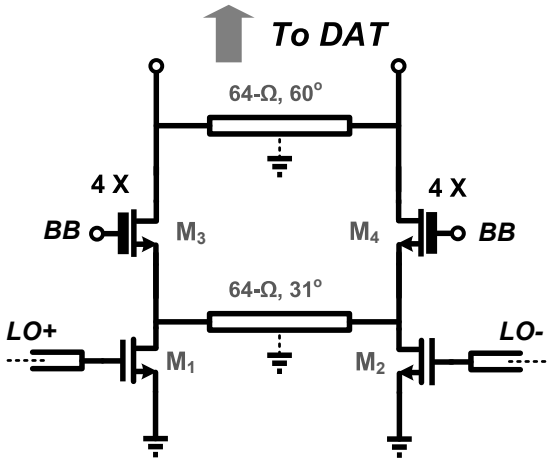


Figure 3.43: Schematic of single power stage.

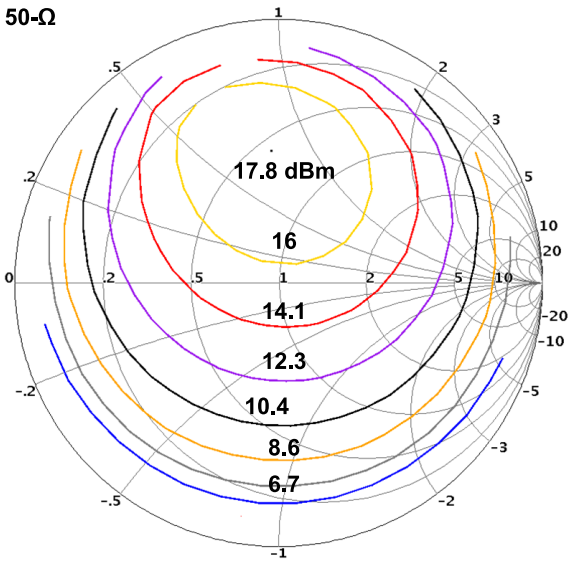


Figure 3.44: Simulated loadpull contours of a single power stage.

Out of a total of four segments, three are digitally controlled (through cascode voltage) and one is controlled through an on-chip DAC to ensure power continuity. Simulated load-pull contours of the one power stage is shown in Figure 3.44.

3.3.4 Driver Stages Design

The driver stage is composed of two cascode amplifiers with a 50μm buffer driving a 200μm stage through an interstage matching network. The driver operates off a V_{DD} of 1.2 V. Schematics of the stages are shown in Figure 3.45.

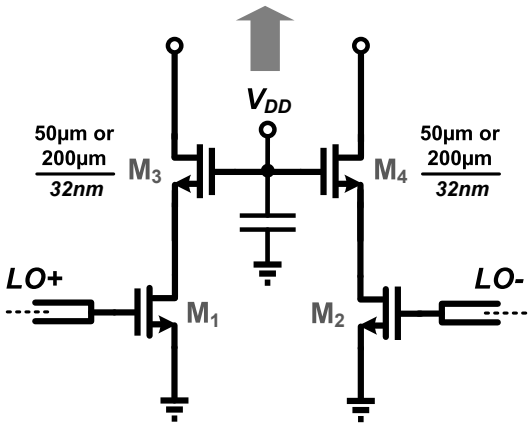


Figure 3.45: Schematics of cascode buffers.

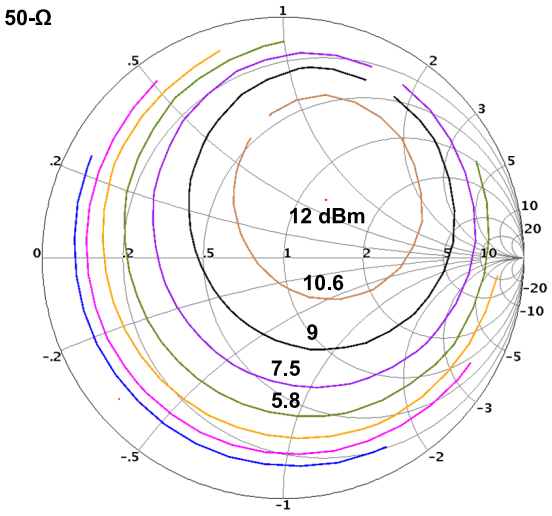


Figure 3.46: Loadpull contours for the 50-μm buffer.

The stages are matched through a large signal load pull match. Figure 3.46 shows simulated load pull contours of the 50-μm buffer and that of the 200-μm is depicted in

Figure 3.47. The two drivers (50 & 200 μm) draw 40 mA and 148 mA from a 1.2 V supply, respectively.

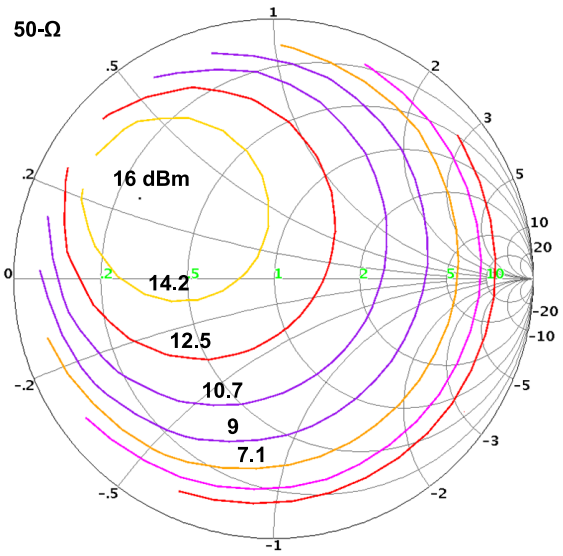


Figure 3.47: Loadpull contours for the 200-μm buffer.

3.3.5 IQ Generation & Phase Interpolator Design

On-chip I & Q signals are generated using a transmission line delay which is approximately 90° at 60 GHz, as shown in Figure 3.48. Coupled transmission lines of a lower characteristic impedance of 40-Ω odd mode impedance are utilized in this structure primarily for lowering the loss.

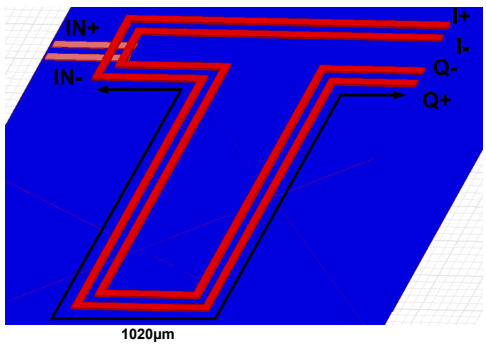


Figure 3.48: On-chip delay based IQ generation.

The exact phase shift is not very critical since the phase can be calibrated by sweeping through the phase rotator settings once. Electromagnetic simulations were performed in HFSS to verify the operation of the structure. Simulation results in Figure 3.49 show voltage waveforms at the outputs when the structure is matched (at the input) to the transformer and (at the output) to the input of the phase rotator. Note that due to the additional quarter wavelength long path ($\lambda/4 \sim 650\mu$m at 60 GHz), the Q-path has additional loss which can again be compensated and calibrated by the phase rotator.

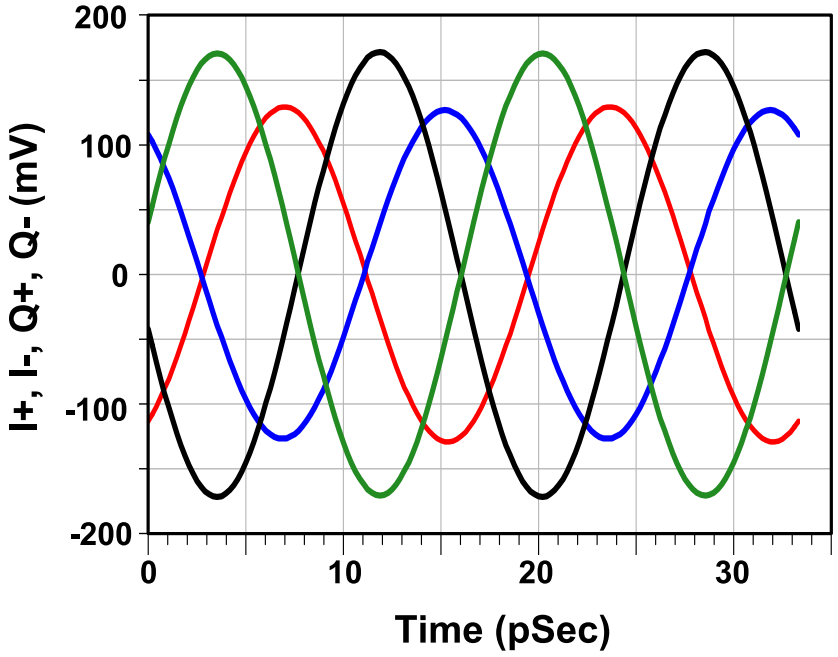


Figure 3.49: Simulated voltages at the output of the IQ generator structure.

The phase interpolator is based on current addition of weighted I and Q components, as shown in Figure 3.50. Each control voltage is operated through a 4-bit R-2R ladder based DAC.

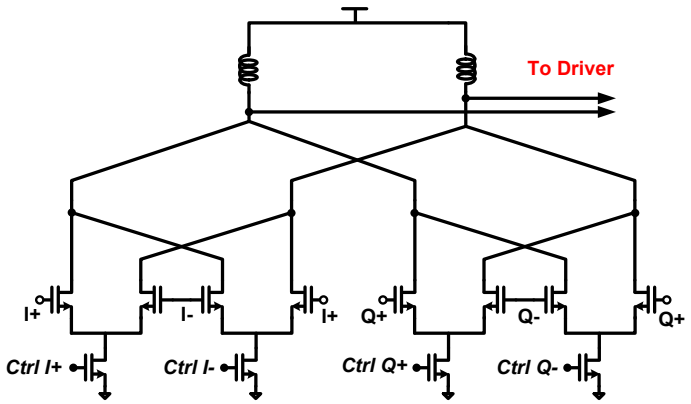


Figure 3.50: Schematics of the phase interpolator.

Figure 3.51 show simulated voltages (magnitude and phase) of several representative points having identical amplitude but varying phases, showing the ability of the phase interpolator to generate constant power variable phase signals across all four quadrants.

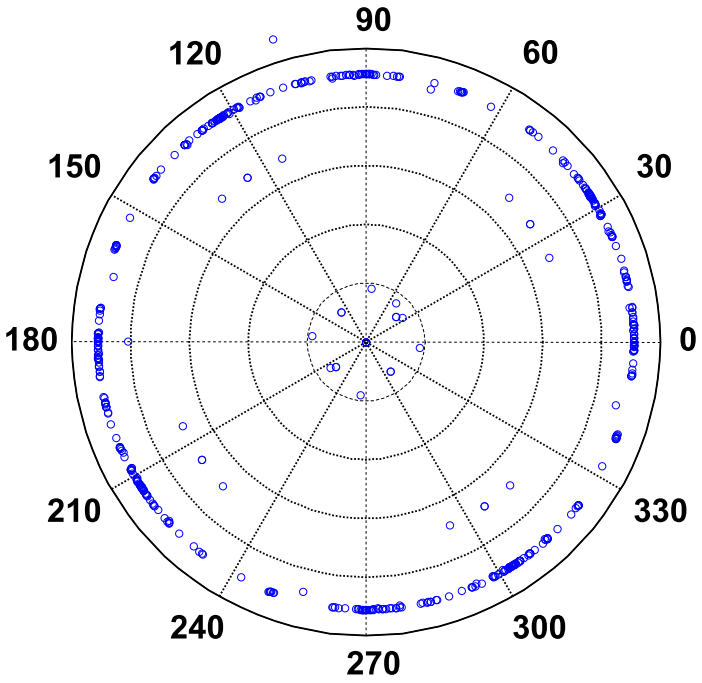


Figure 3.51: Simulated output phases of the phase interpolator.

3.3.6 Measurement Results

The chip was fabricated in IBM's 32nm SOI CMOS process. Figure 3.52 shows the die micrograph of the fabricated chip showing critical circuit blocks and electromagnetic structures. The micro-processor and the RAM occupies a relatively large area, however, in a full system these same elements will be utilized for self-healing all other blocks of the full transmitter/PLL.

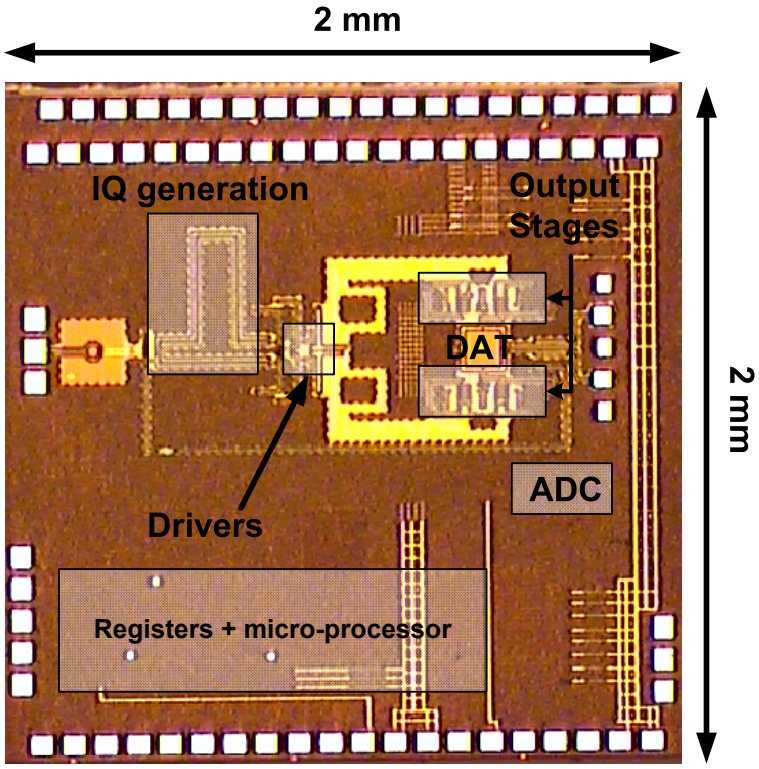


Figure 3.52: Die micrograph of self-healing power mixer chip.

3.3.6.1 Bias Actuation Measurements

The chip was first tested to verify operation of the digital bias actuators. The register widget was programmed through the serial interface using a Xilinx ZynQ FPGA board and an FMC cable. Figure 3.53 shows measured DC current drawn by the four power stages versus DAC

bias setting. Results show good matching between transistors and the controlling DACs.

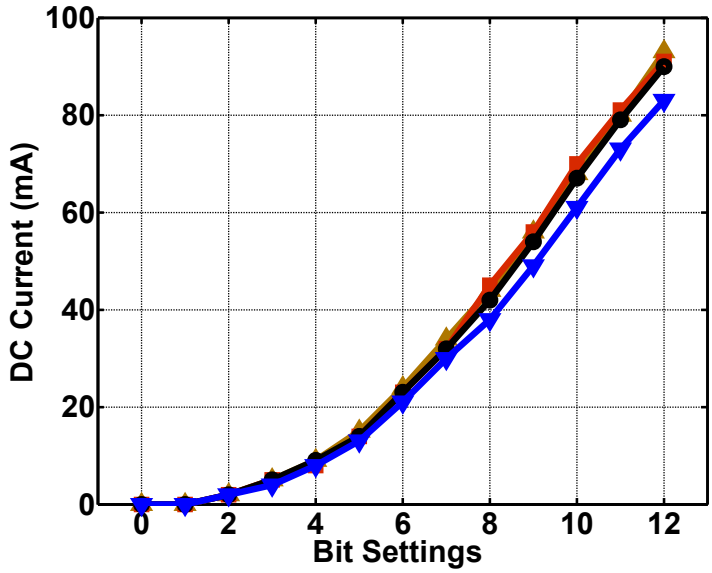


Figure 3.53: DC current versus DAC setting for all four power stages.

Similar measurements were performed on both the 50μm and the 200μm buffers and the results are shown in Figure 3.54.

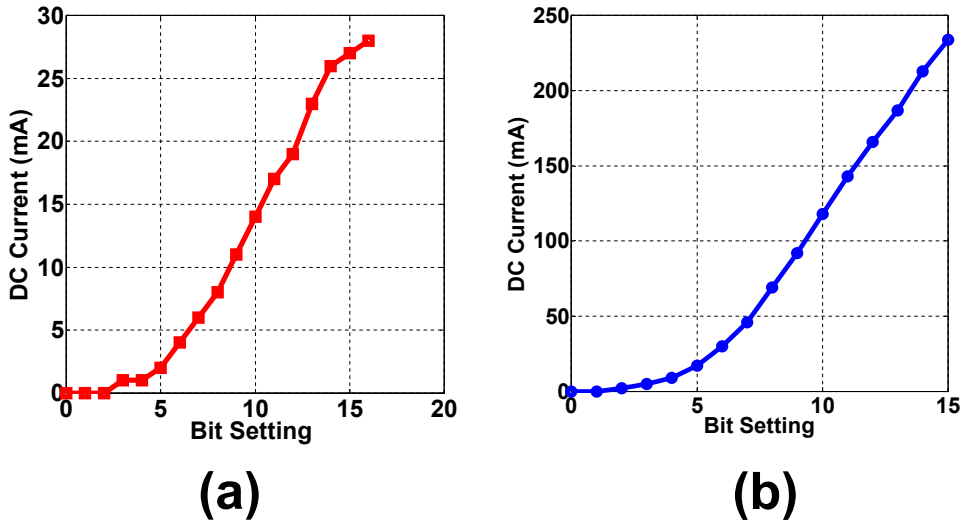


Figure 3.54: DC current versus DAC setting for (a) 50μm and (b) 200μm buffers.

3.3.6.2 CW Measurements

The input and output were then probed and the power measured in a single-ended fashion. Maximum output power of about 17 dBm was achieved at 55 GHz without any laser trimming of interstage inductors. Optimizations were performed for gate bias voltages of all the four power stages together and the last driver stage. Figure 3.55 shows the contour plot of output power versus the two bias voltages.

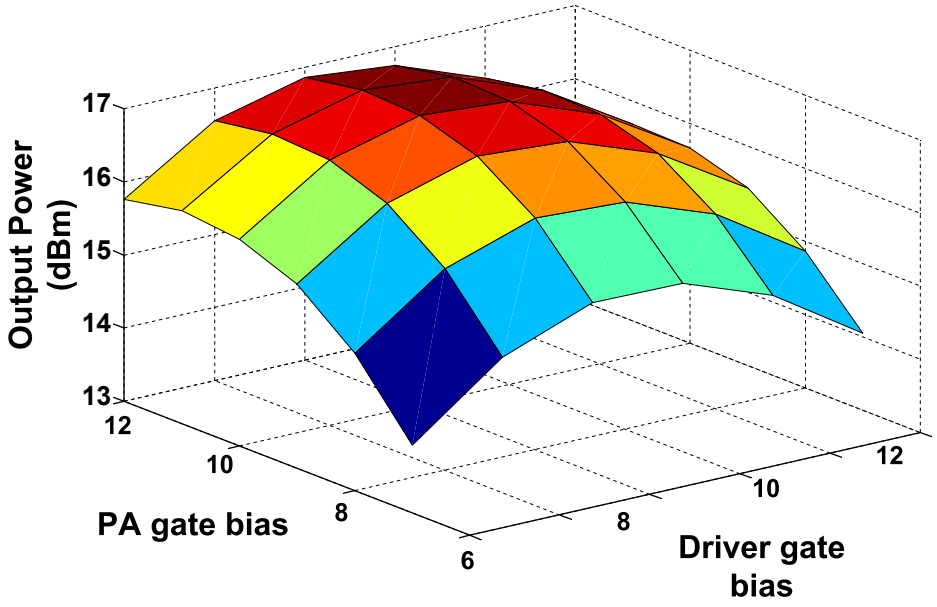


Figure 3.55: Output power versus gate bias voltage of power stages and that of first driver stage.

Frequency sweeps were performed with optimizations at each point to identify the center frequency of the transmitter chip. Plot of saturated output power versus frequency is shown in Figure 3.56, confirming the center frequency as 55 GHz. The discontinuities in power is due to the calibration factor of the power sensor.

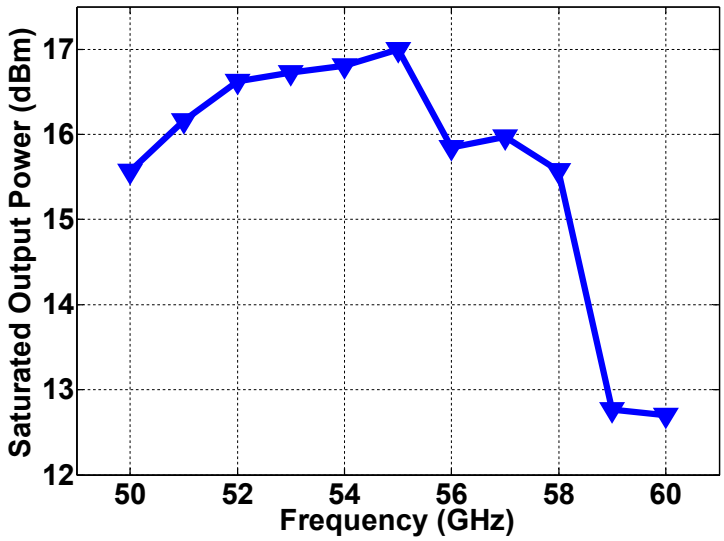


Figure 3.56: Saturated output power versus RF frequency.

Measured output power and gain are also shown in Figure 3.57. Due to drive limitations from the source as well as due to cable losses, it was not possible to saturate the transmitter fully, as seen in the figure. The full transmitter has a small-signal gain of about 12.5 dB with a 1-dB output compression point of 10 dBm.

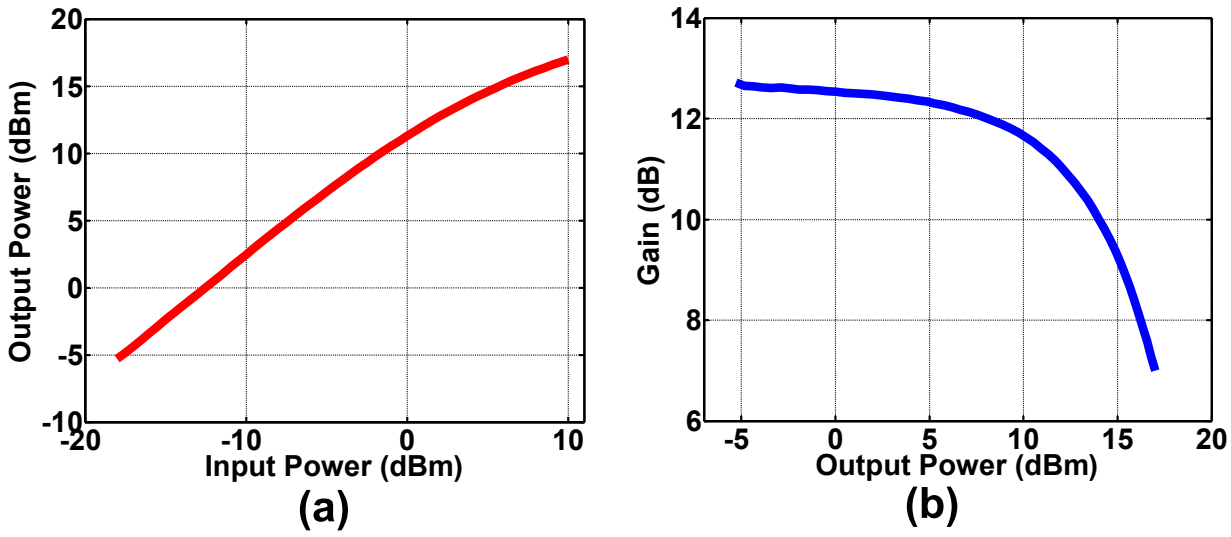


Figure 3.57: (a) Measured output power versus input power and (b) Gain versus output power.

To test functionality of the phase rotators, only the $I+$ and the $Q+$ DACs were kept ON while the other two DACs were set to zero. These DACs are 4-bit R-2R ladder based high speed DACs to enable modulations. Figure 3.58 depicts measurement results showing operation of the phase rotator where the $Q+$ phase rotator DAC was swept for two settings of the $I+$ DAC. The discontinuity from setting 7 to 8 can be attributed to mismatch between the resistors and can easily be calibrated out.

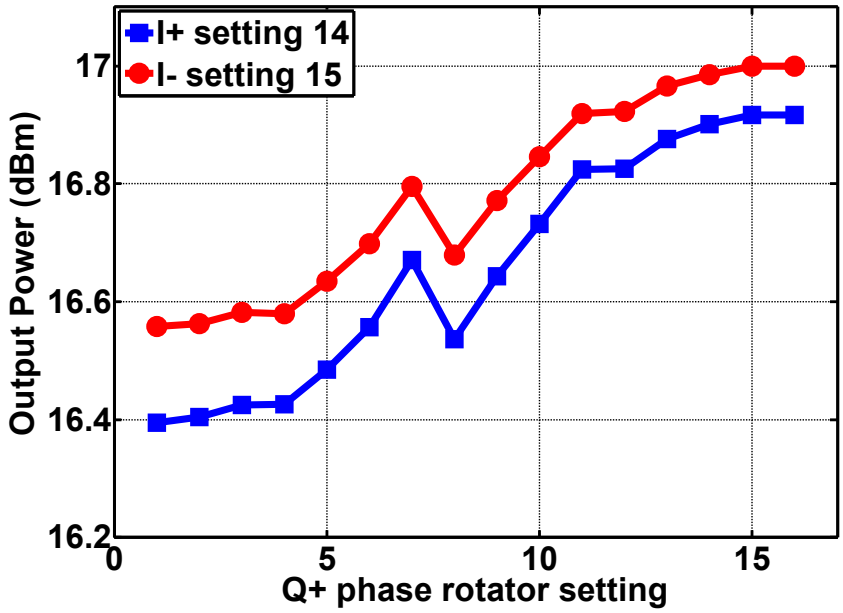


Figure 3.58: Measured output power versus phase rotator setting.

3.3.6.3 Modulation Measurements

Because of the full 360° phase generation capability of the phase rotator, it is possible to directly modulate the transmitter using pure digital phase control through the DACs. Two modulation schemes were tested : (i) OOK, where all phase rotators other than I+ were kept off and a PRBS sequence modulates the I+ phase rotator, and (ii) BPSK, where the Q phase rotators were kept off and I+ & I- were driven in a complementary digital fashion.

In addition to the digital controls of the DAC settings through the serial interface, the I and Q path phase rotators can directly be controlled by high-speed digital bits off-chip.

This enables complete bypassing of the low-speed path through the ZynQ interface. An on-chip MUX selects which path to operate. To minimize off-chip high-speed digital interface overheads, a serial to parallel converter was implemented for each of the data bits (I+, I-, Q+ and Q-). Figure 3.59 shows the schematic of the high-speed serial to parallel converter, where four clock edges are generated using a clock divider and inverters which are then used to sample the serial data line at various instants to generate a 4-bit parallel data. Note that due to the serial nature of data assertion, the clock frequency must be significantly higher than the data rate desired.

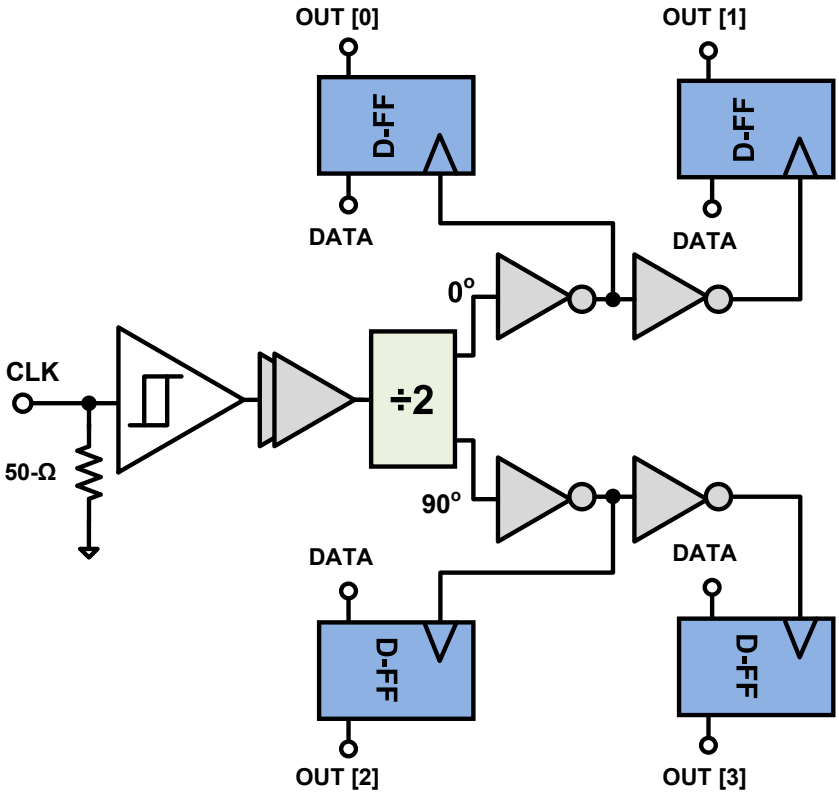


Figure 3.59: On-chip high-speed serial-to-parallel conversion.

Figure 3.60 shows the test setup for modulation based measurements. First, an RF signal at 55 GHz is split using a V-band splitter and then amplified by two amplifiers. One signal is used directly as input/LO to the chip, while the other is used as the down-conversion LO for the external mixer. The high-speed clock as well as the PRBS data sequences are generated

using an AWG.

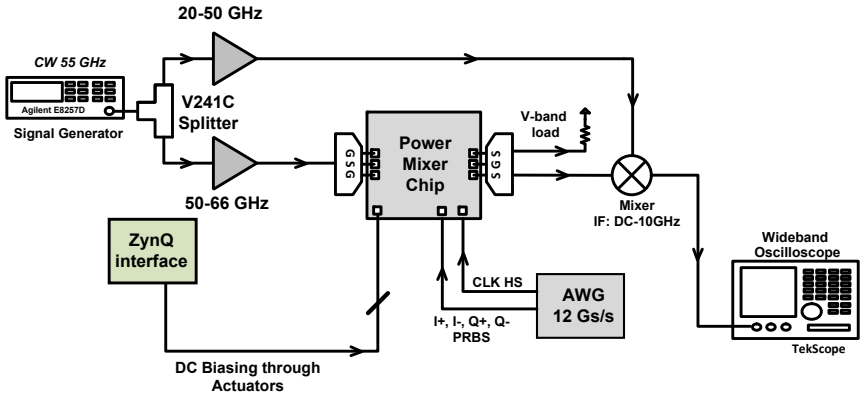


Figure 3.60: Test setup for modulation based measurements.

Measurements were performed with the chip mounted on an FR-4 substrate with relatively long wire-bonds due to limitations of the PCB fabrication. OOK modulations were performed first at low speeds and then at high data rates. No external amplifier was used after the down-conversion so as not to corrupt the modulated signal. To accomodate proper amplifiying operation of one of the external amplifiers, a carrier frequency of 49 GHz was chosen. Figure 3.61 shows demodulated eye diagrams acquired using a real-time scope for OOK modulations at 50 Mb/s where the clock frequency was 500 MHz.

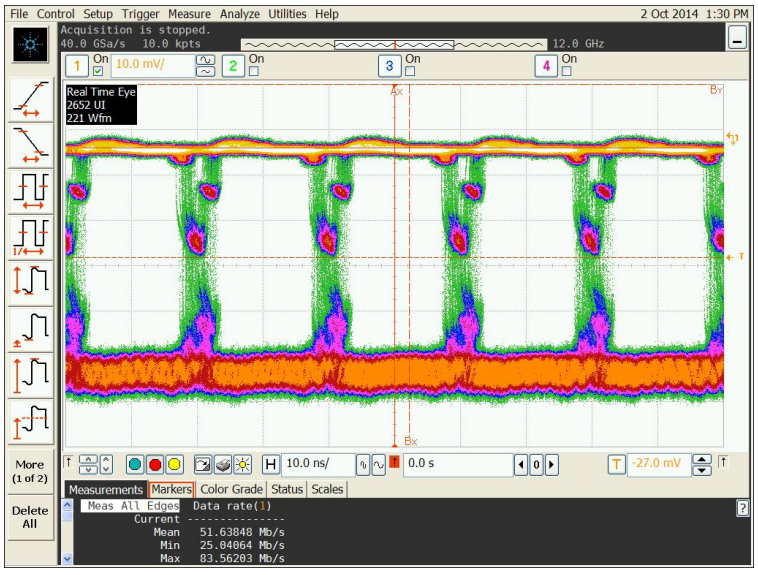


Figure 3.61: Demodulated OOK eye diagram at 50 Mb/s.

At a clock frequency of 5 GHz (limited by the PCB board and wirebonds), Figure 3.62 shows the demodulated eye diagram at 500 Mb/s.

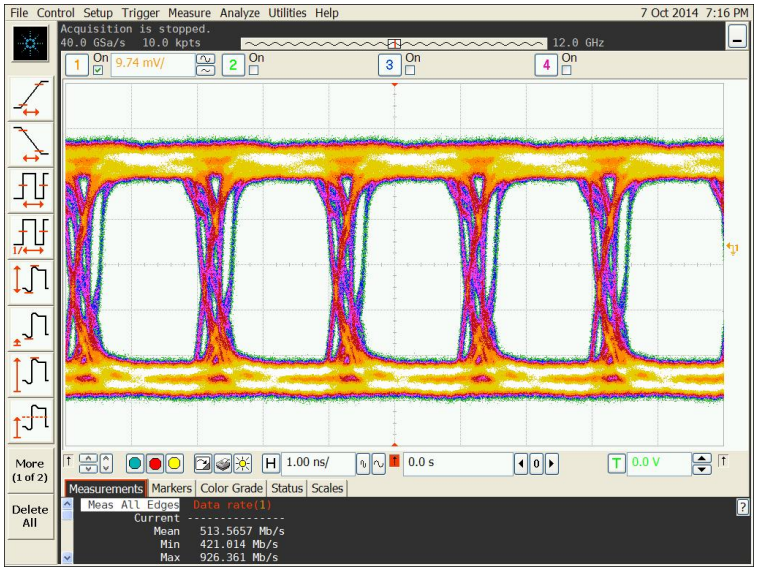


Figure 3.62: Demodulated OOK eye diagram at 500 Mb/s.

Similar measurements were performed for BPSK modulations where I+ and I- were fed with complementary PRBS sequences generated by the AWG. The demodulated eye diagrams are shown in Figure 3.63 for a data rate of 430 Mb/s.

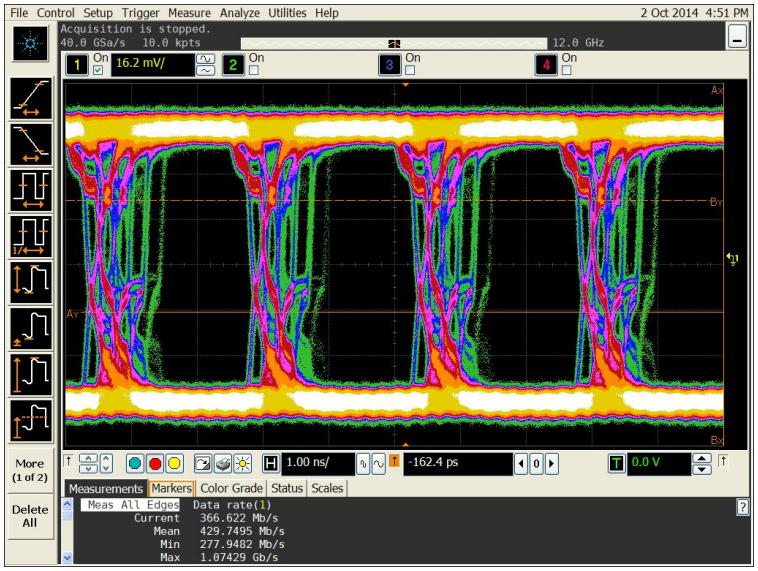


Figure 3.63: Demodulated bpsk eye diagram at 430 Mb/s.

For each of the above measurements, the high speed clock is about 10 times faster than the data rate leading a longer time duration where the parallel data is valid. Measurements were performed by doubling the data rate at the same clock rate, thereby reducing the data valid time, but increasing the modulation speed. Demodulated eye diagram for OOK modulation at 1 Gb/s is shown in Figure 3.64.

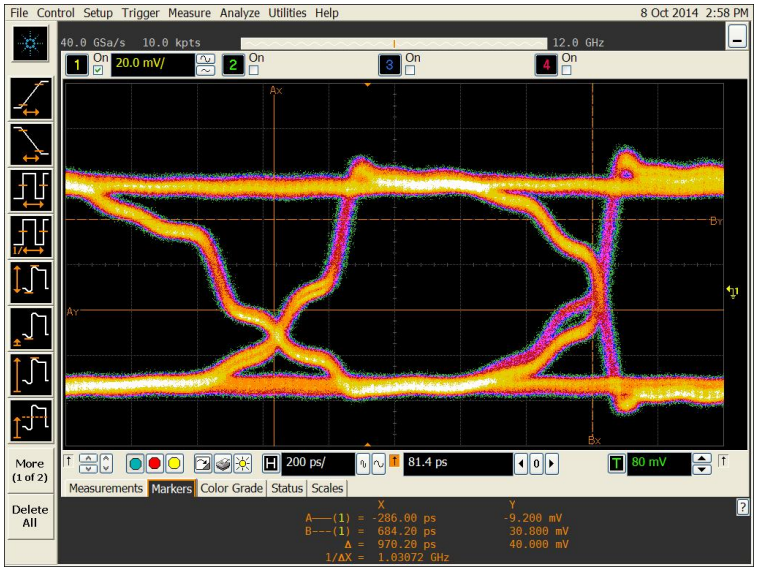


Figure 3.64: Demodulated OOK eye diagram at 1 Gb/s.

3.3.6.4 Closed-loop Healing Measurements

The transmitter chip was found to have an issue reading the on-chip ADC. A test structure was fabricated in the same run to verify the operation of the ADC. The structure was found to be operational at least until a clock frequency of 200 MHz. However, in the main-chip the ADC could not be read possibly due to interfacing/metastability issues in checking for a Data Valid signal. To circumvent the issue, an off-chip standard power sensor was used in place of the on-chip ADC+Sensor and its output digitized in MATLAB to be used during closed-loop healing. Figure 3.65 shows measurement setup of these healing measurements. Both the output power level and the DC currents are read through MATLAB using GPIB control, and then MATLAB increments or decrements the actuation levels depending on the performance requirement. The PC writes a file which is then transferred to the ZynQ board

(SCP protocol) and is then executed in a TCL shell.

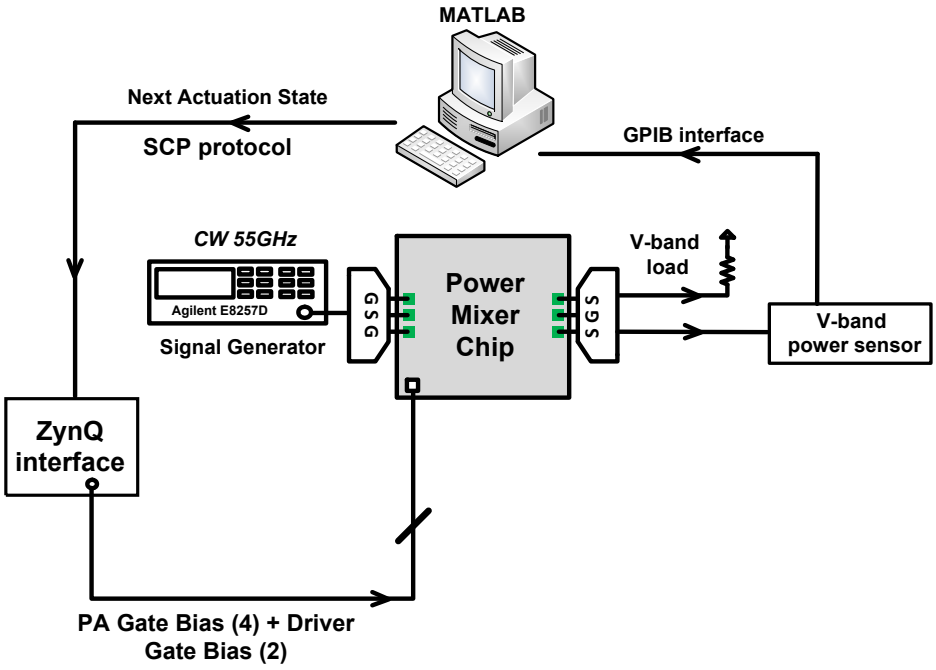


Figure 3.65: Measurement setup for closed-loop healing.

The handshaking protocol between the MATLAB and the ZynQ components can be summarized as follows:

- MATLAB writes the new actuation states depending on the requirement onto a new *.tcl file
- SSH connection to ZynQ and "put" the file
- Write a flag into a file signaling ZynQ that the computation is done and "put" the flag file into ZynQ
- ZynQ executes the tcl file loading the new state and responds back to MATLAB via another DONE flag
- MATLAB reads the flag and then evaluates the RF power and the DC current through GPIB before computing the next actuation state.

A bulk optimization was performed over the gate bias DACs of all the four power elements as well as the gate biases of the two driver stages. The phase rotator settings were set to all $I+$ for this measurement.

First, input power of 7 dBm was applied and the whole space was searched for maximum output power at 55 GHz. Once the optimum settings are found, the input power was swept to generate a P_{out} versus P_{in} plot. A similar plot was also measured using default settings, which are based on simulation settings. Identical measurements were performed at a small signal input power level. Results, shown in Figure 3.66, show an interesting trend. First, the improvement due to healing is about 0.5 dB throughout the power levels almost into saturation. Close to saturation, the advantage is lower, since the PA was originally designed for maximum saturated output power.

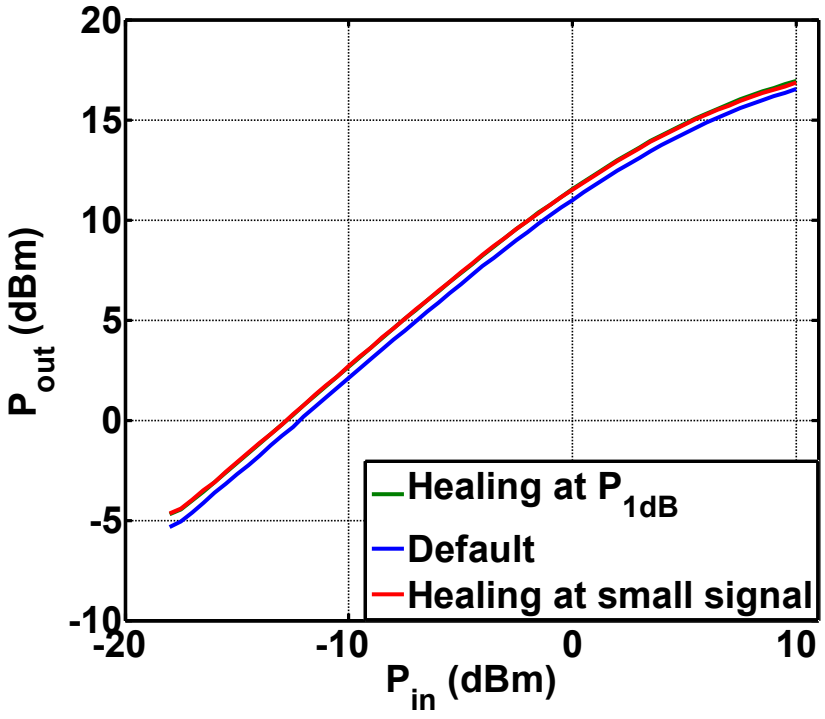


Figure 3.66: Healing for two cases, one at large signal and the other at small signal power levels, compared against the default case.

Although post-healing, power levels are close for the two cases of optimization, the

zoomed in Figure 3.67 clearly shows that the healing at small signal provides slightly higher power than the other case at low power levels. However, the trend is reversed at high power levels, where the optimum state when healed at saturation outperforms the one which was healed at small signal.

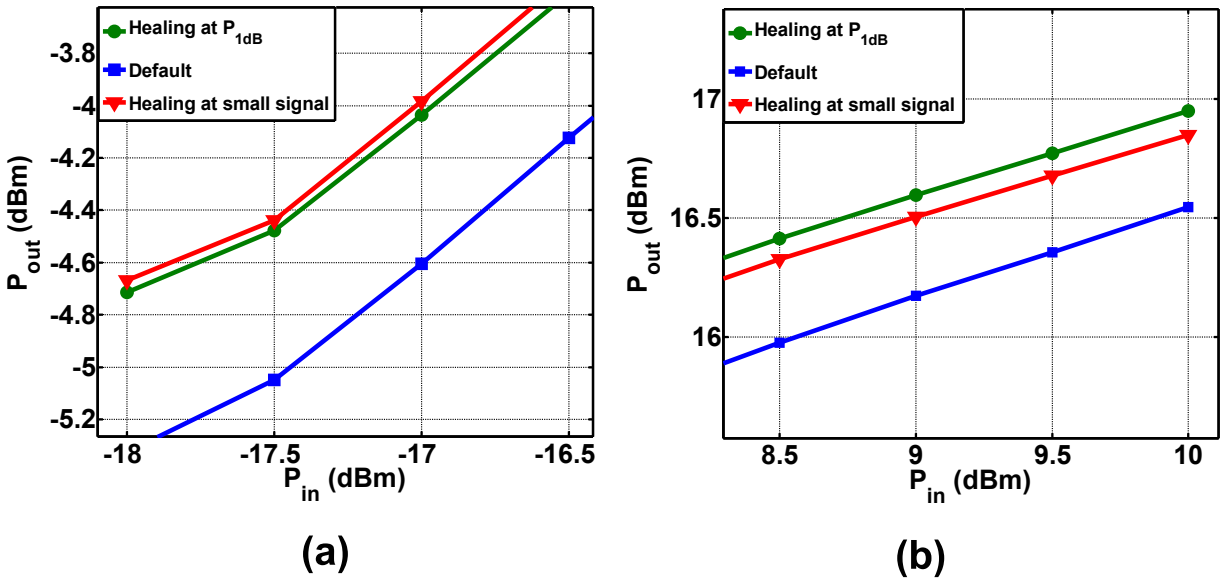


Figure 3.67: Zoom in of Figure 3.66 at small signal and large signal power levels.

Next, healing was performed at all input power levels from 2 to 8 dBm in 0.5 dB increments. At each desired power level, the actuation state with the lowest PA DC current consumption was chosen. Each such chosen point was also required to have a minimum gain of at least 10 dB. Measurement results shown in Figure 3.68 clearly highlight the benefits of closed-loop healing where significant reduction in DC current consumed by the PA is observed before and after healing compared to the default case. A 36% reduction in DC current was observed with and without healing at a 13 dBm output power level (near the 1-dB compression point).

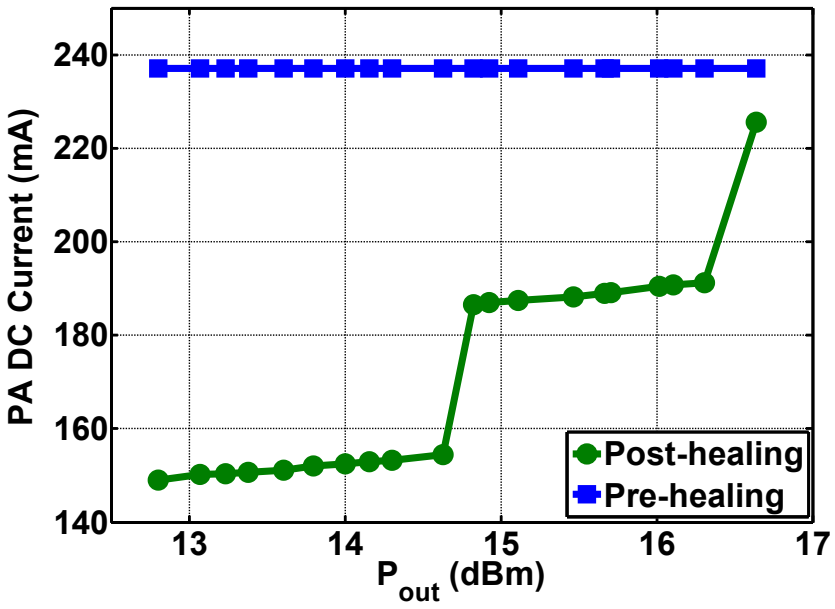


Figure 3.68: PA current versus output power before and after healing.

To test the chip's ability to heal itself when presented with unforeseen circumstances a series of tests were performed. First, the PA was operated from a 1-V supply voltage by setting the reference votlage of the DC sensor regulator. The closed-loop healing algorithm was run to maximize output power at large signal, and then using the optimum settings the input power was swept. Measurement results (Figure 3.69) show an improvement in overall outpu power by about 0.5 dB.

The chip was also healed against output stage mismatch. First, two differential power stages driving one of the two primaries was turned off using the bias actuator DACs. Closed-loop healing was then deployed to try and improve the output power. Next, both primaries of the DAT were driven completely asymmetrically by having each primary driven only by one differential stage. Results before and after healing are presented in Figure 3.70. Measurements show up to a 0.7 dB improvement in both cases when closed-loop healing is deployed.

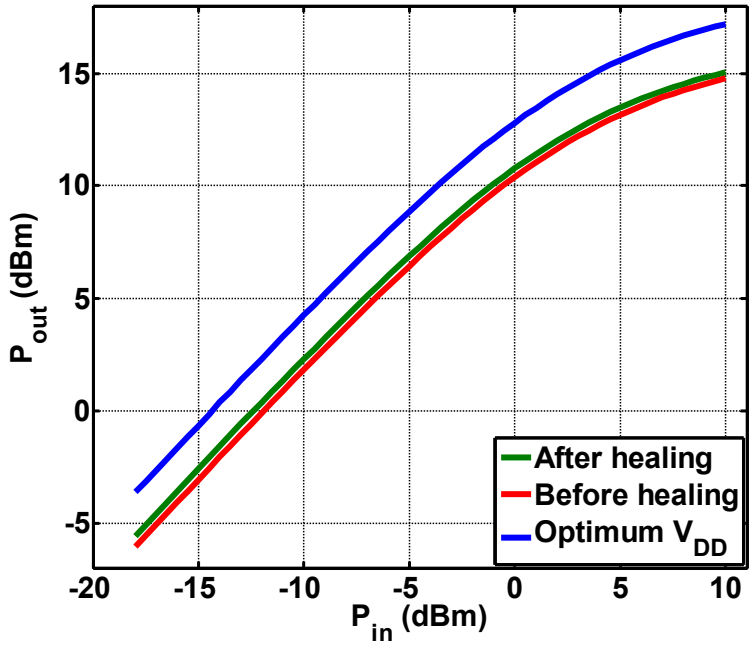


Figure 3.69: Closed-loop healing when PA is operated off a 1-V supply.

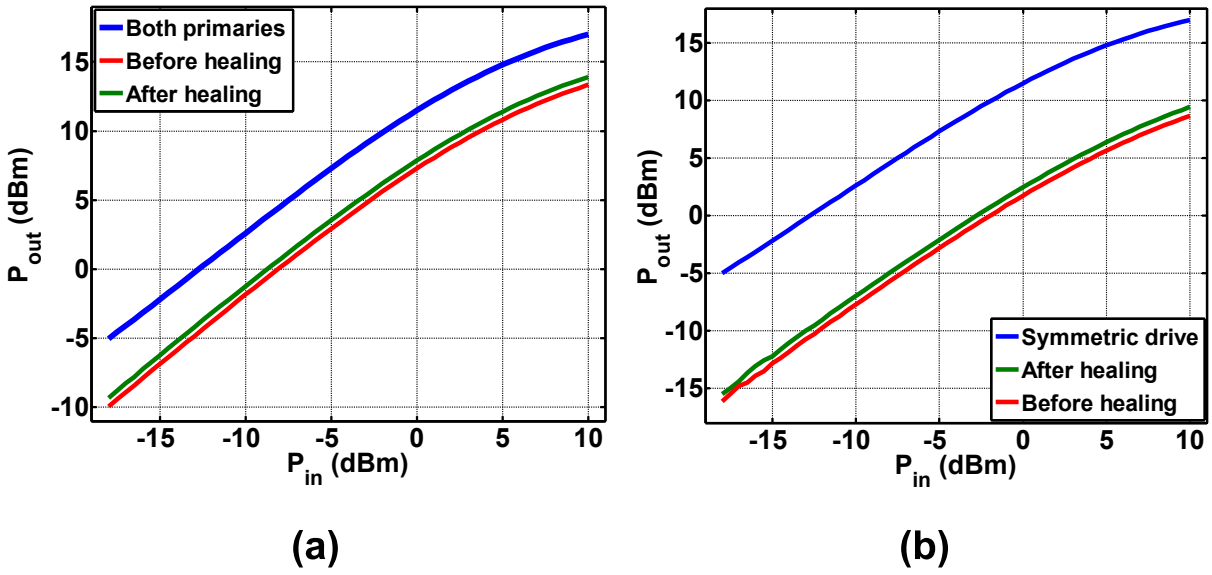


Figure 3.70: Closed-loop healing for (a) one primary OFF and (b) both primaries driven asymmetrically.

3.4 Conclusion

In this chapter, a high-power fully integrated mm-wave power mixer transmitter has been presented, capable of generating constant as well as non-constant envelope modulation at high-speeds. A dual-primary based distributed active transformer power combiner has been shown to be a compact, broad-band, and efficient power combiner for four differential power stages at mm-wave frequencies. The transmitter deploys segmentation as a technique to combine powers from several smaller, digitally controllable output stages, thereby scaling DC current consumption against desired output power, leading to higher efficiencies during non-constant envelope modulations.

High-speed modulations at 4Gb/s for BPSK and QPSK show the broadband nature of the transmitter. High power non-constant envelope modulations have also been demonstrated at 1Gb/s for binary ASK and 500Mb/s for 3-ASK. By combining modulations in the mm-wave LO path and the baseband amplitude path, an example generation of a 16-QAM signal has also been shown. A calibration scheme has been demonstrated to remove systematic non-idealities due to the test setup and significant improvements in EVM have been obtained for high-speed phase modulations. Measurement results over long time periods at higher than nominal supply voltages have been performed to verify reliability of the segmented transmitter for worst case segmentation.

In the second part of the chapter we have shown that variations both in operating points as well as process and mismatch can cause significant degradation in both output power and efficiency. Several sensors and actuators have been added to the existing transmitter along with an integrated on-chip digital control block. Closed loop healing has been demonstrated to improve output power over a wide variety of artificially induced non-idealities. Significant DC power savings have been measured over a wide range of output power levels for the entire transmitter with and without self-healing. In addition, full digital modulation has been demonstrated both for OOK and BPSK schemes with data rates up to 1 Gb/s, limited by test setup.

Chapter 4

Self-healing mm-Wave Power Amplifier in 45nm CMOS

4.1 Introducton

Continuing from the previous chapter, here we will utilize self healing to reduce the adverse affects of process and environmental variation for an example mm-wave power amplifier. After a brief description of the overall architecture, system level measurements of the example PA will be presented showing significant performance and yield improvement against a variety of operating conditions and process variations.

4.2 Design Considerations and Architecture

The fully integrated self-healing PA is designed to operate at 28 GHz implemented in a standard 45 nm SOI CMOS process (Fig. 4.1) [66]. It is a 2 stage, 2-to-1 power combining class AB PA matched to 50-Ω at the input and output. The interstage matching network and output power combining matching network are designed to provide the optimum impedance for maximum saturated output power. The first stage is half the transistor size of the output stage, to ensure that the output stage can be fully driven into saturation. Class AB design was chosen in order to enable linear operation and to allow for non-constant envelope modulation schemes to be implemented.

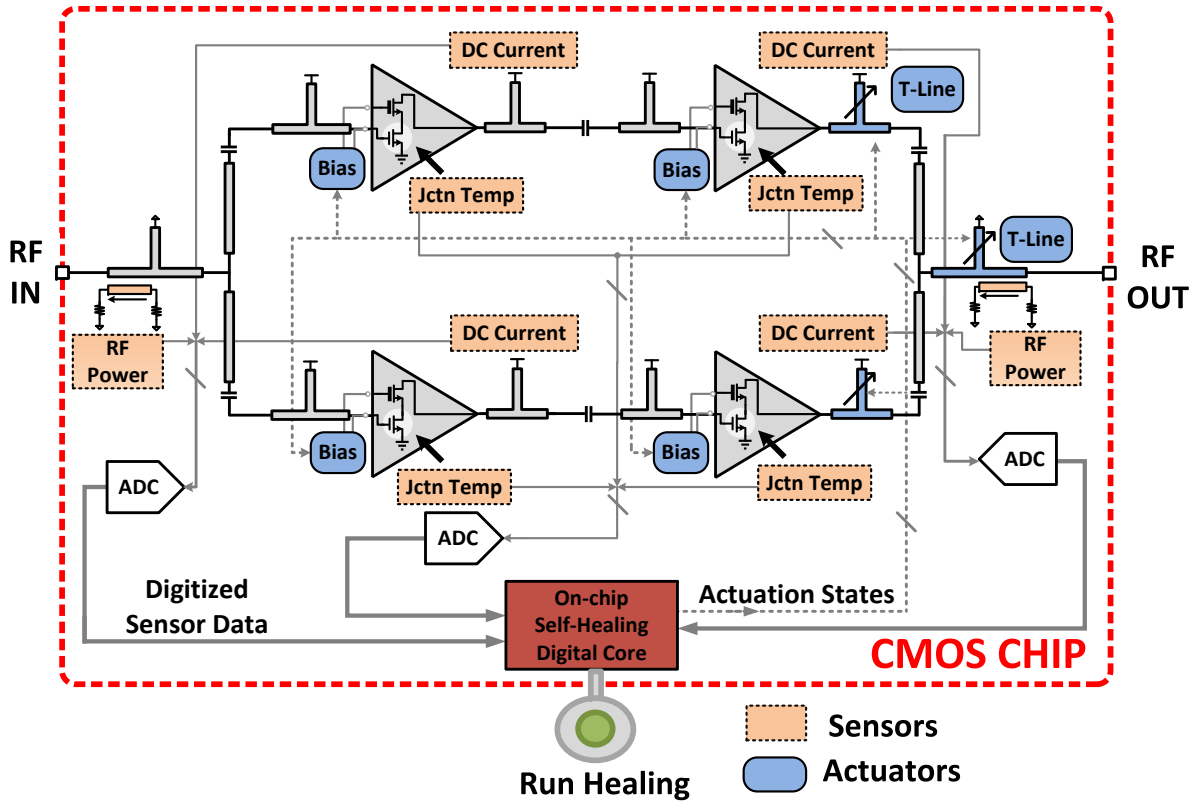


Figure 4.1: Block level architecture of the example integrated self-healing PA. Data from three types of sensors is fed through ADCs to an integrated digital core. During self-healing, the digital core closes the self-healing loop by setting two different types of actuators to improve the performance of the power amplifier.

The power amplifier along with the associated matching network was designed by Steven Bowers. To increase the gain of each amplifying stage, each stage is a cascode amplifier, and two stages are used to further increase the gain. The common source transistors are 56 nm analog transistors with body contacts, while the cascode transistor is a 112 nm thick gate oxide transistor to increase the voltage breakdown limit of the amplifier. The three matching networks use a 2-stub matching technique, and biasing is done through the AC short circuits at the end of the stubs. A metal AC coupling capacitor is used to allow for independent biasing of the inputs and outputs of the amplifying stages. Full 3D electromagnetic simulations of the matching networks, including the capacitors and pads, were performed to ensure proper functionality.

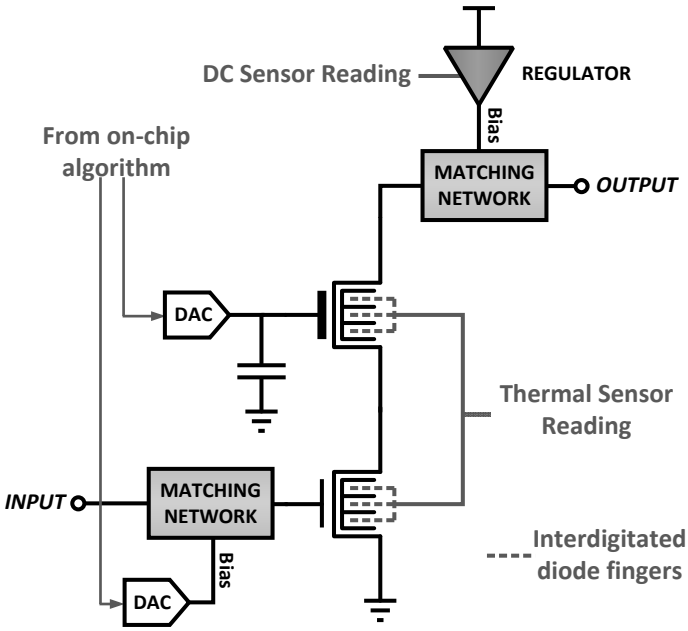


Figure 4.2: Schematic of a single cascode amplifying stage showing connections to matching networks, gate bias actuators, DC sensor, and temperature sensor.

As discussed in Chapter 2, to enable self-healing, several sensors and actuators have been incorporated into the system. Input and output RF power sensors, DC current sensors as well as thermal sensors have been designed and integrated into the PA. On the actuator side, gate bias actuators are implemented on all amplifying stages, and the stubs of the output power combining matching network are tunable to enable tuning of the output network. The digital core is designed to fit within the area between two PA paths, which would have been vacant otherwise. The thermal diodes are closely packed with the PA core so as to reduce any stray parasitic effects, and the RF sensors are kept short. The total sensor overhead for the power amplifier is less than 6%.

The example self-healing PA utilizes tunable transmission line stubs on all three of the output power combining matching network stubs. They were not used on the input matching network or the interstage matching networks because the expected variation in impedances of the loads for those networks (gates of the first and second amplifying stages) was small enough that the advantage of the matching networks did not warrant the additional complexity and

loss the actuators come with.

On-chip 8-bit SAR ADCs (Chapter 2) were implemented to digitize output voltage of the sensors. The digital algorithm was implemented as a custom digital core which was coded in VHDL and synthesized. The self-healing digital was built with a set of instruction sets corresponding to several different modes of operation such as fully automated self-healing, reading sensor data without actuation or step-by-step healing with off-chip control etc. Once the desired instruction set was chosen, and the global state machine controlled all the necessary communication between the various component blocks for actuation loading, sensor reading, and optimization. Due to this modular code setup, many different types of complex optimization algorithms can be incorporated into this general fully integrated self-healing framework. Two modes of fully automated healing algorithms were implemented within the digital core. This is illustrated in Figure 4.3. All the possible modes of operation start with an automated offset calibration step, which measures the DC offset setting of the sensors when the PA is turned off and subtracts it from all future measurements. The algorithm was designed and coded by Kaushik Sengupta.

The first self-healing mode optimized the actuation settings (both bias and t-line combiner settings) for the highest output power. This was an exhaustive search among all the possible 262,144 states. The algorithm starts with the lowest bias settings (lowest DC current) and then continues to increase the bias actuation 1 bit at a time, iterating through all possible combiner settings for each DC current setting. The settings of the driver and output stage are varied independently. The second mode of automated healing tries to find the most efficient state of the PA which can deliver at least a given amount of output RF power. As shown in Fig. 4.3, this mode also starts with the lowest bias setting, reads the sensor data through the shared ADC in a time-multiplied manner, and checks for the desired output power condition for all combinations of the t-line combiner setting. If the output power requirement is not reached, the bias current settings are incremented until the performance goal is reached. As before, the bias setting for the driver stage and the output stage are varied independently. In the current implementation, the digital core uses a test setup limited clock of 25 MHz (though the on-chip core is verified to operate without timing errors until 500

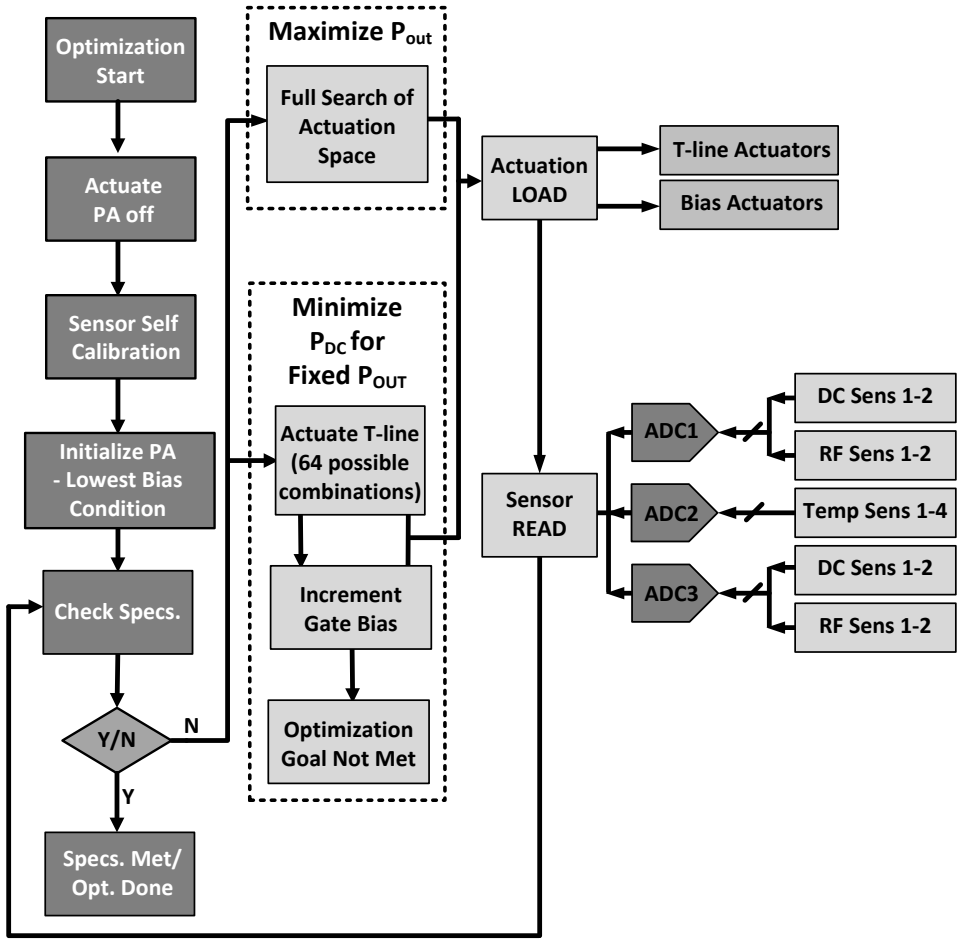


Figure 4.3: Flowchart showing details of self-healing digital core and the possible modes of fully automated self-healing.

MHz) and requires 3 μS per optimization iteration (set actuators, read all sensors, decide on next actuation state). This results in a maximum healing time of 0.8 s when the algorithm is an exhaustive search visiting all possible actuation states.

4.3 Measurement Results

4.3.1 Sensor Measurement Summary

The sensor performances are summarized in Table 4.1.

Table 4.1: On-chip sensors implemented for the self-healing PA.

Sensors	Measured Entities	Responsivity	Range	Sensor 1-bit resol.
True RF Power	In. Power	54 mV/mW	0-10 mW	55 μW
	Op. Power	8.3 mV/mW	01-100 mW	300 μW
DC Sensor	DC drawn by			
	Ip. Stage	8.5 mV/mA	0-60 mA	280 μA
	Op. Stage	4.2 mV/mA	0-120 mA	560 μA
Thermal Sensor	Power dissipated			
	Ip. Stage	4.0 mV/mA	0-130 mW	0.75 mW
	Op. Stage	2.0 mV/mA	0-260 mW	1.5 mW

4.3.2 System Level Measurements

The example PA was fabricated using the self-healing blocks presented in the previous sections, and the system level measurements will be presented in this section. The measurement setup of the PA is shown in Figure 4.4. The PA was mounted on a PCB and probed at 28 GHz. It was driven by an Agilent 83650B signal generator, and probed with Cascade Z-probes. The output went through a calibrated network that included a mm-wave load tuner to an Agilent 8487D power sensor.

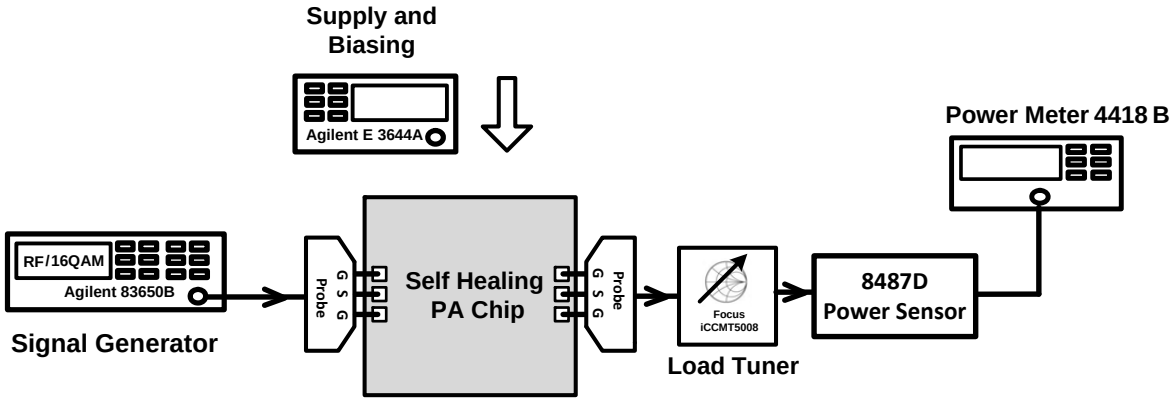


Figure 4.4: Measurement setup for the fully-integrated self-healing power amplifier.

A comparison to an unhealed PA is useful to look at the benefits of the self-healing system, so a default actuation state must be selected as the 'default' state. The default

state is chosen to be the state that had the best saturated performance in simulation and represents the PA that would have been taped out if the self-healing system was not being used. For all of the measurements without self-healing, the DC power consumption of the self healing blocks were omitted when looking at the performance of the default state.

The entire self-healing system was integrated on a single chip, which means that the only external input given to the chip was the mode of operation, which algorithm to run, the desired output power (if needed) and then a go command, and no external calibration or external performance information is used during healing.

4.3.3 Healing process variation with a nominal 50-Ω load

The healing ability of the amplifier for 50 Ω loads is presented first. Figure 4.5a shows the output power verses input power for an amplifier in its default state, as well as one that has been healed for maximum output power at low input power levels, and one that has been healed for maximum output power at the 1 dB compression point, the point where the gain has compressed 1 dB compared to the small signal gain. This plot shows the improvement in output power that can be achieved using self healing, but also shows that there is no one optimal state for all input powers. The optimum load impedance for the maximum output power from the output amplifying transistors varies based upon the input power levels, and thus by tuning the matching network for the current power level, the corresponding optimal matching network can be found. This is an added benefit of the self healing system, as it can be healed for the desired power level, and if that desired power level changes, it can be healed again. Near saturation where the default state was designed, it is close to the optimum, and thus there is not as much room for improvement. However, the default match at small signal is farther from the optimum, and thus there self-healing can provide larger improvements to the performance.

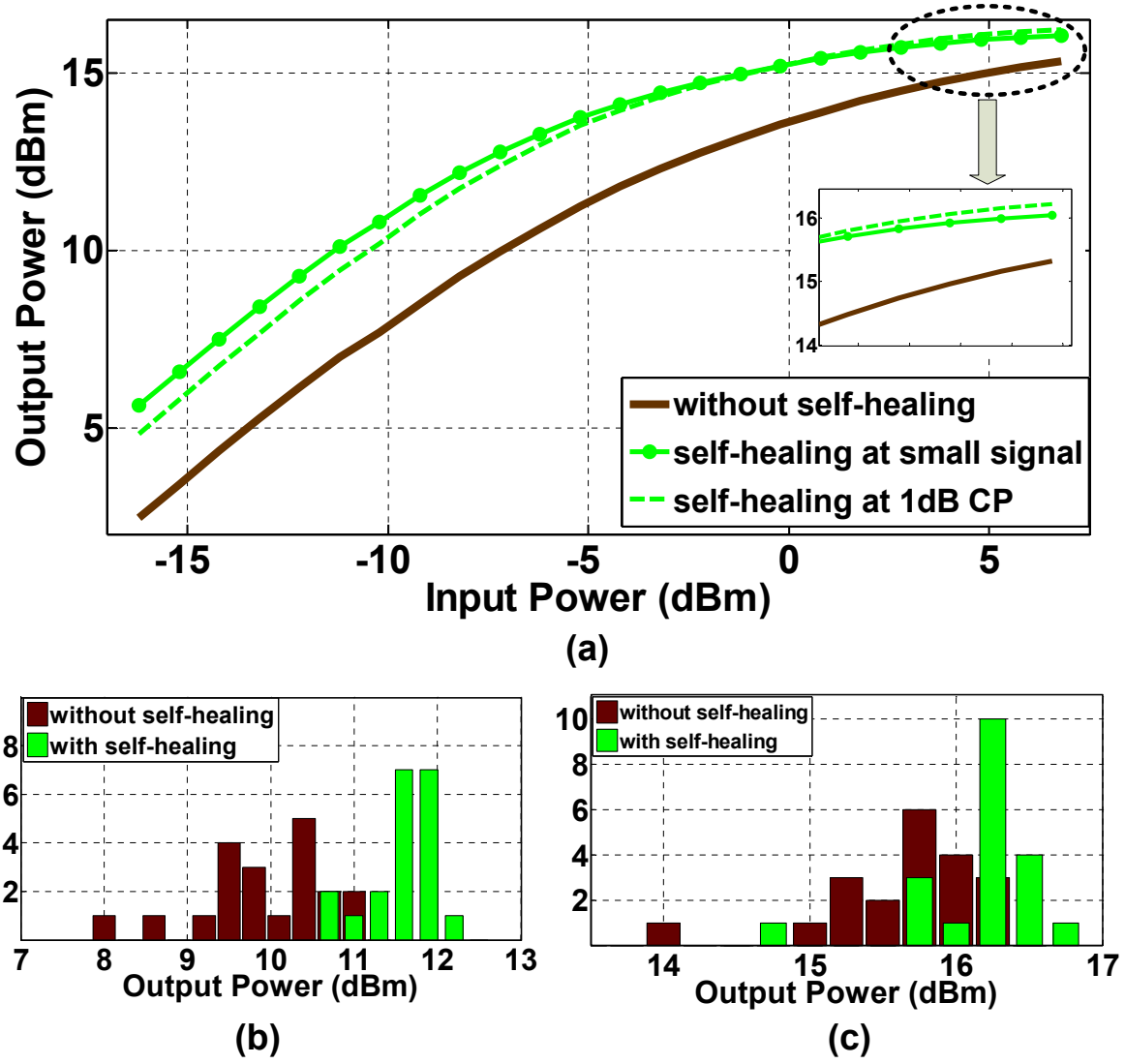


Figure 4.5: Measured output power before self healing, and after self-healing for maximum output power, both for healing done at small signal and at the 1 dB compression point (a), and histograms of 20 measured chips before and after self healing at small signal (b), and at the 1 dB compression point (c).

To show improvement for process variation, 20 chips were measured and histograms showing with and without self healing for maximum output power are shown in Figures 4.5b and 4.5c, respectively. The small signal gain after healing is 21.5 dB, with a saturated output power of 16 dBm and a 1 dB compression point of 12.5 dBm while consuming 520 mW of

DC power at 28 GHz.

The second algorithm to minimize the DC power while maintaining a desired output power is shown next. Figure 4.6 (a) shows the DC power consumption of 20 chips for various output power levels, with a histogram cross section of that plot for 12.5 dBm output power near the 1 dB compression point as shown in Figure 4.6 (b).

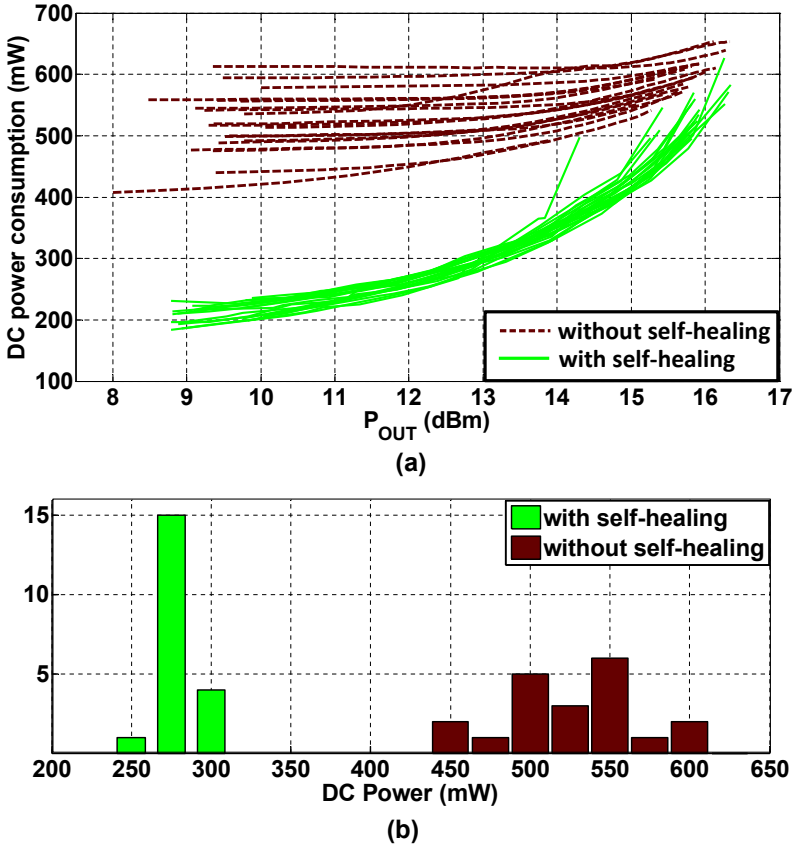


Figure 4.6: Measured DC power consumption for 20 chips before and after self-healing for minimum DC power while maintaining a desired RF power level is used (a), and a histogram cross section of 20 chips (b) of the DC power consumption before and after self-healing to maintain an output power of 12.5 dBm, near the 1 dB compression point.

Because the state is not changing, the DC power levels without self healing for each chip are relatively flat until the transistors really start to saturate and the power increases slightly. Once self healing is turned on, there is still high DC power required to achieve very high output powers, but once the desired output power becomes even a couple dB below the

saturated power, significant reduction in DC power consumption is observed.

The DC power required to produce 12.5 dBm output power sees a 47% reduction in average power level over 20 chips, with a 78% decrease in the standard deviation between chips. This means that self healing is both improving the performance, but also making it much more consistent across chips than in the default case. Once the power is near small signal levels, reductions of greater than 50% for every single chip measured are achieved.

4.3.4 Healing VSWR Environmental Variation with Load Mismatch

The load impedance was varied using a focus microwaves mm-wave load tuner that produces loads within the 12-1 VSWR circle at the tuner, which when calibrating for the loss of the cable and probes becomes a load variation within the 4-1 VSWR circle at the probe tips, which on the real impedance axis corresponds to a resistance from 25 Ω to 200 Ω. The ability of the RF power sensors to detect actual power going to the load, not just the voltage, enables the self-healing system to know the delivered power to the load even under load impedance mismatch. This means that the algorithm can still heal and doesn't require knowledge of the load impedance to run the optimization, since the metric of interest, the output RF power is already known.

The results of healing for maximum output power when the load impedance is swept within the 4-1 VSWR circle are shown as contour plots on the Smith chart in Figure 4.7, and show an improvement in output power across the entire 4-1 VSWR circle. Ten chips were measured, and the results of self-healing for maximum output power for two representative load impedances are shown in Figure 4.8. The first is near the maximum output power, and the second on the edge of the 4-1 VSWR circle. Again both show improvement overall in the output power, as well as a reduction in the variation between chips.

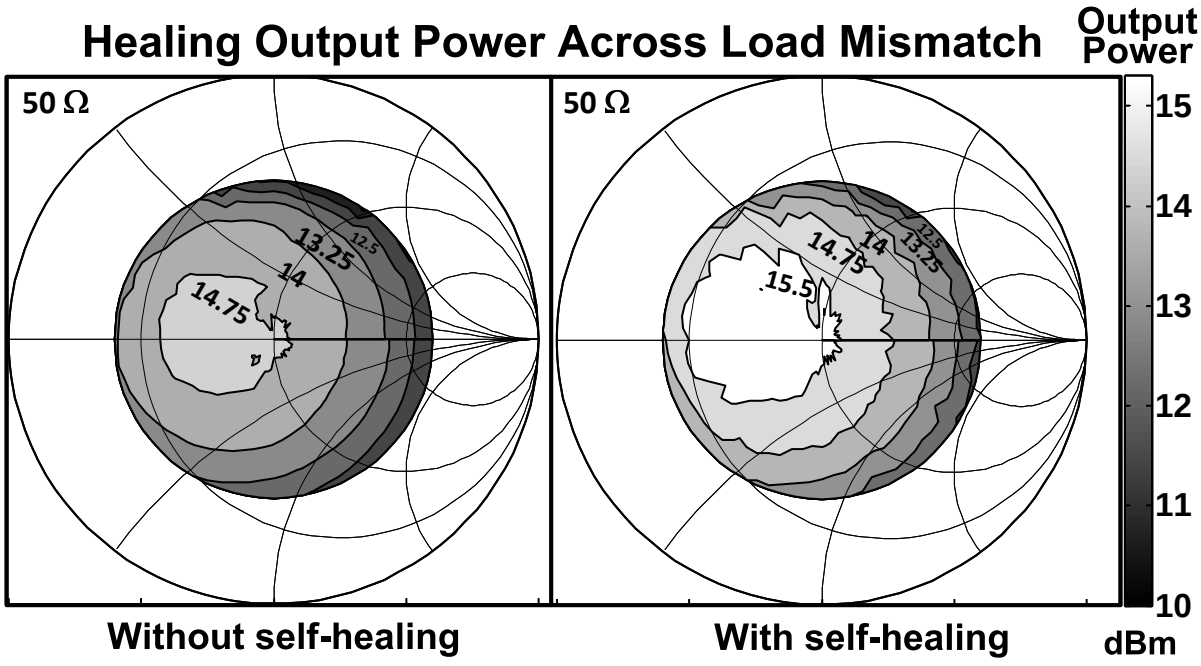


Figure 4.7: Contour plots before and after self-healing for maximum output power for load impedance mismatch show improvement in output power over the entire 4-1 VSWR impedance circle.

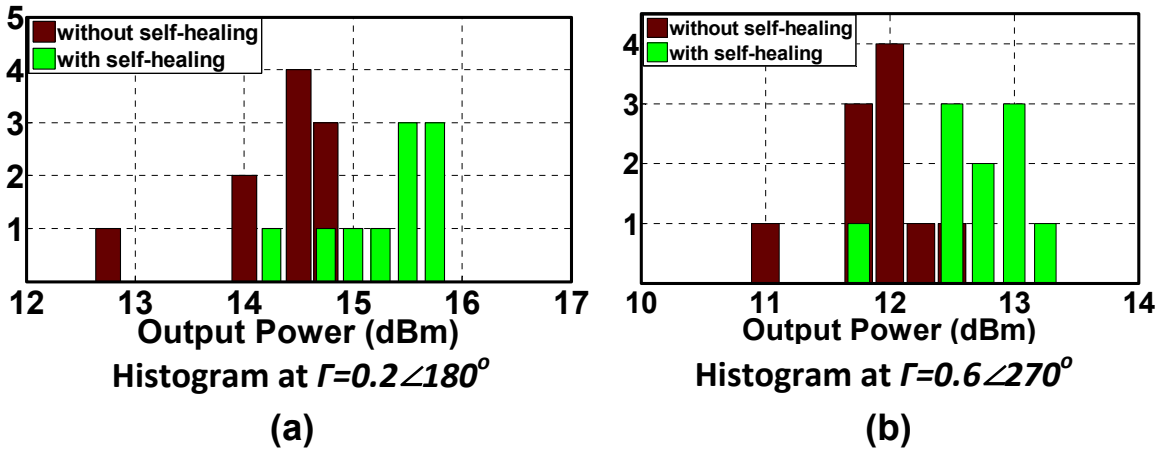


Figure 4.8: Histograms of 10 measured chips showing output power before and after self-healing two representative load impedance points, one near the maximum output power (a), and the other on the edge of the 4-1 VSWR impedance circle (b).

The second algorithm to minimize DC power for a desired output power was also tested under load impedance mismatch, and the results for a desired output power of 12.5 dBm are shown as contour plots in Figure 4.9. The outer-most contour represents the loads where 12.5 dBm was achieved, with the shading and all subsequent contours representing the DC power consumed. For the default state, the DC power consumption remains constant regardless of the load impedance mismatch. With self-healing however, the power can be substantially reduced by up to 35% at impedances near 50 Ω, where the PA was designed, while still maintaining the desired output powers at the more extreme impedance mismatches.

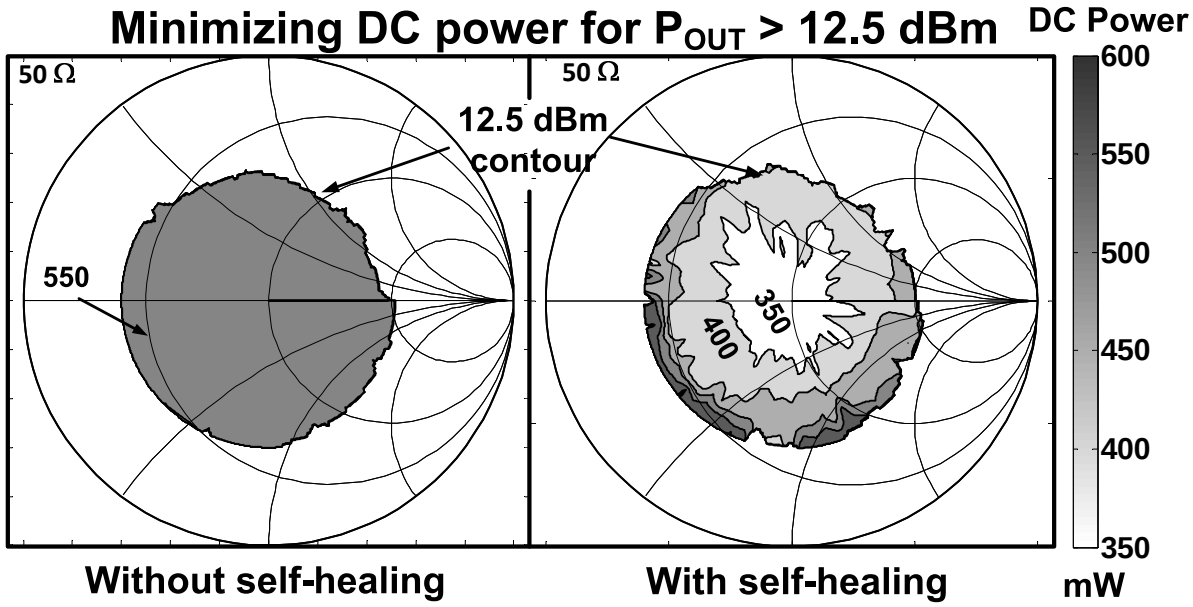


Figure 4.9: Contour plots before and after self-healing for minimum DC power consumption while maintaining 12.5 dBm desired output RF power for load impedance mismatch show improvement in output power over the entire 4-1 VSWR impedance circle.

4.3.5 Healing for Linearity

The PA is designed as a linear amplifier to enable the use of non-constant envelope modulation schemes. The linearity of the PA has been verified using a 100 ksps 16 quadrature amplitude modulation (QAM) signal to measure the error vector magnitude (EVM) for 10 chips, shown for 12.5 dBm output power as a histogram in Figure 4.10. While the self-healing

system does not specifically attempt to improve linearity, a reduction in average EVM from 5.9% to 4.2% is observed when self-healing for maximum output power was applied. The healed chips are able to provide higher output powers while not being pushed as far into saturation, and thus the linearity for a given output power is improved.

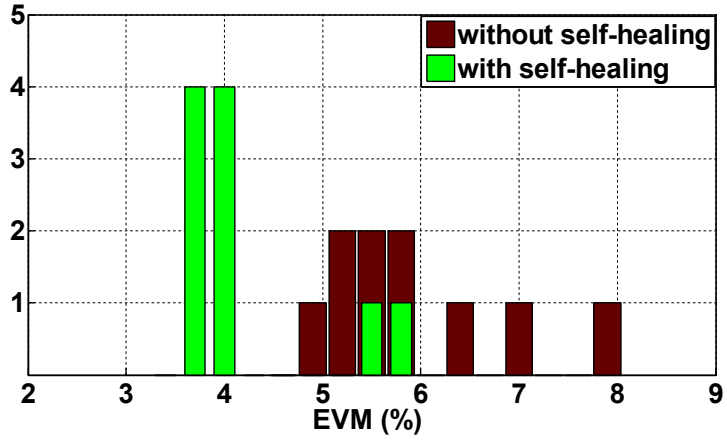


Figure 4.10: Error vector magnitude of 10 chips before and after self-healing for maximum output power show an improvement in linearity after self-healing.

4.3.6 Healing for Partial and Total Transistor Failure

Partial or total transistor failure can be caused by aging, transistor stress such as from voltage spikes, or other phenomenon. To show the self-healing system's ability to heal for partial and total transistor failure, a laser trimmer was used to blast away various parts of one of the output stage transistors. The output stage was chosen as the transistors to cut, as they are the ones that are pushed closest to breakdown and are likely going to be the first ones to fail. Only one of the two output stages was cut to cause a worst case scenario from a mismatch standpoint. As a reference, the amplifier before any laser blasting is shown in Figure 4.11a. Measurements taken in the default state and after healing for maximum output power with half of the common source transistor cut out are plotted in Figure 4.11b. Figure 4.11c shows the results with and without self healing when half of the cascode stage was additionally cut out. Finally, Figure 4.11d shows the results once the entire output

stage is blasted away. This means that the matching network that was expecting to have two similar drives at both inputs now has only a single input, and then a large stub where the other input used to be, destroying the original match.

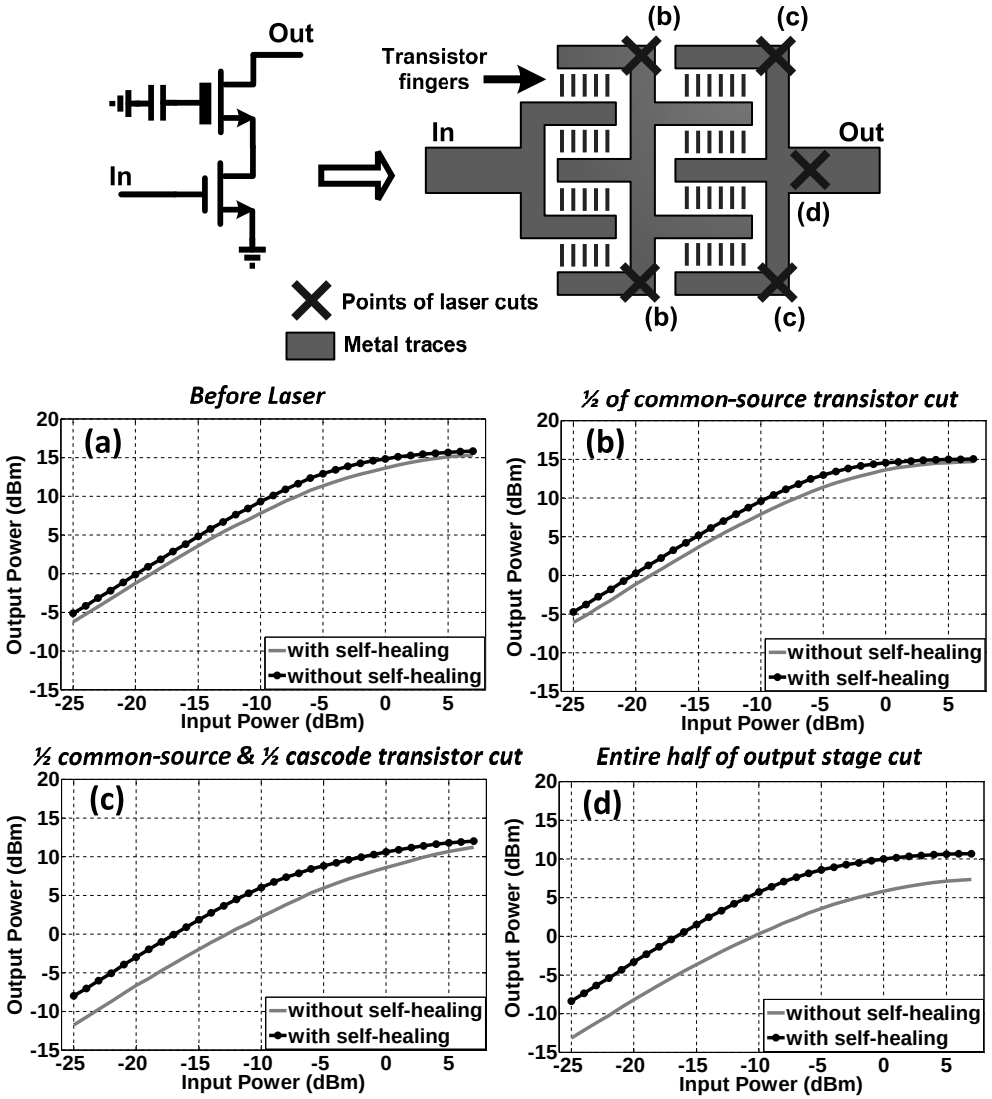


Figure 4.11: Schematic and layout location of laser trim points, and measurements before and after self-healing for maximum output power at various stages of transistor failure due to laser blasting show more than 5 dB improvement when self-healing is used in the worst case scenario of an entire output stage failing.

The default case at small signal loses 7.2 dB in output power from when the output stage is whole to when it is completely cut out. 3 dB of that is due to having only one of

the two output stages providing power, but another 4.2 dB is caused by mismatch of the matching network. Once healing is applied, the loss due to cutting out the output stage is only 3.3 dB, which when taking into account the 3 dB loss from only having a single output stage means that the tunable matching network was able to heal back to the point where there was only 0.3 dB additional loss from this catastrophic event. This is one of the very strong points of self-healing, as under nominal conditions, where the default design is close to optimum, self-healing can only possibly improve the circuit at most the deviation from the optimum, but in cases such as this where the default falls far from the optimum point, and would normally register as a total failure of the entire circuit, self-healing can provide very significant gains and keep the circuit operational even under these types of extreme conditions. A die photo of the entire chip with closeup images before and after the laser blasting is shown in Figure 4.12.

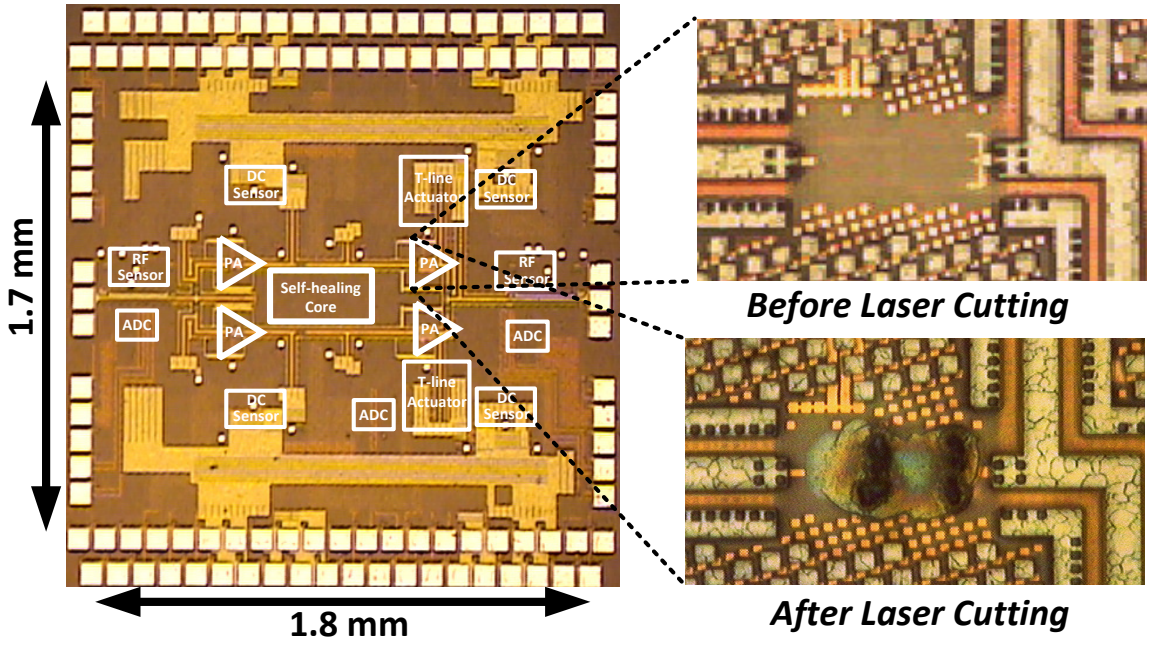


Figure 4.12: Die photo of the self-healing PA with closeup views of one output stage before and after laser blasting.

Table 4.2: Post-healing yield improvement.

Metric	Unit	Specification	Achieved	Baseline Yield	Post-healing Yield
P_{1dB}	dBm	>12.5	13.75	20%	80%
Gain	dB	>20	21.6	30%	100%
PAE	%	>6	7.2	0%	100%
Bandwidth	%	>15	16.4	100%	100%
4-1 VSWR Tolerance	dB	<2.5	2.33	0%	100%

4.3.7 Yield Improvement

An effective self-healing system should improve the yield of the design, and to do that, yield specifications must be defined. For this design, the specifications included a saturation output power >15.5 dBm, gain >20 dB, a power added efficiency >6%, and a 4-1 VSWR tolerance <3 dB, defined in dB as the worst case output power falloff within the 4-1 VSWR circle compared with a nominal 50 Ω load. The PA was able to achieve best case metrics of 16.5 dBm saturated output power, 23.7 dB gain, 7.2% efficiency, and 2.28 dB 4-1 VSWR tolerance. Across 20 chips, the yield of the saturated output power improved from 20% to 90%, the gain improved from 20% to 100%, and the efficiency improved from 5% to 100%. For the 10 chips measured under load impedance mismatch, the yield improved from 0% to 80%, with an overall aggregate yield for all performance specifications improving from 0% to 80%. Results are summarized in Table 4.2.

4.4 Conclusions

A 28 GHz power amplifier was presented as another case study of how an integrated self-healing PA could be implemented. Measurements of multiple chips demonstrate the viability of an integrated self-healing system that requires no external calibration of any kind. Integrating the sensors, actuators, digital algorithm, and data converters on a single chip allows

for a completely automated healing system that improves aggregate yield. Both output power as well as efficiency were improved post-healing compared to default settings obtained from simulation. Consistent improvements were also demonstrated across a 4-1 VSWR load mismatch showing robustness of the system to antenna load variations. To emulate transistor aging, various parts of the output stage were blasted with a high power laser and significant improvements in output power were measured upon depoyment of closed-loop healing.

Chapter 5

Self-healing LO Generation in Receivers

5.1 Background

Phased-array receivers have been known to present significant advantages over their single-element counterparts due to the beam forming capability which leads to improvements in received signal-to-noise ratio as well as directional communication. As the number of elements increases, significant improvements in Equivalent Isotropically Radiated Power (EIRP) as well as interference rejection can be obtained. In recent years, Si-based phased arrays have been demonstrated for high-speed, high-frequency wireless links, particularly those for 60GHz links and automotive radar as well as multi-band multi-beam applications. In these applications, random device mismatches and P.V.T variations directly degrade the accuracy and repeatability of the beam forming. In addition to these, environmental effects like temperature and aging can negatively affect performance of such systems.

A majority of such errors can be attributed to the LO generation particularly the I-Q amplitude and phase mismatch in the phase synthesis/rotator network. In conventional receiver systems, such phase errors are dealt with using off-chip system level calibration schemes. Such one-time calibration schemes are often ineffective in real world situations where external auxiliary test tones often require significant setup and cost overheads. In this chapter we present an example broadband RF receiver system implemented in 65nm CMOS

where self-healing has been applied to heal against IQ phase and amplitude mismatch to improve accuracy of the phase synthesis network.

5.2 Phased Array Systems

CMOS phased array systems have been widely reported in recent literature particulary those covering multiple frequency bands [67]. The principle of operation of a phased array receiver is illustrated in Figure 5.1, where an LO phase shifting approach has been demonstrated. The system has N elements which are spaced at a distance d apart. For an incident wave

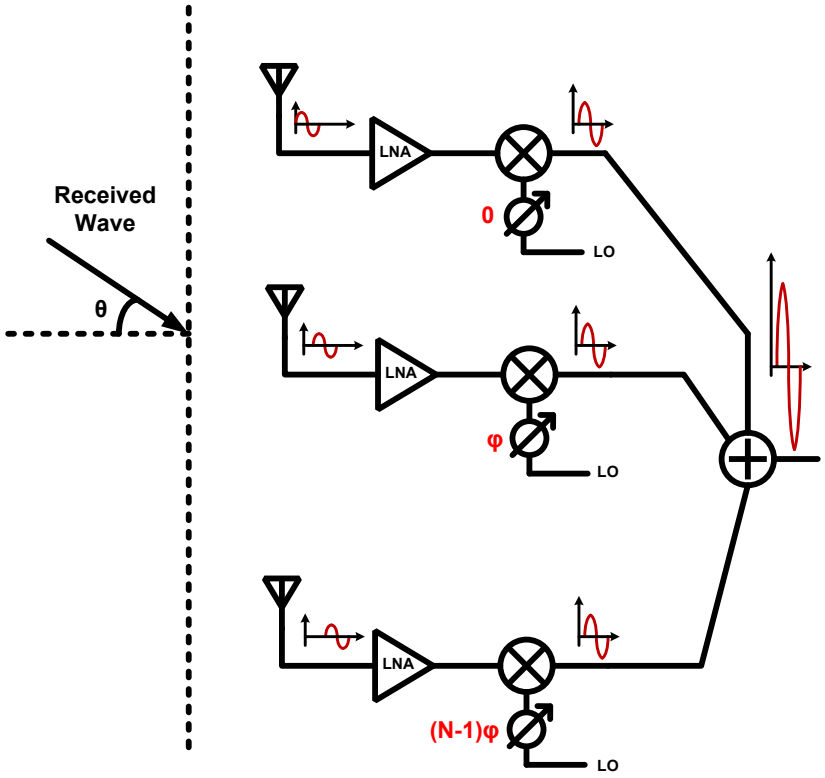


Figure 5.1: Conceptual phased array receiver system with LO phase shifting.

at an angle of θ, it can be easily shown that the signal arrives at each element with a fixed delay difference (ϕ), which can be approximated by Equation 5.1.

$$\phi = \omega_{RF} \frac{d \sin\theta}{c} \tag{5.1}$$

By adjusting each receiving path with this fixed delay difference with respect to the adjacent one, all the signals received by this N element receiver can be added in-phase which will lead to significant improvements in signal-to-noise ratio, since the noise contributions of each of these elements are uncorrelated to the first order. If no coupling is assumed beween the receiving elements, the SNR at the output of an N element array can be expressed by Equation 5.2 where SNR_0 represents the signal-to-noise ratio at the output of one receiving element.

$$SNR_{out} = \frac{(N.V_{out})^2}{N.V_{noise}^2} = N.SNR_0 \tag{5.2}$$

A typical figure of merit of the beam-forming capability of a phased-array system is its peak-to-null ratio (PNR), which can be defined simply as the ratio of the amplitudes of the received signals in the desired direction and the undesired/null direction. This ratio can be significantly degraded by phase and amplitude mismatch and variation between different receive elements. It can be shown, for example, to achieve a PNR of 25dB, the standard deviation of amplitude and phase errors in the receiver should to be less than 0.5dB and 3^o, respectively [68]. These specifications translate to extremely robust and variation intolerant designs in the LO path which are often very challenging in nanometer CMOS. Additionally, these random delay mismatches lead to frequency dependent array performance degradation and thus become exceedingly critical to mitigate in broadband phased array systems. With these facts in mind we demonstrate a self-healing phase synthesis scheme utilizing a combination of self and inter-mixing betwen I and Q signals to extract mismatch information which leads to accurate interpolation of the phases without the need for any external calibration signal. This low-overhead self-healing scheme has been demonstrated to be extremely efficient in reducing phase interpolation errors over a wide range of frequencies.

5.3 LO Generation

The example receiver operates from 6-to-18GHz and can be split into two bands, each containing a separate Rx chain as well as a PLL. The basic architecture consists of a two-step down-conversion system where a 2/3-1/3 LO frequency scheme has been implemented.

5.4 Self-healing Phase Synthesis

5.4.1 Overview

The self-healing scheme was designed by Hua Wang. We will briefly discuss the various steps associated with this technique. Figure 5.2 depicts a simplified LO-phase shifting scheme for a phased-array system. As mentioned in the previous section, the divide-by-2 generates I and Q signals to be used in the second down-conversion mixer. The phase rotators scale and combine the I/Q signals to achieve an interpolated target phase using Cartesian combining. It can be shown that both amplitude and phase imbalances affect the synthesized phase. Moreover, these phase errors may also depend on the target phases. In typical systems, this phase errors can be as large as $10 - 20^o$ before calibration [69].

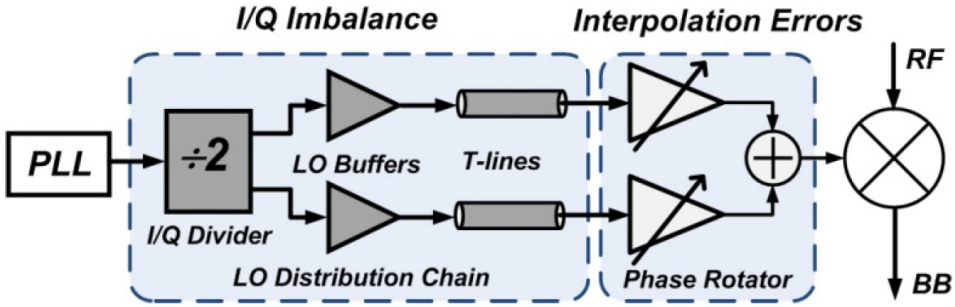


Figure 5.2: Simplified LO-based phase-shifting architecture.

The architecture of the implemented on-chip self-healing phase synthesis scheme is depicted in Figure 5.3. After interpolation of the phase by the variable-gain amplifiers in the Cartesian system, an automatic gain control (AGC) loop equalizes the amplitude variation.

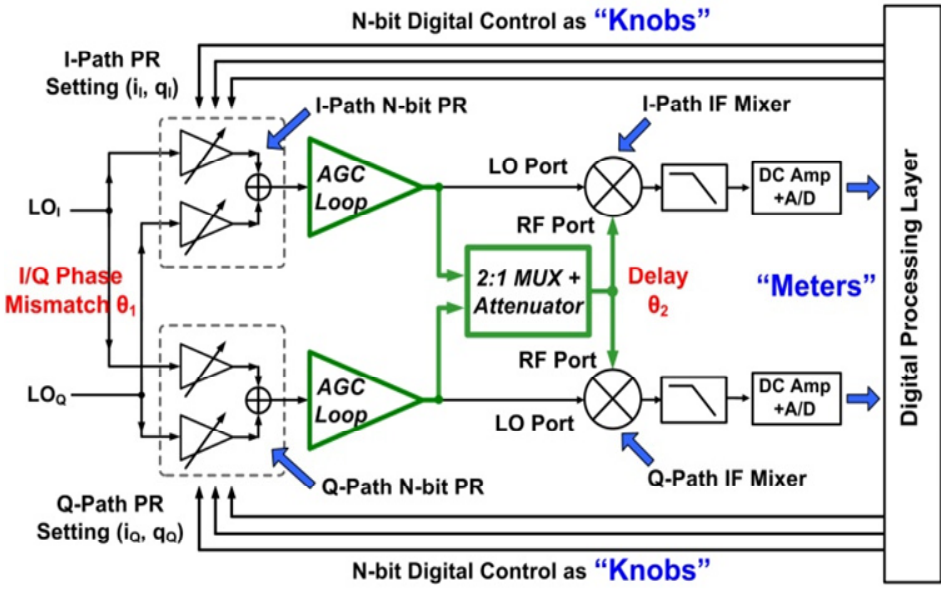


Figure 5.3: Self-healing phase synthesis architecture.

The phase of this fixed amplitude signal can now be estimated through mixing with the quadrature LO components. The same mixers of the signal chain are re-utilized in this scheme to reduce are and power overheads due to self-healing, as well as to enable accurate characterization of the mixer non-idealities. A 2:1 multiplexer selects which LO signal (I or Q) goes through to the RF port of the mixer during self-mixing. To ensure that the mixer does not saturate, the MUX also attenuates the RF port signal appropriately. Once the phase estimation has been completed for all settings of the in-phase and quadrature-phase VGAs, the results can be stored in a lookup table. Thus, given a target phase, the self-healing digital core (not implemented in the present design) can automatically select the optimum VGA gain settings and completes the phase healing functionality. Note that no additional test tone is required in this scheme, only the I and Q LO signals themselves are utilized.

5.4.2 Estimation of I/Q Phase Mismatches and MUX/Attenuator Delay

In addition to the random/systematic I/Q phase mismatches, the additional blocks due to self-healing, namely the MUX/attenuator, add their own phase error which must be estimated before generating a lookup table of the interpolated phases. This calibration step can be described in three steps as follows:

Step 1 I-path phase rotator is set to only pass the in-phase LO signal and the MUX selects it so that self-mixing occurs at the I-path mixer, as depicted in Figure 5.4. The

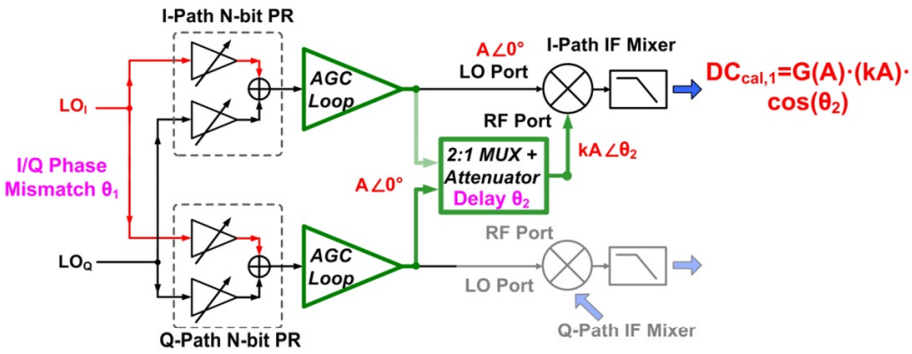


Figure 5.4: Calibration step 1.

output is a DC signal given by Equation 5.3.

$$DCOut_{cal1} = G(A).(kA).\cos\theta_2 \tag{5.3}$$

Step 2 I-path phase rotator is set to only pass the in-phase and Q-path is also set to pass the quadrature-phase LO signals. The MUX selects the signals so that intermixing occurs at the output of the I-path mixer, as shown in Figure 5.5.

The output is a DC signal given by Equation 5.4.

$$DCOut_{cal2} = -G(A).(kA).\sin(\theta_1 + \theta_2) \tag{5.4}$$

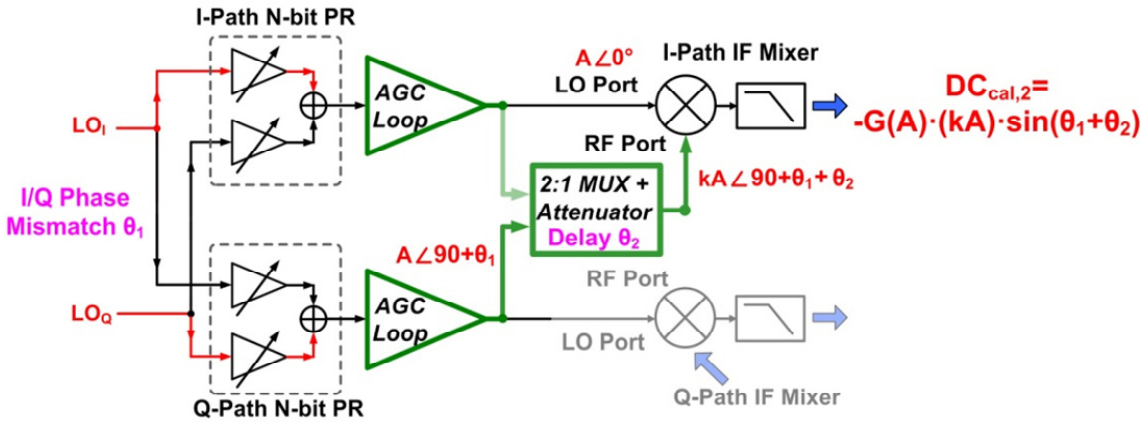


Figure 5.5: Calibration step 2.

Step 3 In the final step of the calibration, the I-path phase rotator is set to only pass the quadrature-phase and Q-path is also set to pass the in-phase LO signals. The MUX now selects the Q-path output signal so that intermixing occurs at the output of the I-path mixer, as shown in Figure 5.6.

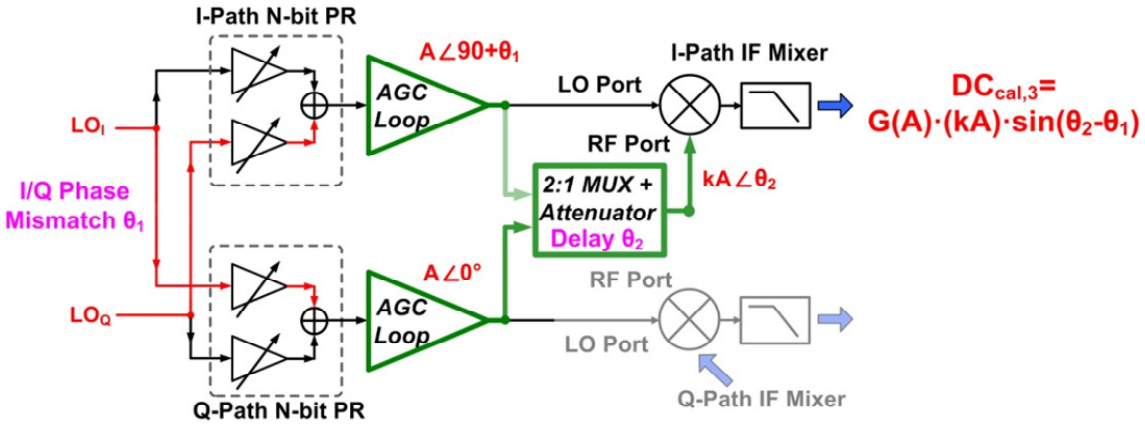


Figure 5.6: Calibration step 3.

The output is a DC signal given by Equation 5.5.

$$DCOut_{cal3} = G(A).(kA).\sin(\theta_2 - \theta_2) \tag{5.5}$$

With this information we can now compute θ_1 and θ_2 by using the following equations (5.6 & 5.7).

$$\theta_1 = a\sin\left(-\frac{DC_{cal2} + DC_{cal3}}{2.DC_{cal1}}\right) \tag{5.6}$$

$$\theta_2 = a\tan\left(-\frac{DC_{cal2} - DC_{cal3}}{2.DC_{cal1}.\cos\theta_1}\right) \tag{5.7}$$

Because typically, these errors θ_1 and θ_2 are much smaller than 90^o in practice, these inverse trigonometric functions are single-valued. An identical set of calibration steps can be performed to characterize the Q-path mixer as well.

5.4.3 Phase Measurement and Lookup Table Generation

With θ_1 and θ_2 determined from the previous steps, we can now proceed to estimate the exact phase shift produced by a given phase rotator setting. Assume that the I-path generates a phase shift of $\theta_x(i_I, q_I)$, where i_I and q_I are the weightings of the I and Q.

*Step I*The Q-path phase rotator is set to all in-phase LO and selected by the MUX to mix the interpolated signal at the I-path mixer. This provides a DC output:

$$DCOut_{meas1} = G(A).(kA).\cos(\theta_x(i_I, q_I) - \theta_2) \tag{5.8}$$

Step II Now the interpolated signal is mixed with the quadrature phase LO by setting the Q-path phase rotator to all Q and selecting it with the MUX. The DC output is given by:

$$DCOut_{meas2} = G(A).(kA).\sin(\theta_x(i_I, q_I) - \theta_1 - \theta_2) \tag{5.9}$$

With these two equations (5.8 and 5.9), the phase shift $\theta_x(i_I, q_I)$ can be uniquely deter-

mined by using equations 5.10 and 5.11. After sweeping all settings for (i_I, q_I), a complete look-up table can be automatically generated. A similar set of measurements can be performed for the Q-path phase lookup table.

$$\sin\theta_x(i_I, q_I) = \frac{\cos\theta_2}{\cos\theta_1}.\left[\frac{DC_{meas1}}{DC_{cal1}}\sin(\theta_1+\theta_2) + \frac{DC_{meas2}}{DC_{cal1}}\cos\theta_2\right] \tag{5.10}$$

$$\cos\theta_x(i_I, q_I) = \frac{\cos\theta_2}{\cos\theta_1}.\left[\frac{DC_{meas1}}{DC_{cal1}}\cos(\theta_1+\theta_2) - \frac{DC_{meas2}}{DC_{cal1}}\sin\theta_2\right] \tag{5.11}$$

5.4.4 Implementation and Measurement Results

The implemented test structure occupies an area of $1.4mm^2$ which is pad limited, as shown in Figure 5.7. The IF mixers are implemented as Gilbert cell based mixers with resistive source degeneration for improved linearity. Switches at the gates of the bottom current transistor control the MUX functionality. The attenuator is implemented using capacitive dividers for broadband attenuation. 10-bits control the VGA settings for the phase interpolator. The

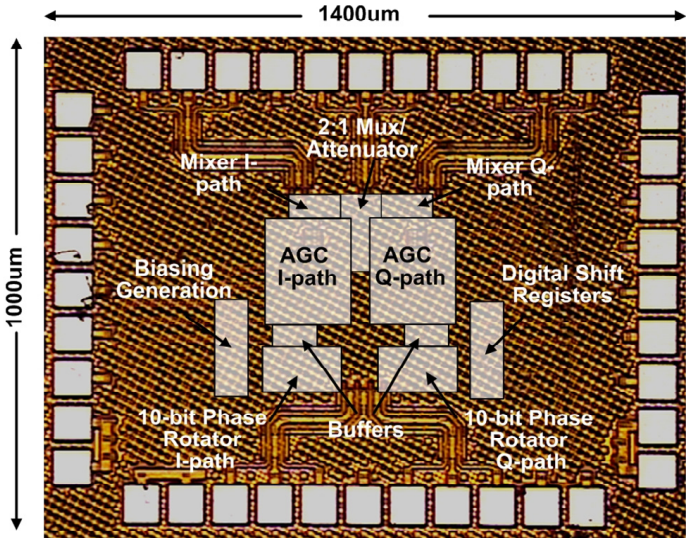


Figure 5.7: Test structure die-photo.

chip was mounted on a PCB with brass backing for thermal connections. Figure 5.8 shows the test-setup used to measure the chip. Digital controls are provided externally through a

Xilinx FPGA chip. Note that although the self-healing itself uses LO signals from the PLL directly to generate the look-up table, a test signal has been used here only to characterize the AGC loop and validate operation of the self-healing technique.

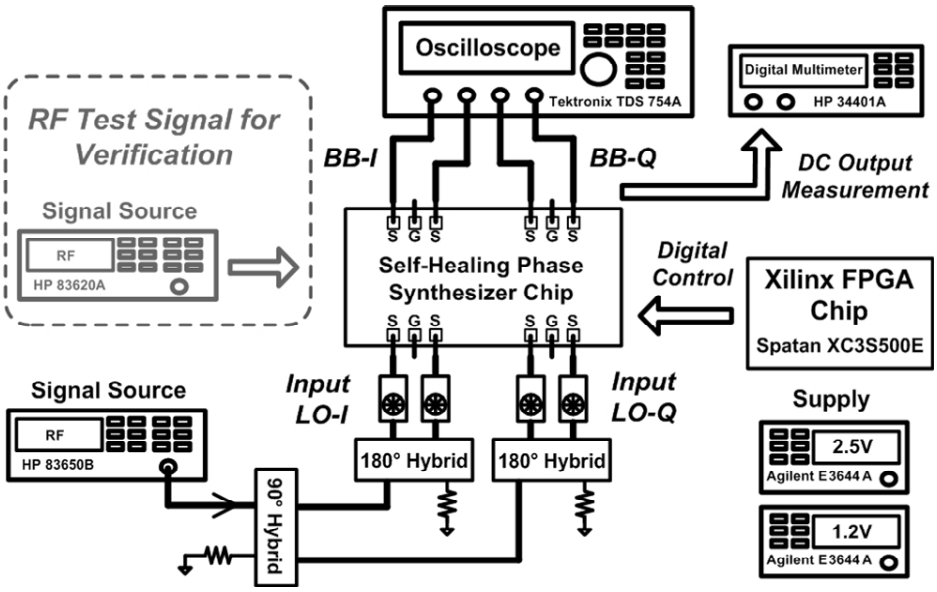


Figure 5.8: Test structure die-photo.

First, operation of the AGC loop was verified to ensure constant LO amplitude to the IF mixer. Exact same I/Q weightings were chosen for both I-path and Q-path mixers. The AGC loop was enabled only in the I-path and not in the Q-path. The MUX is disabled to avoid any inter-mixing and/or self-mixing. A test signal is sent to the RF ports of the two mixer and the downconverted baseband amplitudes are measured across a wide range of LO input levels. Figure 5.9 show operation of the AGC loop in the I-path where over an LO power from -4.65 to +3.15dBm, the I-path down-converted output is flat within ±0.04dB. The Q-path, without the AGC loop, shows the regular saturated behavior for higher LO powers.

Three measurements were performed to validate the self-healing functionality. The ability of the system to "sense" the phase was first verified. Quadrature LOs are sent to the phase rotator for interpolation. The phase shift is characterized using the approach mentioned in

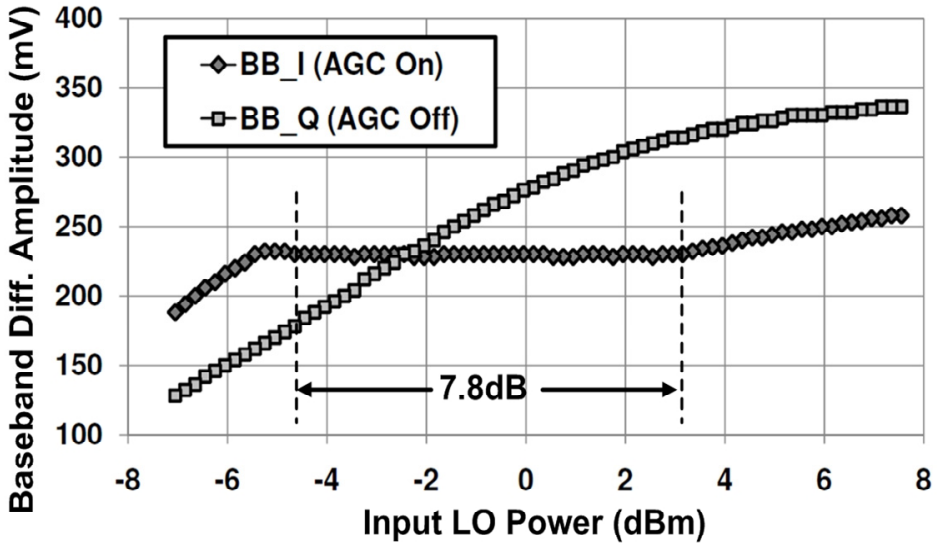


Figure 5.9: AGC loop response for f_{LO}=3GHz and f_{RF}=2.9GHz.

the previous sections. Then, with the same settings, an RF test tone is sent to the mixer and down-converted signal phase is measured using a real time oscilloscope. Results are shown in Figure 5.10 showing a maximum discrepancy of 1.63^o.

When all settings of the I and Q path phase interpolator VGA were varied, a large number of closely spaced phases can be generated, as shown in Figure 5.11. The figure also shows operation of the AGC where most constellation points fall onto the normalized unit circle.

Finally, once the lookup table was generated, the interpolated closest matching phases were compared to the target phases, where intentional mismatch was introduced into the input quadrature LO signals. Figure 5.12 shows that the synthesized phases closely track the target ones.

Over the entire frequency range of operation, 2-to-6GHz, the self-healing technique reduces the RMS phase error between the target and interpolated phases to below 0.6^o, as shown in Figure 5.13.

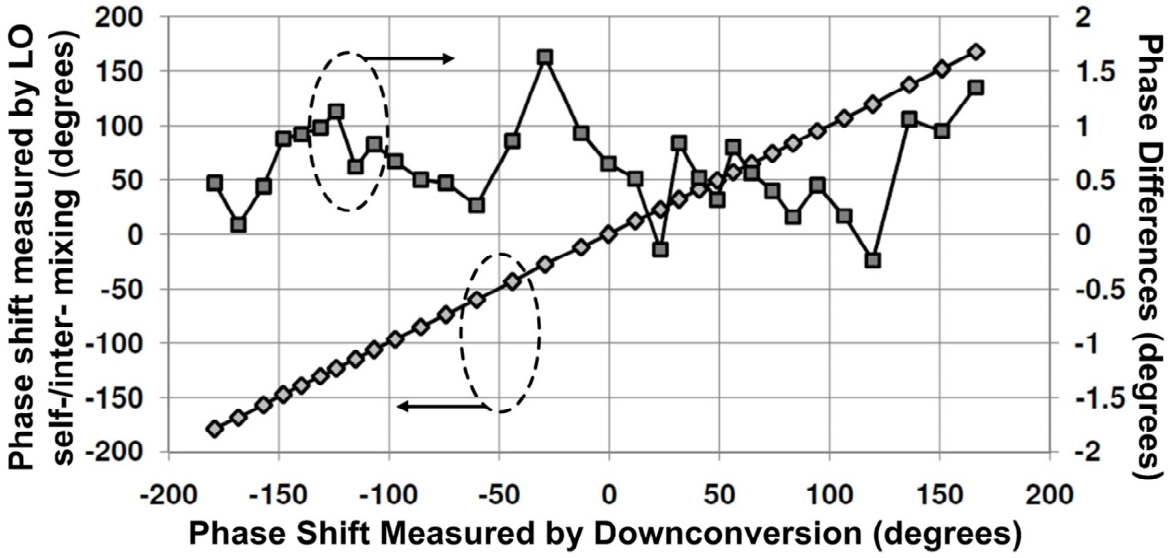


Figure 5.10: Phase measured by LO self/inter-mixing and by RF test tone at f_{LO}=5GHz.

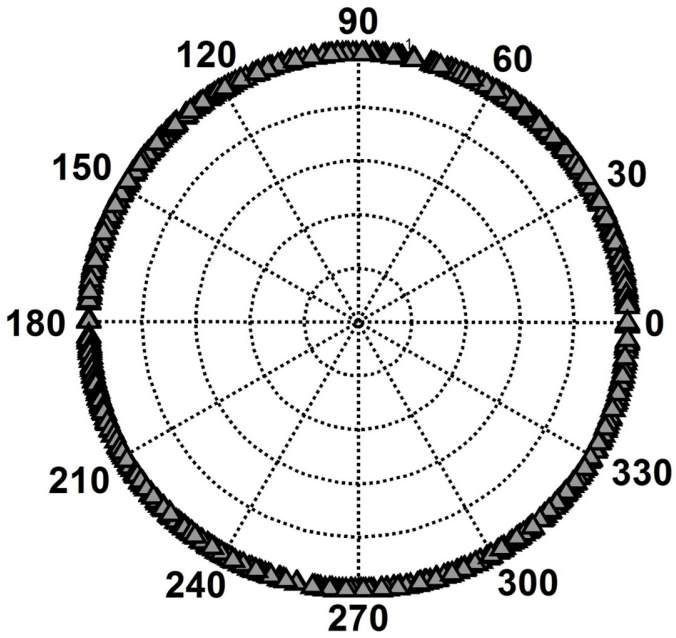


Figure 5.11: Full-range phase interpolation at f_{LO}=6GHz.

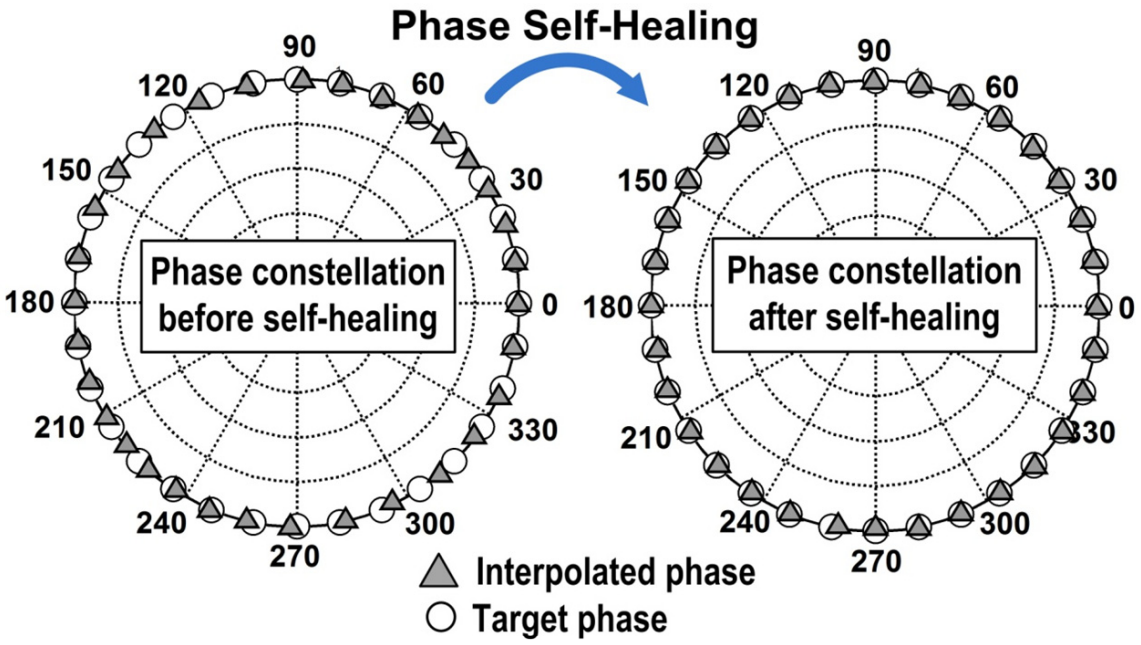


Figure 5.12: Phase constellation with 11.25^o phase steps before and after self healing.

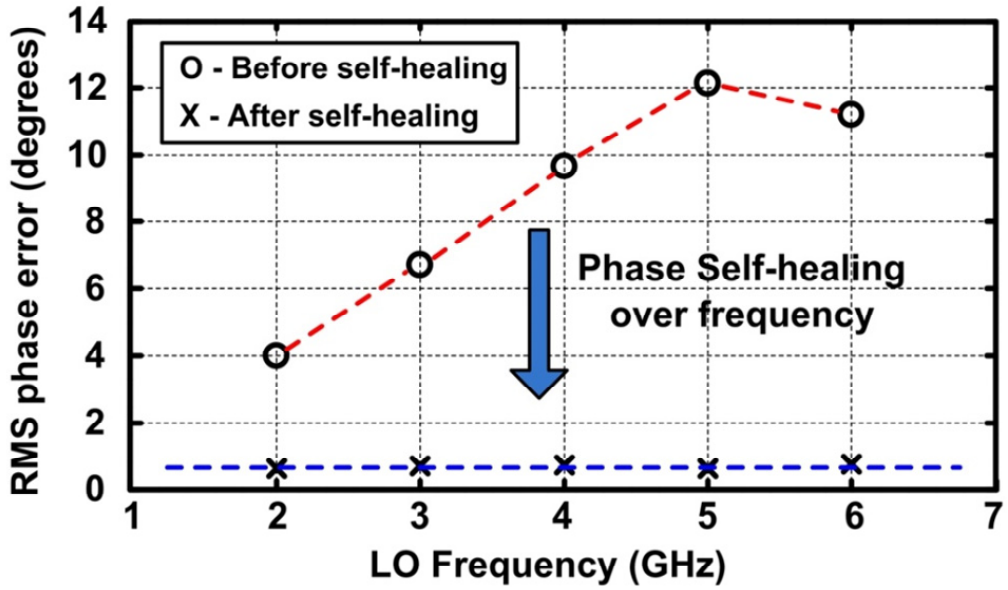


Figure 5.13: RMS phase errors before and after healing versus frequency.

5.5 Wide-band VCO Design

5.5.1 Background

A critical block in the design of wide-band phased array receivers is the voltage controlled oscillator. In the present implementation, the receiver has two PLLs, one for the low-band LB (6-10GHz) and another for the high-band HB (10-18GHz). The $2/3^{rd}$ frequency scheme translates to a VCO frequency range of 4-6.67GHz for the LB and 6.67GHz to 12GHz for the HB. Inherent trade-offs between wide-bandwidth and maintaining high tank quality factor over the frequency range, make design of such wideband VCOs challenging in scaled CMOS processes. In addition, particularly for self-healing systems, the VCO itself needs to be robust enough to be able to cover the entire frequency range in the presence of process/mismatch variations. Another issue in such designs is the stringent requirement of phase noise over such a wide frequency range. Keeping these trade-offs in mind, the following sub-sections will present design details, as well as simulation and measurement results of these two wideband VCOs.

5.5.2 Design Considerations

The choice of the topology of the wideband VCO was governed by a variety of trade-offs. A cross-coupled pair based negative resistance oscillator toplogy was chosen because of its typically lower power consumption as well as less stringent startup conditions. Although a PMOS based cross-coupled pair presents lower $1/f$ noise, for the same g_m its needs a bigger device, leading to additional parasitic capacitance. Thus, for this design, an NMOS based cross coupled pair was chosen. The detailed schematic is shown in Figure 5.14.

Frequency tuning is achieved by 3-bit control using binary weighted capacitors. The design kit had the option of high-Q MIM capacitors with high self-resonance frequencies which were used in the implementation. It must be noted that the switch size must also scale with the coarse tuning capacitors. For fine tuning the VCO frequency, NMOS varactors were incorporated having a control voltage ranging from 0 to 1.2V. Typically in modern CMOS

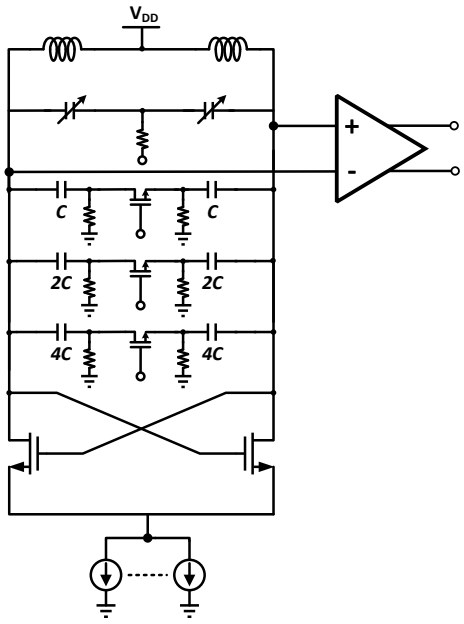


Figure 5.14: Schematic of the VCO core.

processes, varactor Q's are severely limiting and Figure 5.15 shows simulated quality factor of an example varactor where optimizations were performed on the dimensions to maximize Q for a given nominal capacitance value.

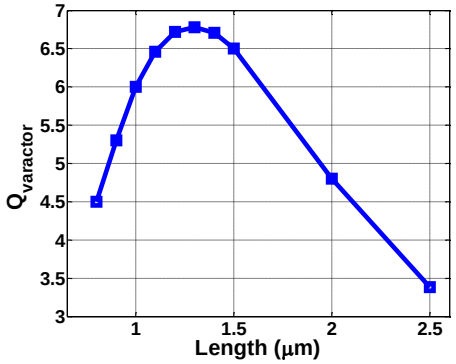


Figure 5.15: Varactor Q-factor versus channel length.

As the VCO is operated close to its lowest frequency, i.e., more capacitors are switched on, the overall Q of the tank degrades, which may lead to startup condition violations. To ensure reliable startup across all frequencies, additional switchable current sources were

implemented, as shown in Figure 5.16. The VCO was fabricated as part of a PLL (designed by Florian Bohn). The die-photo of the implemented VCO (HB) is shown in Figure 5.17.

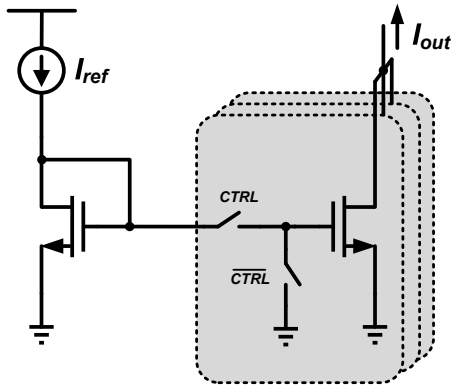


Figure 5.16: Switchable current sources for reliable startup.

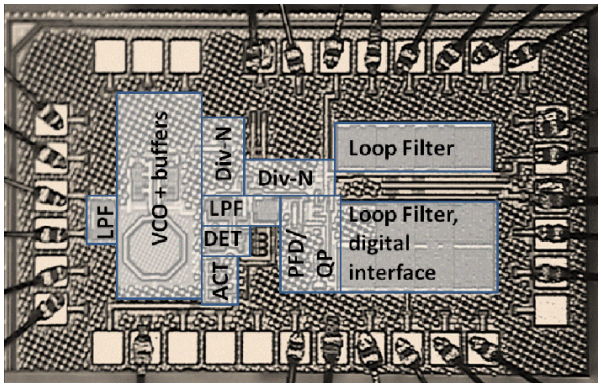


Figure 5.17: Die micrograph of entire HB PLL showing VCO.

5.5.3 Simulation Results

The two VCOs were designed using library elements, namely RF triple well transistors and library MIM capacitors and spiral inductors. Figure 5.18 show the layouts of the two VCOs with integrated buffers and bias circuitry. The cross-coupled NMOS pair and the MIM capacitors are placed symmetrically across the center tapped inductor. The inductor for the low-band has an additional metal-1 ground shield to reduce substrate losses. For the

high-band however, this ground shield may lead to additional capacitance which lowers the self-resonance frequency of the inductor. The frequency tuning characteristics of the two

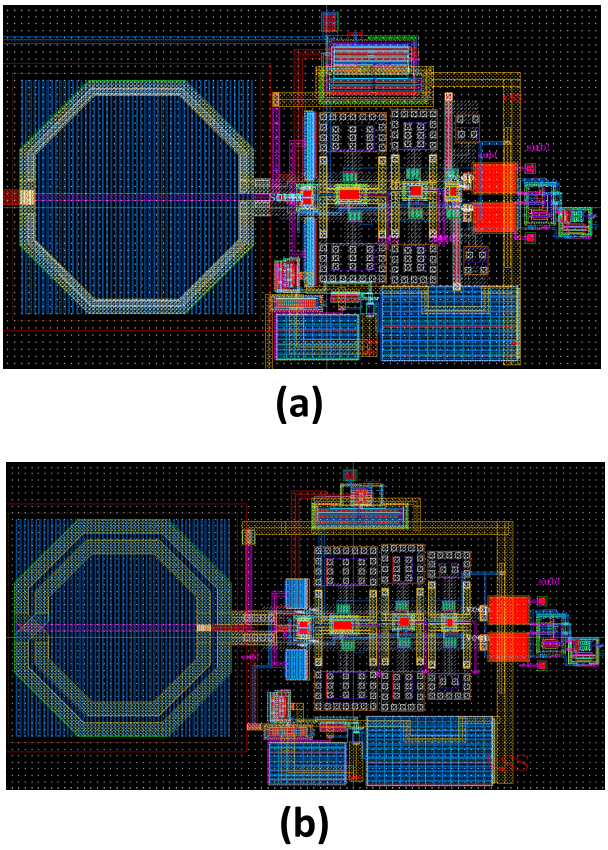


Figure 5.18: Layout for (a) high-band and (b) low-band VCOs.

designed VCOs are depicted in Figure 5.19. Efforts were made to ensure sufficient post-layout overlap between the various coarse tuning bands. The simulated frequency ranges for the two VCOs were 7.12 to 12.65GHz for the high-band and 3.92 to 7.13GHz. The high-band VCO was slightly tuned high in simulation to include effects of possible parasitic inductances.

To evaluate the sensitivity of the VCO to process and mismatch, the highest and lowest frequencies were simulated over 100 Monte Carlo runs. Simulated results for the high-band (Figure 5.20)and low-band (Figure 5.21 VCOs show the robustness of the design. The lowest frequency for the high-band VCO had a μ= 7.14GHz and a standard deviation σ=124.6MHz.

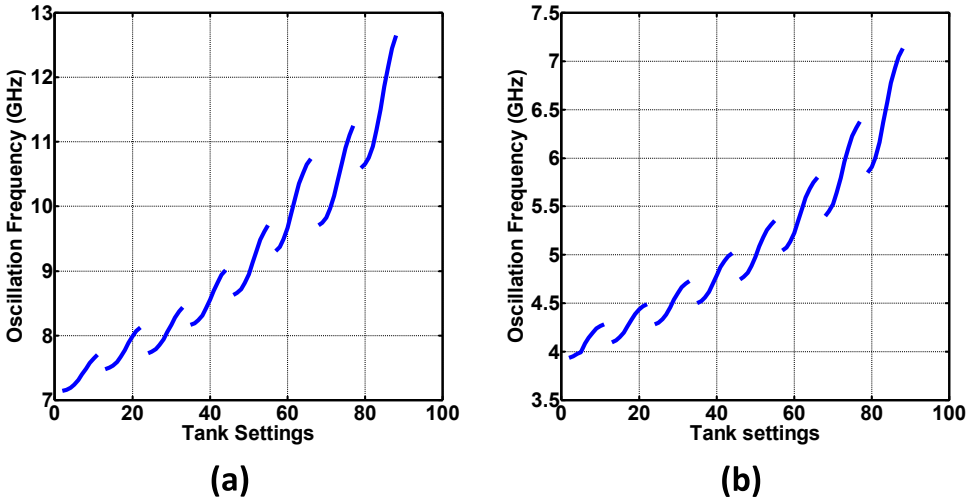


Figure 5.19: Tuning for (a) high-band and (b) low-band VCOs.

The highest frequency showed a μ= 12.93GHz and a standard deviation σ=171.2MHz. Same numbers for the low-band VCO were μ= 3.93GHz and σ=70.8MHz for the lowest frequency and μ= 7.1GHz and σ=98.6MHz.

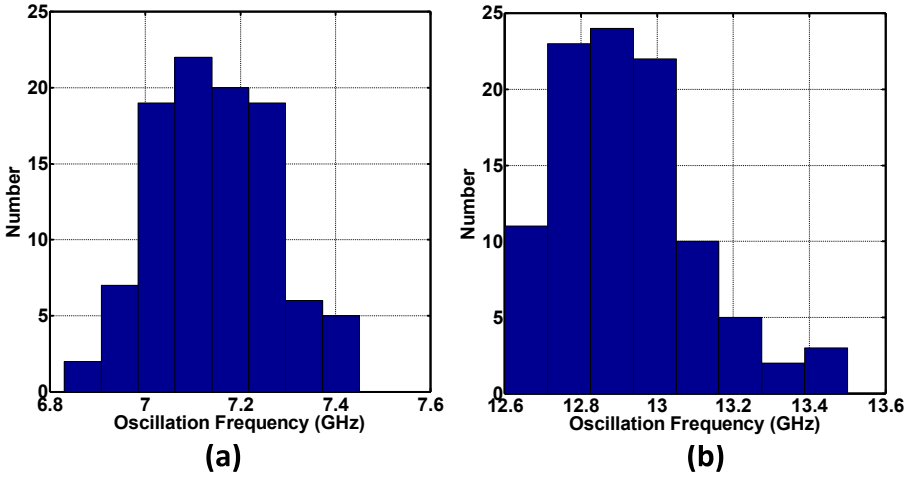


Figure 5.20: Monte Carlo simulations showing variation of (a) lowest and (b) highest frequencies for the high-band.

In addition to process variations and mismatch, short and long term temperature variations may also limit the operating range of the VCO. Temperature simulation results are

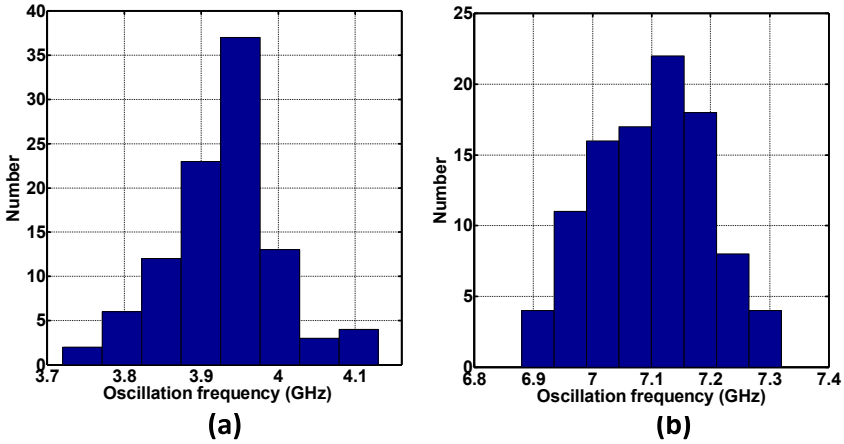


Figure 5.21: Monte Carlo simulations showing variation of (a) lowest and (b) highest frequencies for the low-band.

shown in Figure 5.22. Over a wide range of temperatures -40^o to +120^o the highest frequency of oscillation for the high-band VCO varies by less than 400MHz.

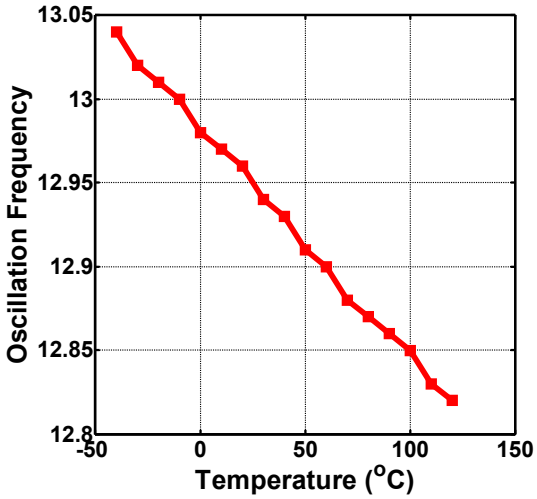


Figure 5.22: Variation of highest oscillation frequency for the high-band VCO versus temperature.

Another important consideration for wideband VCOs is phase noise. Because of the coarse and fine tuning, the tanks in these VCOs contribute more phase noise than their narrowband counterparts. The results of close in phase noise for both VCOs are shown in

Figure 5.23. As expected the phase noise gets worse as frequency increases. In addition, the added current sources at the tail for reliable startup across the entire band also contribute to the phase noise.

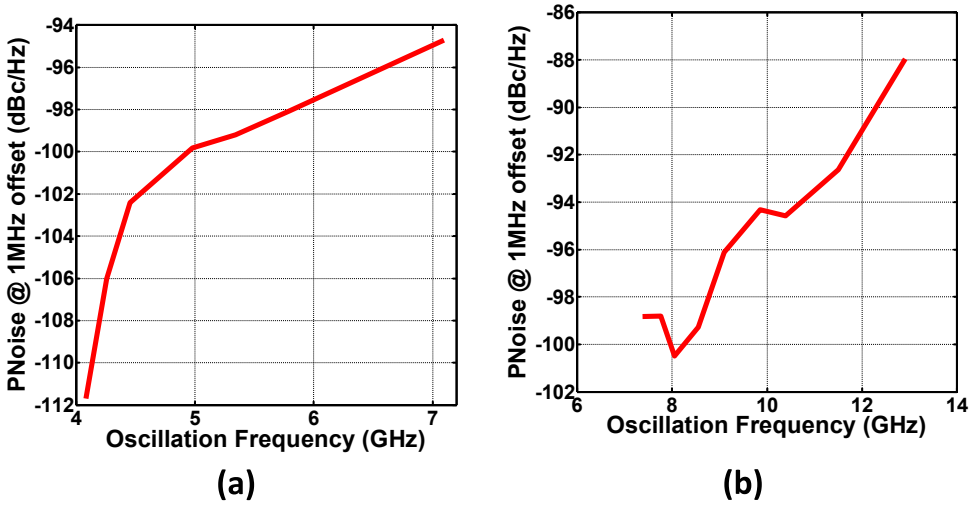


Figure 5.23: Phase noise for (a) low-band and (b) high-band VCOs.

5.6 Conclusion

In conclusion, in the first part of this chapter we have presented a self-healing scheme which re-utilizes building blocks of the same receiver to reduce I/Q phase and amplitude mismatch. The scheme relies on self-mixing between the non-ideal I and Q components of the LO to estimate the mismatch, then sweeps through all phase rotator settings to create a lookup table of target phases. As an example, a 2-to-6GHz self-healing down-converter has been demonstrated, which leads to accurate phase interpolation in the presence of process and modeling uncertainties.

In the second part of the chapter, we have discussed in detail the design of two wide-band tunable voltage controlled oscillators towards the same receiver application. Trade-offs between loop gain, tunability, and startup have been investigated and robustness of the designed VCOs against process variation and mismatch has also been demonstrated.

Chapter 6

Other Works

6.1 Stacked SOI CMOS Power Amplifier

Despite the rapid advancement of CMOS technology, high-power PAs in such processes have remained some sort of a holy grail. One of the major reasons as to why CMOS PAs have been limited to mostly low to moderate power levels is the low breakdown voltage which continues to scale down with technology. Because of this voltage limitation, the only other way to achieve high output powers is to operate at high current regimes. Obviously the PA devices then need to be large to be able to handle such high currents. The optimum impedance of such a large device is usually low and when such an impedance is matched to a typically high impedance of an antenna, for example, the matching network loss increases manifold. Let us evaluate two examples where this breakdown voltage limits the output voltage of a PA.

Figure 6.1 (a) shows the drain voltage waveform associated with a traditional common-source PA. For maximum efficiency, the drain node must swing from 0 to $2V_{DD}$. Let us denote the optimum impedance for this PA as $Z_{opt} = V_{DD}/I_{max}$, where I_{max} is the maximum AC current drawn by the PA. Assuming the transistor can sustain an AC voltage swing of $2V_{DD}$ across its terminals, which is typically true in modern CMOS processes, a cascode PA can potentially be operated at twice the V_{DD}, thereby doubling the drain voltage swing as shown in Figure 6.1. Interestingly, the optimum impedance for such a cascode PA also doubles ($Z'_{opt} = 2.Z_{opt}$) when compared to a common-source PA. However, in such a situation,

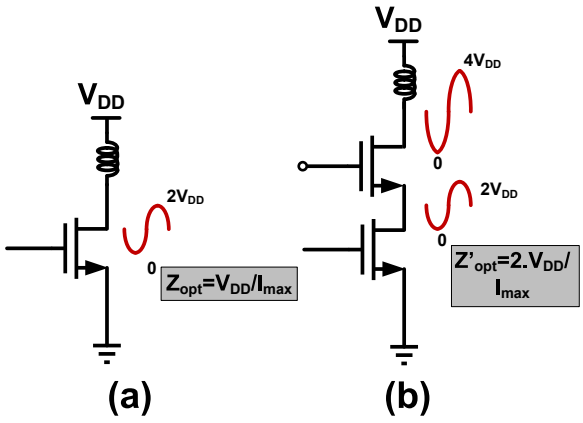


Figure 6.1: Voltage swings and optimum impedance for (a) common-source and (b) cascode PA.

although the individual devices themselves do not have to sustain a large voltage between any of their two terminals, the drain of the cascode device still swings to high enough voltages that could case drain-to-bulk breakdown. Clearly in traditional CMOS processes, where the common bulk terminal is usually connected to the source of the common-source device or ground, this approach cannot be extended to more than two devices, thereby limiting the output voltage and the optimum impedance.

SOI technology has seen rapid development over the last few years with the advantages of reduced junction capacitances leading to higher switching speeds for nm range CMOS devices as well as no wells or latchup effects. An SOI process comprises of three layers: a thin surface layer of silicon, where transistors are fabricated, an insulating layer referred to as buried oxide (BOX), and a support silicon wafer. Transistors fabricated on the top most layer have lower parasitic capacitances, can run at lower voltages, and are less susceptible to noise from background cosmic radiation. Also, each transistor is separated from its neighbor by silicon-dioxide. The advantage of this technology immediately becomes clear in PA design, due to the non-existent common bulk terminal in the devices. These individual floating body devices may now be stacked on top of each other similar to a cascode device without running into drain-bulk breakdown limitations.

How far can we keep stacking devices this way? The limit of such transistor stacking is

the breakdown voltage of the BOX, which is typically of the order of 10 Volts for modern processes. This entails that one could stack up to 5 devices operating off a supply voltage of 5 Volts and at the same time benefit from 5 times the optimum impedance when compared to a single device. The schematic is shown in Figure 6.2.

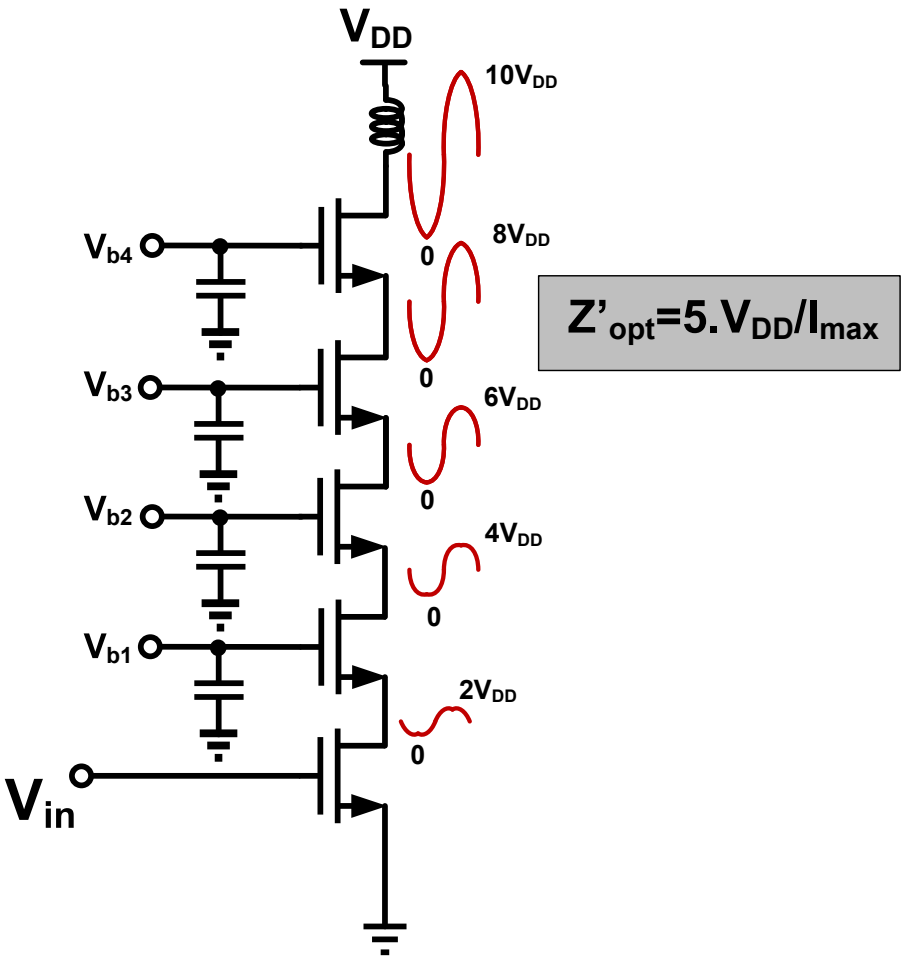


Figure 6.2: 5 transistors stacked while operating off a 5-V V_{DD}.

To ensure that all stacked transistors experience equal voltage stresses, the gates of these devices must not be kept at AC ground. This is especially important for the upper devices where gate-drain or gate-source voltage swings may be too large for reliable operation. Capacitors must be placed at all gates in decreasing values towards the top of the stack to allow sufficient voltage swings at the gates. For an optimized stage, all V_{GS}s as well as all

V_{DS} should be identical for each device to contribute equally to the final output voltage. It can be shown that the n^{th} gate capacitor ($n \geq 2$) can be expressed as a function of the C_{gs} and C_{gd} of the transistors (Equation 6.1 [51]).

$$C_n = \frac{C_{gs,n} + 3C_{gd,n}}{2n - 3} \tag{6.1}$$

6.1.1 Simulation Results

Extensive simulations were performed on an example 4-stacked PA where each transistor was 200μm in width with minimum channel lengths to validate the advantages associated with stacking. Figure 6.3 shows simulated output voltage and current magnitudes versus number of stacks. No additional matching network was inserted between the stacked transistors. It is evident that while the voltage scales linearly with the number of stacks, the drain current remains constant which thereby scales the optimum impedance.

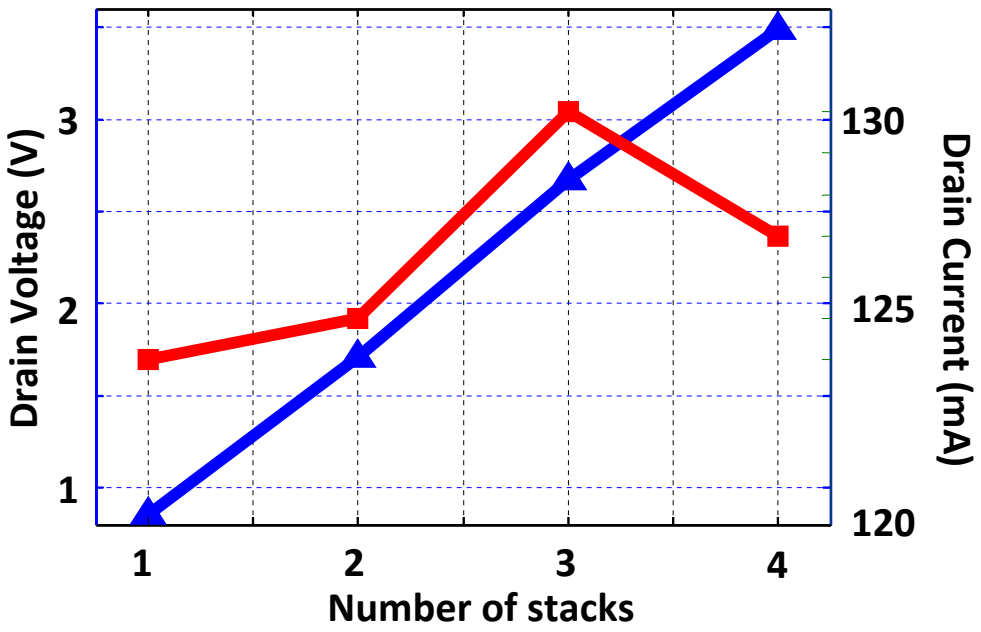


Figure 6.3: Simulated voltage and currents versus number of transistors in a stack.

As expected, the output power also grows linearly with the number of stacks, as shown

in Figure 6.4.

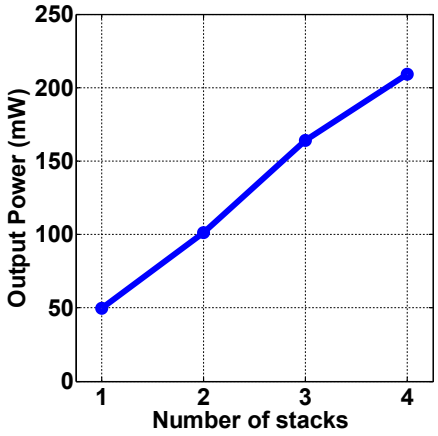


Figure 6.4: Simulated output power versus number of transistors in a stack.

The designed chip had two 5-stacked PAs operating at a center frequency of 60GHz, each having transistors of size 800μm/32nm. A 2-to-1 splitter/combiner architecture was utilized where the input splitter and the output combiner are composed of coupled transmission line combiners which were simulated and optimized using HFSS. At higher frequencies, the interconnection between several parts of the same transistor needs to modeled accurately. Layout of one 800μm transistor was done in 16 blocks of 50μm each. Figure 6.5 shows the simulation testbench for the transistor connections in HFSS. Electromagnetic simulations for the transistors were performed by Steven Bowers.

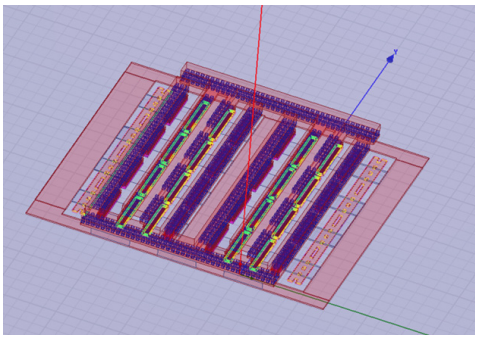


Figure 6.5: HFSS simulation of 800μm transistor.

Figure 6.6 shows simulated drain voltages of one of the 5-stack PAs operating off a V_{DD} of

5-volts. To ensure reliable operation of the transistors and to enable optimum operation, the

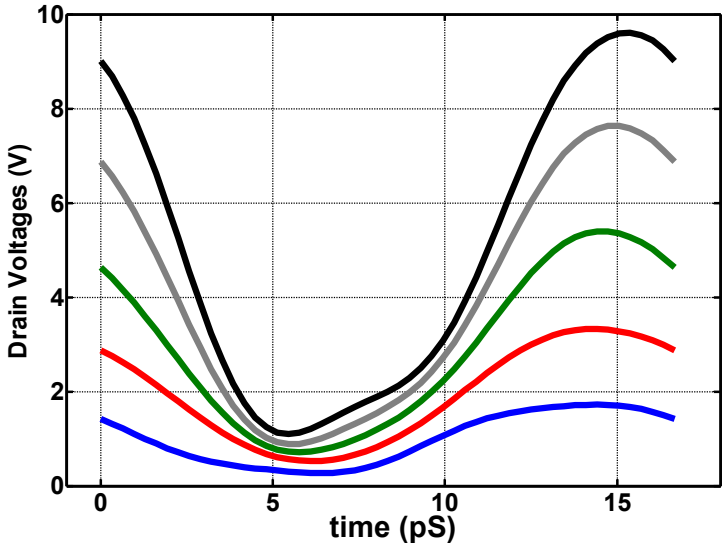


Figure 6.6: Simulated drain voltages for each transistor in a 5-stack PA.

gate capacitors were scaled accordingly for each transistor. Figure 6.7 (a) depicts simulated gate-source voltages of each transistor, confirming the in-phase optimum operation of the stack as a whole. Similarly, the drain-source voltages (Figure 6.7) (b) demonstrates optimum operation for each transistor.

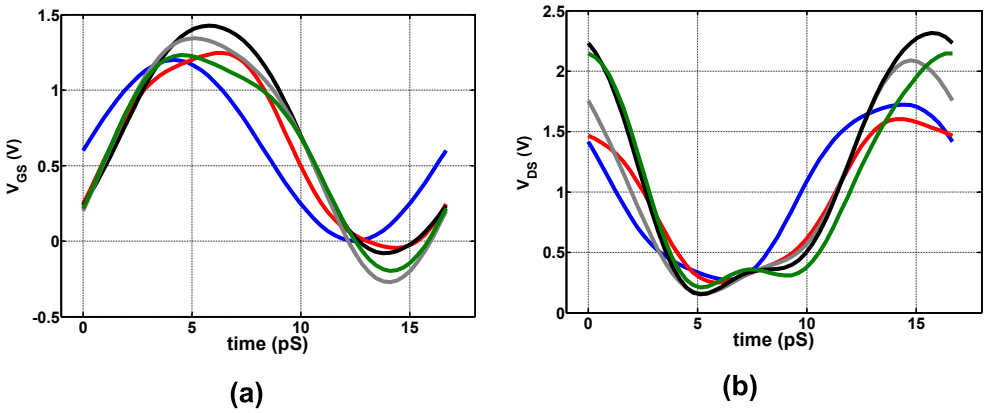


Figure 6.7: Simulated voltages for each transistor in a 5-stack PA.

Figure 6.8 shows the simulated P_{out} versus P_{in} of each stacked PA as well as the efficiency.

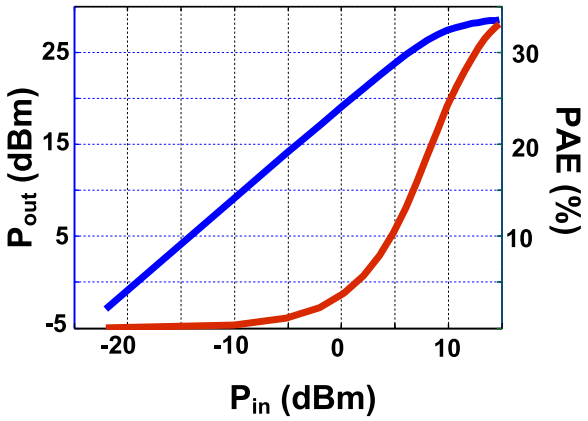


Figure 6.8: Output power and PAE versus input power.

The chip layout is depicted in Figure 6.9 showing the 2:1 split, amplify, and combine architecture. Figure 6.9 also shows various elements associated with the self-healing aspect of the PA. 5-bit DACs similar to the ones used in the previous chapters were used in conjunction with resistive networks for actuating the gate bias points of the five transistors in each PA independently. Output power of the PA is sensed through coupled lines (30dB coupling) terminated by 50-Ohms similar to the previous chapters. The sensed voltage is then rectified and amplified using RF sensors. An on-chip 8-bit SAR ADC is used to digitize these sensor outputs which then are fed to an on-chip synthesized digital core. Due to the relatively high gain and high output power, it is also important to ensure that the PA remains stable for all input and output reflection coefficients. The simulated stability factor (K) is shown in Figure 6.10 where it remains greater than 1 (0 dB) for the entire frequency range.

The Digital Healing Infrastructure allows the power amplifier to compensate for process variation, defects, and environmental changes. It is composed of five core components (Scan Chain, Controller, and SARADC Handshaking, DACs, and SARADC), and three of which (Scan Chain, Bias Healing Controller, and SARADC State Machine) are automatically synthesized and routed from standard cells. The infrastructure interfaces with our sensors and actuators to optimize the output power of our power amplifier. The first block is the scan chain. The scan chain programs the algorithm healing parameters. There is a gradient of modes for the algorithm ranging from coarse to fine healing. Coarse modes provide a faster

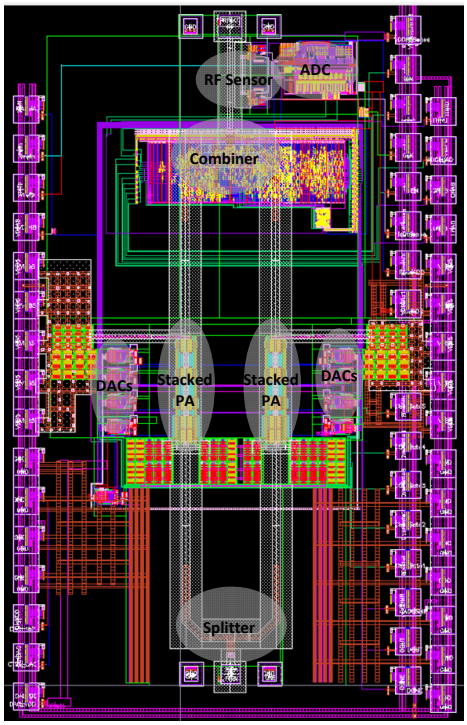


Figure 6.9: Chip layout for the dual 5-stack PA with integrated self-healing.

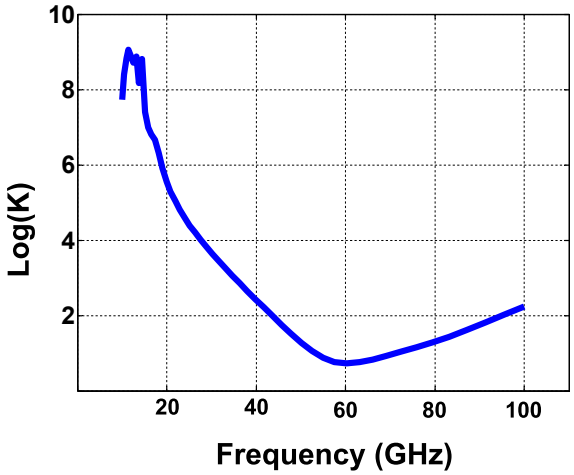


Figure 6.10: Simulated output power and efficiency versus output power.

healing time using a reduced search space. A fine mode maximizes potential output power by enlarging the search space but increases the healing time. In addition, the scan chain also contains bypasses of the healing algorithm to implement a more sophisticated healing algorithm off chip. The bypasses include bypassing programming of the actuators and receiving sensor data. The second block is the controller. The controller initiates an output power healing algorithm that can be tuned from faster, coarse-mode and slower, fine-mode healing. The controller communicates with the SARADC handshaking block and the actuators. The SARADC handshaking multiplexes various sensors to receive information on the output power of the power amplifier. The actuator sets various transistor biases in the power amplifier.

6.2 Polarization Control and Modulation

Traditional wireless communication links consist of a transmitter, transmitting antenna, receiving antenna, and receiver. The two antennas are usually coupled electromagnetically through a single polarization, which is usually linear polarization with a fixed polarization angle, or circular polarization that is either left- or right-handed. The data that is sent through such a link is encoded within the phase and/or amplitude of the signal. One issue with such a setup is that if the antennas move, they can lose polarization match, and the level of coupling can decrease. One proposed solution to this problem is to provide polarization diversity by using two antennas that have orthogonal polarizations and to send either two streams of data or to select the antennas at any given moment that have the highest level of coupling [70] [71] [72]. However, this configuration can still suffer from polarization mismatch if the matched polarization falls between the polarizations of the two antennas. A similar method that uses a single antenna that can switch between two polarizations was presented in [73] but again can experience polarization mismatch if the desired polarization falls between the two switchable polarizations. A third solution, dynamic polarization control (DPC), has been proposed that can enable a single antenna to radiate with any desired polarization, and is controlled completely electronically without any mechanical reconfiguration [74].

Electromagnetically, there is another aspect of the signal that can be modulated: the polarization. With both a transmitter with controllable polarization, as well as a receiver that can detect the polarization of the incoming electromagnetic (EM) field, polarization modulation, or pol-mod, can be implemented by encoding the data to be transmitted within the polarization itself, such as the polarization angle. One example of this is to use the amplitudes of the horizontal and vertical linear polarizations to create a two dimensional constellation, akin to their duals in quadrature signals, such as quadrature phase shift keying (QPSK) or quadrature amplitude modulation (QAM). In addition, rotation of either the transmitter or receiver along the axis of propagation will simply result in a rotation of the constellation, and can be corrected in a similar fashion to constellation rotations due to time delay of IQ encoded data, such as training sequences. This is very beneficial, as this rotation will not result in any loss in power due to polarization mismatch due to the dual polarized nature of the receiver. Here we present polarization modulation, with a scaled proof of concept printed circuit board (PCB) based 2.4 GHz polarization modulation link provided as a demonstration of the concept.

6.2.1 System Architecture

Any linear polarization of an electric field E, can be expressed as a vector sum of two perpendicular linear polarizations E_x and E_y. A polarization modulation transmitter/receiver architecture utilizes this orthogonal decomposition by spatially combining two perpendicular polarizations with different gains to generate the desired electric field polarization. Figure 6.11 shows a conceptual diagram of the architecture for the transmitter (a) and receiver (b). The baseband signal, after proper weighting in two independent paths (A_1 and A_2), is up-converted using phase shifted LO signals which then feed two ports of a dual-port antenna which can simultaneously radiate orthogonal polarizations E_x and E_y. The far-field electric field can now be expressed as a superposition of the two radiated polarizations:

$$\mathbf{E} = k_1 A_1 \cos(\omega_0 t)\hat{\mathbf{u}}_\mathbf{x} + k_2 A_2 \cos(\omega_0 t)\hat{\mathbf{u}}_\mathbf{2}, \tag{6.2}$$

where the parameter $k$includes frontend gain as well as the radiation parameters and is assumed to be the same for the two polarizations. Depending on the choice of A_1 and A_2, the polarization angle of the transmitted signal can be controlled such that:

$$angle(\mathbf{E}) = \tan^{-1}(A_2/A_1). \tag{6.3}$$

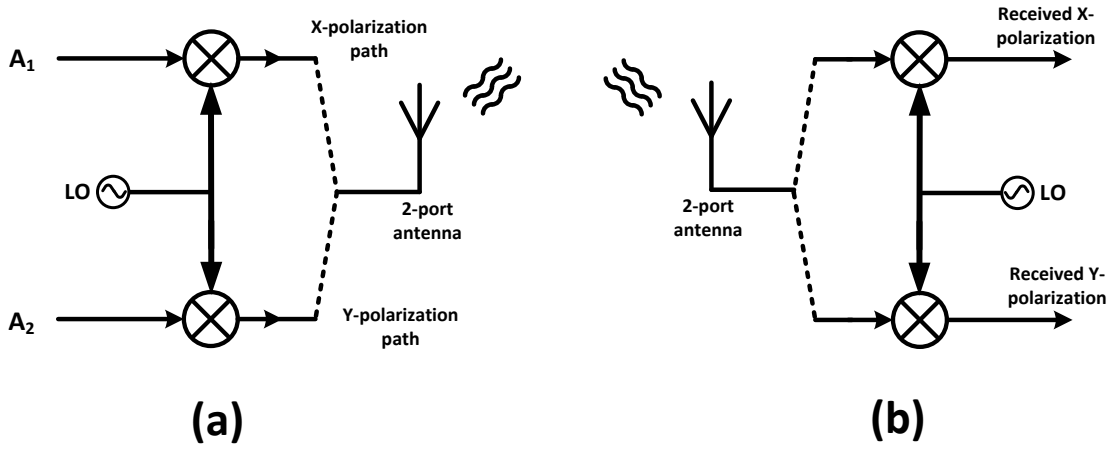


Figure 6.11: System architecture for polarization modulation (a) transmitter and (b) receiver.

Polarization control can thus be achieved by choosing different weights, A_1 and A_2, for a desired polarization angle.

This ability to dynamically adjust the polarization also opens up the opportunity (possibility) for data transmission (receiving) through polarization, where polarization angles and their relative magnitudes can be treated as data symbols. Linear polarization modulation is implemented in the present system by using symbol-dependent weights A_1 and A_2. It may also be noted that this architecture is also capable of transmitting/receiving circular polarization if the LO phases are offset by $\pi/2$. A similar analysis can be performed for the receiver.

6.2.2 System Implementation

6.2.2.1 Antenna Design

Electromagnetic radiation and reception in a polarization modulation transceiver can be achieved by a dual port antenna which is capable of transmitting/receiving two orthogonal polarizations (E_x and E_y) corresponding to the two driving/receiving ports (Port-X and Port-Y). Furthermore, it should provide sufficient isolation between the ports to minimize cross-polarization in both transmitting and receiving modes. Such isolation is also necessary to avoid input impedance variation at Port-X based on how strong the signal Port-Y is and vice versa.

A dual port patch antenna can be designed to achieve these goals, as shown in Figure 6.12(a) [75]. The patch is designed to be resonant at 2.4 GHz and quarter-wave transmission lines are used to match the input impedance of each port to 50Ω. Figure 6.12(b) shows the simulated radiation pattern and antenna gain while the patch is driven only at Port-X to radiate Ex. The same radiation pattern and antenna gain would result when the antenna is driven only at Port-Y instead of Port-X to radiate E_y. Figure 6.12(c) shows the isolation between the two ports versus frequency. The simulated isolation of 37 dB ensures that in the transmitter each port can be driven almost independently with arbitrary amplitude and phase to send various polarizations and in the receiver each port only picks up part of the polarized E-field which is aligned with it.

6.2.2.2 Transmitter and Receiver Design

Figure 6.13(a) shows implementation details of the transmitter at 2.4 GHz. This work was done in collaboration with Amirreza Safaripour. Variable gain amplifiers (VGAs) (PGA870) provide the necessary weights for polarization control over a 30 dB range. The LO phase shifts are obtained through a weighted summation of in-phase and quadrature-phase LO signals (AD8341). The up-converted signal in each path is then amplified and radiated using the on-board antenna. It must be noted that the response time of the VGA gain-control determines the limit of symbol rate for linear polarization modulation. The receiver block

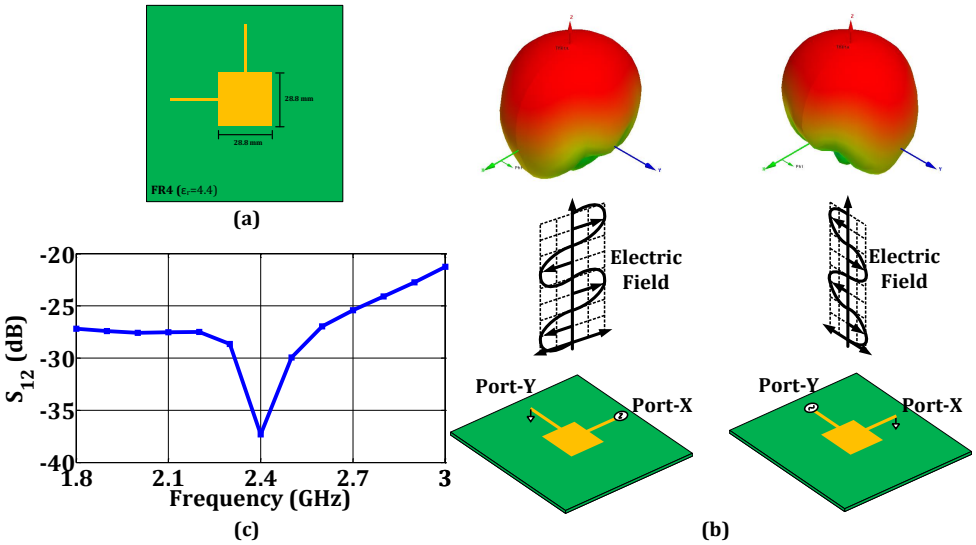


Figure 6.12: Patch antenna simulations showing port isolation as well as gain patterns for X and Y polarizations, maximum gain 2.7dB.

diagram is depicted in Figure 6.13(b) where a TriQuint TQP3M9037 low noise amplifier (LNA) amplifies the received signal in the corresponding polarization which is then down-converted using a MAX2042 mixer.

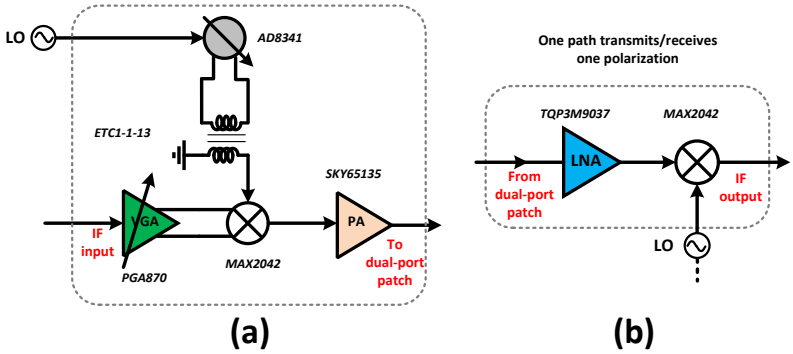


Figure 6.13: Implementation details of polarization modulation (a) transmitter and (b) receiver.

6.2.3 Measurement Results

The transmitter and receiver systems were first measured as stand-alone systems to verify polarization control and detection. Figure 6.14 shows the measurement setup for the system at 2.4 GHz. The X and Y polarization transmitter paths were first measured separately

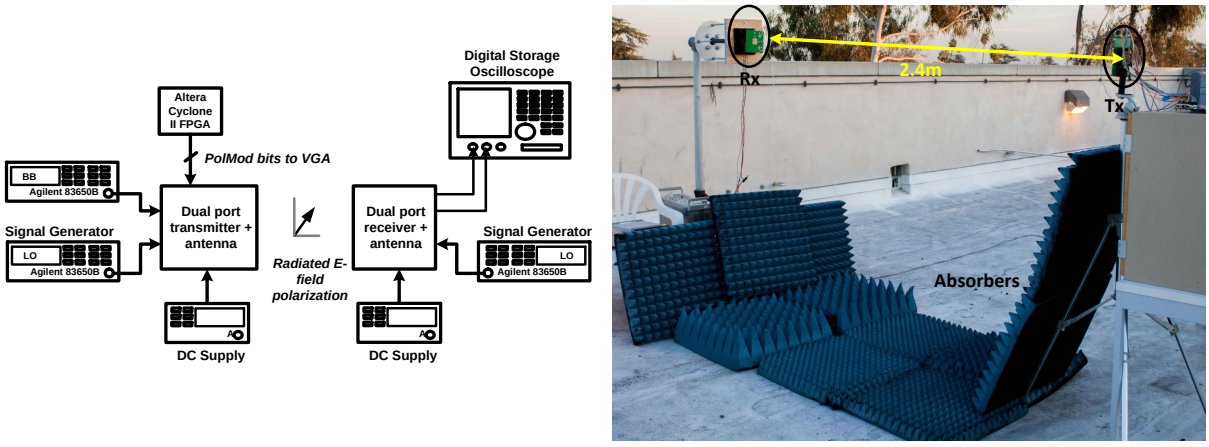


Figure 6.14: Measurement setup for polarization modulation.

using a linearly polarized horn antenna. Figure 6.15(a) shows output power variations of each path versus orientation of the horn.

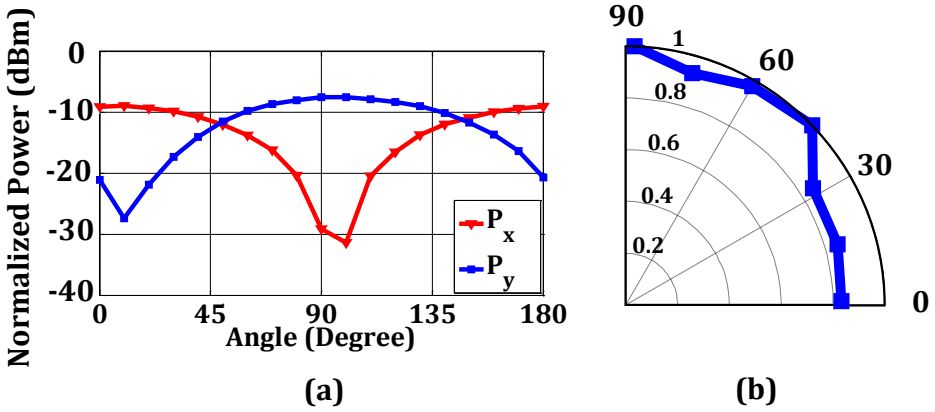


Figure 6.15: Transmitter stand-alone measurements showing (a) power variation across angle of the receiving horn and (b) dynamic polarization control over the first quadrant.

As expected, the angles for maximum received power by the horn are 90° apart for the two

polarizations. Measured isolation between polarizations was 20.1 dB. To verify the ability to rotate the transmitted polarization, both paths were operated with varying VGA settings to ensure constant total output power (P_x+P_y). Figure 6.15(b) shows total output power variation of <1 dB over varying polarization angle.

The receiver sub-system was also measured using the same horn as a transmit antenna across different transmitted polarizations. An isolation of 23.4 dB was measured with a maximum variation in total captured power of <1.5dB. Similar to the transmitter, the maximum signal from one receiver path corresponds to the minimum signal in the other (Fig.6.16(a)). It is interesting to note that unlike conventional linearly polarized receivers, this receiver is able to capture the entire power across both polarizations. A dual of a QPSK signal in E_x and E_y plane was also generated using the transmit horn antenna and resolved into the two receiver paths. The recovered signal points in one of the quadrants are shown in Figure 6.16(b).

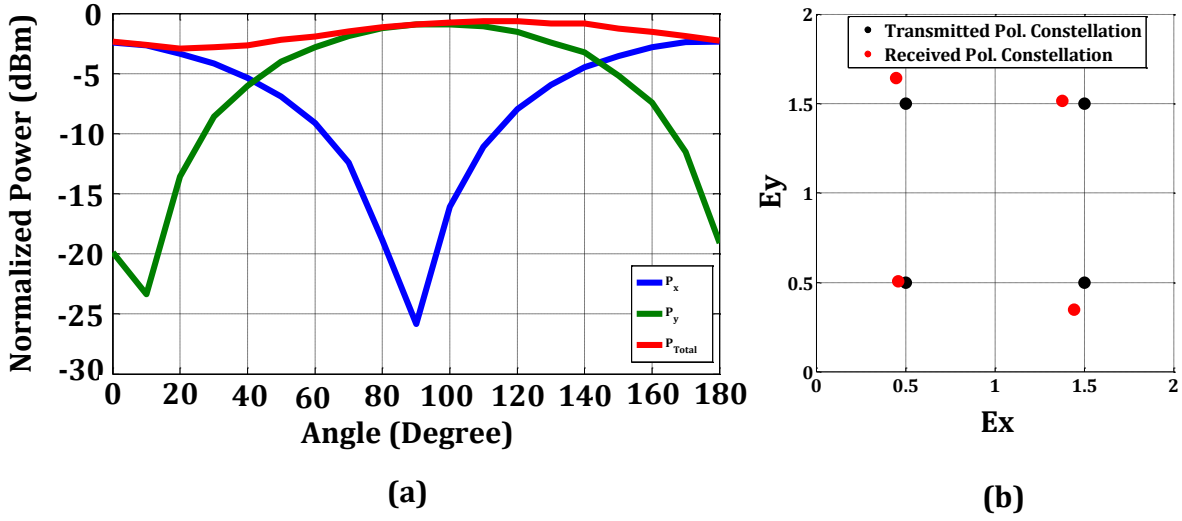


Figure 6.16: Receiver stand-alone measurements showing (a) received power variation across angle of the transmitting horn and (b) received polarizations having different magnitudes as well as angles.

6.3 Dynamic Manipulation of Magnetic Beads

6.3.1 Introduction

Lately there has been a lot of interest in lab-on-chip systems biological / microfluidic systems which are integrated together with electronic control / readout circuitry within the same chip. A lot of recent works in this area have focused on nano-sized magnetic particles which can be easily used for tagging cells, DNA, etc. Compared to pre-existing optical systems, these nano-particles do not require external expensive lasers to operate; neither do they affect normal cell / chemical behavior in most cases. As such, magnetic nano-particle based systems are promising candidates for Point-of-Care (POC) diagnostics. One of the few biological systems with native magnetic response is the Magnetotactic Bacteria (MTB). These bacteria (e.g. Magnetospirillum Magneticum) can have potential applications for sensing magnetic fields of materials as well as drug carriers using passive magnetic field based guidance systems. One such platform has been discussed here Dynamic, integrated micro-manipulation platform for magnetic nano-particles. This platform enables direct and precise generation of forces and fields affecting movement of nano-particles and/or MTBs as well as dynamic control of these forces for observing effects of time-varying fields on these particles. Also, as an added advantage, the platform is directly compatible with thin-film / integrated Circuit fabrication technologies with no additional post-processing.

6.3.2 System Description

The micro-manipulation platform for magnetic nano-particle system is based on magnetic fields generated by multiple, controllable current carrying conductors. It is important to note it is the gradient of the magnetic field and not the field magnitude itself which affects the motion of these particles. While previous magnetic manipulation systems have almost exclusively focused on movement of magnetic particles towards regions of zero-field gradient, precise control over particle speed and direction, especially over time, has not been addressed. The proposed platform attempts to achieve micron-level control of magnetic forces experi-

enced by a nano-particle as well as allow dynamic control of these forces whenever required so as to enable time-varying magnetic field guidance in one integrated system. An important advantage of this platform is scalability in the sense that more current carrying conductors can easily be incorporated into the system to allow even more degrees of freedom as well as increase in the control area.

A simplified example implementation of this generic platform is shown in Figure 6.17. It

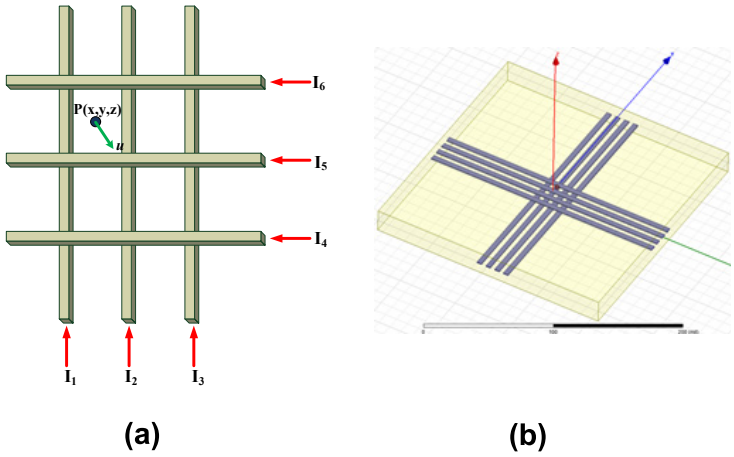


Figure 6.17: Simple 6-wire magnetic manipulation platform and Maxwell simulation model for a similar 8-wire platform.

depicts a grid structure consisting of six current carrying conductors arranged in perpendicular directions on two different levels. Magnetic fields generated by these wires can be easily computed using 3D Electromagnetic (EM) solvers such as Maxwell 3D. It can also be shown that simple MATLAB based calculation of magnetic field gradients approximates the actual behavior closely. Therefore dynamic computation of currents can be performed much faster without the need for extensive electro-magnetic or FEM solvers. A Quasi-Newton based optimization was then run on this system to achieve the following:

- Generate precise magnetic fields (X, Y, Z components) at specific locations (x, y, z) akin to existing magnetic manipulation systems.
- More importantly, generate precise force components (X, Y, Z components at specific locations.

This project was done together with Alex Pai, in collaboration with City of Hope. A PCB was fabricated on Rogers RT Duroid material with 10 mil thickness and with metal traces on one side forming a cross structure with a total of eight traces (four in each direction). Magnetic beads of diameter 4 μm were used in the experiment where currents of up to 1 A were flown through the copper wires. Gold wirebonds were utilized for the cross structure. Currents were flown through two perpendicular traces at any point of time, ensuring maximum magnetic field at the junction or the cross (current controller was designed by Jeff Sherman). Figure 6.18 shows screenshots for four such example cases, where a cluster of beads are transported across various grid points by dynamically turning ON or OFF different current carrying wires, thereby changing the location of the maximum magnetic field.

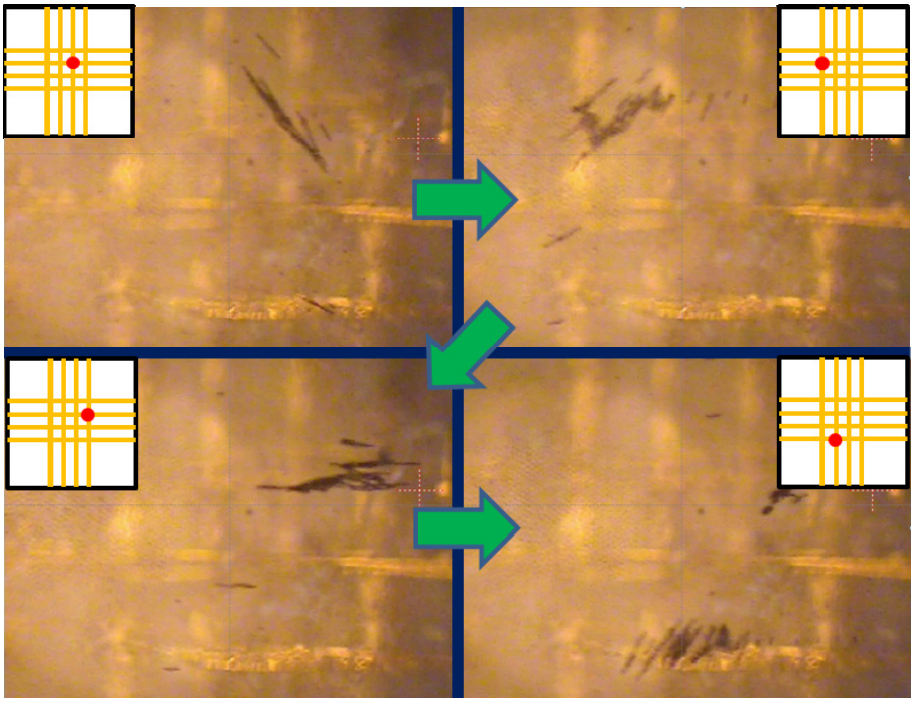


Figure 6.18: Systematic bead movement by dynamically controlling location of maximum magnetic field.

This concept was then extended for cells with magnetic nano-particles inside them for possible applications in targeted drug delivery. Cell based experiments and further experi-

mental validation are being performed.

Appendix A

Synthesizing On-Chip Self-Healing Core: VHDL to Layout

This appendix is meant to serve as a guideline towards enabling digital synthesis and place and route for a standard CMOS process. IBM's 32nm SOI process is used as an example.

A.1 Basic Flow

Figure A.1 shows the step-by-step flow in detail. Note that the flow starts off with verilog, but should work in the same way if the source code is VHDL.

1. We start off with a behavioral verilog code containing input, output, and behavioral description of what the block is supposed to do.

2. Synopsys Design Vision is invoked to synthesize the verilog block. A technology dependent file (usually in .db format) is required in this step. This file contains information on:

   (a) Which standard cells are available for use

   (b) Timing information about these standard cells

   The synthesis tool generates another verilog file which is structural and contains instantiations of logic gates and their connectivities. It also produces a timing constraint file which will be used by the place and route tool.

3. At this point we are ready to place and route. The tool used for this is Cadence Encounter. There is also the option of scripting this entire process. Information or files needed to proceed with this step are:

   (a) Technology dependent .db file same as the one used in Step 2.

   (b) Technology dependent .lef file. This file contains information about pin locations as well as DRC rules for metal routing.

   (c) Technology dependent captable — contains information on metal to metal routing capacitance.

   (d) Timing constraint file (.sdc) generated in Step 2.

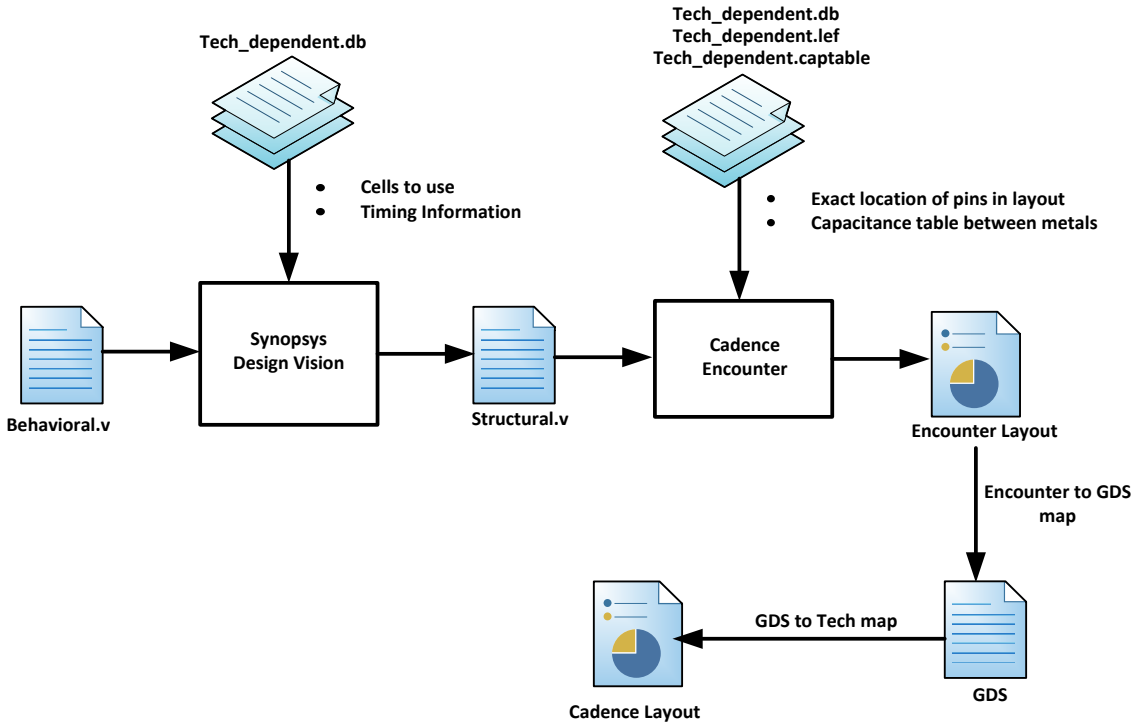


Figure A.1: Basic flow starting from behavioral code all the way to Cadence layout.

The details of Step 3 are outlined later.

4. The next step after place and route is to export the GDS from Encounter to Cadence using two map files.

5. Importing of the schematic is done through the icfb import options and will be outlined in detail later on.

6. With the schematic and the layout properly imported, the design should be ready for LVS.

A.2 Step-by-step Description of the Flow

Notes:

- Most commands used in this description are general commands, but some of them may be specific to the CHIC cluster.

- Some example files are being used in this tutorial, some of them are placeholders for actual technology files which are not available at the time of creation of this document. Details of these files will be discussed in the Appendices.

- The timing library files used in this example are purely for demonstration purposes and do not contain accurate information about timings of the logic gates.

- A very basic set of gates are used in the synthesis (contained in the example files). As more and more gates are added, the synthesis process becomes more efficient both in terms of area as well as timing.

- Functional verification is assumed to be already performed on the verilog/vhdl codes.

- It is assumed that you have a Cadence library containing the standard cells with layouts, schematics, and symbols for each.

The synthesis as well as the place and route tools produce a host of files which may flood your working directory. In practice, the following directory structure works well:

```
/myDesign                       - design root directory
   /verilog                     - original source code directory
   /synth                       - synthesis directory
   /pnr                         - place and route directory.
```

A.2.1 Verilog Example

Let us start off with an example of a 4-to-16 decoder which is coded in verilog behaviorally as follows:

```
module decoder4 (in,out);
input [3:0] in;
output [15:0] out;
reg [15:0] out;
always @ (in)
begin
  out = 0;
  case (in)
    4'b0000 : out = 16'b0000_0000_0000_0001;
    4'b0001 : out = 16'b0000_0000_0000_0010;
    4'b0010 : out = 16'b0000_0000_0000_0100;
    4'b0011 : out = 16'b0000_0000_0000_1000;
    4'b0100 : out = 16'b0000_0000_0001_0000;
    4'b0101 : out = 16'b0000_0000_0010_0000;
    4'b0110 : out = 16'b0000_0000_0100_0000;
    4'b0111 : out = 16'b0000_0000_1000_0000;
    4'b1000 : out = 16'b0000_0001_0000_0000;
    4'b1001 : out = 16'b0000_0010_0000_0000;
    4'b1010 : out = 16'b0000_0100_0000_0000;
    4'b1011 : out = 16'b0000_1000_0000_0000;
    4'b1100 : out = 16'b0001_0000_0000_0000;
    4'b1101 : out = 16'b0010_0000_0000_0000;
    4'b1110 : out = 16'b0100_0000_0000_0000;
    4'b1111 : out = 16'b1000_0000_0000_0000;
  endcase
end
endmodule
```

The file is named *decoder4.v* and should be placed in your \verilog directory.

A.2.2 Synthesis

To synthesize the above code, we need a script file. A sample script file is provided below for reference:

```
# set TOP to the name of the module to be synthesized
set TOP "decoder4"
set edifout_netlist_only "false"

# paths for libraries used in synthesis
set search_path [concat {. /opt/CDS/cadence2005/ic5141/tools.lnx86/dfII/etc/dci/synlibs} -
$search_path]

# paths specific to technology
set target_library [list "/home/kaushikd/Digital_Synthesis/synth/reduced2_typical.db"]
set link_library [list "/home/kaushikd/Digital_Synthesis/synth/reduced2_typical.db"]

# Read in the design
define_design_lib control -path ./CONTROL
analyze -work control -f verilog ../verilog/$TOP.v
elaborate -lib control $TOP

# Define clock (period in ns)  Highlighted for a combinational design
#create_clock clk -name clk -period 1E-9
#set_propagated_clock clk

# Define output load (in pF)
set_load 0.01 [all_outputs]

# Set the optimization constraints
set_max_area 0
set_max_dynamic_power 0

# Map and optimize the design
current_design $TOP
uniquify
```

```
set_flatten true -effort medium -minimize multiple_output -phase true

# Run synthesis
compile -ungroup_all -map_effort high -incremental_mapping

# Run lint on the design
check_design

# Create some reports
redirect [format "%s%s" $TOP ".area"] { report_area }
redirect [format "%s%s" $TOP ".power"] { report_power }
redirect [format "%s%s" $TOP ".timing"] { report_timing -max_paths 25 }
redirect [format "%s%s" $TOP ".clock"] { report_clock }

# write output
write -hierarchy -format verilog -output [format "%s%s" $TOP ".v"]
write -hierarchy -format ddc -output [format "%s%s" $TOP ".ddc"]
```

Save this file as *.scr in the /synth directory. To run synthesis we need to invoke Synopsys Design Vision:

```
$ module load  syn-dc-F-2011.09-SP3
$ cd myDesign/synth
$ design_vision
```

When the Design Compiler GUI starts, use *File ->Execute Script* and point to the .scr file you just created. The bottom command line shows you any warnings or errors encountered in the process. The resulting schematic can be viewed by right-clicking on the module name in the *Logical Hierarchy* and selecting *Schematic View*.

Our *decoder4.v* synthesized using a limited number of gates (will be discussed later) will look like the one in Figure A.2:

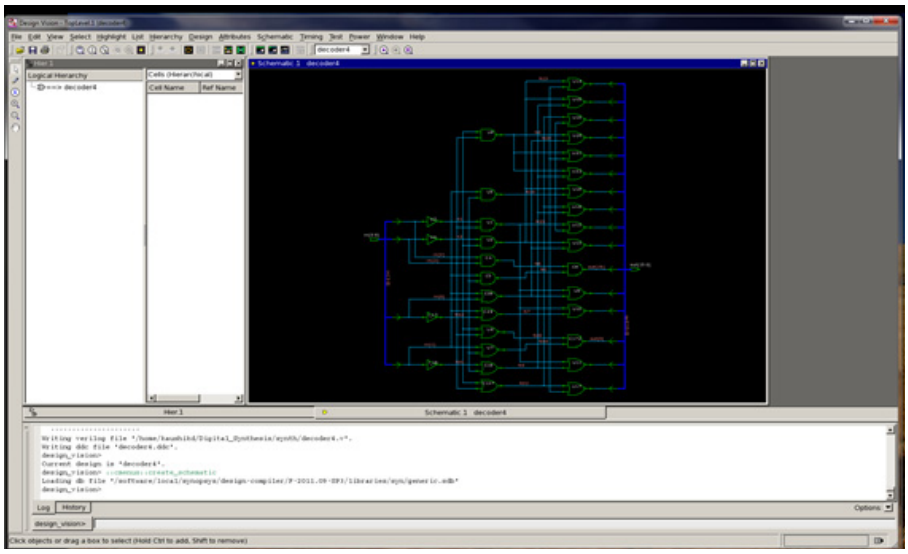


Figure A.2: Synthesized 4-to-16 decoder.

There are a lot of options you can play around with. Particularly useful are the *Timing Analysis*, which tells you the amount by which your design meets the timing margins at the clock frequency specified. Also, for complicated designs *Path Inspector* is a useful tool. We are not going to go into the details of this synthesis step, although for complicated and timing sensitive designs, the tool will tell you a lot about timings you can expect post layout.

The synthesis tools generates a bunch of files. Of particular interest are the */synth/decoder4.v* file (Note that this is not the original .v file you wrote, which is in the */verilog* directory) and the *decoder4.sdc* file, which is a timing constraint file to be fed into the place and route tool. Look into the */synth/decoder4.v* file, note that the synthesis tool has generated a verilog file containing actual logic gates, the ones you have in the Cadence Library.

A.2.3 Place and Route

Once we have the synthesis up and running, the next step is Place and Route. We will use Cadence Encounter to do this. Keep in mind, this is not just a matter of placing the gates and creating interconnections; the tool also takes care of timing while routing, meaning metal-to-metal routing capacitance is taken into account while the tool routes signals. As

a result, depending on the area constraints or the density, it may take several iterations for the tool to converge upon a routing strategy.

To invoke Cadence Encounter:

```
$ cd myDesign/pnr
$ /software/opt/CDS/cadence2007/edi91/tools.lnx86/bin/encounter
```

Use *File->Import Design* and fill up the fields as shown in Figure A.3:

Click the *Advanced* tab and click *Power*. Put *vdd!* in *Power Nets* and *gnd!* in *Ground Nets*. Click *RC Extraction*. Under *Typical Capacitance Table File*, point to the captable for the process technology. Click *OK*. You can also choose to save the configuration for future use by clicking *Save*. The main Encounter window should now look something like in Figure A.4. Note that the input and output ports are all clustered in the bottom left.

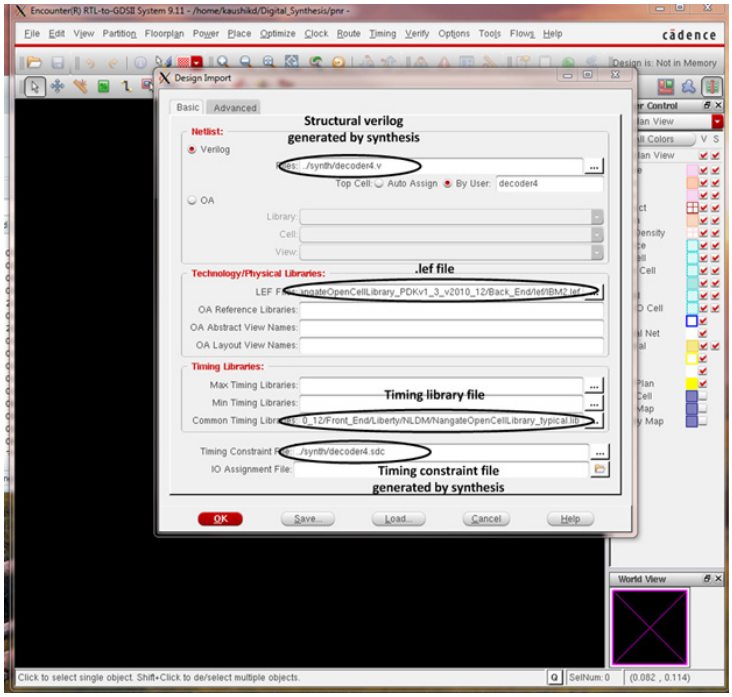


Figure A.3: Importing design to Encounter - 1

We can now specify the floorplan we want including the aspect ratio and other details. To do this, use *Floorplan->Specify Floorplan*. Most of the fields are self-explanatory, for example the aspect ratio specifies the ratio of the height to the width of the design. The

core margins specify how much room is to be left on all sides for power and ground rings. The values in Figure A.5 usually work well for smaller designs, for bigger designs they will need to be changed. The core margins are also dependent on the number of inputs and outputs the design has.

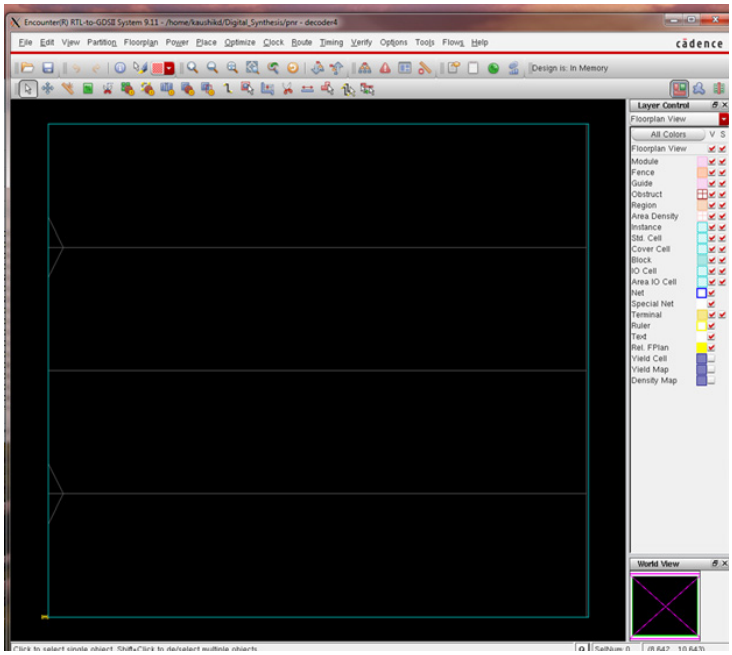


Figure A.4: Main Encounter window after importing design

The final floorplan window will now look like Figure A.5.

Figure A.5: Specifying floorplan details for the design.

Power and ground rings will be added now using *Power->Power Planning->Add Ring*.

Note that in this example we are using M5 and M4 as power metals. Depending on the size of the design, higher or lower metal levels may be used. Click the *Via Generation* tab. At the time of creation of this document, the metal layers are read from a template .lef file. In IBM 32nm SOI, the higher metal levels are not named as Mx. As a result, for simplicity, we are going to use only metals 1 through 5 for the current design. Select metal5 as the Top stack via layer and metal1 as the bottom stack via layer. The window should now look like Figure A.6.

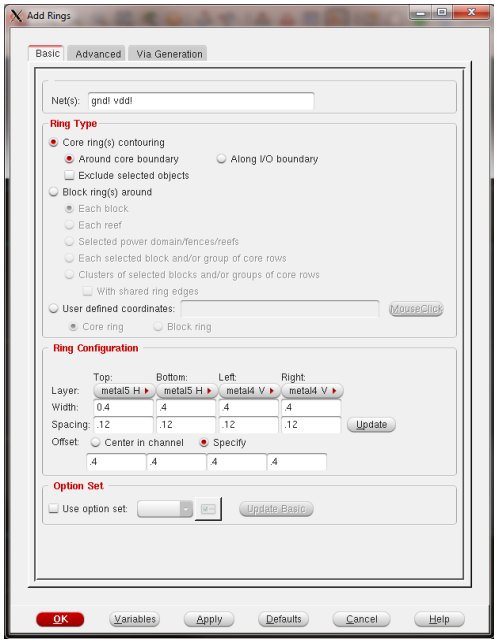


Figure A.6: Power ring settings in detail.

Press *OK*. You can see that the power rings have been added around the periphery. Now we need to add power stripes to create some sort of a power grid. This is particularly useful for bigger design. The number of stripes as well as their spacing can be specified. Click *Power->Power Planning->Add Stripe* and follow sample settings in Figure A.7. Again, click Via generation and select metal5 as the top metal layer and metal1 as the bottom metal layer. After adding the stripes and the ring, the encounter window should look like in Figure A.8.

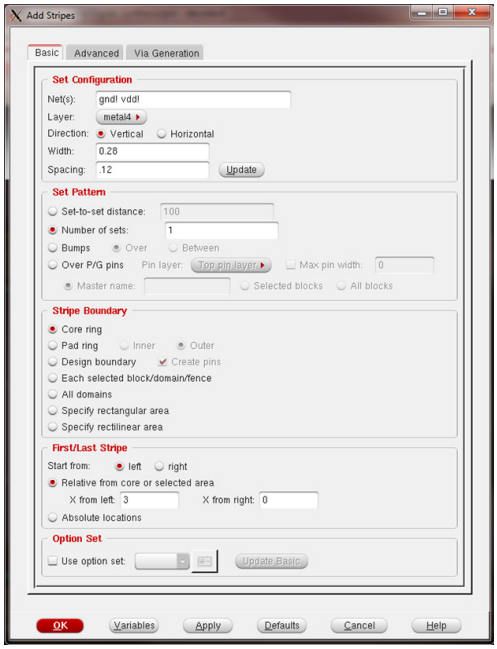


Figure A.7: Adding power stripes to the design.

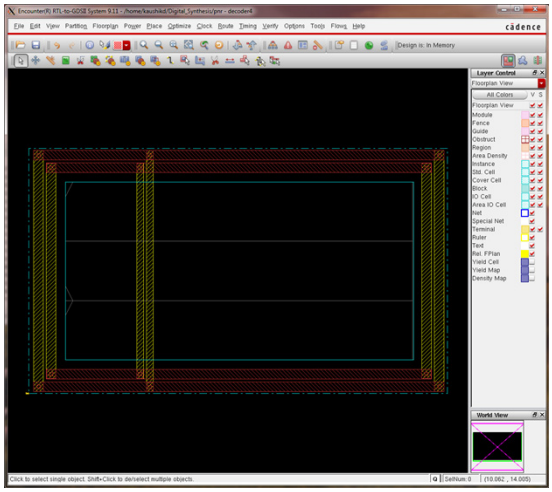


Figure A.8: Encounter window after adding power ring and stripes.

Now we can specify locations of the I/O pins. To do this go to *Edit->Pin Editor*. Finish assigning pin locations as shown in Figure A.9. Note that multiple I/O pins can be assigned by selecting them through shift+click. You should see the pins being spaced out along the edge you specified in this step. If there are too many pins along one edge, the tool pops up a

warning if it violates DRC. Now we are ready to place the standard cells. Click *Place->Place Standard Cell*.

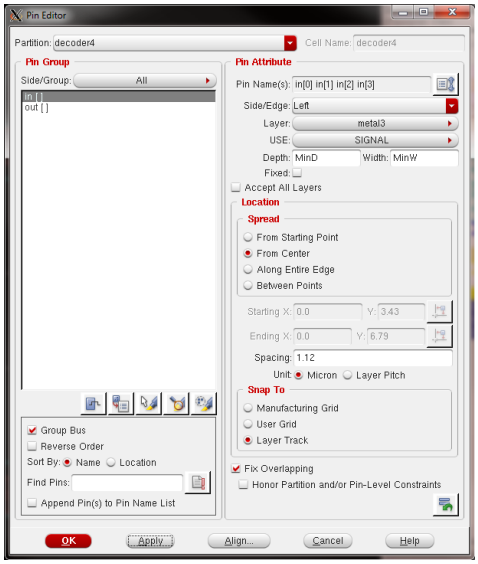


Figure A.9: Specifying pin locations and spacing.

Open up a detailed option box using the *Mode* button. Since we have already placed the pins, uncheck the *Place IO Pins* box. Make sure *Run Timing Driven Placement* is checked. Remember, keep checking the terminal window for warnings or errors that may have occurred. After running the placement, click the button shown in Figure A.10.

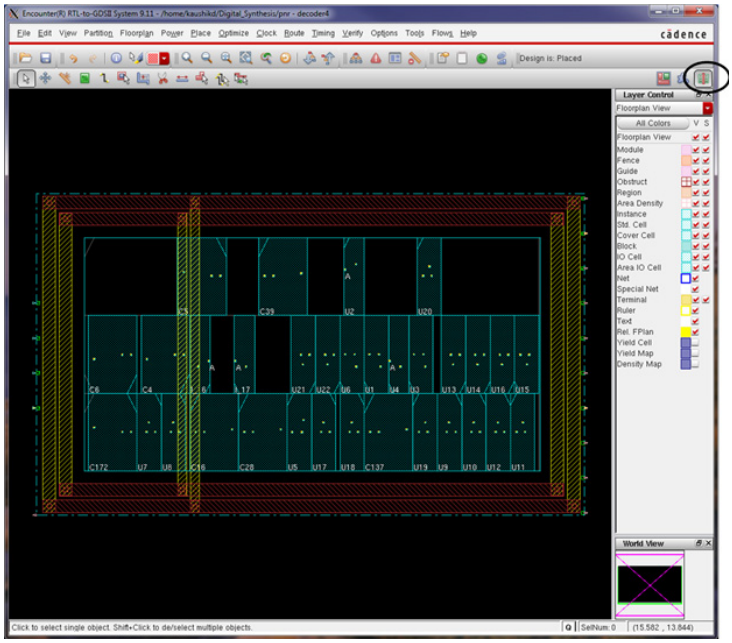


Figure A.10: All standard cells placed in the design.

We will now hook up the power and ground nodes of the standard cells. For this, *Power->Connect Global Nets*. For the standard cells in our present design, the ground and supply connections are named VDD & VSS. Under *Connect*, select *Pin* and enter the standard cell pin name, then enter the corresponding global net name (e.g., vdd and gnd) in *To Global Net*. Also, we need to specify *Tie Hi & Tie Low* options under *Connect*, selecting *Tie High* (or *Tie Low*) instead of *Pin*, making sure that the *Pin Name*(s) field is empty, then entering the corresponding global net name under *To Global Net*. These settings are shown in Figure A.11.

Once this step is done, go to *Route->Special Route*. Again, change the *Top Layer* option to metal5. And in the *Via Generation* tab, specify the top stack via layer as metal5 at the two places. Click *OK* and you should see the supply connections routed for each cell. For bigger designs, we need to design a clock tree at this step, but for simplicity we are skipping this step.

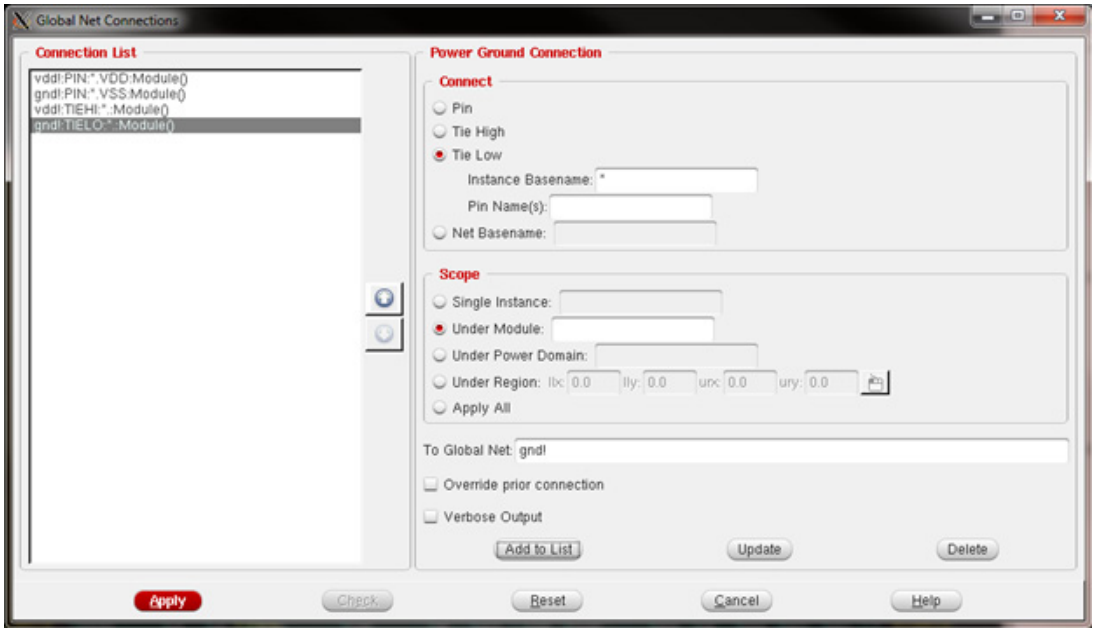


Figure A.11: Adding global net connections.

We are now ready to do a detailed route. Select *Route->NanoRoute->Route*. Select *Mode*, change the routing metal layers to metal5 and metal2. Note that in our standard cells, all the pins are on metal2. So we are making sure the tool does not route using metal1. This is to prevent possible occurrences of DRC errors within the cell, once the tool routes using metal1. You can set multiple options here in the various tabs. These govern how strict the routing will be, how many vias per connections, antenna violations, etc. One important option is under the *AdvDRC* tab, where you check the *Enclose Via Completely in Standard Cell Pin*. This prevents a lot of DRC errors due to vias sticking out of small metal islands. Set the number of cores and start the process by clocking *OK*. The tool will take several iterations to converge. It will show an X box where it thinks there is a DRC error. The encounter window now should look something like in Figure A.12.

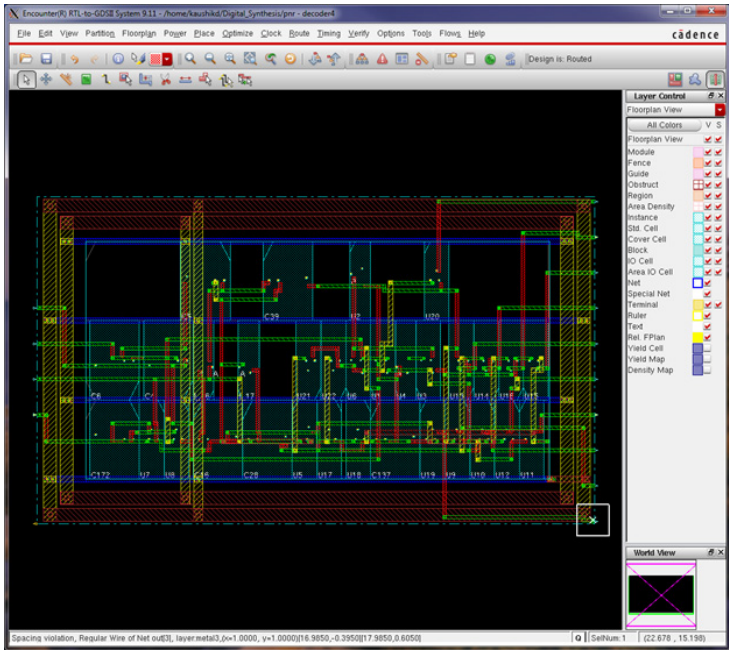


Figure A.12: After nanoroute.

You may want to add filler cells at this point, buy choosing *Place->Physical Cell->Add Filler*. Select the filler cells and they should be placed in the vacant spaces.

We are now ready to export the GDS to Cadence. *File->Save->GDS*. Leave the options default except the map file. This is a custom encounter map file not to be confused with the normal gds map file you use with virtuoso. For IBM32nm SOI, the map file is in

```
/home/kaushikd/Digital_Synthesis/dno/NangateOpenCellLibrary_PDKv1_3_v2010_12/custom-
Encounter_ibm.map
```

or

```
/home/kaushikd/Digital_Synthesis/dno/NangateOpenCellLibrary_PDKv1_3_v2010_12/custom-
Encounter_ibm_mod.map
```

You also need to export the schematic netlist using the following in the encounter command line:

```
saveNetlist decoder4.pnr.v -excludeLeafCell
```

Now we can import the layout and the schematic into virtuoso.

A.2.4 Importing GDS and Netlist

1. Once you have exited Encounter, you should now import the GDS first.

2. For this, make a local copy of the Cadence Library, which contains the standard cells you used. This ensures every digital directory is self contained for every design. You may choose not to do this as well.

3. In the ICFB Window, import Stream and point to the GDS file, the newly copied Library and the map file (discussed in the previous section).

4. To import the netlist, use the ICFB Window and Import>Verilog. Import the schematic under the name used in the saveNetlist command used in Encounter

A.3 Generating Timing Information for Standard Cells

Timing information for the standard cells is critical in high-performance digital design, where each gate contains delay and power information for all possible inputs and outputs which the synthesis tool utilizes to synthesize a design. The typical format of the file is a *.lib file which can easily be converted to a database file *.db.

The tool which performs this characterization is Encounter Library Characterizer which is part of Cadence Encounter Timing Systems. The tool performs Spectre or HSPICE simulations on each cell and outputs the date in the library format. Inputs to the tool are:

1. Spectre netlist : *.scs is a file which contains instances of all the standard cells but without any connection.

2. Spectre model files or HSPICE model files for simulations.

3. Gate functionality : normally the tool generates this information directly, if not, we need to manually modify the generated gate file.

4. Configuration file : contains simulation settings — tolerances, etc, output format (ECSM power or timing, etc) and which standard cells to characterize, etc.

5. Simulation setup file : High and low voltage levels, threshold voltages for different corners, FF, SS, TT, rise and fall times.

Depending on the process, the tool may or may not recognize the functionality of the gates. In that case, we need a gate.org file specifying inputs and outputs and functionality. A sample gate.org file is provided below:

```
DESIGN ( AND2_X1 );
// =================
//  PORT DEFINITION
// =================
SUPPLY0 VSS ( VSS );
SUPPLY1 VDD ( VDD );
INPUT A1 ( A1 );
INPUT A2 ( A2 );
OUTPUT ZN ( ZN );
// ===========
//  INSTANCES
// ===========
AND ( ZN, A1, A2 );
END_OF_DESIGN;

DESIGN ( INV_X1 );
// =================
//  PORT DEFINITION
// =================
SUPPLY0 VSS ( VSS );
SUPPLY1 VDD ( VDD );
INPUT A ( A );
OUTPUT ZN ( ZN );
// ===========
//  INSTANCES
// ===========
NOT ( ZN, A );
END_OF_DESIGN;

DESIGN ( NAND2_X1 );
// =================
```

```
//  PORT DEFINITION
// =================
SUPPLY0 VSS ( VSS );
SUPPLY1 VDD ( VDD );
INPUT A1 ( A1 );
INPUT A2 ( A2 );
OUTPUT ZN ( ZN );
// ===========
//  INSTANCES
// ===========
NAND ( ZN, A1, A2 );
END_OF_DESIGN;

DESIGN ( NOR2_X1 );
// =================
//  PORT DEFINITION
// =================
SUPPLY0 VSS ( VSS );
SUPPLY1 VDD ( VDD );
INPUT A1 ( A1 );
INPUT A2 ( A2 );
OUTPUT ZN ( ZN );
// ===========
//  INSTANCES
// ===========
NOR ( ZN, A1, A2 );
END_OF_DESIGN;

DESIGN ( OR2_X1 );
// =================
//  PORT DEFINITION
// =================
SUPPLY0 VSS ( VSS );
SUPPLY1 VDD ( VDD );
INPUT A1 ( A1 );
INPUT A2 ( A2 );
OUTPUT ZN ( ZN );
// ===========
```

```
//  INSTANCES
// ===========
OR ( ZN, A1, A2 );
END_OF_DESIGN;
```

Types of simulation information generated by Encounter Library Characterizer:

1. Pin-to-pin delay with state dependency.
2. Input setup/hold constraints.
3. Power consumption.
4. Input and output pin capacitance.
5. ECSM (effective current source model) Timing
   (a) Voltage waveform at output pin.
   (b) Multi-piece ECSM capacitance.
6. ECSM Power
   (a) Current waveform at power-grid pins for different combinations of slew and load.
7. ECSM Noise
   (a) Voltage in Voltage out waveform on an input or output pin of a cell.
8. Statistical ECSM
   (a) Sensitivities to device parameters at arc level for all load, slew, delay, waveform and timing checks.

A sample ELC configuration file is provided below.

```
EC_SIM_USE_LSF="1";
EC_SIM_LSF_CMD="";
EC_SIM_LSF_PARALLEL="1";
EC_SIM_NAME="spectre +csfe";
EC_SIM_TYPE="SPECTRE";
#EC_IGN_XTG=1;
#EC_REDUCT_DELAY_FLAG=0;

#SYNLIB="/home/kaushikd/NangateOpenCellLibrary_typical_ecsm.lib";

EC_SPICE_SIMPLIFY="1";
EC_CHAR="ECSM-TIMING";
#EC_ECSM_NUM=15;

#skips checking the existance of nodes in the subcricuit file while reading gate file

EC_SIM_SUPPLY0_NAMES="VSS";
EC_SIM_SUPPLY1_NAMES="VDD";

# Transient simulation options (set EC_CASE_SENSITIVITY=1 for spectre)
#EC_TR_OPTIONS=accurate;
#EC_TRAN_OPTIONS="errorpreset=conservative";

# Improves analysis of sequential circuits (FFs)
#EC_HALF_WIDTH_HOLD_FLAG=1;

# Specifies the spice subcircuit file
SUBCKT="/home/kaushikd/Digital_Synthesis/timing/all_cells3.sp";

# The model file name
#MODEL="/home/kaushikd/Digital_Synthesis/timing/models/design.scs";
MODEL="/home/kaushikd/Digital_Synthesis/timing/HSPICE/models/design.inc";
DESIGNS= "INV_X1 NAND2_X1 NOR2_X1 AND2_X1 DFFR_X3 OR2_X1";

# Specify the simulation setup file -- lots of params in this file that determine
SETUP="/home/kaushikd/Digital_Synthesis/timing/elc_simulation.setup";
#SETUP="setup_file";
```

```
# Specify the process corner in the simulation setup file.
PROCESS="typical";

# These two used together. LIB is the name of the spice library file, while CORNER
# is the spice corner in the library file.
#LIB="xxx.lib";
#CORNER="ss";

# Name of the subcircuit that is hierarchical and needs to be expanded. Mandatory
# for hierarchical design.
#EXPAND="cell_name";

# The name of the statistical configuration file. See user guide for description
#STAT_CONFIG="stat.cfg";

# For incremental characterization -- .lib file to use
#SYNLIB="incr.lib";

# BOOL file that describes macros and their corresponding functions.
#BOOL="bool_file"
```

The simulation setup files dictates the transient simulation parameters like duration, step size, as well a corner threshold voltage data and gate output load capacitances. A sample file is provided below. Information about best and worst case threshold voltages can be found in the design manual.

```
// default setup defines for SignalStorm
// 2000 - 2001 (c) Simplex Solutions, Inc
// Version : 1.2 March, 2001
// Example given at end of Appendix A in doc.
// Sets the levels of the input or output signals for generating the simulation input
// waveforms, initializing the output voltage level, and measuring the simulation waveforms
// for delay and power in cell library characterization
Signal  DEFAULT_SIGNAL {
unit = REL ;
vh   = 1.0 1.0 ;
vl   = 0.0 0.0 ;
vth  = 0.5 0.5 ;
```

```
vsh  = 0.9 0.9 ;
vsl  = 0.1 0.1 ;
tsmax= 3n    ;
} ;
Simulation DEFAULT_SIMULATION {
transient = 10n 50n  1p  ;
dc        = 0.1  0.9  0.1  ;
bisec     = 8.0n 8.0n 1p  ;
incir     = "" ;
outcir    = "" ;
resistance = 10000K ;
} ;
Index      DEFAULT_INDEX {
slew = 0.1n  0.4n  0.8n  1.5n  2.0n  3.0n;
load = 0.1p  0.2p  0.4p  0.6p  0.8p  1.0p;
} ;
margin DEFAULT_MARGIN {
 cap    = 1.0 0.0 ;  // gate cap
wcap    = 1.0 0.0 ;  // wire cap
wresist = 1.0 0.0 ;  // wire resistance
delay   = 1.0 0.0 ;  // cell delay

ecap    = 1.0:1.0:1.0 0.0:0.0:0.0 ;  // effective cap
power   = 1.0:1.0:1.0 0.0:0.0:0.0 ;
current = 1.0:1.0:1.0 0.0:0.0:0.0 ;
slew    = 1.0:1.0:1.0 0.0:0.0:0.0 ;
interconnect  = 1.0:1.0:1.0 0.0:0.0:0.0 ;
iopath  = 1.0:1.0:1.0 0.0:0.0:0.0 ;
setup   = 1.0:1.0:1.0 0.0:0.0:0.0 ;
hold    = 1.0:1.0:1.0 0.0:0.0:0.0 ;
release = 1.0:1.0:1.0 0.0:0.0:0.0 ;
removal = 1.0:1.0:1.0 0.0:0.0:0.0 ;
recovery= 1.0:1.0:1.0 0.0:0.0:0.0 ;
width   = 1.0:1.0:1.0 0.0:0.0:0.0 ;
 } ;
// Nominal factors to calculate typical (average) values in cell characterization.
nominal DEFAULT_NOMINAL {
cap     = 0.0:0.5:1.0 0.0:0.5:1.0 ;  // rise / fall
```

```
check   = 1.0:1.0:1.0 1.0:1.0:1.0 ;  // rise / fall
current = 0.0:0.5:1.0 0.0:0.5:1.0 ;  // rise / fall
power   = 0.0:0.5:1.0 0.0:0.5:1.0 ;  // rise / fall
slew    = 0.0:0.5:1.0 0.0:0.5:1.0 ;  // rise / fall
delay   = 0.0:0.5:1.0 0.0:0.5:1.0 0.0:0.5:1.0;  // rise / fall / Z
 } ;
Load       DEFAULT_LOAD {
pin(*)  =  0.001p 0.001p;
} ;
slew    DEFAULT_SLEW {
pin(*)  =  0.01n 0.01n;
} ;
vth    DEFAULT_VTH  {
pin(*)  =  0.3 0.3;
} ;
// Vtn and Vtp voltages obtained from design manual (Regular-Vt, Electrical parameters -
section)
Process typical {
        voltage = 1.00 ;
        temp = 25 ;
        Corner = "TT" ;
// Used to characterize 0->Z and 1->Z arcs.
        Vtn = 0.319 ;
        Vtp = 0.332 ;
} ;
Process best {
        voltage = 1.00 ;
        temp = 0 ;
        Corner = "FF" ;
        Vtn = 0.286 ;
        Vtp = 0.282 ;
} ;
Process worst {
        voltage = 1.00 ;
        temp = 125 ;
        Corner = "SS" ;
        Vtn = 0.396 ;
        Vtp = 0.381 ;
```

```
} ;
set process(typical,best,worst){
simulation = DEFAULT_SIMULATION;
signal = DEFAULT_SIGNAL;
margin = DEFAULT_MARGIN;
nominal = DEFAULT_NOMINAL;
index = DEFAULT_INDEX;
} ;
```

A.4 Specific Steps for IBM's 32nm SOI CMOS Process

These steps are based on co-integration of synthesized block with main circuit. Some of these points are specific to IBM32nm SOI CMOS process.

1. All standard cells have VDD! and VSS! as global nets for supply and ground respectively.

2. There should not be explicit pins for these global nets in the symbol or cell description.

A.4.1 Place and Route

1. Importing the design

   (a) As shown in Figure A.13 *File ->Import Design*

   (b) Switch to the *Advanced* tab. Select *Power* and put in VDD! and VSS! for *Power Nets* and *Ground Nets* respectively.

   (c) Select *RC Extraction ->Typical Capacitance Table* file. Point to

   `/home/kaushikd/32soi_12LB_5L1x_..._2T16X_nm.CapTbl`

2. Specifying floorplan

   (a) Click *Floorplan ->Specify Floorplan*

   (b) You can specify the floorplan by aspect ratio or actual size.

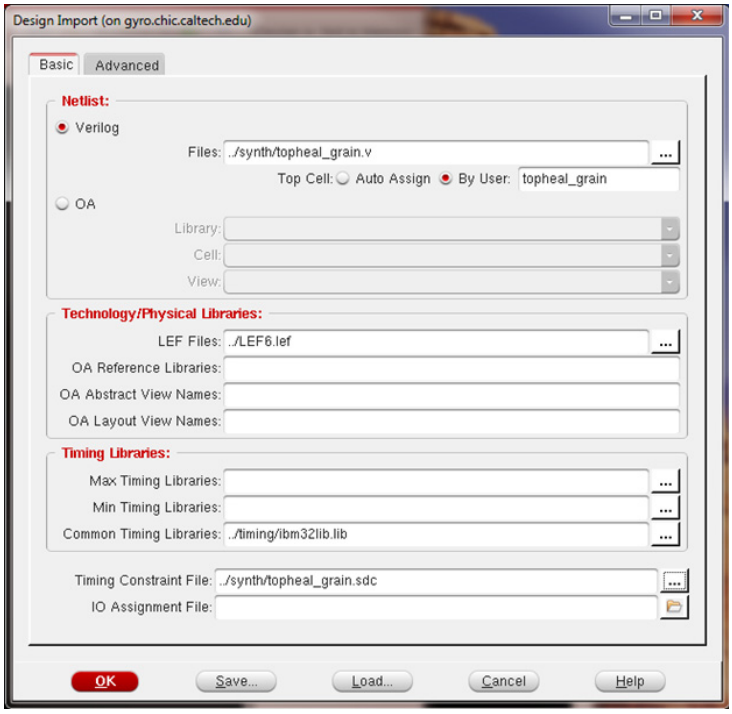


Figure A.13: Importing design in IBM32nm.

(c) Select a big enough Core Boundary and low enough Core Utilization to spread out your design if area is not that big of a concern.

3. Adding Power Rings

   (a) Click *Power ->Add Rings*

   (b) Add power rings on metals M5 and M4 as explained in the sections above.

   (c) For big designs you may need to add additional Power Stripes

4. Adding Pins

   (a) Click *Edit ->Pin Editor*

   (b) Place all the pins at the required locations using the Pin Editor.

5. Placing Standard Cells

   (a) Click *Place ->Place Standard Cell*

(b) Select *Mode* and check the following:

i. *Congrestion Effort* - Low

ii. Uncheck *Place IO Pins*

iii. Check *Specity Maximum Density* and select a low number 0.2 (20% fill factor while placing cells)

(c) Place the standard cells now and check if they are spaced out enough. If not, then the routing may not be complete and this will lead to wrong designs being routed.

6. Connecting Global Nets

(a) Click *Power ->Connect Global Nets*

i. Tie all Pins named VDD! to VDD! net and all Pins named VSS! to VSS! net

ii. Tie High to VDD! net

iii. Tie Low to VSS! net

iv. Your Connection List on the left should now have 4 entries. Press *Apply* after highlighting each entry in the *Connection List*.

7. Special Route

(a) Click *Route ->Special Route*

(b) Select Top Layer as M5 and bottom layer as M1

(c) Change the Crossover and Target Connection Layer Ranges in the *Via Generation* tab also to M5 and M1

(d) Press *OK* to see the VDD! and VSS! nets connected

8. Nano Route

(a) Click *Route ->Nano Route*

(b) Top Routing Layer M5, Bottom Routing Layer M2 [**If you select M1, there may be some additional DRC and conflicts since the cells have internal routing at M1]

(c) Uncheck *Automatically Stop If There Are Too Many Violations*

(d) In the *Antenna* tab, uncheck *Repair Process Antenna Violations* , uncheck *Delete and reroute nets with violations during advanced search and repair*

(e) In the *AdvDRC* tab, check *Enclose Via completely in Standard Cell Pin*

(f) In the *Misc* tab, increase the *Maximum Number of Warning Messages to Report* to a high value

(g) Press *OK* and get back to the original dialogue box

(h) At the *End Iteration* field, enter a value >40 to ensure complete routing

(i) Press *OK* and Encounter will try and route the design. Always check the terminal. If you see a line like "Routing stopped due to too many Violations", it means your design has not been fully routed. You should then go back and place cells in a less congested way so that routing succeeds.

** Note that even though the final messages after NanoRoute has completed may say "Total Number of Fails" as 0, it may not mean your design was fully routed. Scroll up to see more messages and look for violations or stopped routing related messages. If you do not see this, it means you are all set for LVS.

Now we need to export the layout as a GDS and the netlist as a verilog Netlist. Steps are already discussed in the previous section.

A.5 Measurement Results

With the steps outlined in the previous sections, a test chip was fabricated in IBM's 32nm SOI CMOS process. The Verilog code was synthesized for a half-rate serial to parallel converter with bit redundancy checking. Figure A.14 shows the die photo of the designed chip.

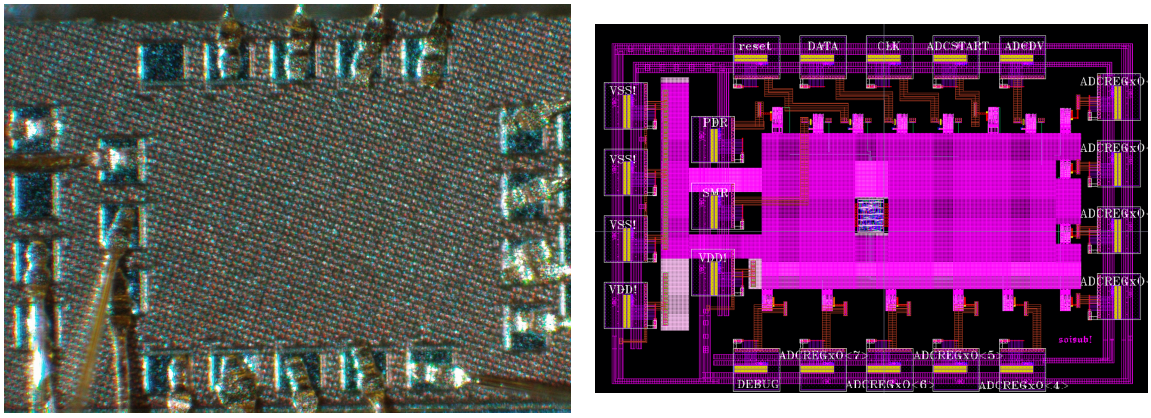


Figure A.14: Die-photo of the fabricated chip.

The chip was functionally verified with the Altera DE2 board with the in-built FPGA up to a clock rate of 125 MHz, limited by the FPGA itself. The initial handshaking between the Altera board and the chip is shown in Figure A.15.

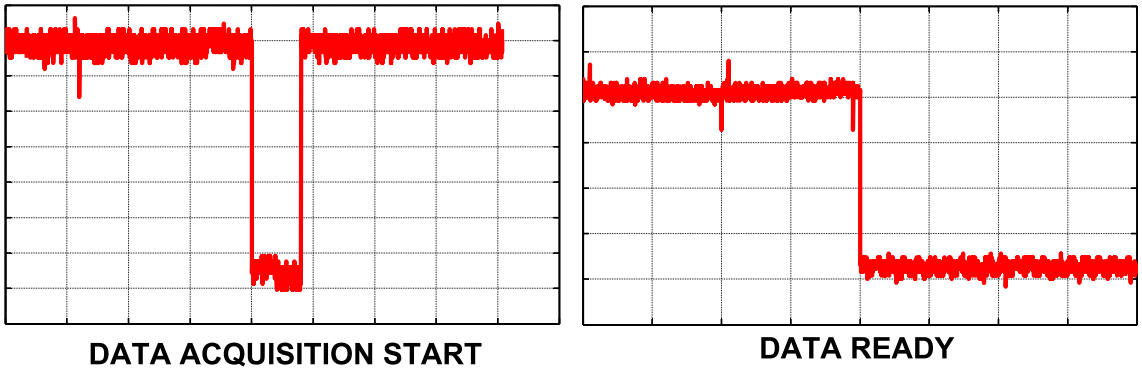


Figure A.15: Measured voltage waveforms showing handshaking between Altera board and chip.

Figure A.16 show measured voltage waveforms at the parallel data outputs.

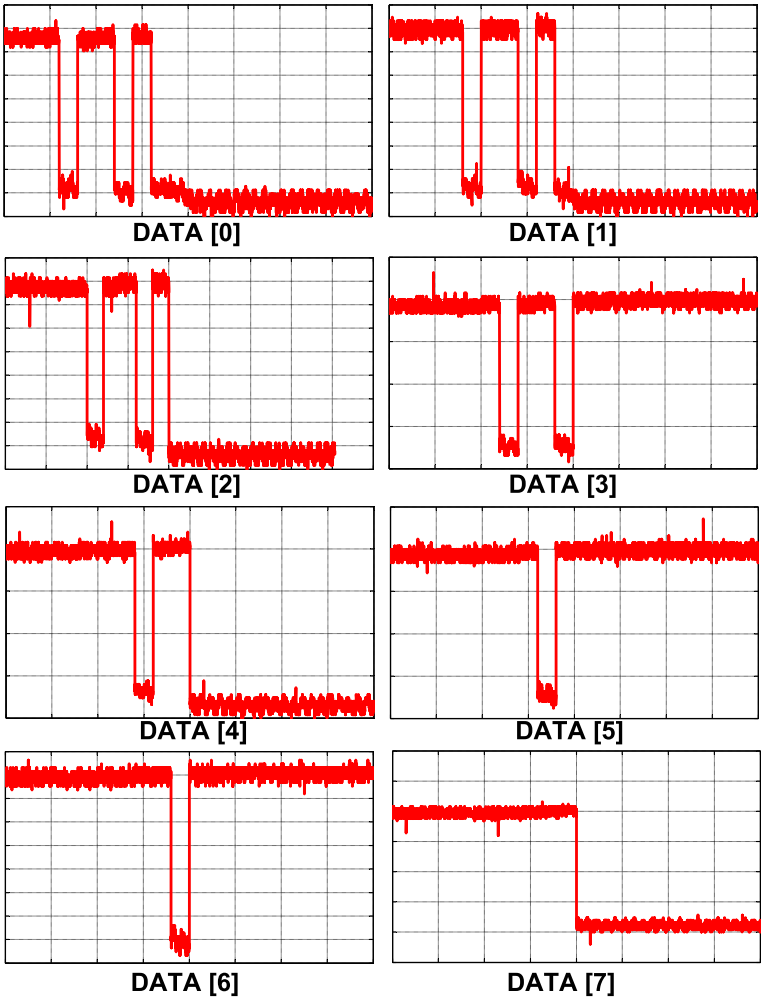


Figure A.16: Measured voltage waveforms showing parallel data bits getting set.

A.6 Scripting the Encounter Place and Route

Scripting utilizes three files fe.tcl(main scripting file), env.tcl (sets relevant paths), and encounter.conf (loads the verilog, timing library, timing constraints, and captable).

The following examples were used in our March 2014 tapeout, and located in /home/apai/Digital_Synthesis2/scripts:

Example fe_topheal.tcl

```
### This is a comment.
# Note that some of the commands are commented below and may be used
```

```
# based on the design to be placed and routed.

#set the relevant paths
     source env.tcl

#load the verilog, timing library, timing constraints, captable
     loadConfig $designDir/encounter_topheal.conf

#load the floorplan information
# if you do not specify the floorplan information, the tool
# will itself determine the die size based on the design read.
# Many times designer will like to specify the routing blockages,
# placement blockages etc. in the floorplan to meet their specific requirements.
# Also, Design may also include soft IP block, blackboxes and hard blocks.
# Their location and orientations are usually specified by the designers manually
# in the floorplan.
# Power routing structures are also included in the floorplan file as .spr file
#loadFPlan $designDir/s35932.fp

#size by ratio/core utilization
#floorplan -r 1.0 0.7 6 6 6 6

#size by exact dimension
     floorplan -d 110 110 6 6 6 6

addRing -spacing_bottom 0.5 -width_left 1 -width_bottom 1 -width_top 1 -spacing_top 0.5
-layer_bottom M5 -width_right 1 -offset_bottom 0.4
-offset_top 0.4 -around core -layer_top M5 -spacing_right 0.5 -offset_left 0.4 -offset_right 0.4
-spacing_left 0.5 -layer_right M4 -layer_left M4
-stacked_via_bottom_layer M1 -stacked_via_top_layer M5 -nets { gnd! vdd! }

#addStripe  -set_to_set_distance 100 -spacing 5 -xleft_offset 50 -layer metal4 -width 5 -nets
{ gnd vdd }
#set place mode, if needed
# Parameters: -lowEffort, -mediumEffort, -fp, -highEffort, [-timingdriven][-noTimingDriven]
[-assignIoPins]
#setPlaceMode -lowEffort
setPlaceMode -mediumEffort -fp false -timingDriven false -placeIoPins false
```

```
# do placement
placeDesign

#save placed design
saveDesign $outDir/topheal.placed.enc
defOut -placement -routing -floorplan -netlist $outDir/placed.def

#globalNetConnect vdd! -type tiehi
#globalNetConnect vdd! -type pgpin -pin VDD -override
#globalNetConnect gnd! -type tielo
#globalNetConnect gnd! -type pgpin -pin VSS -override

#check timing now.
# The tool will run a fast router internally to quickly estimate the
# routing and the timing. This is crucial to get an estimate of the design
# feasability by quickly estimating the wire length and topology early
#in the design cycle
editPin -cell topheal -pin {ADCREG[0] ADCREG[1] ADCREG[2] ADCREG[3] ADCREG[4]
ADCREG[5] ADCREG[6] ADCREG[7]  PDR ADCSMR}
-spreadType CENTER -side LEFT -spacing 2
editPin -cell topheal -pin {CLK reset LOAD GRAB ADCLOAD EN IADCSMR DOUT DONE
THRU DIN} -spreadType CENTER -side TOP -spacing 8
editPin -cell topheal -pin { B1[0] B1[1] B1[2] B1[3] B1[4] B2[0] B2[1] B2[2] B2[3] B2[4]
 B4[0] B4[1] B4[2] B4[3] B4[4] A4[4] A4[3] A4[2]
A4[1] A4[0] A2[4] A2[3] A2[2] A2[1] A2[0] A1[4] A1[3] A1[2] A1[1] A1[0] ADCMUX}
-spreadType CENTER -side BOTTOM -spacing 3
#timeDesign -preCTS
sroute -crossoverViaBottomLayer M1 -crossoverViaTopLayer M5
setNanoRouteMode -routeWithTimingDriven false -drouteFixAntenna false -routeBottomRoutingLayer
2 -routeTopRoutingLayer 5 -routeWithViaInPin true
#setNanoRouteMode -routeWithTimingDriven false -drouteFixAntenna false -routeWithViaInPin true
globalDetailRoute
#report_timing after_place.timing.viol.rpt

# To get routing reports use
#reportRoute
#reportWire -detail $outDir/wire_report.txt 1.0
```

```
#reportCongestArea -outfile $outDir/congestion.rpt

# Check the sanity of the routes
# you will find that there may be short/open violations at this point
# as the fast router (but quality may not be good) was used to do quick routing
# and estimate the wire length and topology  at this point in the design.
#verifyGeometry -report $outDir/after_initial_route_viol.rpt

# try optimizing the design at this stage to see the chances of
# converging the design
#optDesign -preCTS

#report_timing after_opt.timing.rpt

# If the timing reports are satifactory, then we can proceed to do the
# final routing with sign-off router. This is done using Nanoroute
# integrated with SocEncounter. Please note that this is one of the market
# leaders in this space.

# use sign-off router now
#globalDetailRoute

#save routed design
#saveDesign $outDir/SARADC_SM3.routed.enc
#defOut -placement -routing -floorplan -netlist $outDir/routed.def

#check timing
#timeDesign -postRoute

# if there are still timing violations in the design
# you can run the below command
#optDesign -postRoute -si

#check connectivity
#verifyConnectivity -report $outDir/final_conn.rpt

#check geometry violations
#verifyGeometry -report $outDir/final_geom.rpt
```

```
saveNetlist topheal.pnr.v -excludeLeafCell
streamOut /home/apai/Digital_Synthesis2/scripts/topheal.gds -mapFile /home/apai/
Digital_Synthesis2/
dno/NangateOpenCellLibrary_PDKv1_3_v2010_12/
customEncounter_ibm_mod.map -libName DesignLib -structureName topheal -stripes 1
-units 1000 -mode ALL
exit
```

Example encounter_topheal.conf

```
#################################################################
#  Generated by:   Cadence Encounter 09.11-s084_1
#  OS:             Linux x86_64(Host ID gyro.chic.caltech.edu)
#  Generated on:   Mon Mar 18 03:51:09 2013
#################################################################
global rda_Input
set cwd /home/apai/Digital_Synthesis2/pnr
set rda_Input(import_mode) {-treatUndefinedCellAsBbox 0 -keepEmptyModule 1 }
set rda_Input(ui_netlist) "../synth/topheal.v"
set rda_Input(ui_netlisttype) {Verilog}
set rda_Input(ui_rtllist) ""
set rda_Input(ui_ilmdir) ""
set rda_Input(ui_ilmlist) ""
set rda_Input(ui_ilmspef) ""
set rda_Input(ui_settop) {1}
set rda_Input(ui_topcell) {topheal}
set rda_Input(ui_celllib) ""
set rda_Input(ui_iolib) ""
set rda_Input(ui_areaiolib) ""
set rda_Input(ui_blklib) ""
set rda_Input(ui_kboxlib) ""
set rda_Input(ui_gds_file) ""
set rda_Input(ui_oa_oa2lefversion) {}
set rda_Input(ui_view_definition_file) ""
set rda_Input(ui_timelib,min) ""
set rda_Input(ui_timelib,max) ""
set rda_Input(ui_timelib) "/home/apai/Digital_Synthesis2/timing/ibm32lib.lib"
set rda_Input(ui_smodDef) ""
set rda_Input(ui_smodData) ""
```

```
set rda_Input(ui_locvlib) ""
set rda_Input(ui_dpath) ""
set rda_Input(ui_tech_file) ""
set rda_Input(ui_io_file) ""
set rda_Input(ui_timingcon_file,full) ""
set rda_Input(ui_timingcon_file) "../synth/topheal.sdc"
set rda_Input(ui_latency_file) ""
set rda_Input(ui_scheduling_file) ""
set rda_Input(ui_buf_footprint) {}
set rda_Input(ui_delay_footprint) {}
set rda_Input(ui_inv_footprint) {}
set rda_Input(ui_leffile) "/home/apai/Digital_Synthesis2/LEF6.lef"
set rda_Input(ui_cts_cell_footprint) {}
set rda_Input(ui_cts_cell_list) {}
set rda_Input(ui_core_cntl) {aspect}
set rda_Input(ui_aspect_ratio) {1.0}
set rda_Input(ui_core_util) {0.7}
set rda_Input(ui_core_height) {}
set rda_Input(ui_core_width) {}
set rda_Input(ui_core_to_left) {}
set rda_Input(ui_core_to_right) {}
set rda_Input(ui_core_to_top) {}
set rda_Input(ui_core_to_bottom) {}
set rda_Input(ui_max_io_height) {0}
set rda_Input(ui_row_height) {}
set rda_Input(ui_isHorTrackHalfPitch) {0}
set rda_Input(ui_isVerTrackHalfPitch) {1}
set rda_Input(ui_ioOri) {R0}
set rda_Input(ui_isOrigCenter) {0}
set rda_Input(ui_isVerticalRow) {0}
set rda_Input(ui_exc_net) ""
set rda_Input(ui_delay_limit) {1000}
set rda_Input(ui_net_delay) {1000.0ps}
set rda_Input(ui_net_load) {0.5pf}
set rda_Input(ui_in_tran_delay) {0.0ps}
set rda_Input(ui_captbl_file) "/home/kaushikd/32soi_12LB_5L1x_3L2x_1L4x_2T16x_nm.CapTbl"
set rda_Input(ui_preRoute_cap) {1}
set rda_Input(ui_postRoute_cap) {1}
```

```
set rda_Input(ui_postRoute_xcap) {1}
set rda_Input(ui_preRoute_res) {1}
set rda_Input(ui_postRoute_res) {1}
set rda_Input(ui_shr_scale) {1.0}
set rda_Input(ui_rel_c_thresh) {0.03}
set rda_Input(ui_tot_c_thresh) {5.0}
set rda_Input(ui_cpl_c_thresh) {3.0}
set rda_Input(ui_time_unit) {none}
set rda_Input(ui_cap_unit) {}
set rda_Input(ui_oa_reflib) {}
set rda_Input(ui_oa_abstractname) {}
set rda_Input(ui_oa_layoutname) {}
set rda_Input(ui_sigstormlib) ""
set rda_Input(ui_cdb_file,min) ""
set rda_Input(ui_cdb_file,max) ""
set rda_Input(ui_cdb_file) ""
set rda_Input(ui_echo_file,min) ""
set rda_Input(ui_echo_file,max) ""
set rda_Input(ui_echo_file) ""
set rda_Input(ui_xtwf_file) ""
set rda_Input(ui_qxtech_file) ""
set rda_Input(ui_qxlayermap_file) ""
set rda_Input(ui_qxlib_file) ""
set rda_Input(ui_qxconf_file) ""
set rda_Input(ui_pwrnet) {vdd!}
set rda_Input(ui_gndnet) {gnd!}
set rda_Input(flip_first) {1}
set rda_Input(double_back) {1}
set rda_Input(assign_buffer)   {1}
set rda_Input(use_io_row_flow) {0}
set rda_Input(ui_gen_footprint) {0}
```

Example env.tcl

```
# set the following parameter to point to the
# location where the design data is located.
# Design data includes gate level netlist, design floorplan,
# design configuration file (.conf file), timing constraint file (.sdc)
set designDir /home/apai/Digital_Synthesis2/scripts
```

```
# set the following variable to point to the location where the
# libraries are located.
# Timing Library - .lib
# LEF file
# Capacitance Table (.cap)
set libDir $designDir

# set the following variable to point to the directory
# where you want to store the intermediate results.
# The intermediate results include the design saved at
# various stages of the flow. e.g You may want to save the
# design after placement or routing or clock tree synthesis
# or in between while doing timing optimizations.
set outDir ./outputDir
```

To run the script, use the following command from the scripts directory:

```
/software/opt/CDS/cadence2007/edi91/tools.lnx86/bin/encounter -init fe_topheal.tcl
```

Bibliography

[1] G. Moore, "Cramming more components onto integrated circuits," *Proc. IEEE*, vol. 86, no. 1, pp. 82–85, Jan 1998.

[2] "Advancing Moore's law in 2014," http://download.intel.com/newsroom/kits/14nm/pdfs/Intel_14nm_New_uArch.pdf, 2014.

[3] J. Hoyt, H. Nayfeh, S. Eguchi, I. Aberg, G. Xia, T. Drake, E. Fitzgerald, and D. Antoniadis, "Strained silicon MOSFET technology," in *IEDM Techn. Dig.*, Dec 2002, pp. 23–26.

[4] T. Mizuno, S. Takagi, N. Sugiyama, H. Satake, A. Kurobe, and A. Toriumi, "Electron and hole mobility enhancement in strained-Si MOSFET's on SiGe-on-insulator substrates fabricated by SIMOX technology," *IEEE Electron Device Lett.*, vol. 21, no. 5, pp. 230–232, May 2000.

[5] K. Mistry and .et.al., "A 45nm logic technology with high-k+metal gate transistors, strained Silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging," in *IEDM Techn. Dig.*, Dec 2007, pp. 247–250.

[6] B. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully-depleted tri-gate CMOS transistors," *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 263–265, April 2003.

[7] R. Dennard, F. Gaensslen, H.-N. Yu, R. Leo, E. Bassous, and A. R. Leblanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Solid-State Circuits Soc. Newslett.*, vol. 12, no. 1, pp. 38–50, Winter 2007.

[8] S. Nassif, N. Mehta, and Y. Cao, "A resilience roadmap," in *Design Automat. Test Europe Conf. Exhib.*, 2010, pp. 1011–1016.

[9] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High-performance CMOS variability in the 65-nm regime and beyond," *IBM J. Research Develop.*, vol. 50, no. 4.5, pp. 433 –449, July 2006.

[10] T. Mizuno, J. Okumtura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 2216 –2221, Nov 1994.

[11] P. Stolk and D. Klaassen, "The effect of statistical dopant fluctuations on MOS device performance," in *Int. Electron Devices Meeting*, Dec. 1996, pp. 627 –630.

[12] K. Kuhn, C. Kenyon, A. Kornfeld, M. Liu, A. Maheshwari, W. Shih, S. Sivakumar, G. Taylor, P. VanDerVoorn, and K. Zawadzki, "Managing process variation in Intel's 45nm CMOS technology," *Intel Tech. J.*, vol. 12, no. 2, pp. 93–109, June 2008.

[13] A. Asenov, A. Brown, G. Roy, B. Cheng, C. Alexander, C. Riddet, U. Kovac, A. Martinez, N. Seoane, and S. Roy, "Simulation of statistical variability in nano-CMOS transistors using drift-diffusion, Monte Carlo and non-equilibrium Green's function techniques," *J. Comput. Electron.*, vol. 8, no. 3-4, pp. 349–373, 2009.

[14] V. Constantoudis, G. Kokkoris, E. Gogolides, E. Pargon, and M. Martin, "Effects of resist sidewall morphology on line-edge roughness reduction and transfer during etching: is the resist sidewall after development isotropic or anisotropic?" *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 9, no. 4, pp. 041 209–041 209–11, 2010.

[15] B. Yu, P. Harpe, and N. Van Der Meijs, "Efficient sensitivity-based capacitance modeling for systematic and random geometric variations," in *Asia South Pacific Design Automat. Conf.*, Jan 2011, pp. 61–66.

[16] J. R. Black, "Electromigration - a brief survey and some recent results," *IEEE Trans. Electron Devices*, vol. 16, no. 4, pp. 338–347, Apr 1969.

[17] O. Hammi, J. Sirois, S. Boumaiza, and F. Ghannouchi, "Study of the output load mismatch effects on the load modulation of doherty power amplifiers," in *IEEE Radio Wireless Symp.*, 2007, pp. 393–394a.

[18] A. Keerti and A.-V. Pham, "RF characterization of SiGe HBT power amplifiers under load mismatch," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 2, pp. 207–214, Feb 2007.

[19] R. Kumar and V. Kursun, "Voltage optimization for temperature variation insensitive CMOS circuits," in *Midwest Symp. Circuits Syst.*, 2005, pp. 476–479 Vol. 1.

[20] K. Sundaresan, P. Allen, and F. Ayazi, "Process and temperature compensation in a 7-MHz CMOS clock oscillator," *IEEE J. Solid-State Circuits*, vol. 41, no. 2, pp. 433–442, Feb 2006.

[21] S. Sakurai and M. Ismail, "Robust design of rail-to-rail CMOS operational amplifiers for a low power supply voltage," *IEEE J. Solid-State Circuits*, vol. 31, no. 2, pp. 146–156, 1996.

[22] K. Siwiec, T. Borejko, and W. Pleskacz, "PVT tolerant LC-VCO in 90 nm CMOS technology for GPS/galileo applications," in *IEEE Int. Symp. Design Diag. Electron. Circuits Syst.*, 2011, pp. 29–34.

[23] J. Yang and Y. Kim, "Self adaptive body biasing scheme for leakage power reduction in nanoscale CMOS circuit," in *Proc. Great Lakes Symp. VLSI*, ser. GLSVLSI '12. New York, NY, USA: ACM, 2012, pp. 111–116.

[24] D. Ghai, S. Mohanty, and E. Kougianos, "Design of parasitic and process-variation aware nano-CMOS RF circuits: A VCO case study," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 9, pp. 1339–1342, Sept 2009.

[25] Y. Cui, B. Chi, M. Liu, Y. Zhang, Y. Li, and Z. Wang, "Process variation compensation of a 2.4GHz LNA in 0.18μm CMOS using digitally switchable capacitance," in *IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 2562–2565.

[26] L.-F. Tanguay and M. Sawan, "Process variation tolerant LC-VCO dedicated to ultra-low power biomedical RF circuits," in *Int. Conf. Solid-State Integr.-Circuit Tech.*, Oct 2008, pp. 1585–1588.

[27] D. Sylvester, D. Blaauw, and E. Karl, "ElastIC: an adaptive self-healing architecture for unpredictable silicon," *IEEE Design Test Comput.*, vol. 23, no. 6, pp. 484–490, 2006.

[28] T. Das, A. Gopalan, C. Washburn, and P. Mukund, "Self-calibration of input-match in RF front-end circuitry," *IEEE Trans. Circuits Syst. II, Expr. Briefs*, vol. 52, no. 12, pp. 821–825, Dec 2005.

[29] A. Goyal, M. Swaminathan, A. Chatterjee, D. Howard, and J. Cressler, "A new self-healing methodology for RF amplifier circuits based on oscillation principles," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 10, pp. 1835–1848, 2012.

[30] J.-C. Liu, A. Tang, N. Wang, Q. Gu, R. Berenguer, H. Hsieh, P. Wu, C. Jou, and M. C. F. Chang, "A V-band self-healing power amplifier with adaptive feedback bias control in 65 nm CMOS," in *IEEE Radio Freq. Integr. Circuits Symp.*, 2011, pp. 1–4.

[31] J.-C. Liu, R. Berenguer, and M. Chang, "Millimeter-wave self-healing power amplifier with adaptive amplitude and phase linearization in 65-nm CMOS," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 5, pp. 1342–1352, May 2012.

[32] K. Jayaraman, Q. Khan, B. Chi, W. Beattie, Z. Wang, and P. Chiang, "A self-healing 2.4GHz LNA with on-chip S_{11}/S_{21} measurement/calibration for in-situ PVT compensation," in *IEEE Radio Freq. Integr. Circuits Symp.*, 2010, pp. 311–314.

[33] Y. Huang, H. Hsieh, and L. Lu, "A low-noise amplifier with integrated current and power sensors for RF BIST applications," in *Proc. IEEE VLSI Test Symp.*, 2007, pp. 401–408.

[34] F. Cheung and P. Mok, "A monolithic current-mode CMOS DC-DC converter with on-chip current-sensing technique," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 3–14, 2004.

[35] R. Whatley, T. Ranta, and D. Kelly, "CMOS based tunable matching networks for cellular handset applications," in *IEEE MTT-S Int. Microw. Symp. Dig.*, June 2011, pp. 1–4.

[36] Y. Youngchang, K. Jihwan, K. Hyungwook, H. A. Kyu, L. Ockgoo, L. Chang-Ho, and J. Kenney, "A dual-mode CMOS RF power amplifier with integrated tunable matching network," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 1, pp. 77–88, Jan 2012.

[37] S. Park, Y. Palaskas, A. Ravi, R. Bishop, and M. Flynn, "A 3.5 GS/s 5-b flash ADC in 90 nm CMOS," in *IEEE Custom Integr. Circuits Conf.*, Sept. 2006, pp. 489 –492.

[38] Y.-Z. Lin, Y.-T. Liu, and S.-J. Chang, "A 5-bit 4.2-GS/s flash ADC in 0.13-μm CMOS," in *IEEE Custom Integr. Circuits Conf.*, Sept. 2007, pp. 213 –216.

[39] J. Li and U.-K. Moon, "A 1.8-V 67-mW 10-bit 100-MS/s pipelined ADC using time-shifted CDS technique," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1468 – 1476, Sept. 2004.

[40] S. Mortezapour and E. Lee, "A 1-V, 8-bit successive approximation ADC in standard CMOS process," *IEEE J. Solid-State Circuits*, vol. 35, no. 4, pp. 642 –646, April 2000.

[41] J. Russell, H., "An improved successive-approximation register design for use in A/D converters," *IEEE Trans. Circuits Syst.*, vol. 25, no. 7, pp. 550 – 554, Jul 1978.

[42] K. Okada, K. Matsushita, K. Bunsen, R. Murakami, A. Musa, T. Sato, H. Asada, N. Takayama, N. Li, S. Ito, W. Chaivipas, R. Minami, and A. Matsuzawa, "A 60GHz 16QAM/8PSK/QPSK/BPSK direct-conversion transceiver for IEEE 802.15.3c," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2011, pp. 160–162.

[43] C. Marcu, D. Chowdhury, C. Thakkar, J.-D. Park, L.-K. Kong, M. Tabesh, Y. Wang, B. Afshar, A. Gupta, A. Arbabian, S. Gambini, R. Zamani, E. Alon, and A. Niknejad, "A 90 nm CMOS low-power 60 GHz transceiver with integrated baseband circuitry," *IEEE J. Solid-State Circuits*, vol. 44, no. 12, pp. 3434–3447, Dec 2009.

[44] H. Asada, K. Bunsen, K. Matsushita, R. Murakami, Q. Bu, A. Musa, T. Sato, T. Yamaguchi, R. Minami, T. Ito, K. Okada, and A. Matsuzawa, "A 60GHz 16Gb/s 16QAM low-power direct-conversion transceiver using capacitive cross-coupling neutralization in 65 nm CMOS," in *IEEE A-SSCC Dig. Tech. Papers*, Nov 2011, pp. 373–376.

[45] A. Tomkins, R. Aroca, T. Yamamoto, S. Nicolson, Y. Doi, and S. Voinigescu, "A zero-IF 60 GHz 65 nm CMOS transceiver with direct BPSK modulation demonstrating up to 6 Gb/s data rates over a 2 m wireless link," *IEEE J. Solid-State Circuits*, vol. 44, no. 8, pp. 2085–2099, Aug 2009.

[46] J. Kang, A. Hajimiri, and B. Kim, "A single-chip linear CMOS power amplifier for 2.4 GHz WLAN," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2006, pp. 761–769.

[47] D. Chowdhury, C. Hull, O. Degani, P. Goyal, Y. Wang, and A. Niknejad, "A single-chip highly linear 2.4GHz 30dBm power amplifier in 90nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2009, pp. 378–379,379a.

[48] D. Chowdhury, P. Reynaert, and A. Niknejad, "Design considerations for 60 GHz transformer-coupled CMOS power amplifiers," *IEEE J. Solid-State Circuits*, vol. 44, no. 10, pp. 2733–2744, Oct 2009.

[49] P. Reynaert and M. Steyaert, "A 1.75-GHz polar modulated CMOS RF power amplifier for GSM-EDGE," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2598–2608, Dec 2005.

[50] V. Pinon, F. Hasbani, A. Giry, D. Pache, and C. Garnier, "A single-chip WCDMA envelope reconstruction LDMOS PA with 130MHz switched-mode power supply," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2008, pp. 564–636.

[51] A. Balteanu, I. Sarkas, E. Dacquay, A. Tomkins, G. Rebeiz, P. Asbeck, and S. Voinigescu, "A 2-bit, 24 dBm, millimeter-wave SOI CMOS power-DAC cell for watt-level high-efficiency, fully digital m-ary QAM transmitters," *IEEE J. Solid-State Circuits*, vol. 48, no. 5, pp. 1126–1137, May 2013.

[52] J. Chen, L. Ye, D. Titz, F. Gianesello, R. Pilard, A. Cathelin, F. Ferrero, C. Luxey, and A. Niknejad, "A digitally modulated mm-wave cartesian beamforming transmitter with quadrature spatial combining," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2013, pp. 232–233.

[53] Y. Zhao, J. Long, and M. Spirito, "Compact transformer power combiners for millimeter-wave wireless applications," in *IEEE Radio Freq. Integr. Circuits Symp.*, May 2010, pp. 223–226.

[54] T. LaRocca, J.-C. Liu, and M.-C. Chang, "60 GHz CMOS amplifiers using transformer-coupling and artificial dielectric differential transmission lines for compact design," *IEEE J. Solid-State Circuits*, vol. 44, no. 5, pp. 1425–1435, May 2009.

[55] I. Aoki, S. Kee, D. Rutledge, and A. Hajimiri, "Distributed active transformer-a new power-combining and impedance-transformation technique," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 1, pp. 316–331, Jan 2002.

[56] J. Chen and A. Niknejad, "A compact 1V 18.6dBm 60GHz power amplifier in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2011, pp. 432–433.

[57] D. Zhao and P. Reynaert, "A 0.9V 20.9dBm 22.3%-PAE E-band power amplifier with broadband parallel-series power combiner in 40nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2014, pp. 248–249.

[58] D. Chowdhury, P. Reynaert, and A. Niknejad, "Design considerations for 60 GHz transformer-coupled CMOS power amplifiers," *IEEE J. Solid-State Circuits*, vol. 44, no. 10, pp. 2733–2744, Oct 2009.

[59] U. Pfeiffer and D. Goren, "A 23-dBm 60-GHz distributed active transformer in a silicon process technology," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 5, pp. 857–865, May 2007.

[60] S. Kousai and A. Hajimiri, "An octave-range watt-level fully integrated CMOS switching power mixer array for linearization and back-off efficiency improvement," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2009, pp. 376–377,377a.

[61] K. Dasgupta, K. Sengupta, A. Pai, and A. Hajimiri, "A 19.1dBm segmented power-mixer based multi-Gbps mm-Wave transmitter in 32nm SOI CMOS," in *IEEE Radio Freq. Integr. Circuits Symp.*, June 2014, pp. 343–346.

[62] D. Greenberg, J.-O. Plouchart, and A. Valdes-Garcia, "Electromigration-compliant high performance fet layout," Patent US 2012/0 112 819 A1.

[63] J.-W. Lai and A. Valdes-Garcia, "A 1V 17.9dBm 60GHz power amplifier in standard 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2010, pp. 424–425.

[64] S. Shopov, A. Balteanu, and S. Voinigescu, "A 19 dBm, 15 Gbaud, 9 bit SOI CMOS power-DAC cell for high-order QAM W-band transmitters," *IEEE J. Solid-State Circuits*, vol. 49, no. 7, pp. 1653–1664, July 2014.

[65] H. Wang and A. Hajimiri, "A wideband CMOS linear digital phase rotator," in *IEEE Custom Integr. Circuits Conf.*, Sept 2007, pp. 671–674.

[66] S. M. Bowers, K. Sengupta, K. Dasgupta, B. D. Parker, and A. Hajimiri, "Integrated self-healing for mm-wave power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 3, pp. 1301–1315, 2013.

[67] S. Jeon, Y. Wang, H. Wang, F. Bohn, A. Natarajan, A. Babakhani, and A. Hajimiri, "A scalable 6-to-18GHz concurrent dual-band quad-beam phased-array receiver in CMOS," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb 2008, pp. 186–605.

[68] H. Wang, "Precision frequency and phase synthesis techniques in integrated circuits for biosensing, communication and radar," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, USA, 2009.

[69] H. Wang, S. Jeon, Y. Wang, F. Bohn, A. Natarajan, A. Babakhani, and A. Hajimiri, "A tunable concurrent 6-to-18GHz phased-array system in CMOS," in *IEEE MTT-S Int. Microw. Symp. Dig.*, June 2008, pp. 687–690.

[70] W. Lee and Y. Yu, "Polarization diversity system for mobile radio," *IEEE Trans. Commun.*, vol. 20, no. 5, pp. 912–923, Oct 1972.

[71] R. Vaughan, "Polarization diversity in mobile communications," *IEEE Trans. Veh. Technol.*, vol. 39, no. 3, pp. 177–186, Aug 1990.

[72] D. Schaubert, F. Farrar, A. Sindoris, and S. Hayes, "Microstrip antennas with frequency agility and polarization diversity," *IEEE Trans. Antennas Propag.*, vol. 29, no. 1, pp. 118–123, Jan 1981.

[73] M. Fries, M. Grani, and R. Vahldieck, "A reconfigurable slot antenna with switchable polarization," *IEEE Microw. Compon. Lett.*, vol. 13, no. 11, pp. 490–492, Nov 2003.

[74] S. Bowers, A. Safaripour, and A. Hajimiri, "Dynamic polarization control of integrated radiators," in *IEEE Radio Freq. Integr. Circuits Symp.*, June 2014, pp. 291–294.

[75] A. Adrian and D. Schaubert, "Dual aperture-coupled microstrip antenna for dual or circular polarisation," *Electron. Lett.*, vol. 23, no. 23, pp. 1226–1228, November 1987.