Transmission Line Transformers for Broadband K/Ka Power Amplifiers in Bulk CMOS

Rohit Braganza

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2025-70
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-70.html

May 15, 2025

Transmission Line Transformers for Broadband K/Ka Power Amplifiers in Bulk CMOS

By

Rohit Julian Braganza

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ali M. Niknejad, Chair
Professor Kirstofer S. J. Pister
Professor Panos N. Patatoukas

Spring 2025

Transmission Line Transformers for Broadband K/Ka Power Amplifiers in Bulk CMOS

Abstract

Transmission Line Transformers for Broadband K/Ka Power Amplifiers in Bulk CMOS

by

Rohit Julian Braganza

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ali M. Niknejad, Chair

The K (18-27GHz) and Ka (27-40GHz) have recently been used for a variety of new applications. The success of Space X has made the space industry a viable commercial (in addition to its traditional military usage) enterprise, finding applications in cellular and wireless communications, GPS, imaging, and more; many satellites rely on the K and Ka frequencies for terrestrial communication. Outside of space, uses of these bands include automotive radar and cellular communications. And with the broad range of spectrum now available commercially, it is highly likely other uses will arise in the coming years. This work demonstrates a single 'plug and place' power amplifier that can operate across the entirety of both the K and Ka bands. Furthermore, it is done in a cost-effective bulk CMOS node, instead of the significantly more expensive III-V, FinFet, or SOI processes. The key enabler of this bandwidth extension is the use of transmission line transformers, which are shown to lend themselves well to CMOS IC design in addition to traditional microwave design. In this thesis we will thus describe the theory behind transmission line transformers, their CMOS implementations, and their usage in octave-bandwidth power amplifiers.

Contents

List of Figures

List of Tables

Acknowledgments

There are many people to whom I have to express my gratitude. In the early days, much thanks has to be given to Lorenzo Iotti and Nima Baniasadi, both of whom taught me a lot about the secrets of RF-IC design. Likewise thanks to Sashank Krishnamurthy for the wonderful conversations, advice, and technical help on early tapeouts. I also enjoyed the discourse with Luya Zhang, Yi-An Li, and Bonjern Yang. Later, I had the privilege of working with new fellow students who likewise helped me immensely, namely Yahia Ibrahim, Hesham Beshary, and Rami Hijab. Also thanks to Amin Torabi, Sunjin Choi, Ali Ameri, Brian Richards, and 'Candy' Columba Corpus Steudeman for the countless conversations we have had.

I've always enjoyed the wonderful conversations with Professors Ali Niknejad, Osama Shana'a, Kristofer Pister, and Panos N. Patatoukas, and thank them for their advice over the years, as well as serving on my qualification exam committee. Panos' course and particularly the hours spent talking on a wide variety of topics was one of the highlights of my time in the PhD program. As for advisor Ali Niknejad, what stands out for me was his support during calls with outside groups/sponsors, with his uncanny ability to clearly articulate tricky concepts a major source of help (that helped get me through tricky questions and situations). He is also easily one of the most genuinely nice persons I have ever met. Back in my undergraduate years Krisfofer Pister's analog circuits course and his open-office hours policy are the reasons I became an analog circuit designer in the first place (instead of a chemist). Osama's RFIC course and his office hours helped me immensely as a practicing RF engineer as he is incredibly effective at tying theory to practice; I am extremely grateful for the time he spent outside his primary job (VP at Mediatek) to teach.

A special thanks to fellow student Ethan Chou. Through both the undergraduate and PhD degrees we have now done, he has always been a great conversationalist, a reliable source of help, and friend for almost a decade now.

And of course, I'd like to thank my good BWRC friends Kunmo Kim, Danielius Kramnik, and Harrison Liew, all of whom helped me through the PhD with enumerable moments of fun, laughter, and camaraderie that I will look back on fondly.

Last, but certainly not least, the biggest acknowledgment must be to fellow BWRC student Yikuan Chen. We worked together on the same projects in the beginning, and continued to help each other throughout our respective PhDs. We worked together on a total of four tape-outs (and all the hours associated with said tapeouts). It is no exaggeration to say I would not have finished the PhD without his help.

Chapter 1

Introduction

1.1 K-Ka Band Power Amplifier Overview

1.1.1 Use Cases

The K (18-27GHz) and Ka (27-40GHz) bands have seen recent utilization for a variety of applications. One of the biggest beneficiaries is the satellite-communication industry, which has used practically all bands from X (8GHz) through E (90GHz), for communication, imaging, GPS, meteorology, and TV, for both military and commercial purposes. Other commercial applications include automotive radar, which utilizes frequency around 24GHz[1]. Notably, efforts are currently underway for "5G NR" to utilize this spectrum for cellular use. In the US, FR2 band n257 and n258 use spectrum between 24.25 to 29.5GHz; in the rest of the world, proposals range from 24GHz (Japan) to 43.5GHz (India) [1]. This work focuses on power amplifiers (PAs) that can be used for these K and Ka applications.

1.2 Current CMOS K-Ka PAs

As a technology, CMOS has always been favored for its ease of digital integration, size, and ostensibly lower cost. At higher frequencies however, RF PA designs in CMOS become challenging due to a variety of well known issues, including lossy routing layers and vias, low device Fmax (the frequency at which the unilateral gain of a transistor is equal to 1), high parasitic capacitance, and low oxide breakdown voltages. Circuit innovation therefore has been necessary to provide acceptable performance from CMOS designs, though the resulting performance inevitably lags other technologies, notably GaN, as seen in figure 1.1. The higher cost of alternative technologies however means that the cheaper CMOS will remain of notable interest and relevance. Note in this work we will be utilizing a bulk, 28nm CMOS processes, as bulk CMOS nodes retain a competitive cost advantage against their

[1]The industry has more recently moved on to 77 GHz for civilian automotive radar, which is outside the frequency range of this work, but still worth mentioning.

III-V counterparts, and we will base our comparisons to similar bulk CMOS designs. Newer CMOS technologies such as FinFET and SOI provide significantly improved RF performance over bulk CMOS designs, but lack a key feature of CMOS: low cost! [2] Such SOI/FinFET PA designs are therefore targeted for fully integrated systems that interface directly with a digital core; for standalone PAs it makes little sense to use a process more expensive than GaAs or GaN that provides worse RF performance. This emphasis on lower costs is becoming increasingly important due to the increased adoption of massive arrays of antennas used in MIMO systems; these systems inherently rely on a large number of cheap RF front ends. Even in the defense sector, the advent of drone warfare necessitates cost effective, mass produced RF front ends in the K and Ka bands [2].

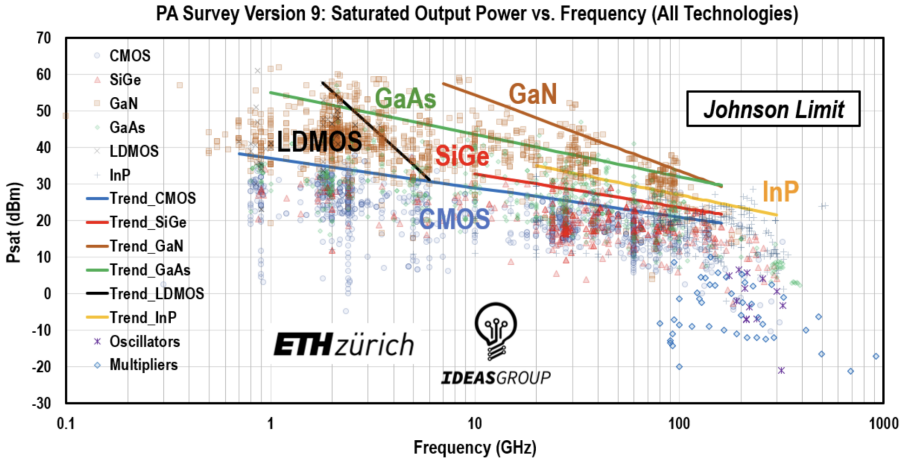

Figure 1.1: Survey of PAs, source:[3]. Note this does not distinguish between SOI and CMOS, despite the significant performance improvement SOI offers for PAs

While many papers have focused on improving CMOS PA efficiency and output power, these designs tend to be narrow band. Few if any CMOS PAs can operate as a "plug and place" part that can work for all of the aforementioned K and Ka band applications. Such parts would be useful for ITAR (e.g. satcom and space) and defense applications where streamlined supply chains are as important, if not more important, than the part itself[3]. In this use case, a broadband CMOS PA offers a compelling alternative to distributed amplifiers or GaAs driver amplifiers. Distributed amplifiers have terrible efficiency (single digit efficien-

[2]The author has often found that the latest nodes advanced nodes are more expensive than their III-V counterparts, certainly for small batch sizes!

[3]This is most apparent in the US military, the largest, and perhaps most sophisticated logistical operation in human history. The military relies heavily on distributed amplifiers — evidenced by the catalog of military firms such as MACOM — since a single part can work for just about every application from C band radar to mmWave communication. This interoperability is far more important than the terrible efficiency distributed amplifiers offer.

cies), low output power, and are moderately expensive (as most use III-V semiconductors); GaAs drivers, while more efficient, are about an order of magnitude more expensive than their CMOS counterparts. In large arrays these costs add up quickly.

With this in mind, figures 1.2 and 1.3 show recent notable bulk CMOS designs with large (>60%) bandwidths, as well as a few narrow band examples for perspective. To the author's knowledge, the three designs shown are the only bulk CMOS PAs with large bandwidths operating in the K/Ka bands. Additional information for each design, including the supply voltage, are included in Table 1.1. A few things are readily apparent; as the bandwidth increases, unsurprisingly the saturated power and efficiency tend to decrease. Note that the supply voltage in figures 1.3 and 1.2 is not fixed; [4] at first may appear to break the trend of power efficiency/output versus bandwidth, but note that it uses a 1.8V supply, compared to the 1V supply in [5]. Higher supply voltages lead to improved performance and therefore must be taken into account[4]. Note all of the designs have roughly similar linearity measures (class A/B, EVM around -25dBc).

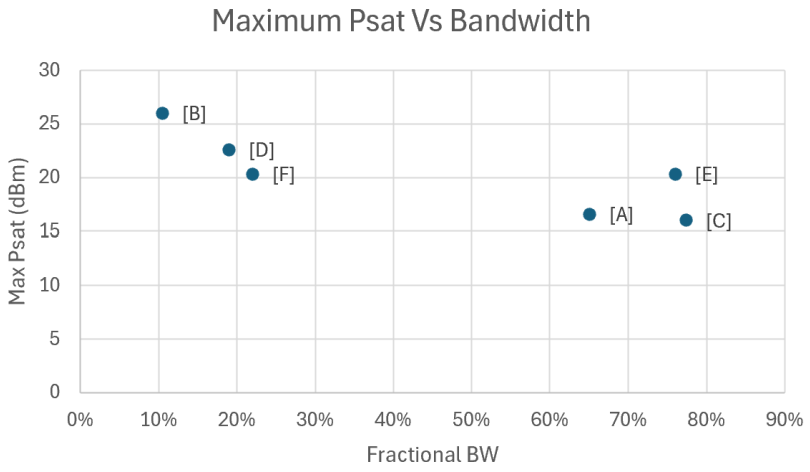

Figure 1.2: Survey of notable recent K and mmWave band CMOS PAs, comparing the highest Psat withing a given bandwidth; Table 1.1 shows the minimum Psat value over frequency.

[4]It must also be noted that it is doubtful these high supply voltage designs would work as a viable commercial product, as the large swings likely violate all lifetime breakdown/electro-migration considerations for bulk CMOS, which are nominally 0.8V processes.

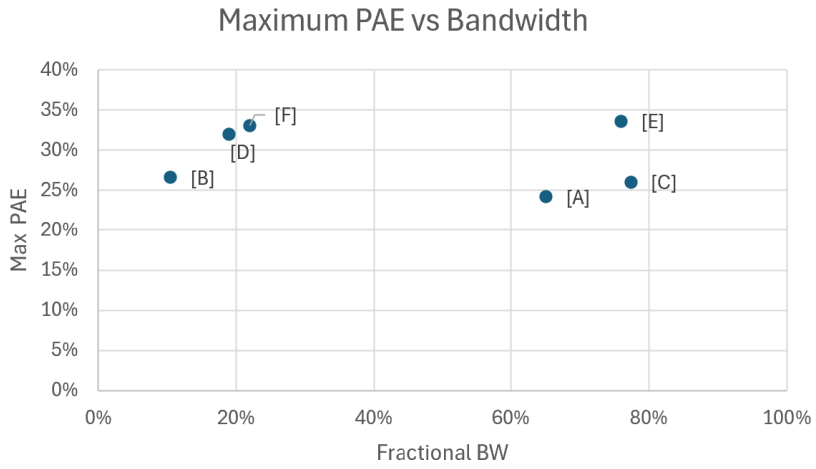

Figure 1.3: Survey of notable recent K and mmWave band CMOS PAs, comparing the highest PAE (at Psat) within a given bandwidth. Note that while the efficiencies may appear to be unchanged, the provided output power is significantly higher for the narrow-band designs.

Table 1.1: PA Overview

Label	Ref.	VDD	Center Frequency (GHz)	Bandwidth (%)	Psat, Max (dBm)	Psat, Min (dBm)	PAE, Max (%)	PAE, Min (%)
[A]	[6]	0.9V	43	65.1%	16.6	13	24.2%	9%
[B]	[7]	2.2V	38	10.5%	26	25	26.6%	22%
[C]	[8]	2V	31	77.4%	16	12	26%	6%
[D]	[5]	1V	35.5	19%	22.6	20	32%	29%
[E]	[4]	1.8V	31.75	76%	20.3	18	33.6%	19%
[F]	[9]	2.2V	27	22%	20.3	19.7	33.1%	30%

All Min/max values are over operating bandwidth. PAE is given at Psat.

1.3 Proposed Design

The proposed PA aims to cover both K and Ka Bands in a bulk CMOS process, with comparable output power and efficiency at a low supply voltage (1V). The simulated performance and a comparison to similar broadband designs is provided in table 1.2. Two PA variations were designed and fabricated, differing by the power combining done (two and four). Both achieve record fractional bandwidths. The four way design has higher output power than the others, and the two-way offers comparatively high output power (greater than [6] and [8], less than [4]). The peak efficiencies for both variations are lower than the comparative designs, however the supply used here is considerably lower than [8] and [4], the output

power delivered is higher than [6] and [8], and the efficiency at the edge of the band (min PAE) is higher than [6] and [8].

Table 1.2: Proposed PA

Label	Ref.	VDD	Center Frequency (GHz)	Bandwidth (%)	Psat, Max (dBm)	Psat, Min (dBm)	PAE, Max (%)	PAE, Min (%)
	This work (sim), 4-Way	1V	25	80%	22	20.5	18%	11%
	This work (sim), 2-Way	1V	28	92%	18.5	17.3	20%	12%
[A]	[6]	0.9V	43	65.1%	16.6	13	24.2%	9%
[C]	[8]	2V	31	77.4%	16	12	26%	6%
[E]	[4]	1.8V	31.75	76%	20.3	18	33.6%	19%

All Min/max values are over operating bandwidth. PAE is given at Psat.

Chapter 2

Transmission Line Transformers And Baluns

2.1 A Little History

Transmission line transformers (TLTs) are not new by any means; the Guanella transformer was first discovered by its namesake in 1944 [10], and Ruthroff would further the field with his seminal paper in 1959 [11]. The motivation behind transmission line transformers is simple: take advantage of the inherently broadband nature of transmission lines. Back then these were typically built with coax cables or wires; although today most designs are planar in nature, the key concepts remain the same, just as they were back in 1944!

2.2 Basic Concepts of Transmission Line Transformers

Unlike standard inductors/transformers that transfer power through magnetic flux coupling, transmission line transformers transfer power through TEM transmission. The simplest TLT can be thought of as simply a coupled line segment, as shown in figure 2.1. From transmission line theory, the impedance Z_{in} for a lossless line can be shown to be 2.1:

$$Z_{in} = Z_o \frac{Z_L + jZ_o \tan(\beta l)}{Z_o + jZ_L \tan(\beta l)} \tag{2.1}$$

Note if $Z_o = Z_L$, then $Z_{in} = Z_L$, implying that this 1:1 impedance transformer ideally is frequency independent[1]. Intuitively this makes sense, as in this configuration the coupler is

[1]Note if not used as a 1:1 transformer, this essentially becomes just a quarter wave matching section. Note that while the 1:1 does not perform any impedance transformation (and therefore it may seem a stretch to call it a "transformer") we will see that this is a building block for baluns and actual impedance transformers.

simply acting as a transmission line, and an ideal TEM line is matched at all frequencies if its characteristic impedance is equal to the boundary impedances[2].

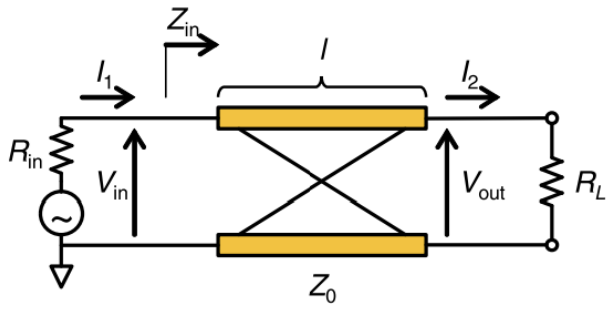

Figure 2.1: The basic 1:1 (Guanella) transmission line transformer. Figure from [12].

Now let us consider the 4:1 Guanella transformer, shown in figure 2.2, as drawn in Guanella's original paper. This can be explained intuitively as follows. If we assume the coils are tightly coupled (and therefore behave ideally), then any current flowing into one end of the coil will be induced into the coupled coil with opposite polarity. If we assume differential input drive, one can see that the resulting output current must be twice that of the input; assuming ideal coils, by conservation of power, we must have half the voltage swing on the output, and therefore we have achieved a 4:1 impedance transformation! This is depicted in figure 2.3. Mathematically, it can be shown for ideal, lossless lines that [13] (we will derive this ourselves in section 4.2):

$$Z_{in} = \frac{1}{2} Z_o [\frac{Z_L/2 + jZ_o \tan(\beta l)}{Z_o + j(Z_L/2)\tan(\beta l)}] \tag{2.2}$$

Similar to the 1:1 transformer, here when $Z_o = Z_L/2$, $Z_{in} = Z_L/4$, once again independent of frequency.

2.3 Even And Odd Mode Impedances and Transformer Bandwidth

We've seen that ideally, when the source and load are purely real, these structures offer infinite bandwidth. Of course in reality such things do not happen; the transformers will inevitably have a finite bandwidth. In transmission line transformers, the primary bandwidth limitation occurs when we consider the existence of a "global" ground plane (or return path). This can be thought of in two ways: the easier visualization of this is as a parallel, parasitic

[2]Note by 'ideal' transmission line, we are referring to not only a lossless line, but an infinite even-mode line as well. The infinite even-mode assumption is a key point that we will discuss in detail shortly!

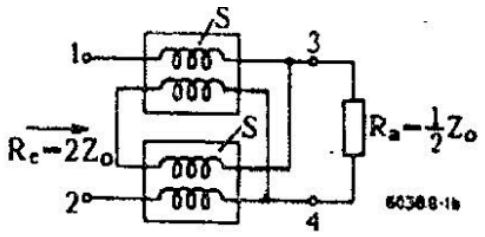

Figure 2.2: The 4:1 impedance transformer, as first conceived by Guanella[10]

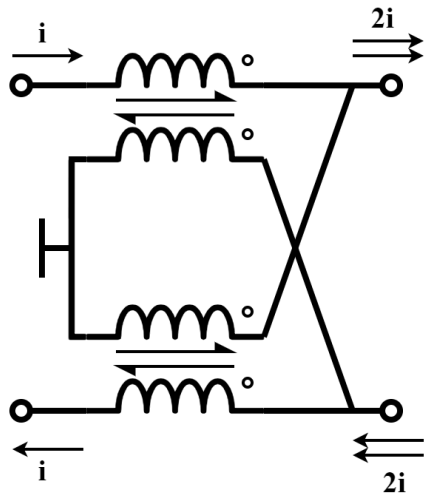

Figure 2.3: The 4:1 Guanella transformer explained with current paths

transmission line between the conductor and the global ground (figure 2.4, (a)). Or more accurately, there exists an additional coupling between the intended conductors and the additional "ground" conductors, akin to a four-line coupler (see figure 2.4, (b)). For balanced to unbalanced conversion this further degrades the performance, as the extra parasitic lines break the symmetry of such structures, as we will see ahead shortly.

2.3.1 Unbalanced Transformers: The Ruthroff 4:1 TLT.

Shown in figure 2.5 is the Ruthroff 4:1 unbalanced to unbalanced (un-un or single ended) transmission line transformer. It operates in the same fashion as the Guanella 4:1 transformer described earlier. Figure 2.6 redraws it in its planar form, and includes the aforementioned parasitic ground transmission lines. The first thing to notice is that unlike the balanced Guanella transformer, we no longer have perfect symmetry between the input and outputs. This results in a critical shunt parasitic transmission line, highlighted in figure 2.6 (b). At

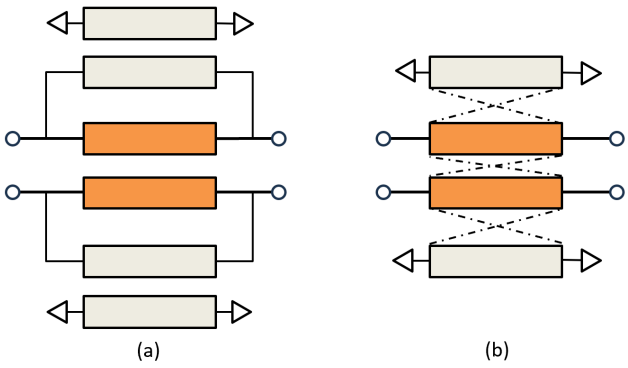

Figure 2.4: The basic 1:1 TLT with drawn parasitics in beige. In (a), the parasitic lines are views as parallel transmission lines. In (b), they are shown as parasitic coupling, or as a four line coupler.

some frequency, this line will become equivalent to a half-wavelength line to ground, and therefore provide a short to the output! This will naturally limit the bandwidth of the structure. We will investigate this further in the proceeding section.

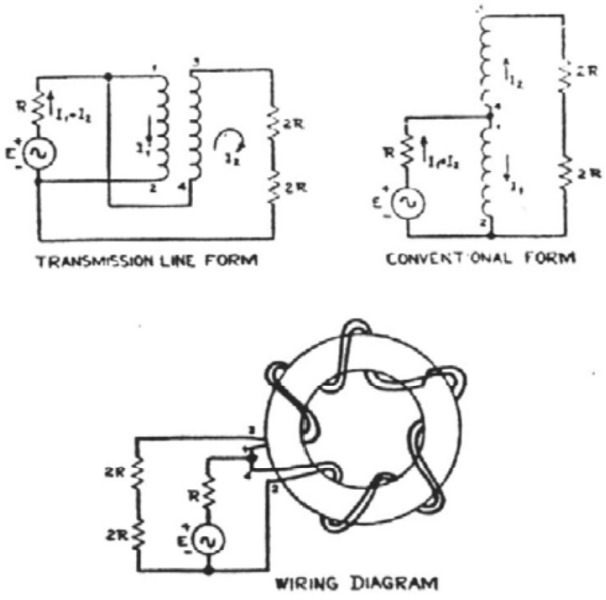

Figure 2.5: Ruthroff's original representation of his 4:1 un-un transformer [11].

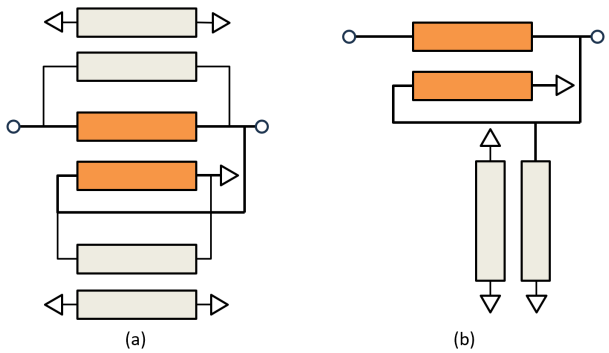

Figure 2.6: In (a), we show the planar version of Ruthroff's transformer, with parasitic transmission lines included in beige. In (b), we redraw the circuit to highlight the bandwidth limiting shunt parasitic line (ignoring the other parasitic line for now).

2.3.2 Even/Odd Modes, and Coupling factor

To maximize the bandwidth of these structures, it has been well documented in literature that one should maximize the coupling coefficient between the coupled lines [14][15]. This is illustrated in figure 2.7. We know this relates directly to the even and odd modes as the coupling factor is related by (for two uniformly coupled backward-wave lines) [14]:

$$k = \frac{Z_{oe} - Z_{oo}}{Z_{oe} + Z_{oo}} \tag{2.3}$$

The characteristic impedance is given by (again for two lines):

$$Z_{oe} Z_{oo} = Z_o^2 \tag{2.4}$$

Put another way, for a given characteristic impedance, to maximize the bandwidth, one should maximize the even mode impedance (and adjust the odd mode impedance lower accordingly). This is equivalent to maximizing the coupling factor, but thinking in terms of odd/even mode impedances is easier for the designer's intuition.

This relates nicely to the parasitic transmission line model mentioned previously. In a coupled line structure the even mode (magnetic wall) is given by the parasitic capacitive coupling to the surrounding ground plane. The odd mode (electric wall) relates to the coupling between adjacent conductors. The goal of the designer thus is to maximize the impedance of the parasitic ground lines, while simultaneously increasing the coupling between the conductors (to maintain the desired characteristic impedance). Historically, this has been done by coiling the conductor wires around a ferrite core, as seen in Ruthroff's original paper (figure 2.5; the resulting transformers typically work from 10MHz-200MHz, an incredible number of octaves). Planar, high-frequency designs however cannot utilize ferromagnetic materials,

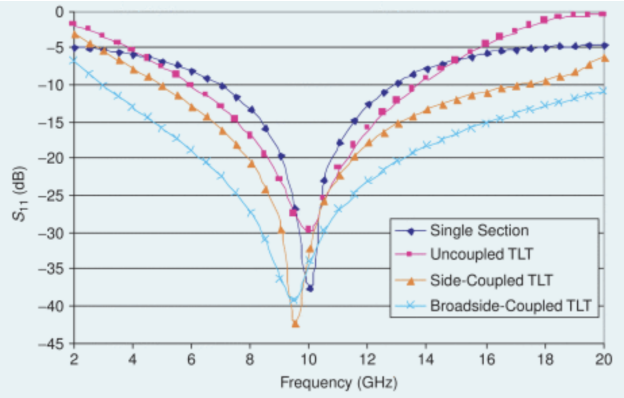

Figure 2.7: For the Ruthroff transformer, structures with tighter coupling coefficients result in larger bandwidths [15]

.

and more obviously, typically have a ground plane right underneath which drastically reduces the even-mode impedance! The "tricks" to counteract this either attempt to shield the conductors from the ground plane- such as multilayer lines, helical lines, three/multi-finger lines, and spiral structures- or to increase the separation to the "ground" return path using for instance a coplanar structure when possible [14] [16]. We will cover this in detail when we go into integrated circuit implementations of these structures.

2.4 Baluns

We shall first define what a balun is. An ideal flux (k=1) center-tapped transformer provides voltage balance on the balanced side, and current balance on the unbalanced side. It's worth noting that such transformers are not practically realizable outside of the low MHz range as the achievable coupling factor decreases significantly in the GHz range (when k drops even a small amount, from 1 to 0.9, there is a marked change in performance [17]. This is illustrated in figure 2.8. As the frequencies approach the GHz range, very quietly the word "transformer" is replaced by "balun," although the desired outcome remains the same. At K band, flux transformers with high coupling coefficients are practically impossible on chip. With transmission line baluns however, we can design broadband, realizable structures that mimic in many ways the ideal flux transformer.

2.4.1 The Coaxial Balun

A simple coaxial balun is shown in figure 2.9. At the unbalanced end, the outer shield is grounded, and the inner wire is terminated/launched. If we terminate the other end of

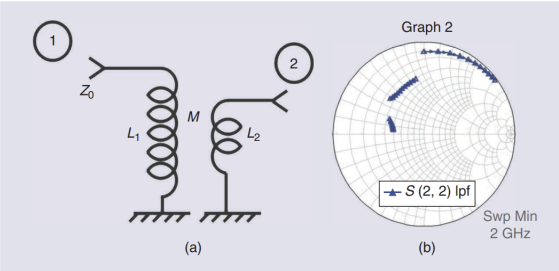

Figure 2.8: Input impedance of a flux transformer for K=1, 0.9 and 0.5 shown on the Smith chart. From [17].

the coax, then the voltage between the inner conductor and outer conductive shield must remain constant (for a TEM line); therefore, if instead of grounding the balanced output, we tap it instead with a split-matched termination (with the center of the split to ground, akin to a center-tapped flux transformer), we get a differential output! Notice here that we satisfy both criteria of a balun: we achieve voltage balance on the balanced side, and on the unbalanced side, in a TEM line the inner and outer (return) currents must be equal and opposite if the voltage inputs are equal and opposite (current balance). In an ideal world, the structure would work for all frequencies, but once again that pesky parasitic ground plane comes into play. Because there is in fact a second transmission line in this figure, between the outer conductor and what we nebulously call "ground." This extra line breaks the symmetry on the balanced output, and leads inevitably to phase mismatch and bandwidth limitations. Uncompensated, this structure will provide 180 degrees of phase difference only at the quarter wavelength frequency- at this center frequency the shorted parasitic line will look like an open circuit and therefore will not affect the performance. At the other frequencies, this parasitic line will not be open, with the worst-case occurring at the half wavelength, where it is a short to ground! The performance of this 'basic' balun is shown in figure 2.11, which illustrates the poor phase balance.

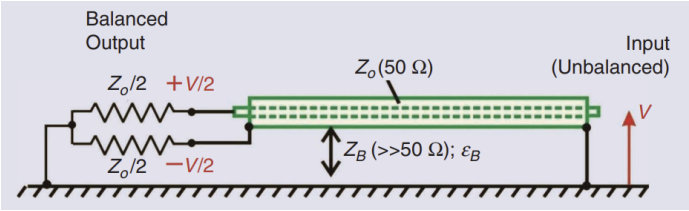

Figure 2.9: The basic coaxial balun [17], with the pesky parasitic ground plane transmission line of impedance 'Z_B'.

2.4.2 The Marchand Balun

Nathan Marchand's seminal 1944 paper [18] was the first to tackle the phase balance issue, and serves as a useful instructional case. The simplest, and for many purposes a very effective solution, is to simply add a short stub with the same impedance as the parasitic outer transmission line to the inner conductor of the coaxial balun at the balanced output. This restores the symmetry of the structure as now both of the balanced ports have an additional shunt short-stub, instead of just one of them. This is shown in figure 2.10. A comparison of the performance of the basic cable balun and the short stub compensated balun is shown in figure 2.11- note how it fixes the phase imbalance of the basic coax balun and improves the amplitude balance near the band edges.

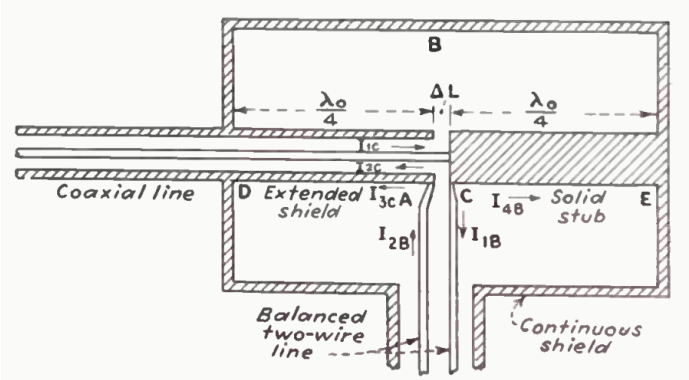

Figure 2.10: The coaxial balun with a compensating short stub, from Marchand's original paper [18].

Marchand further improved his balun design by replacing the shunt short stub with instead a series quarter-wave open stub, creating what is now known as the Marchand Balun [18], shown in figure 2.12. In [19] Marchand derives a formula for the input impedance of the balun, and shows that to improve the bandwidth (maximally flat), the impedance of the open stub should be designed as follows, where Z_{op} is the characteristic impedance of the parasitic ground transmission line, assumed to be equal for both the compensating stub and the primary coax line.

$$Z_{Stub} = \frac{Z_o^2}{2Z_{op}} \tag{2.5}$$

It is worth noting that the impedance of the stub is *not* the same as the balun itself, unlike many published designs with symmetrical transmission lines that claim to be Marchand baluns [3].

[3] In [16], the authors show that the symmetrical "Marchand" balun is actually inferior to the simple short stub compensated balun.

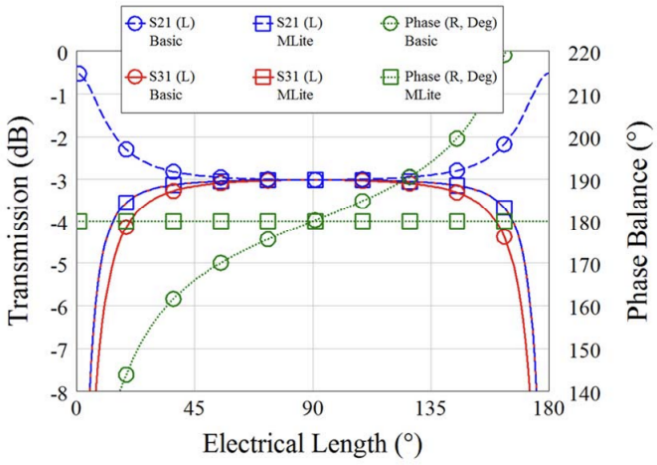

Figure 2.11: The short stub compensated coax balun ("MLite") fixes the amplitude and phase balance, and improves overall bandwidth compared to the basic cable balun. Source [16].

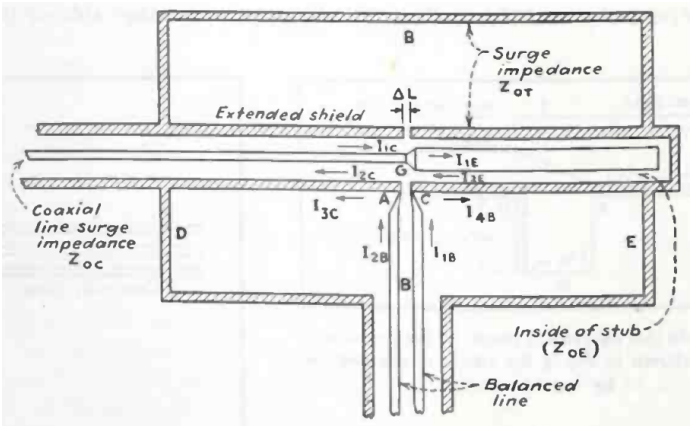

Figure 2.12: The Marchand Balun [18], with a compensating series open stub

It is also worth mentioning a major difference between cable or coaxial baluns and their planar counterparts. In a coax balun, the inner conductor is shielded from the ground plane by the coaxial sleeve; in a planar, coupled line style of balun, the "inner" conductor is exposed to the ground plane beneath it, and therefore has a parasitic transmission line as well. This extra parasitic line helps explains the markedly worse performance of planar structures, most notably the increase in amplitude imbalance present throughout the band [16].

2.5 Summary

In this chapter we have seen that transmission lines can be used as transformers and baluns. While inherently broadband, these structures' bandwidth is inevitably limited by the presence of parasitic ground plane transmission lines. In the next section we will go over designs that cater specifically for CMOS integrated circuits and power amplifiers.

Chapter 3

Transmission Line Transformers for CMOS Power Amplifiers

3.1 CMOS Design

Integrated circuits bring constraints that affect the design of passives. First, for a digital CMOS process (no ultra-thick metals, relatively small height from the substrate to top metal, and a high loss, medium impedance silicon substrate) used here, proper attention must be made to metal and substrate losses; long traces have simply too much conduction (skin effect) and substrate losses to be used here. Pseudo-coaxial ("rectax") structures are difficult, if not impossible to make (unless you have four thick metal layers), and we therefore need to remember the coupled line even/odd mode considerations mentioned in the previous chapter (as we are forced into planar type structures). And lastly we need to be aware of the return or "ground" paths, as on-chip this is unclear at best unless explicitly designed for.

3.2 Guanella Design

As mentioned in section 2.3, for a coupled-line Guanella transformer, we wish to maximize the even mode impedance. One benefit of integrated circuits over PCBs is that we do not have to have a ground plane underneath our structure, and the height between adjacent metal layers is typically quite small[1]. All of this lends itself well to broadside coupled lines, which offer relatively tight coupling coefficients (since the separation between metal layers is minimal), and therefore a low odd-mode impedance. We can further enhance the even mode impedance by using coplanar feeds, where the return or ground path is coplanar to the signal and not through a plane underneath. It is important to distinguish the difference between removing the ground shield for transmission lines versus inductors. For CMOS inductors at K/Ka frequencies, the use of patterned ground shields is not recommended as the

[1]This is not always a benefit, since very thin dielectrics mean the height in structures like microstrip lines are very small

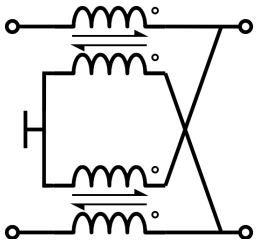

Figure 3.1: Guanella Scehmatic, differential currents shown.

combination of ultra-thin metals and the skin effect makes the conductive losses prohibitive. While patterned ground shields do attempt to reduce eddy current (magnetic) losses, the general shield loss (electric, simply capacitive coupling to the shield) becomes overwhelming. It is therefore simply better to expose the more resistive (undoped) substrate to reduce both the magnetic and electric coupling currents (and therefore the i^2R loss) at these frequencies. For our transmission line transformers, we are primarily focused on increasing the even mode impedance, and therefore would not want a ground plane underneath even if the lower metals were near perfect conductors.

An example CMOS Guanella balun is shown in figure 3.2. The light blue ring represents a defined common mode ground/return path; the substrate itself is left undoped and exposed. Note the even mode cannot arbitrarily be increased by expanding this ring because the substrate itself can act as a (lossy) return path. The coupled coils are implemented via broadside coupled lines. The connection between the 'inner' coils can be seen by the vias at the bottom of the broadside structure (purple); this is where a center tap could be provided as well if desired. Note the shape of the structure is irrelevant; this is not a flux inductor where circular forms are optimal. As transmission lines, they can be meandered as one wishes; mitering is not necessary as the stray corner capacitance here is negligible for such small features. The structure transforms twelve and a half ohms to fifty ohms; as per equation 2.2, the characteristic impedance of the coupled lines was designed to be 25 ohms. The total length of each line is approximately 1/8 to 1/10 of the center wavelength.

The resulting S parameters, simulated in EMX, are shown in figure 3.3. From S21 we see relatively low loss, and the Smith Chart shows broadband matching (swept from 14GHz to 50GHz) for the 12.5:50 ohm transformer. Note the resulting Smith Chart contour is mostly tracing the inevitable extra parasitic shunt capacitance, indicating excellent matching overall.

3.2.1 The Guanella as an Autotransformer

Redrawing the 4:1 Guanella schematic such that all coils are arranged linearly, as shown in figure 3.4 reveals an interesting insight: the 4:1 Guanella looks identical to that of an auto-transformer! This means that by changing the relative lengths of each coupled coil (or equivalently moving the output tap points, akin to an autotransformer), we can achieve

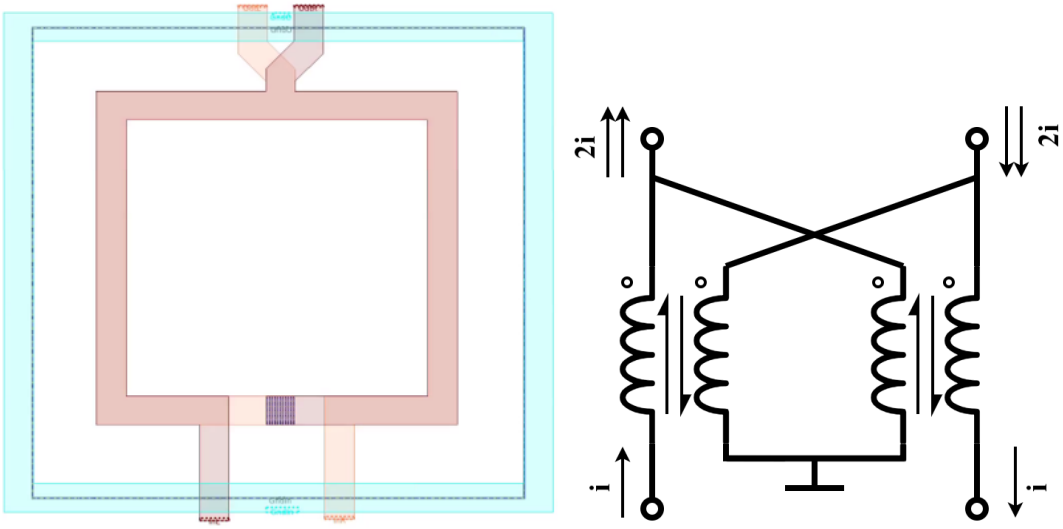

Figure 3.2: A 12.5:50 ohm broadside coupled Guanella transformer; a center tap can be provided at the vias on the bottom. The corresponding schematic is shown on the right.

lower impedance ratios if so desired. This will be utilized in later in our PA design.

3.3 Guanella Baluns

In RF integrated circuit design, differential topologies are preferred as the lack of a well-defined global ground plane makes simulating the return path of single-ended designs (and therefore inductance) difficult. However antennas are typically single ended, thus necessitating the use of a balun to interface with the antenna. For a PA, it would be beneficial to incorporate the balun into the output matching network to save both area and insertion loss. As mentioned previously, a balun must provide voltage balance on the balanced side, and current balance on the single-ended side. Looking at the Guanella transformer, it is clear that the center tap provides voltage balance (it is a virtual ground for odd-mode signals). However, there is nothing that enforces only odd-mode currents to exist (a common mode current can flow through through the Guanella and into the center tap). To remedy this, as suggested in [20], one can use a simple transmission line section (or 1:1 transformer), as shown in 3.5. In a TEM transmission line, the return current have opposite phase of the transmitting current, and therefore we enforce current balance (odd mode currents only) at the unbalanced output. It is worth mentioning that Ruthroff had implemented the same configuration back in 1959 in a wire and ferrite form, shown in figure 3.6 [11].

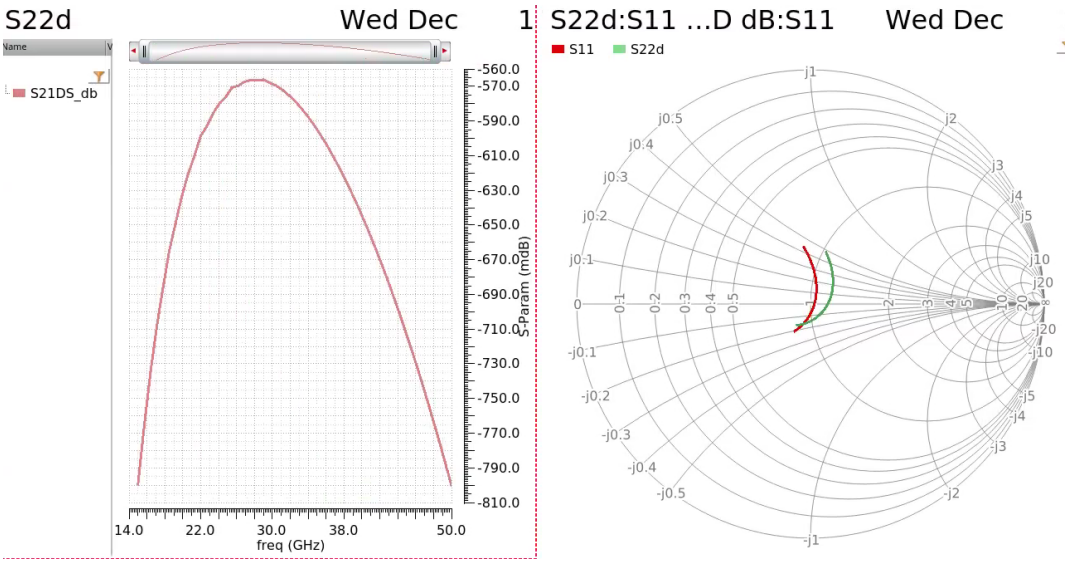

Figure 3.3: S-Parameters for 3.2. Source and load impedance are 12.5 and 50 ohms differential.

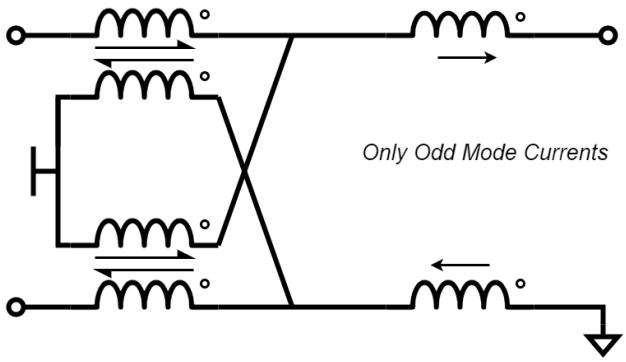

Figure 3.5: The addition of a transmission line section to the 4:1 Guanella enforces current balance at the unbalanced output and therefore creates a balun.

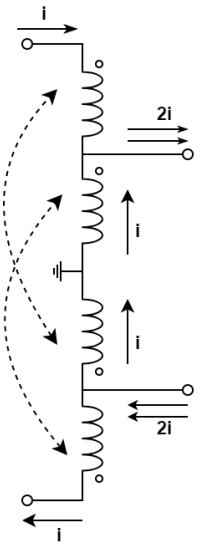

Figure 3.4: The 4:1 Guanella drawn 'unfolded' as an autotransformer.

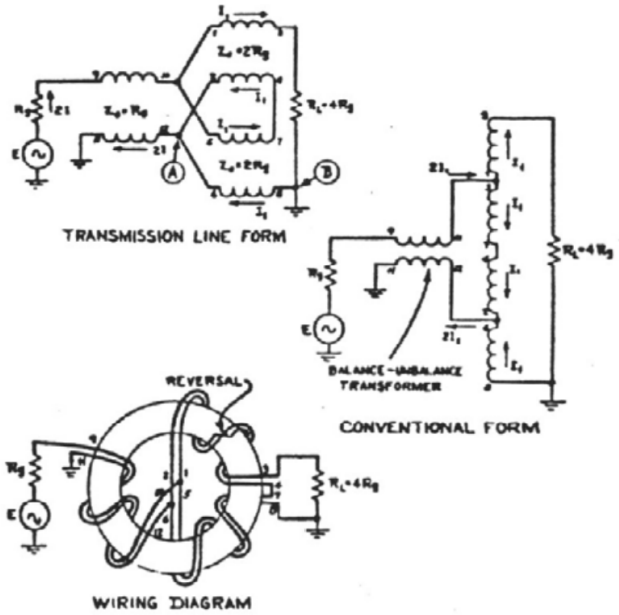

Fig. 4—4:1 Impedance transformer. Unbalanced—symmetrical.

Figure 3.6: Ruthroff used the same modification to Guanella's transformer back in 1959 to create a balun [11].

An example design is shown in figure 3.7. The top octagon shape is the 4:1 Guanella transformer, with a center tap provided to feed in the supply voltage. A meandered transmission line followed by an ac coupling capacitor complete the design. Note the Guanella uses the top two (relatively thick) copper metal layers, M9 and M10 for both performance and electro-migration considerations (as they have to handle the large DC bias current of a PA). The output transmission line (50 ohms) uses M9 and AP, the aluminum redistribution layer. Note that because DC current does not need to flow through the transmission line section, we can use the aluminum top-metal layer (aluminum has about three times less current handling capability for this process). The structure was simulated in Cadence EMX Designer (EMX) and found to have a useful matching bandwidth (S11<-10dB) of 16-40GHz, amplitude mismatch <0.47dB (nominal 0.3dB) across the band; phase mismatch <3 degrees; and insertion loss of approximately 0.9-1dB. The area used is 260um x 200um. Both the Guanella and transmission line sections use broadside coupled lines. Note unlike the basic cable balun, we do not need add an explicit compensating short stub (the 'MLite' balun) here for two reasons: first, as a broadside coupled line with coplanar return paths, both of the differential paths have approximately identical parasitic stub impedances; second, the high even-mode impedance from the broadside lines also improve the overall balance, regardless of symmetry.

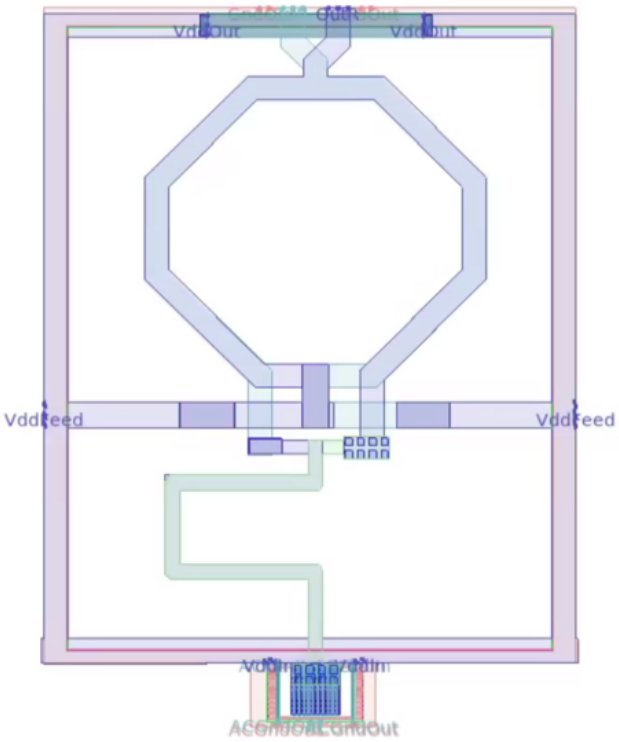

Figure 3.7: An example 50ohm single ended to differential 12.5 ohm balun, with an output coupling capacitor to block DC signals.

As the signal paths are transmission line based, and not flux based, we can of course happily meander the lines to create a smaller footprint, as seen in figure 3.8. The area here was shrunk by 30%, (235um x 155um) with marginal degradation to the electrical performance.

An important practical consideration when designing these structures is setting up the ports correctly in EMX. First, it is imperative to have a defined ground or return path, here given by the outer ring. This ring needs to have low loss, and should match exactly how it will be utilized in the final layout. In this design, separate ports were used to define the grounds at both the input, output, and supply feeds (each port is referred to a "global" ground EMX, which is essentially modeled as an infinite ground plane underneath the chip); the resulting balun performance was then determined by properly connecting the ports in schematic simulations, defining the sources between the signal ports and the locally defined ground ports. An example illustrating the importance of the return path is shown in figure 3.9, where the lack of symmetry in the return path (from where the grounded transmission line is) causes the phase mismatch to worsen to +/-12% over the bandwidth.

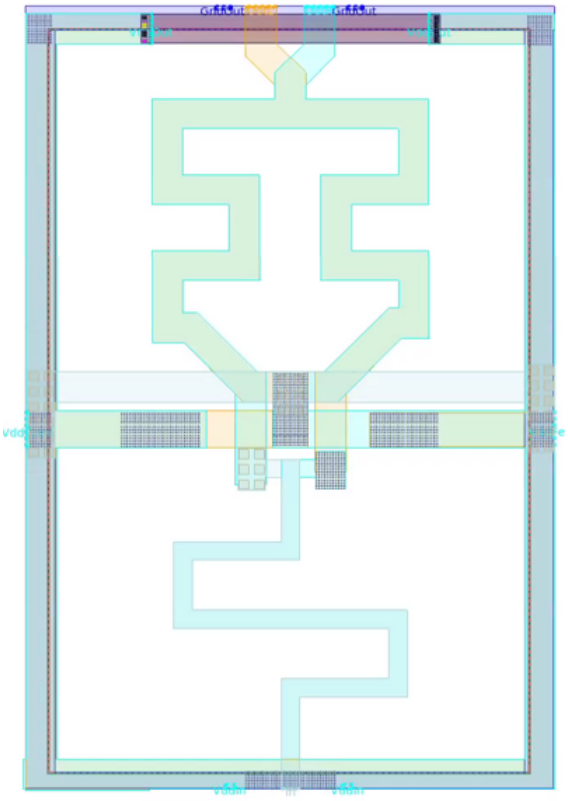

Figure 3.8: Meandered version of the Guanella Balun

3.4 Transmission Line Power Combining

In PA design, increasing the transistor size in an effort to increase the output power delivered unfortunately results in degraded efficiency due to the increased routing losses and parasitic reactances. In CMOS design especially, for real designs with layout parasitics included this effect worsens non-linearly (notice doubling the cell count does not double the achievable output power) as the transistor increases in size. This is shown in table 3.1, where the number of cells represents increasing transistor sizes; the corresponding drop in performance is readily evident. This reality leads the designer to consider the two types of power combining, series and parallel combining.

In series combining, the voltage on the primary is the sum of the voltages on the secondaries (the current has to remain the same); if each terminal impedance is the same, then the input impedance is simply the sum of all the terminal impedances (see figure 3.10). At first glance, this seems perfect for power amplifiers. As an example, imagine a fifty-ohm antenna interfacing with four identical amplifiers combined in series; in this scenario

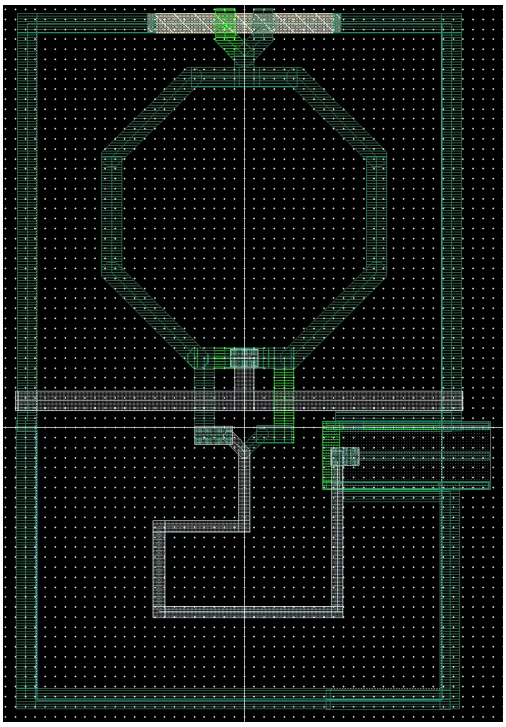

Figure 3.9: An example of a poor Guanella balun, where the lack of symmetry in the return path provides worsens the phase balance.

Table 3.1: Maximum achievable PAE and output power vs transistor size in a bulk 28nm CMOS process, at 40GHz

Number of Cells	Pout (dBm)	Pout (mW)	PAE (%)
9	11	12.6	37
12	12.2	16.6	35
18	13	20	30
25	13.8	24	26

everything is matched when each amplifier is matched to a 12.5Ω impedance, meaning we have simultaneously power-combined and provided impedance transformation. A significant problem arises however when we consider lossy (real) transformers. Notice that current (and therefore power) at the primary (antenna or input side) of the transformer flows through all the primaries of the N transformers. This means that as we increase N, the IR drop (from undesired conductive losses) across the primary increases as well, rapidly diminishing the transformer efficiency as additional stages are added, making it an impractical method to increase the overall output power after a certain number of transformers (for more details,

see [21])[2]. This is quite important in digital CMOS designs, where the transformer quality factors are quite low (around 10).

The alternative to series combining is naturally parallel combining (figure 3.11). Here, notice that the efficiency of the overall combined transformer is equivalent to the efficiency of each individual transformer section (as they are all simply combined in parallel- they do not share the same primary). Therefore there are no additional passive losses when increasing N. However, the downside with parallel combining is that the impedance transformation goes against what a PA designer typically wants, as the input impedance (impedance seen by the antenna) decreases as we add more stages. To match to a 50 ohm antenna with $N = 4$ for instance requires the impedance of each branch to be 200 ohms- good luck getting any power out of each amplifier with that!

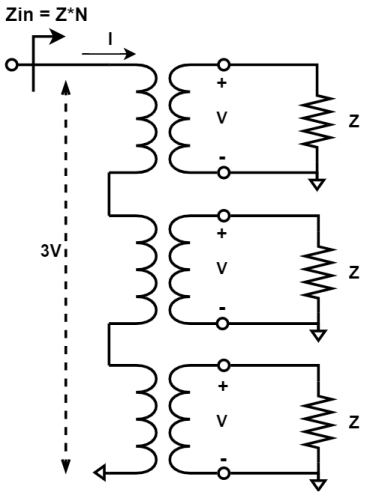

Figure 3.10: Illustration of series combining. Note how the voltages add on the primary.

[2]As described in [21], the choice between series and parallel designs will depend on the design specification of the PA and the process used, as there is a trade-off between the efficiency of each individual amplifier and the efficiency of the transformer-combiner. Unfortunately in practice this is not a readily apparent decision for the designer to make, and often requires both designs to be made and compared. In this work we will simply show a method to make parallel combining a viable option for designers.

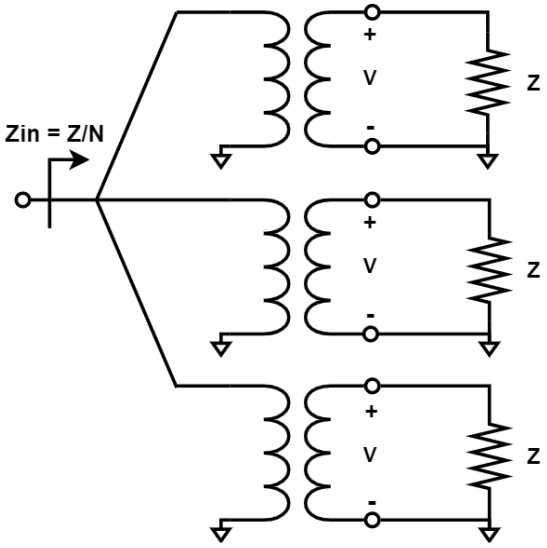

Figure 3.11: Illustration of parallel combining

3.4.1 Guanella Transformers for Parallel Combining

To make parallel combining feasible, we need to counteract the unfavorable impedance transformation. For flux inductors the simplest solution is to simply increase the turns ratio of each individual transformer, and increase the impedance seen at each primary. An example is shown for $N = 3$, with a fifty ohm source, in figure 3.12. While this is certainly realizable in CMOS, it comes with several drawbacks. First, adding additional turns increases the loss of each transformer, reducing the overall efficiency. And second, as mentioned before, flux transformers already have low bandwidth due to their low coupling coefficients in IC design; adding additional turns only exacerbates this. Also note further increasing the turns ratio above two is practically impossible for K and Ka band designs (and the coupling coefficient k typically will not exceed 0.7). To remedy these issues, we can simply replace each flux transformer with their transmission line equivalents! A four to one Guanella can simply be thought of as the n=2 transformer in figure 3.12, except with a considerably larger bandwidth.

An example Guanella parallel combiner is shown in figure 3.13. Note using transmission line transformers provides some additional practical benefits for IC designs. The connecting transmission lines (which again enforce current balance) can be used to tie all the branches together (the total length of each line is roughly $\lambda/10$, enough to act as a distributed and not a lumped element). This allows for the entire structure to be meandered/condensed into a relatively small area, here of 450um x 376um. As expected of a transmission line based structure, it has a wide bandwidth of 18-40GHz, approximately 0.9dB insertion loss, worst case four degrees of phase mismatch, and 0.3dB amplitude mismatch.

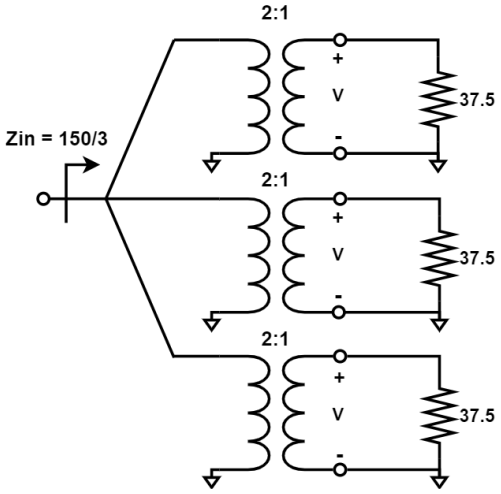

Figure 3.12: An example of a matched parallel combined case with transformer turns. Here the transformer turns ratio causes a 4:1 increase in the impedance presented by each transformer to the antenna. If $Z = 37.5\Omega$ then, we have a match to a 50Ω antenna.

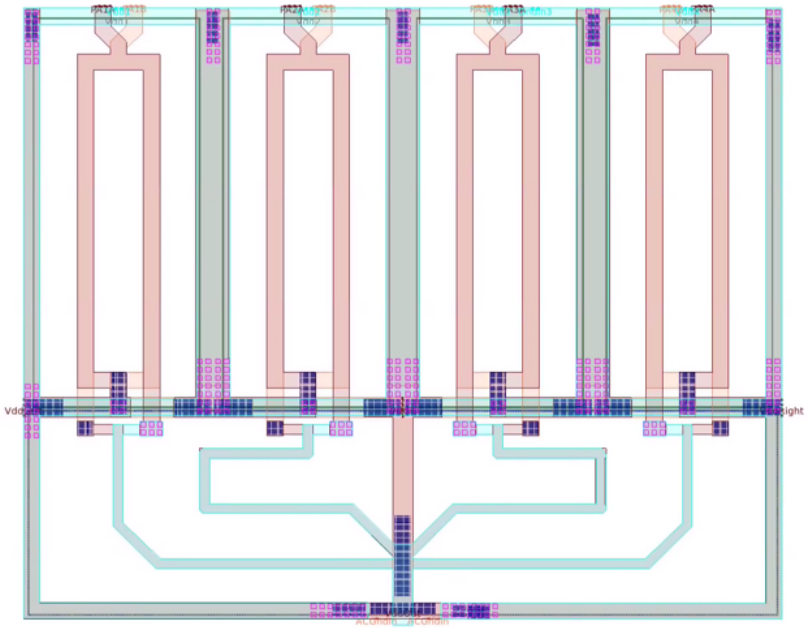

Figure 3.13: A four way combiner with 4:1 Guanella transformers

Chapter 4

K and Ka Band Power Amplifiers

This chapter details the design of several variations of power amplifiers that span the entire K (18-27 GHz) and Ka (27-40 GHz) spectrum.

4.1 50 : 12.5Ω Guanella PA

One useful application for a Guanella balun is as the balun and impedance transformer for a PA. Assuming a fifty ohm antenna, the 4:1 Guanella, when designed properly, will provide a 12.5 ohm, real impedance to the amplifier; note parasitic capacitance from the wirebond/probe pads and from fringing capacitances (in the passives) will also provide a small reactive component to this impedance. The transistors were thus sized such that their optimum load impedance (for PAE or Pout) matches the 12.5 ohms provided by the Guanella. The loadpull of a standard cascode configuration sized for roughly 12.5 ohms in 28nm CMOS is shown in figure 4.1, at 40GHz. Note at lower frequencies (20GHz), the loadpull is fairly similar, with the optimal impedances closer to 15 ohms, though naturally the maximum achievable values are much greater. This suggests that octave bandwidth designs are certainly feasible if we keep the load impedance presented to the transistors roughly constant over frequency.

To this end, a 12.5 to 50Ω Guanella balun was designed, shown in figure 4.3, (metrics provided in the caption). The impedance seen by the PA with this Guanella balun is plotted in figure 4.2. We can see that with the Guanella on its own, the impedance is close to real axis (as expected, mentioned in the previous paragraph). However glancing at the load pull (figure 4.1), we can see that both the output power and PAE prefer a slightly inductive load (commonplace for most CMOS PA designs). It therefore makes sense to add a small amount of series inductance, shifting the impedance as seen on the right side of figure 4.2.

The resulting structure is shown in figure 4.4. The cascode transistor cell is on the top. From here, we have the small series inductance (20pH) and Guanella Balun which provides the impedance shown in figure 4.2. However there are two more important practical considerations to note: first, the Guanella Balun is DC coupled, meaning we need to provide

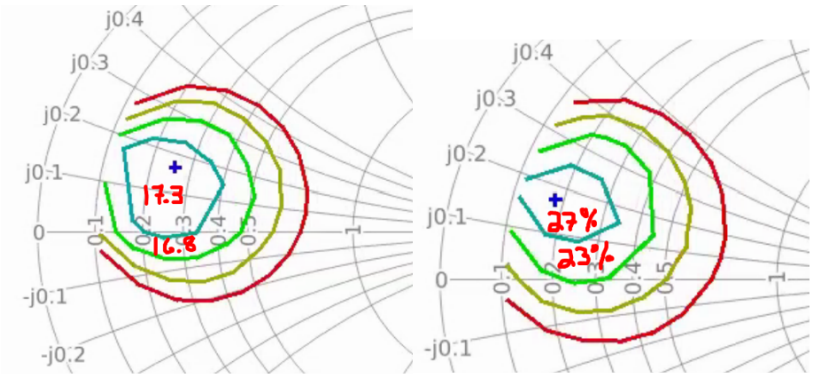

Figure 4.1: Sample 40GHz loadpull; left is saturated output power (in dBm), right is PAE. Both for a 1V supply, with a cascode transistor.

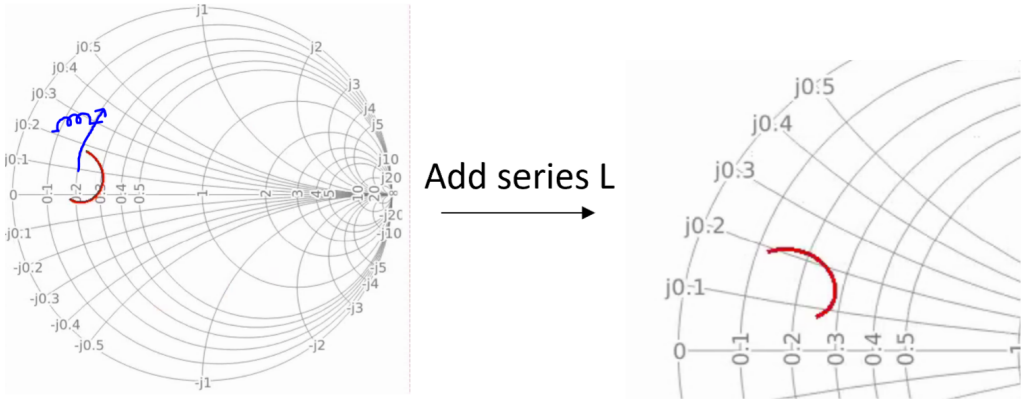

Figure 4.2: Walkthrough of the output matching. Smith chart sweeps are from 14 to 40GHz. The left image shows the differential impedance seen with the GSG pad and Guanella Balun. The right shows the resulting impedance after a series impedance is provided prior to the Guanella.

an AC coupling capacitor to break the DC short circuit from VDD to ground. This is placed directly after the balun. Then, to get off chip, there needs to be some sort of pad, which provides a significant amount of unwanted capacitance (roughly forty picofarads); this needs to be resonated out, provided by a second series inductor placed right before the pad.

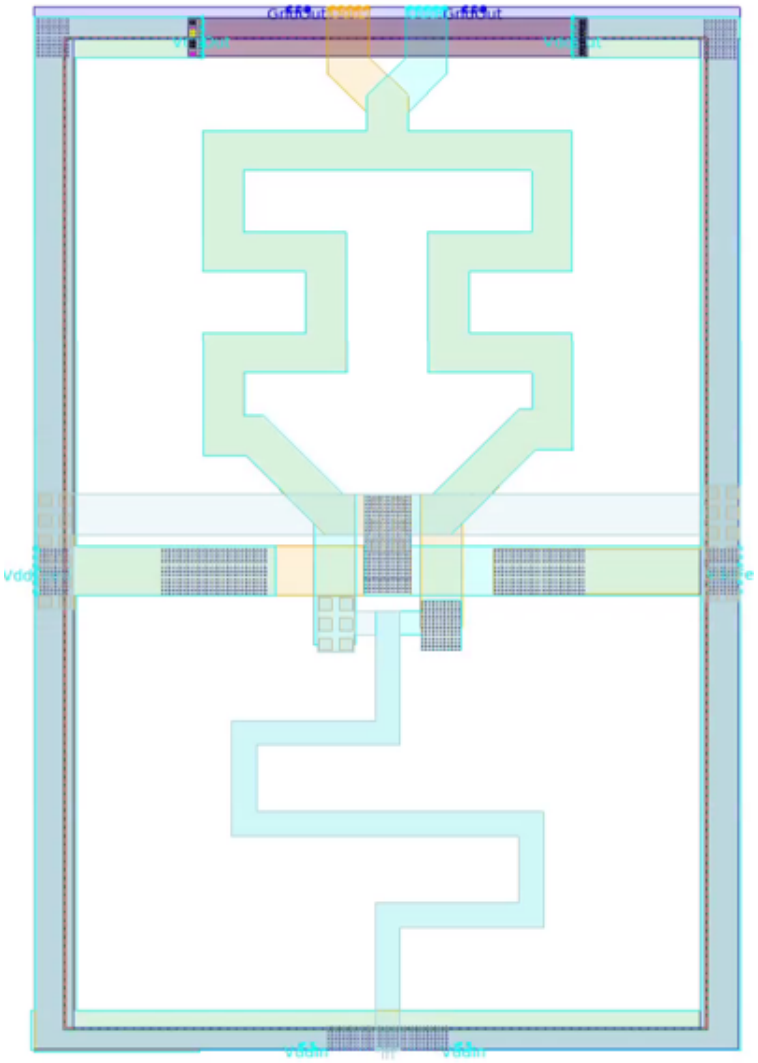

Figure 4.3: Meandered 12.5 : 50Ω Guanella balun. Provides < 0.6dB amplitude mismatch (typical 0.4dB), 175-178 degrees phase difference, and approximately 1dB insertion loss from 18 to 40GHz.

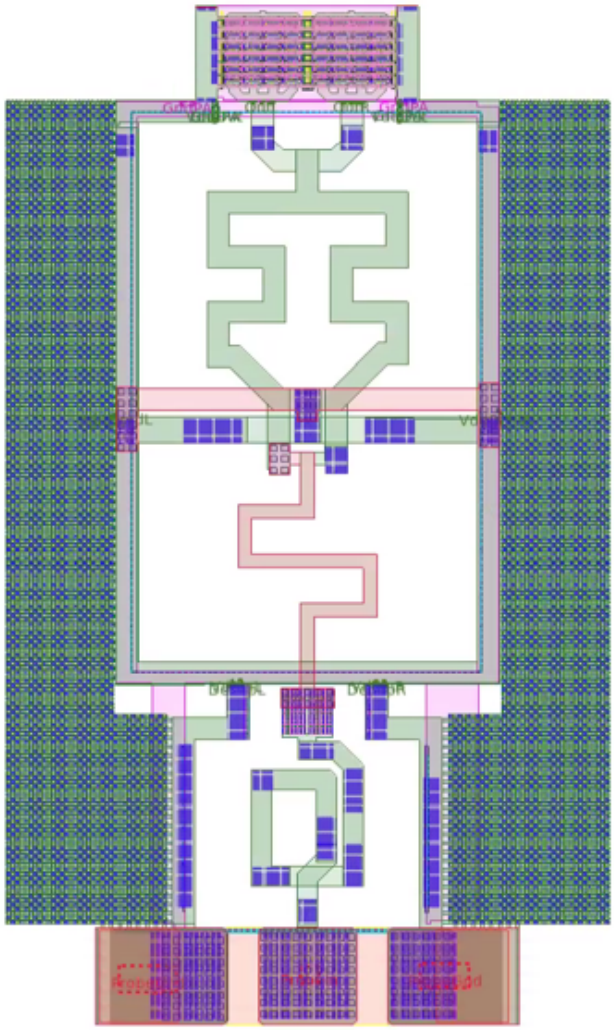

Figure 4.4: Output matching structure, containing, from top to bottom, the cascoded differential transistors, a small (20pH) series inductor, the Guanella balun, an AC coupling capacitor, and an additional series inductance to resonate out the bottom probe pad.

4.1.1 Practical Implementation and Simulation Details

Here we will talk about setting the EM simulation correctly. In figure 4.5, we show the location of all the ports in EMX. First note we include both the RF ground ring as well as the power grid to make sure we include the return path in simulation. All the ports are referenced to a 'global ground;' none of the ports are considered to be local grounds in EMX- this will be done in schematic. We are using GSG probes, so both ground pads are electrically shorted to each other, as the probe is presumably a coaxial-like structure. VDD

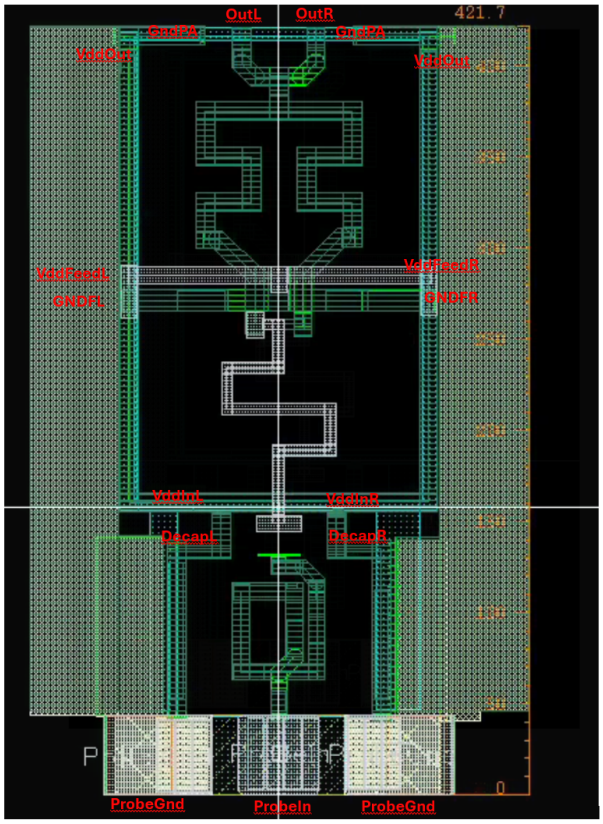

Figure 4.5: Location of all the EMX ports. Note that we have included the power grid on the sides

and GND are made separate to model the decoupling capacitance ('decap') requirements (once there is plenty of decap, this turns out to be inconsequential, but it's worth figuring out what 'plenty' entails).

In figure 4.6, we show a correct setup of the Virtuoso schematic simulation. We need to first pick a reference (ground) location; this is arbitrary, but because Virtuoso's ideal balun contains a global ground in its schematic, we must put our ground with the that component's ground. All other ports then are referenced to their local 'ground' (the proper EMX port). In figure 4.7, we show an incorrect setup. Here we see that we have ground at both input and output ports, (remember the Virtuoso balun is referred internally to the global ground), meaning we have shorted out the return path! As a result, the impedance trajectory on the Smith Chart is incorrect (and rather strange looking).

The additional inductor (called 'L2' here) added to resonate with the GSG pad improves the efficiency and output power at higher frequencies, though at lower frequencies the efficiency is slightly degraded. This is shown in Figs. 4.8 and 4.9.

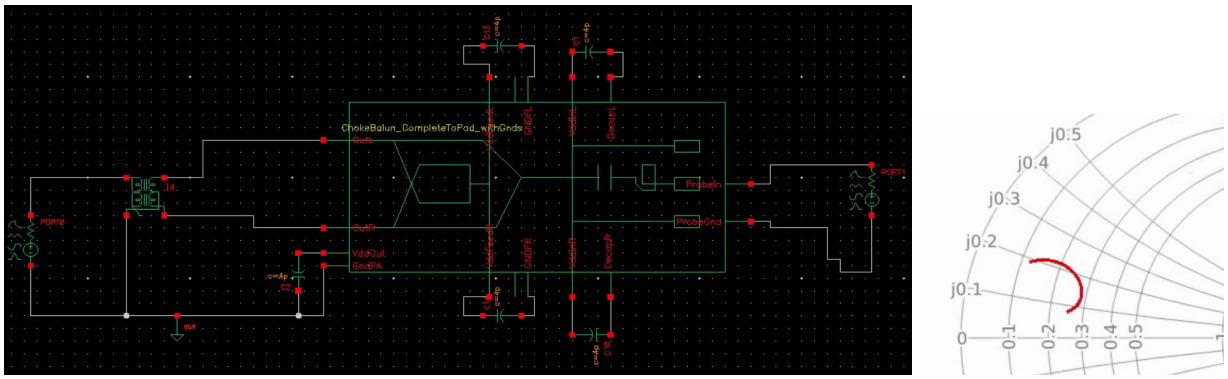

Figure 4.6: Example of a correct schematic setup, and the resulting smith chart plot.

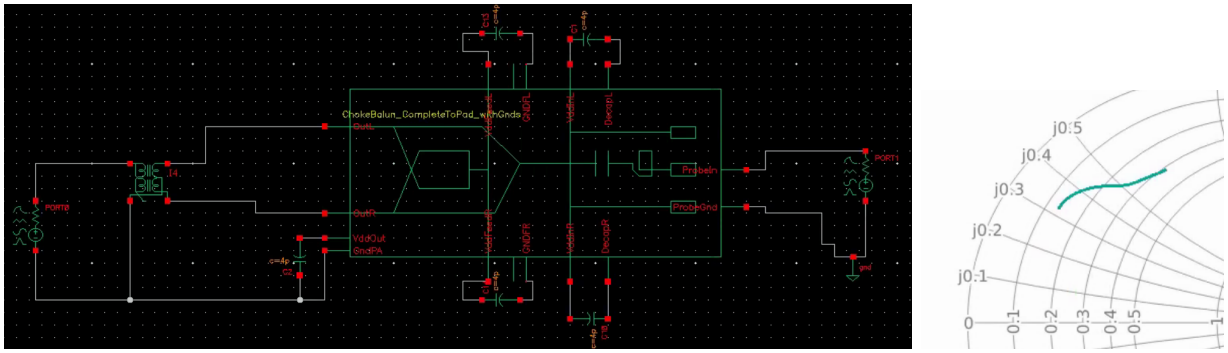

Figure 4.7: An incorrect schematic setup, as Virtuoso's balun contains a ground, meaning the grounds on both sides are electrically shorted. The resulting Smith Chart trajectory is thus off.

4.2 Interstage Matching

Providing a broadband interstage matching network is challenging on-chip. Due to the low power gain of CMOS output stages (at these frequencies), the efficiency of the preceding stage (and the interstage match) is critical; one cannot simply just cascade matching networks for bandwidth because the passive losses will severely degrade the overall PAE. Exacerbating the problem is that the last stage of a power amplifier typically has a large transistor size (to provide more power), and therefore provides a large capacitance (with ideally low series resistance) to match to, a rather difficult task for a broadband match. Even when adding a shunt load resistance to the gate, at high frequencies the gate capacitance impedance is considerably lower than that of a (reasonable) shunt resistor. In this section we will explore how the Guanella transformer provides an alternative method for matching.

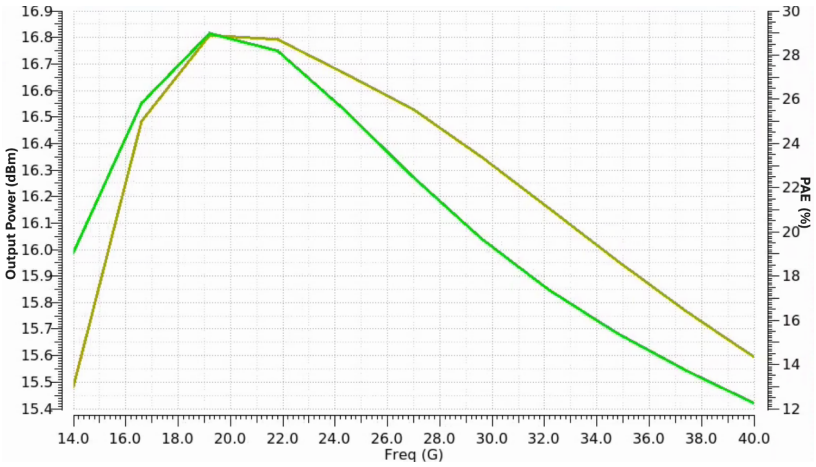

Figure 4.8: Output power (yellow) and PAE (green) of the output stage with additional (L2) inductor.

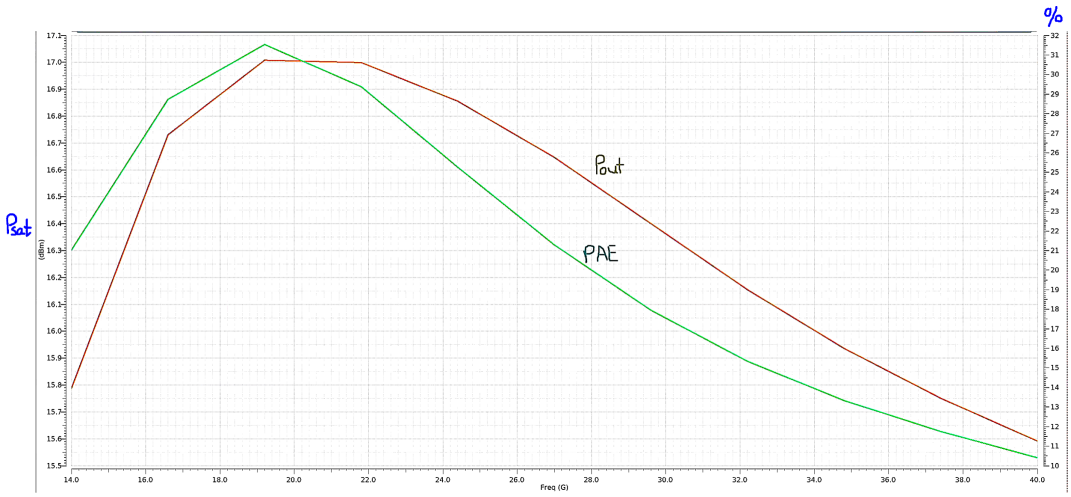

Figure 4.9: Output power and PAE of the output stage without additional (L2) inductor. The primary Psat tick marks start at 15.5dBm and increment by 0.1dBm. PAE starts at 10% and increments by 1.

4.2.1 Guanella Low-Side Impedance

In the next two sections we will derive the input and output impedances of the Guanella 4:1 transformer, assuming finite (and lossless) transmission length. We know for a simple transmission line, illustrated in figure 4.10, the following equations (from the telegrapher equations):

$$V_2 = V^+ e^{-j\beta l} + V^- e^{j\beta l}$$
$$I_2 = (V^+/Z_o)e^{-j\beta l} - (V^-/Z_o)e^{j\beta l}$$

Here the line is assumed to go from the left (l=0) to the right. We can apply Euler's theorem

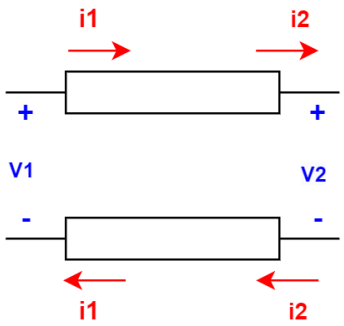

Figure 4.10: A general transmission line

to re-write this in terms of sine/cosine, substituting for V^+ and V^- as well.

$$V_2 = V_1 \cos(\beta l) - jI_1 Z_o \sin(\beta l) \tag{4.1}$$
$$I_2 = I_1 \cos(\beta l) - j(V_1/Z_o) \sin(\beta l) \tag{4.2}$$

To find V_1 and I_1 we flip the reference point and the sign of the sine term [22].

$$V_1 = V_2 \cos(\beta l) + jI_2 Z_o \sin(\beta l) \tag{4.3}$$
$$I_1 = I_2 \cos(\beta l) + j(V_2/Z_o) \sin(\beta l) \tag{4.4}$$

It is worth noting these two equations are used in Ruthroff's derivation of his namesake transformer, seen in Appendix A of [11]. We will see the result of the Guanella differs from that of the Ruthroff transformer.

We will first derive the input impedance of the low-impedance side of the Guanella transformer, using the method suggested by [23]. Assume each TLT has length l and characteristic impedance Z_o. We set up our schematic as shown in figure 4.11. By inspection:

$$V_1 = V_3 = V_{in}$$
$$I_2 = I_4$$

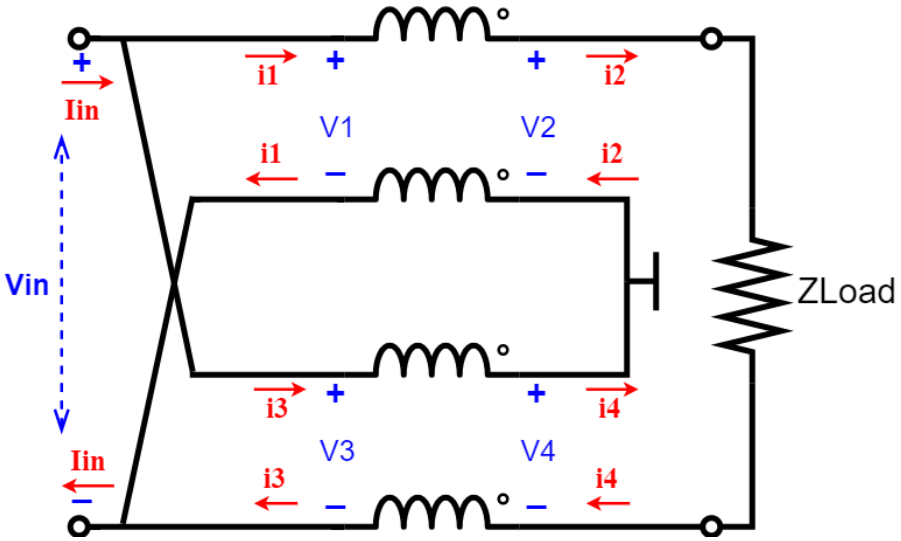

Figure 4.11: Guanella low-side impedance derivation.

By symmetry, we can see as well that:

$$V_2 = V_4$$
$$I_1 = I_3$$

This gives us the simplified diagram shown in figure 4.12. We can now set up our system of equations as follows:

$$I_{in} = 2I_1 \tag{4.5}$$
$$V_{in} = V_1 \tag{4.6}$$
$$I_2 Z_{Load} = 2V_2 \tag{4.7}$$

Now plugging in equations 4.1 and 4.2 into 4.7 and replacing V_1 with V_{in} and I_1 with $I_{in}/2$, we get:

$$Z_{load}(I_{in}/2)\cos(\beta l) - jZ_{load}(V_{in}/Zo)\sin(\beta l) = 2V_{in}\cos(\beta l) - j2(I_{in}/2)Z_o\sin(\beta l)$$

Grouping like terms:

$$V_{in}[2\cos(\beta l) + j(Z_{load}/Z_o) * \sin(\beta l)] = (I_{in}/2)(Z_{load}\cos(\beta l) + j2Z_o\sin(\beta l))$$

Which gives us the input impedance of the low-side:

$$Z_{in} = \frac{Z_{Load}\cos(\beta l) + j2Z_o\sin(\beta l)}{4\cos(\beta l) + j2(Z_{Load}/Z_o)\sin(\beta l)} \tag{4.8}$$

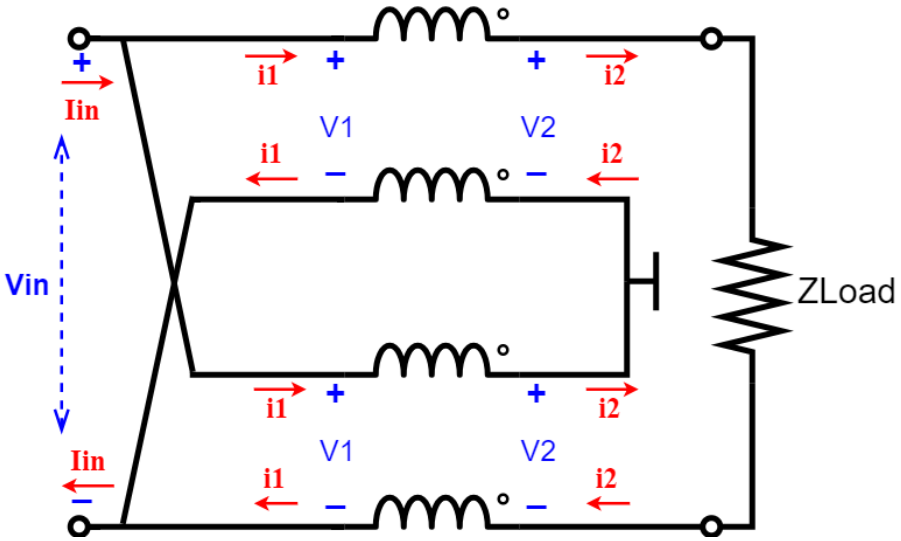

Figure 4.12: Simplified set-up of the low side impedance, after using symmetry observation.

This is equivalent to equation 2.2, re-written below to help the reader's recollection, as $\tan(x) = \sin(x)/\cos(x)$.

$$Z_{in} = \frac{1}{2} Z_o [\frac{Z_L/2 + jZ_o \tan(\beta l)}{Z_o + j(Z_L/2)\tan(\beta l)}]$$

4.2.2 Guanella High-Side Impedance

Let us now derive for ourselves the high-side impedance of the 4:1 Guanella. Our set-up is shown in figure 4.13.

We can write the following system of equations:

$$V_{out} = 2V_2 \tag{4.9}$$
$$I_{out} = I_2 \tag{4.10}$$
$$2I_1 Z_s = V_1 \tag{4.11}$$

This time we substitute equations 4.3 and 4.4 into 4.11.

$$2Z_s[I_2 \cos(\beta l) + j(V_2/Z_o)\sin(\beta l)] = V_2 \cos(\beta l) + jI_2 Z_o \sin(\beta l)$$

Rearranging, and replacing I_2 with I_{out} and V_2 with $V_{out}/2$ gives us:

$$\frac{V_{out}}{2}\frac{2jZ_s}{Z_o}\sin(\beta l) - \frac{V_{out}}{2}\cos(\beta l) = jI_{out}Z_o \sin(\beta l) - 2Z_s I_{out} cos(\beta l)$$

An intermediate algebra step:

$$I_{out}[2Z_s \cos(\beta l) - jZ_o \sin(\beta l)] = V_{out}[(1/2)\cos(\beta l) - j(Z_s/Z_o)\sin(\beta l)]$$

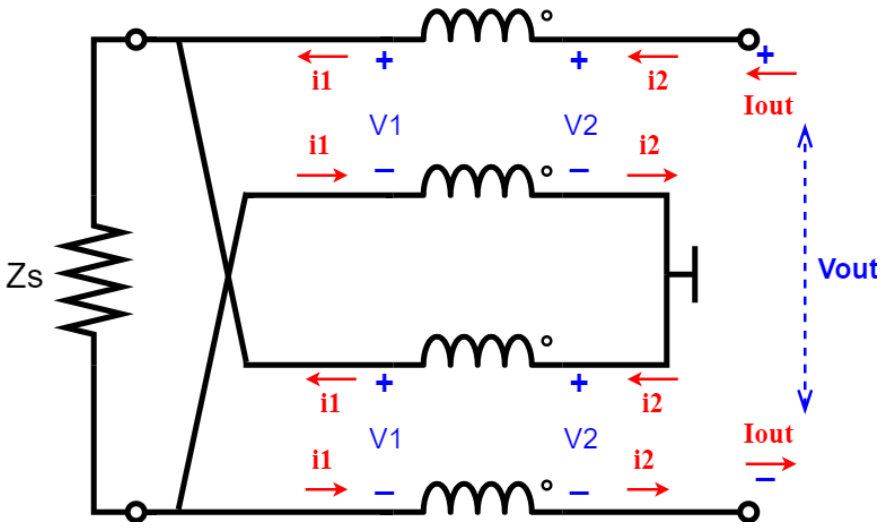

Figure 4.13: Set-up to derive the high-side impedance.

This gives the following impedance relation:

$$Z_{out} = \frac{2Z_s\cos(\beta l) - jZ_o\sin(\beta l)}{(1/2)\cos(\beta l) - j(Z_s/Z_o)\sin(\beta l)} \tag{4.12}$$

This can be simplified to the following:

$$Z_{out} = 2Z_o\frac{2Z_s - jZ_o\tan(\beta l)}{Z_o - 2jZ_s\tan(\beta l)} \tag{4.13}$$

As a sanity check, notice when $Z_o = 2Z_s$, $Z_{out} = 4Z_s$ for all frequencies, as expected. Also of interest is that as $l \to 0$, $Z_{out} \to 4Z_s$ regardless of Z_o.

4.2.3 Comparison of the Ruthroff and Guanella Transformers

The low-side impedance of the Ruthroff transformer is given by [11]:

$$Z_{in} = Z_o\frac{Z_{Load}\cos(\beta l) + jZ_o\sin(\beta l)}{2Z_o(1+\cos(\beta l)) + jZ_{Load}\sin(\beta l)} \tag{4.14}$$

Notice for the Ruthroff transformer the condition $Z_o = Z_L/2$ does *not* provide matching for all frequencies. Instead one gets a complex impedance that is dependent on the line length (only when $l = 0$, $Z_{in} = Z_L/4$). This is most easily seen when we chose $(\beta l) = \pi/2$.

$$Z_{in} = Z_o\frac{jZ_o}{2Z_o + jZ_{load}}$$

From here we can see that so long as the line length is not infinitesimally short, we will never get a perfect 1:4 match for the Ruthroff transformer, irrespective of the characteristic impedance. Therefore for our PA design, given the choice between the Ruthroff and Guanella transformer, the Guanella transformer is preferable. Luckily in IC design fully differential designs are commonplace, and frankly preferred, making this an easy decision.

4.2.4 Guanella Transformers For Interstage Matching

Let us consider the low-side impedance equation, (eq. 2.2), and assume our load is an ideal capacitance, $Z_L = 1/(j\omega C)$, similar to a large MOSFET gate. We can simplify the equation to then:

$$Z_{in} = \frac{1}{2} Z_o [\frac{-j + j2\omega C Z_o \tan(\beta l)}{2\omega C Z_o + \tan(\beta l)}] \tag{4.15}$$

We can re-write this as follows:

$$Z_{in} = \frac{1}{4\omega C + 2\tan(\beta l)/Z_o} [-j + j2\omega C Z_o \tan(\beta l)] \tag{4.16}$$

We see that the imaginary component will be completely canceled when:

$$2\omega C Z_o \tan(\beta l) = 1 \tag{4.17}$$

Looking at the numerator, we can see that the transformer in a way acts like a series short stub, and certainly Z_o can be chosen such that there is no imaginary part at a specific frequency. If we set the length of the Guanella to be $\lambda/8$, to zero the imaginary component we use equation 4.17 to choose $Z_o = 1/(2\omega_o C)$, where w_o is the frequency at which the Guanella is $\lambda/8$. The input impedance then varies as follows (normalized to Z_o):

$$\frac{Z_{in}(\omega/\omega_o)}{Z_o} = 0.5 * \frac{-j + j(\omega/\omega_o)\tan(\omega/\omega_o * \pi/4)}{(\omega/\omega_o) + \tan(\omega/\omega_o * \pi/4)} \tag{4.18}$$

Equation 4.18 is plotted in figure 4.14.

A good way to see how this can help inter-stage matching is with an example. Let us assume that the transmission line is 450um long, and the velocity is half the speed of light (typical for CMOS transmission lines); βl is then $2\pi f(1.5 * 10^{-12} s)$. Furthermore, assume the gate cap is 400fF, and we design Z_o to be roughly 25 ohms. Figure 4.15 graphs the imaginary part of the impedance, from 10GHz to 60GHz. For reference, the impedance of a 400fF capacitor at 10GHz is -40j, and at 60GHz it is -6.6j.

The takeaway here is that we have drastically reduced the magnitude of the imaginary part across the entire bandwidth, especially at lower frequencies, completely zeroed it out at one frequency, and made it inductive after, all of which is beneficial to PA design.

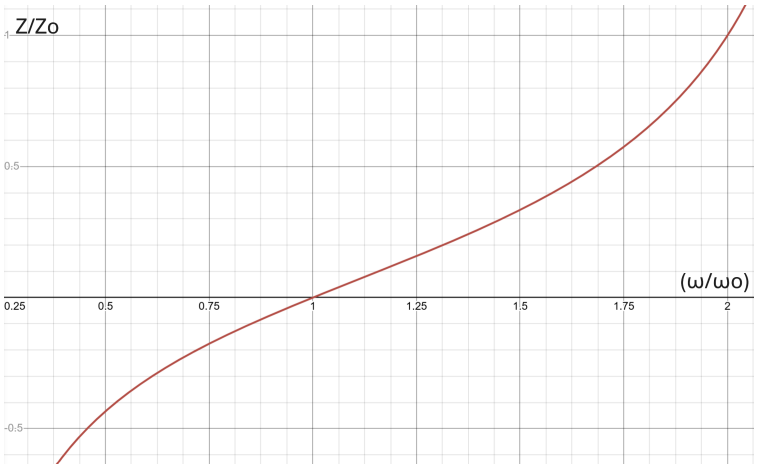

Figure 4.14: Imaginary part of the input impedance for a capacitive load, normalized to Z_o, versus normalized frequency. Z_o is chosen such that the imaginary part is zero at $\omega/\omega_o = 1$.

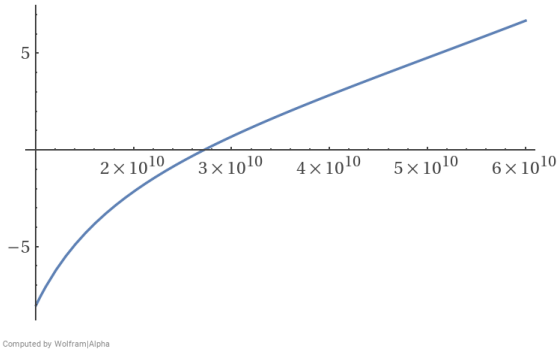

Figure 4.15: Imaginary part of the low-side input impedance for a 400fF cap and a 25 ohm (Z_o) Guanella, ($\lambda/8$ at approximately 27GHz). Note this is *not* normalized.

4.2.5 High Impedance Side of an Interstage Guanella Match

Let us plug in the impedance of a capacitor as the source impedance to the Guanella in equation 4.13. We arrive at the following relation:

$$Z_{out} = 2Z_o \frac{-2j - 2j\omega C Z_o \tan(\beta l)}{\omega C Z_o - 2\tan(\beta l)} \tag{4.19}$$

Once again, we could choose Z_o to eliminate the imaginary part, however this time we require $\tan(\beta l)$ to be less than zero, which requires the line length to be greater than $\lambda/2$.

$$j\omega C Z_o \tan(\beta l) = -1 \tag{4.20}$$

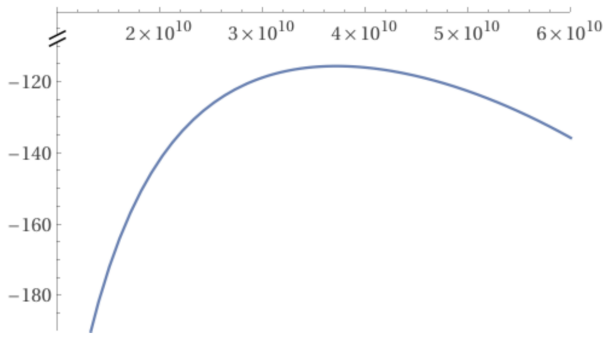

Figure 4.16: Imaginary part of the high side impedance for a 400fF cap and $Z_o = 25\Omega$ Guanella, $\lambda/8 = 27$ GHz.

As a result, one should use a Guanella in reverse (high side) simply to increase the real part, and not to decrease the imaginary part, as longer line lengths lead to increased losses (remember our Guanellas are typically $\lambda/8$ long). Taking a look at an example, using the same 400fF capacitor, 25 ohm Z_o, and $\lambda/8 = 27$ GHz, the imaginary part of the impedance is plotted in figure 4.16. The key here is that while the impedance increases as expected, it actually flattens out over frequency (from the nature of the tangent function), so at higher frequencies we have effectively a smaller equivalent capacitor. By varying the length, Z_o, (and potentially the impedance ratio) then of both a high and low side Guanella, we can potentially create a broadband matching network with a high/low (or vice-versa) in series (and not something that simply 'undoes' the preceding Guanella).

4.2.6 Guanella with Shunt RC Load

For broadband interstage matching, inevitably a shunt resistor is added to the gate (as the network Q of a gate is far too large). We will show here that we do not have to change much, if anything, in our design. Looking at the low-side impedance:

$$\text{For a shunt RC load of } Z_L = \frac{R - j\omega CR^2}{1 + (\omega RC)^2}$$

$$Z_{in} = Z_o/2 * \frac{0.5 * \dfrac{R - j\omega CR^2}{1 + (\omega RC)^2} + jZ_o \tan(\beta l)}{Z_o + \dfrac{j\tan(\beta l)}{2}\dfrac{R - j\omega CR^2}{1 + (\omega RC)^2}}$$

$$Z_{in} = Z_o/2 * \frac{0.5R - 0.5j\omega CR^2 + jZ_o \tan(\beta l)(1 + (\omega RC)^2)}{Z_o(1 + (\omega RC)^2) + jR\tan(\beta l)/2 * (1 - j\omega CR)} \tag{4.21}$$

If we make a key assumption, that $(\omega RC) \gg 1$, we can simplify this mess considerably[1].

$$Z_{in} \approx Z_o/2 * \frac{0.5R - 0.5j\omega CR^2 + jZ_o \tan(\beta l)(\omega RC)^2}{Z_o(\omega RC)^2 + 0.5\omega CR^2 \tan(\beta l)} \tag{4.22}$$

We can now see to zero the imaginary part at a single frequency we need:

$$Z_o \tan(\beta l) = 1/2\omega C \tag{4.23}$$

Just as previously seen in equation 4.17. As an example, let us choose a typical R, of 100 ohms. When we use the same scenario as figure 4.15, with a large gate capacitance, we can plot the imaginary and real parts, and see that the imaginary part behaves in the same manner as the capacitance-only model while the real part is relatively negligible, thereby validating our original approach. This is shown in figure 4.17.

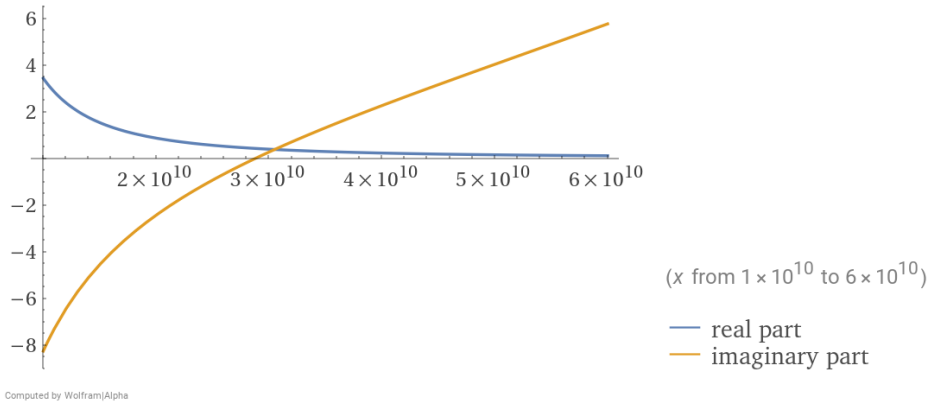

Figure 4.17: Same scenario as figure 4.15, but with a shunt 100Ω as well. Note this is *not* normalized.

4.3 Power Amplifier Interstage Match Walkthrough

In this section we will detail the interstage matching network used for our K/Ka band PA. The input impedance (differential) of the output stage is shown in figure 4.18, from 14 to 40GHz. Unsurprisingly it's primarily capacitive (around 400fF, used in the previous examples).

[1]While not always true, and certainly not considerably much greater than 1, it still gives us useful insight as engineers and works reasonably well for the frequencies of interest and our application. In the example provided for instance, ωRC is 5 at 20 GHz.

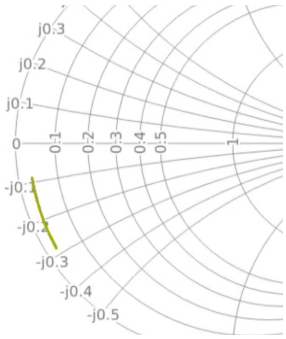

Figure 4.18: Differential input impedance of the PA output stage, 14-40GHz

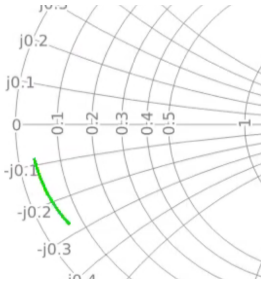

Figure 4.19: Differential input impedance of the PA output stage, but with a shunt 100 ohm resistor as well. Unsurprisingly things do not change much.

When we add a termination resistor (100 ohm differential) in shunt to the gate, we see little change (figure 4.19, again due to the large gate capacitance dominating the impedance. We therefore must take care of the imaginary part. To that end, we will use the low-side end of a Guanella transformer, as seen previously. A plot of the impedance trajectory looking into the lower end of the Guanella with the PA gate (Guanella is the same as the examples, $Z_0 = 25\Omega$, albeit an EMX extracted model of an actual transformer, not just an equation!) is shown in figure 4.20; it corresponds well to what we expect: a low magnitude imaginary component with a crossover from capacitive to inductive mid-band. A comparison to a single series inductor (100pH) is shown in figure 4.21 to give perspective; notice the significantly more broadband nature of the Guanella.

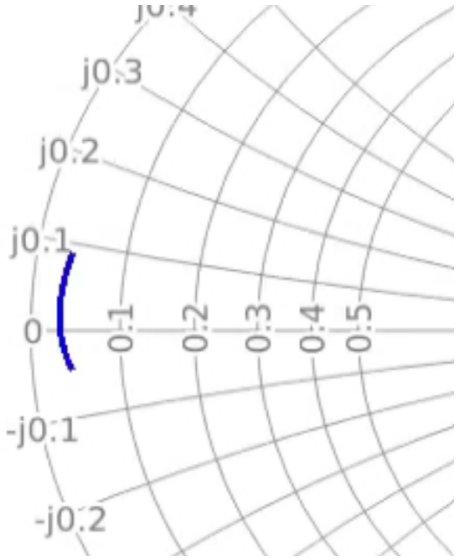

Figure 4.20: S11 when we transform down the gate impedance with a 4-1 Guanella, from 14-40GHz.

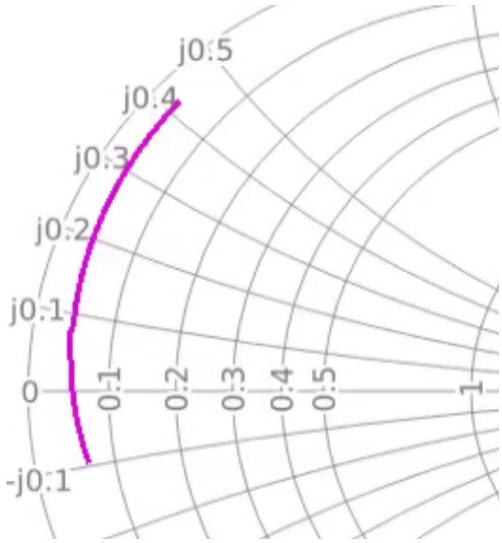

Figure 4.21: A comparison with a simple series 100pH inductor, 14-40GHz.

Note we actually have a choice in how we arrange the back-to-back Guanellas: we can either first lower the impedance and then transform it back up, or vice versa. Either method

works, we simply pick the one that provide better simulation results[2] In this case, it was found that using the high side first followed by the low provided better results. The structure is shown in figure 4.22. The preceding stage feeds in from the top, with the output stage gate at the bottom. From the bottom to the top, we have a small series inductance followed by a high-side 1:4 Guanella, followed by low-side 3:1 Guanella[3]. The 3:1 transformer allows the matching network to provide a higher impedance to the driver stage; the small inductor provides the slight inductance desired by most class A/B drivers.

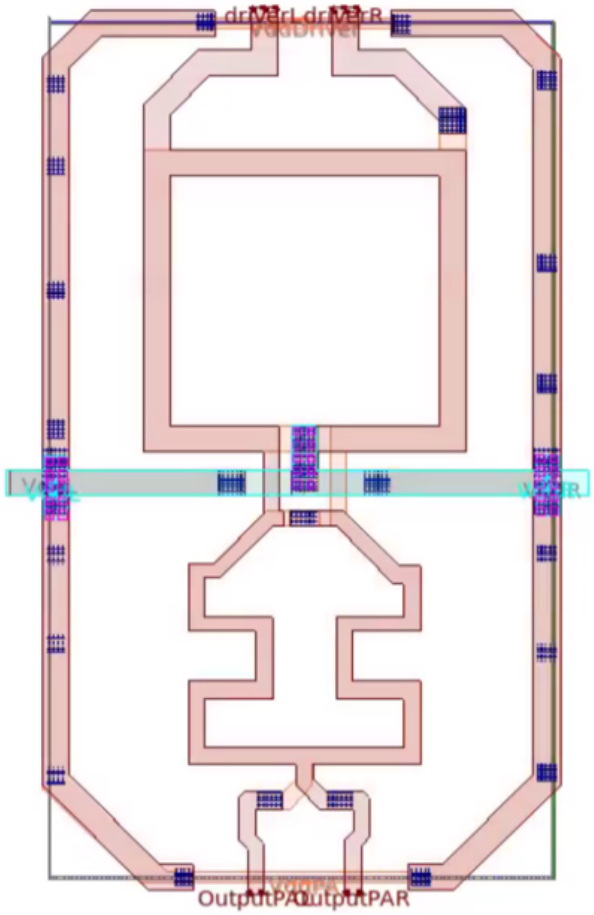

Figure 4.22: Interstage matching network. The output PA is at the bottom.

The resulting impedance presented to the driver is shown below in figure 4.23. While it certainly does not provide the best/tightest matching across the entire octave bandwidth,

[2]The difference here is the characteristic impedance of each Guanella will differ based on their order, to satisfy equation 4.17.

[3]Recall that a Guanella is similar to an autotransformer, and therefore we can adjust the impedance ratio by adjusting the tap points

it is good enough for a driver stage (affecting the overall PAE by a few percentage points across the bandwidth), and obviously much better than the simple resistive termination of figure 4.19. It's important to note that the overall length is slightly over $\lambda/4$ (two $\lambda/8$ Guanellas and a small inductor), so in comparison to typical quarter-wave matching sections, the bandwidth is considerably larger. Note the overall size of the matching structure could have been reduced if meandered or spiraled; this should be done in future work (ran out of time here). A center tap provides the supply voltage feed for the driver stage; an ac-coupling capacitor is thus also needed to set the gate bias voltage of the output stage.

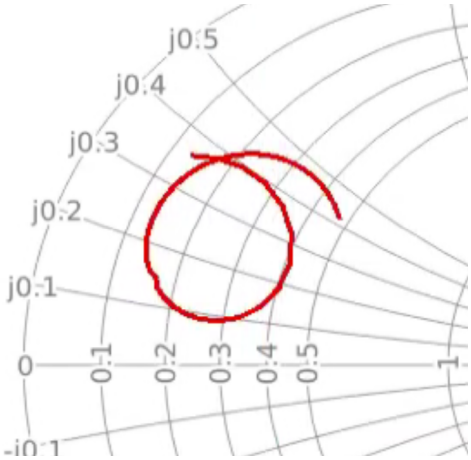

Figure 4.23: Impedance provided to the driver stage from the interstage matching network, from 14 to 40 GHz.

A similar procedure was used to create the input match to the driver; luckily the smaller driver transistor size makes this an easier problem. Note in the variant shown here, we matched to 25 ohms single ended (50 differential), to use a 2:1 balun in the lab (it also makes the matching naturally easier); a standalone product would of course need a 50 ohm reference. The resulting matching network is shown in figure 4.24, and it's S11 is shown in figure 4.25. The lower capacitance and lower reference (50 ohms differential/25 single-ended) makes the match considerably simpler, hence the good matching ($S11 < -15dB$). It again has back-to-back Guanellas with a series inductance, although this time both Guanellas were 4:1 (no 3:1).

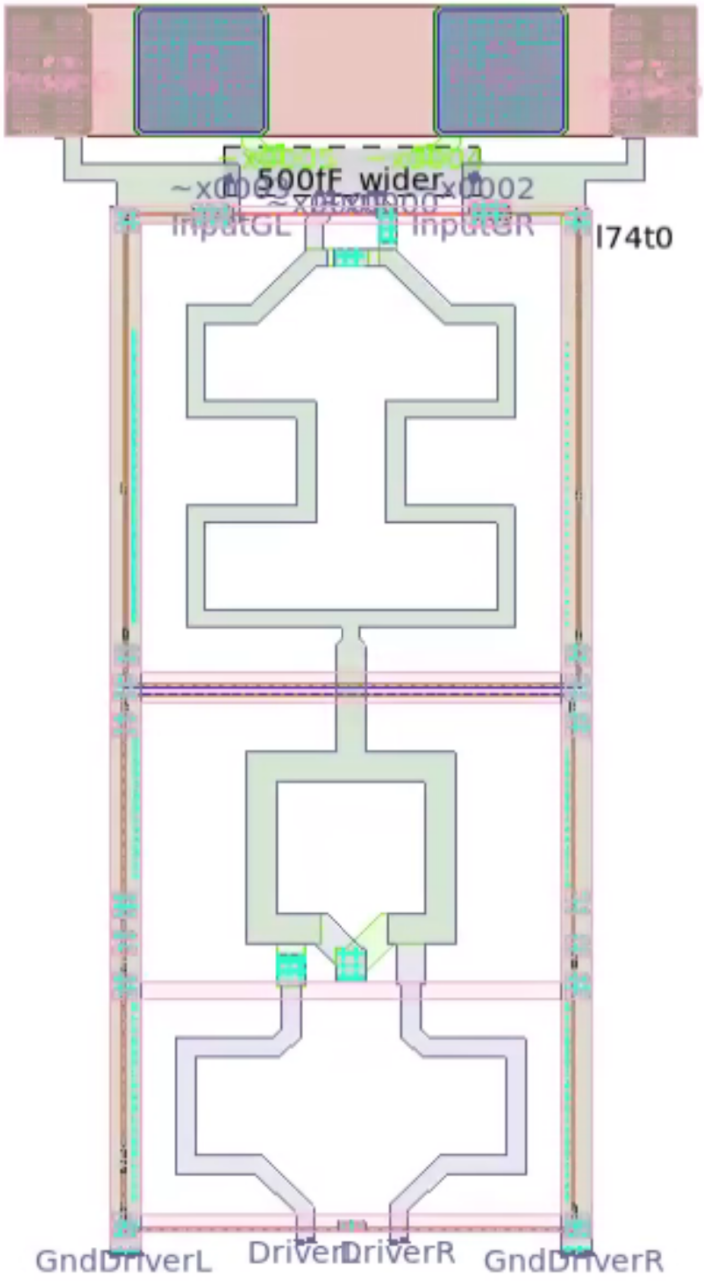

Figure 4.24: Input Matching network with GSSG pad

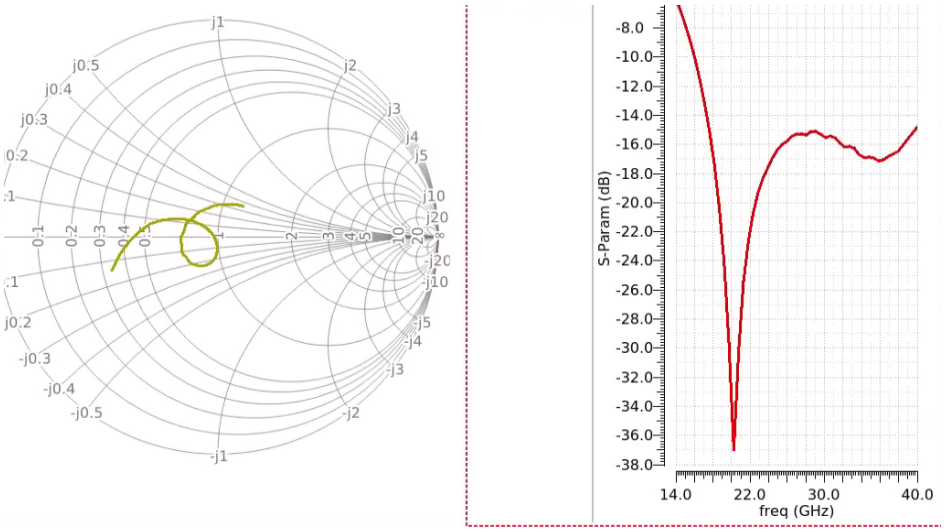

Figure 4.25: Simulated S11 of the input match; note this is referred to 50 ohms differential

4.4 Overall PA

The schematic of the full, two stage K/Ka PA is shown in figure 4.26. It shows from the the top to bottom, the input matching network, first stage transistors, interstage matching network, output stage transistors, and the output matching network to pad. This design was taped out and measured. The measured results of the PA over bandwidth are shown in figures 4.28 and 4.30. The end results are similar to simulation. The frequency is shifted slightly lower (likely due to incomplete dielectric information in our EM models), and the transistors were biased slightly more in class B than simulated (lower output power but greater efficiency than simulated), but otherwise the results matched well. We compare the performance to other papers in table 4.1. Here we can see the power amplifier provides record bandwidth for similar bulk CMOS PAs, with moderate output power and efficiency (with notably only a 1V supply voltage). 64-QAM EVM measurements were done as well, with the results shown in figure 4.33. As this is a broadband amplifier, large symbol rates (up to 4GBd) were achievable, with the main limitation arising from the gain flatness (or lack thereof) of the amplifier. A die photo is shown in figure 4.27, and a detailed layout is shown in figure 4.32.

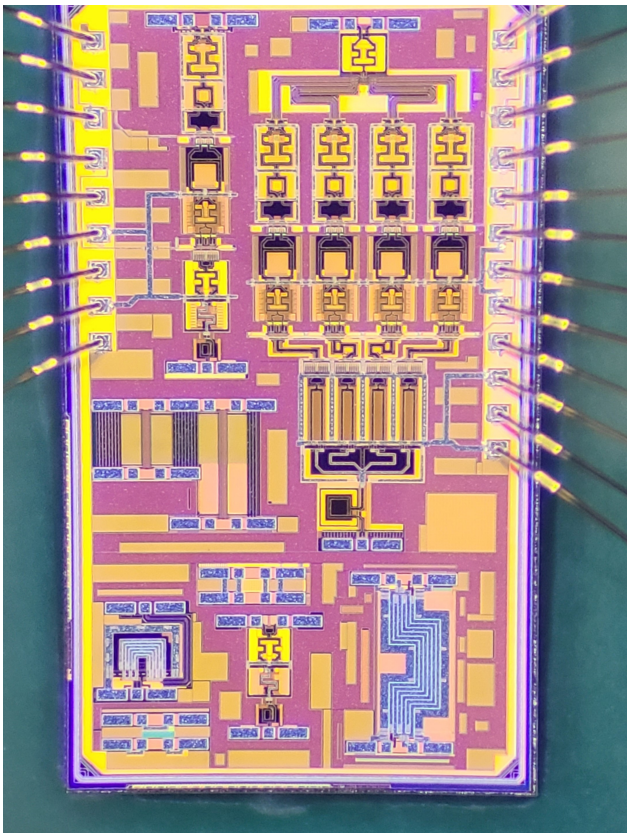

Figure 4.27: Die Photo of the single (no power combining) PA, as well as the first version of the four way PA.

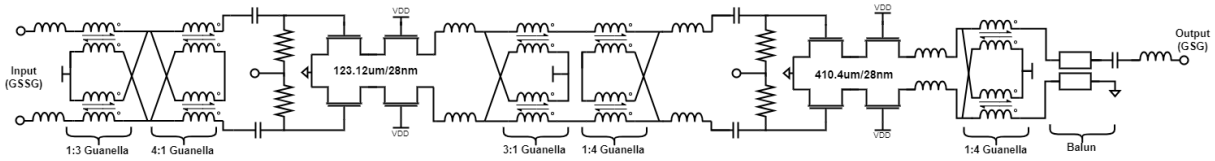

Figure 4.26: Schematic of the K/Ka PA. Transistor sizes are given post-shrink.

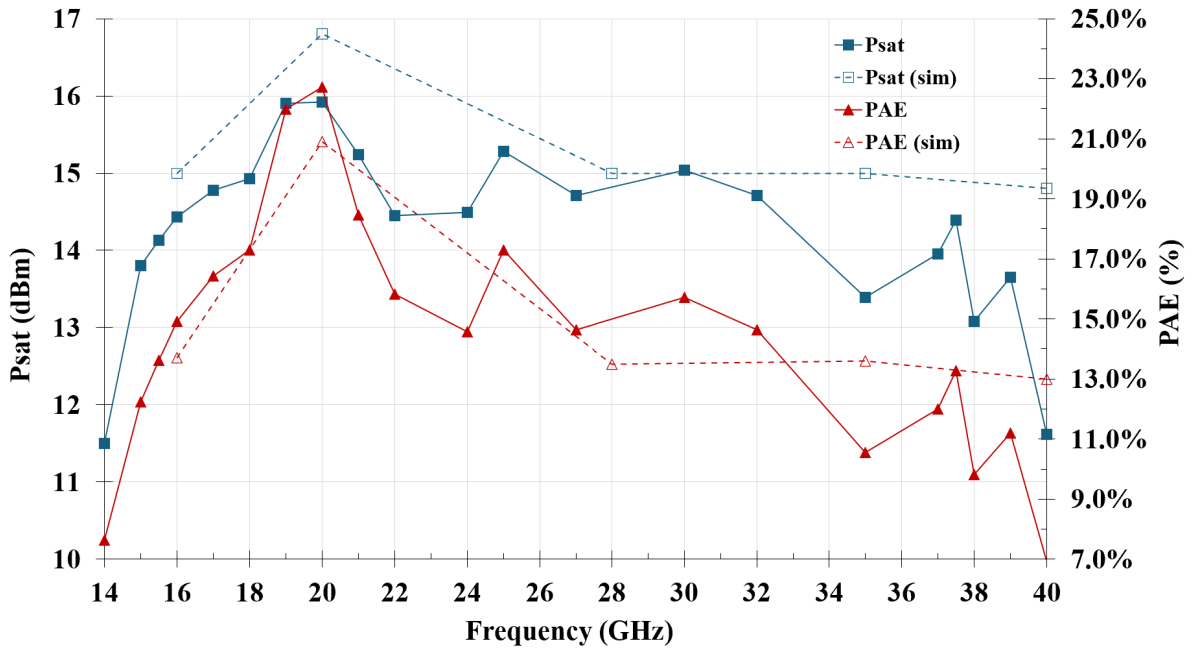

Figure 4.28: Measured PAE & Psat of the single (no power combining) K/Ka band PA.

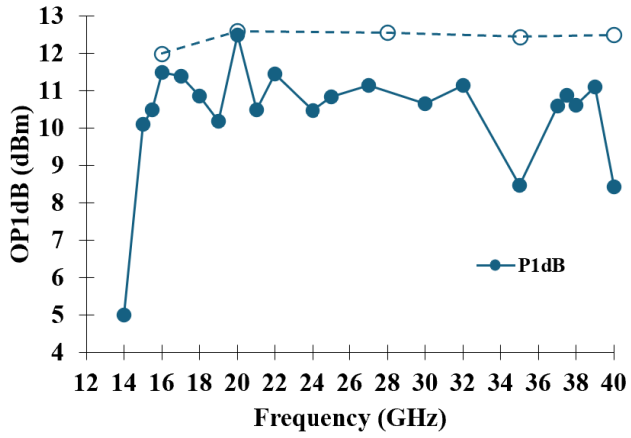

Figure 4.29: Measured P1dB of the non-combined PA.

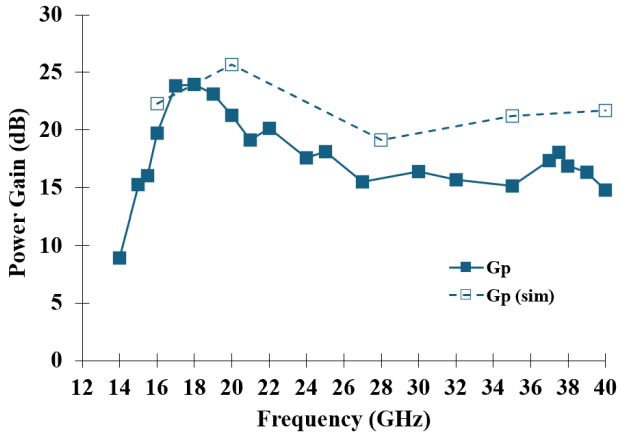

Figure 4.30: Measured Power Gain of the non-combined PA.

Table 4.1: Measured K/Ka PA, No Combining

Ref.	VDD	Frequency	Bandwidth (%)	Psat, Max (dBm)	Psat, Min (dBm)	PAE, Max	PAE, Min	Area (mm^2)
[6] (2018)	0.9V	29-57 GHz	65.1%	16.6	13	24.2%	9%	0.160
[8] (2021)	2V	18-37.5 GHz	77.4%	16	12	26%	6%	0.105
[4] (2023)	1.8V	19.7-43.8 GHz	76%	20.3	18	33.6%	19%	0.106
This work	**1V**	**15-37.5 GHz**	**85.7%**	**15.9**	**13.4**	**24.8%**	**11.7%**	**0.175**

All Min/max values are over operating bandwidth. PAE is given at Psat.

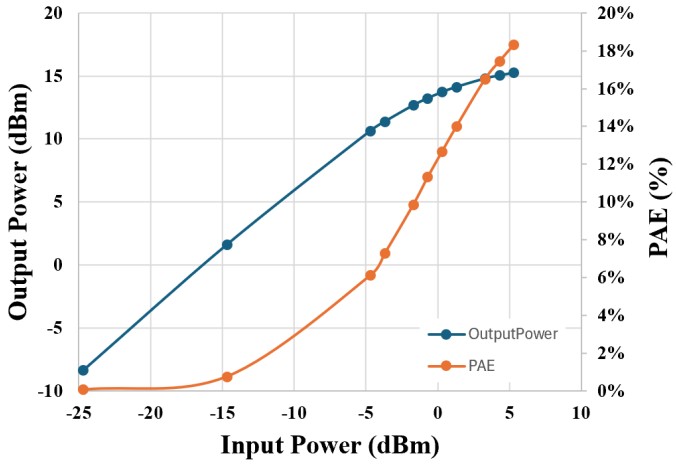

Figure 4.31: Measured (single slice) Output Power and PAE versus Input Power at 30GHz.

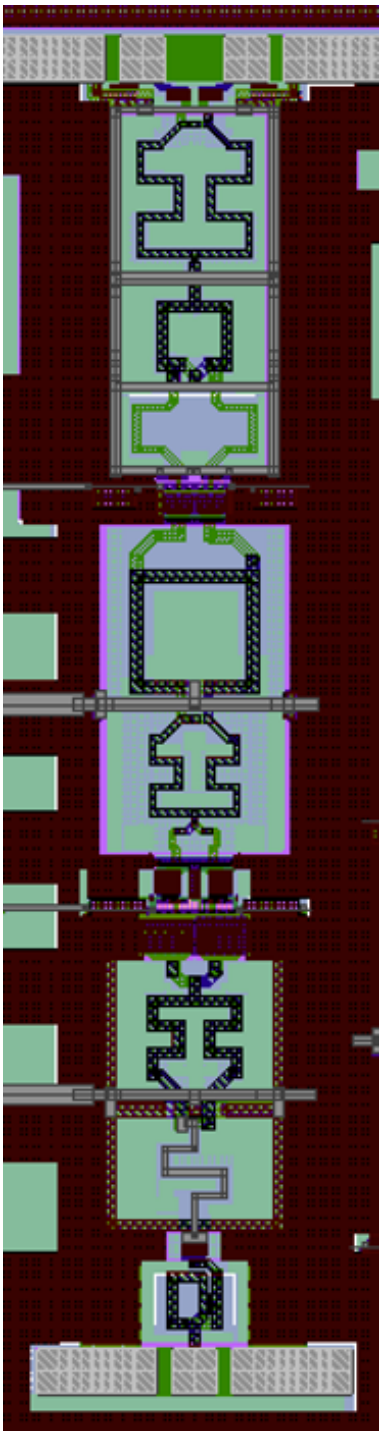

Figure 4.32: Layout of the K/Ka Power Amplifier.

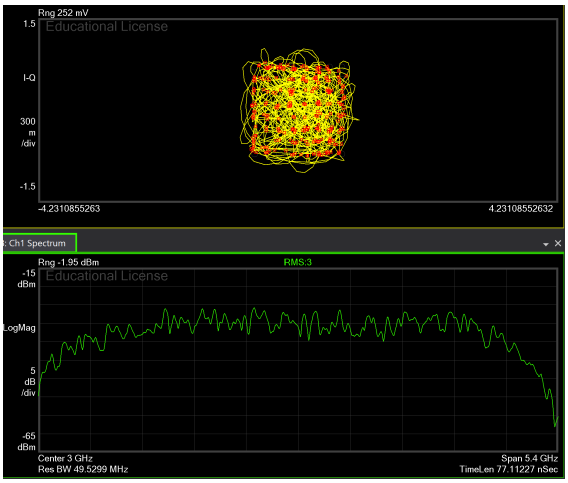

Figure 4.34: Example 64-QAM, 4GBd constellation (carrier at 19GHz) and down-converted frequency spectrum (centered at 3GHz). This is for an output power of 6.8dBm with EVM of -25.8dBm.

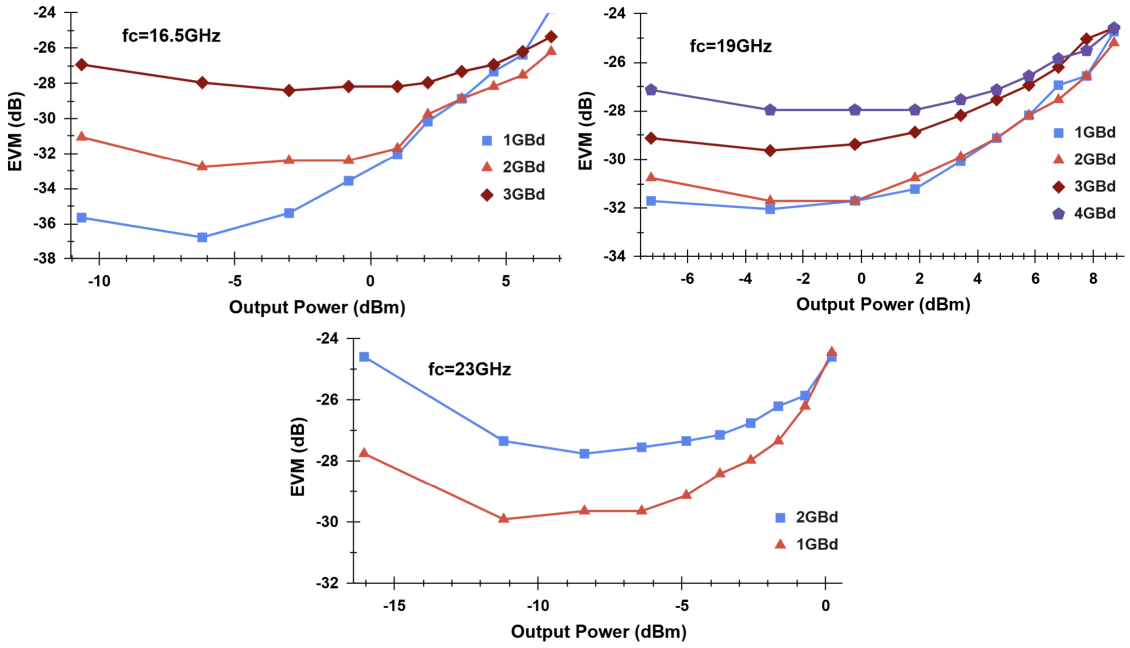

Figure 4.33: 64-QAM EVM of the K/Ka Power Amplifier for various symbol rates and carrier frequencies. Modulated signals were generated with an arbitrary waveform generator (that operates up to 25GHz).

4.5 Four Way Combined K/KA Pa

To further increase the output power of a PA without drastically affecting the efficiency, one inevitably must power combine, as mentioned in section 3.4. To this end, we describe a four-way combined version of the PA. Using the same output stage as the previous (single) PA, if we were merely corporate combine four PAs (each with $Z_{opt} = 12.5\Omega$) we would achieve matching for a 3.125Ω antenna- not realistic. If we use four Guanella transformers however, we can match to a 12.5Ω antenna, a significant improvement. To match to a fifty ohm antenna then, we simply use another 4:1 transmission line transformer, this time a Ruthroff (as the output is single-ended after the balun). The Ruthroff on its own is shown in figure 4.37, and its performance in figure 4.38; its small area (145 x 155 um^2) and low insertion loss (0.45dB) allows for the power combining to remain relatively efficient and broadband. The combined output match is shown in figure 4.35 (Guanella balun, Ruthroff transformer, and a small series inductance as well for better efficiency). Note the performance of each Guanella Balun is essentially the same as that in 4.3, the difference here is that we tie each of the transmission line (balun) sections together in a corporate combining fashion. The performance of the PA is summarized in table 4.2. We can see we increased the Psat by a little over 5dB, while the PAE dropped from (at best) roughly 21% to 18%. The performance over frequency is plotted in figure 4.36. Note the frequency response has a noticeable shift; this is due to the additional series inductance caused from the additional transmission lines in the interstage matching network, seen in figure 4.39. For input matching, 4:1 Guanella is added to 'undo' the four way split.

The 4-way PA was taped out alongside the single slice PA. The final design unfortunately had issues with respect to the power plane (far too lossy), especially for the inner slices of the PA. This flaw was unfortunately critical under measurement; when measured, the overall power was indeed a little over 5dBm greater than that of the single slice. However the chip quickly overheated and caused the threshold voltage of the transistors to slowly drift (as a result of large DC currents flowing through a very constrained area). As a result, no measurements could be made for the 4-way design.

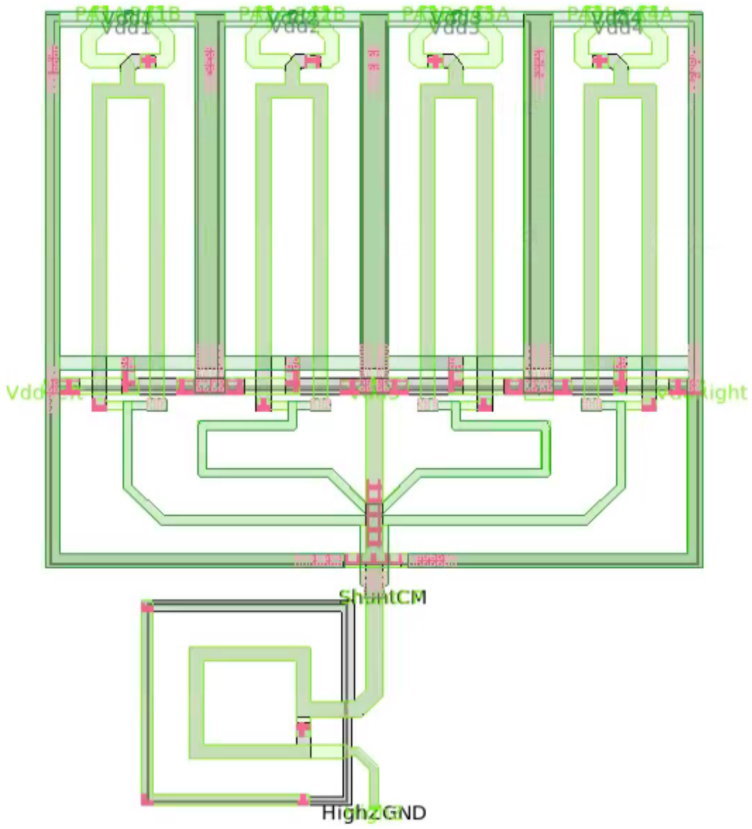

Figure 4.35: Four-way corporate combined Guanella baluns and a Ruthroff transformer.

Table 4.2: Four Way PA Overview

	VDD	Center Frequency (GHz)	Bandwidth (%)	Psat, Max (dBm)	Psat, Min (dBm)	PAE, Max (%)	PAE, Min (%)
This work (sim), 4-Way	1V	25	80%	22	20.5	18%	11%

All Min/max values are over operating bandwidth. PAE is given at Psat.

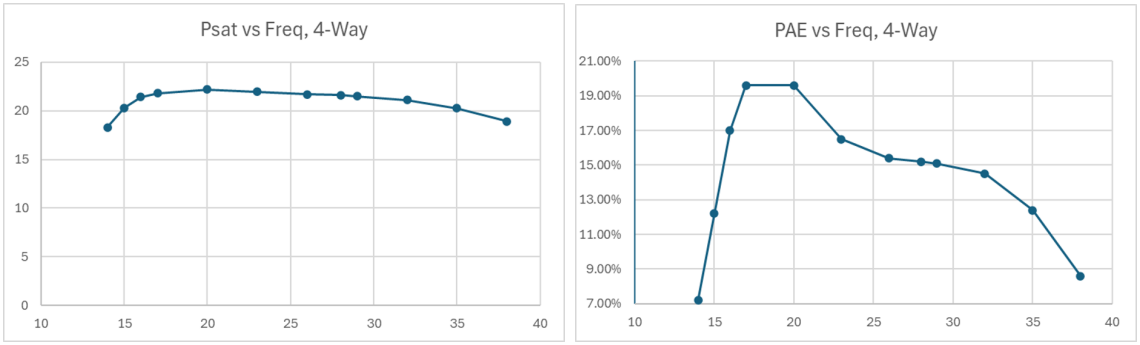

Figure 4.36: Four-way, version 1, simulated Psat and PAE over frequency.

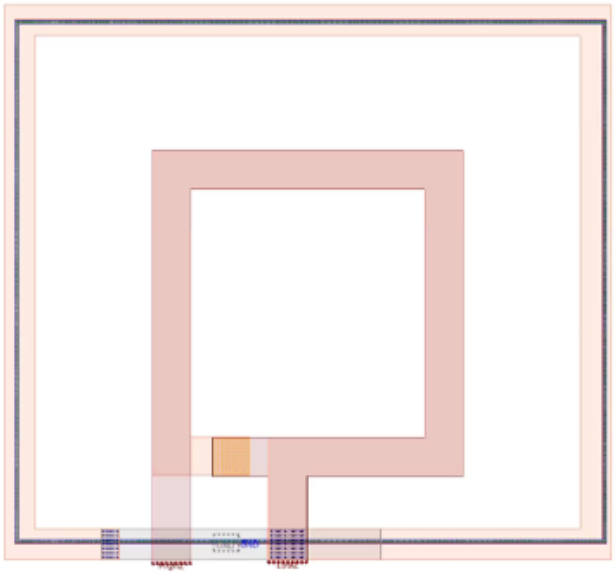

Figure 4.37: Broadside-coupled Ruthroff transformer, dimensions are 145x155um.

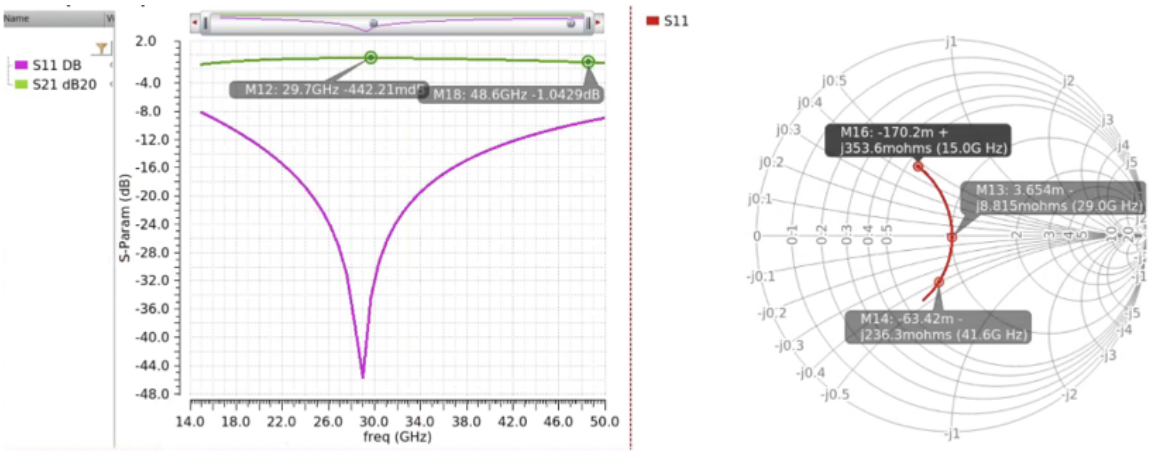

Figure 4.38: Smith chart, matching, and insertion loss of the Ruthroff transformer.

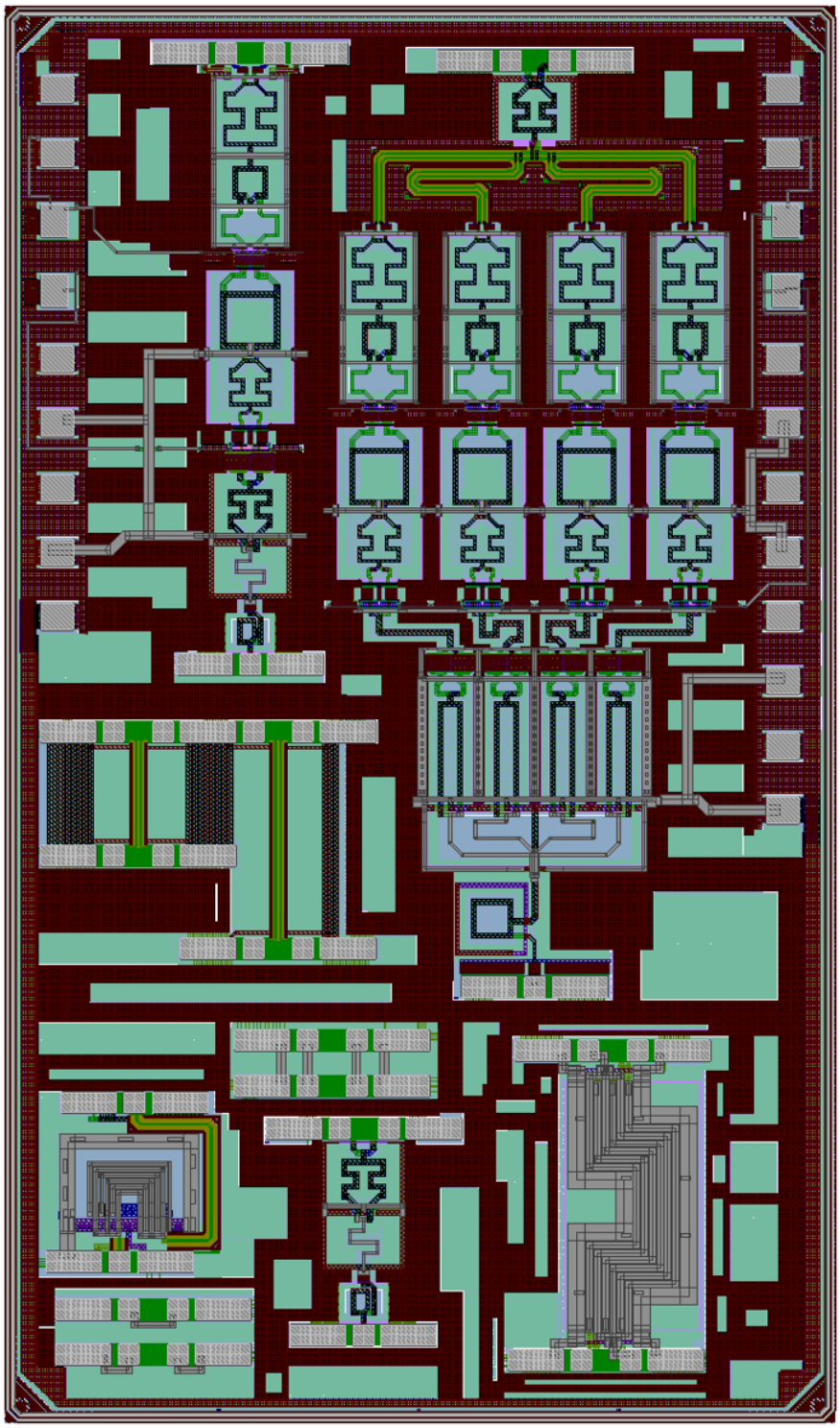

Figure 4.39: Chip top level view of the v1 four-way PA, and the single PA.

4.6 Two-Way Combined

Here we will go over a two-way combined variation. For this process, to achieve an optimum impedance of 12.5Ω (which is provided by the Guanella transformer) requires an extraordinarily large transistor, which degrades the overall efficiency. With this in mind, it makes sense then to corporate combine two 25 : 100Ω Guanella transformers, to provide a higher impedance to the output transistor while still matching to the 50Ω antenna.

The output matching network is shown in figure 4.40, with the two 25 : 100Ω Guanella baluns and an additional series inductance (20pH). Note the outputs were combined before the odd-mode transmission line/balun to allow for a lower-loss 50Ω broadside-coupled transmission line (versus two 100Ω lines). As this was a second tapeout, a few additional improvements were made. Here, the balun itself was spiraled to increase the even-mode impedance, providing both a bandwidth and loss benefit. This is illustrated in figure 4.41, along with the performance of the output stage on its own.

The interstage matching networks use the same topology (two Guanellas), just modified for the changed output stage. Compared to the first version, the matching network area was reduced by meandering the 3:1 stage; the resulting impedance and layout is shown in figure 4.42. Also challenging is the input matching network. To save area, it was decided to first match from the pad, and then power split to the both gates, which unfortunately doubles the effective capacitance we are matching to; the end result isn't particularly great, as seen in figure 4.43, and likely could be improved upon in future generations given more design time. The layout of the full PA can be seen in figure 4.45. The simulated performance of the entire PA is shown in figure 4.44. We compare the three designs' simulated performance in 4.3. Note the two-way design provides a simulated improvement of around 2dB more output power for approximately the same efficiency as the non-combined PA, while also providing a larger bandwidth (15-40GHz). This design was taped out in September of 2024; a die photo is shown in figure 4.50.

The measured Psat and PAE for the two-way PA is shown in figures 4.46 and 4.47. The supply voltage was set at 1V, the output stage gate bias at 550mV, and the driver gate at 600mV (corresponding to roughly class A/B and class A for the output and driver respectively). The measured results for the two-way combined PA ended up significantly better than what was simulated under nominal conditions. This most likely occurred from a 'fast' process corner, specifically in regards to metal conductivity. When the metal resistivity decreases, this provides a major boost to both the PA's power gain and PAE. For these Guanella designs, the PAE is significantly improved when there is lower metal loss as the large DC drain current has to be supplied through the Guanella. The PA operates roughly from 14.5GHz to 39GHz, for a fractional bandwidth of 91.6%, which is slightly lower than simulated, but still an improvement over the single-slice PA (which was already record-breaking in terms of bandwidth). The overall performance of the amplifier is better than that of the non-combined PA, as the smaller required output transistor size significantly improves the efficiency of the overall PA. For instance, the power-combined PA offers at its best a measured 24.7% PAE with 18.4dBm Psat (at 20GHz), whereas the single slice

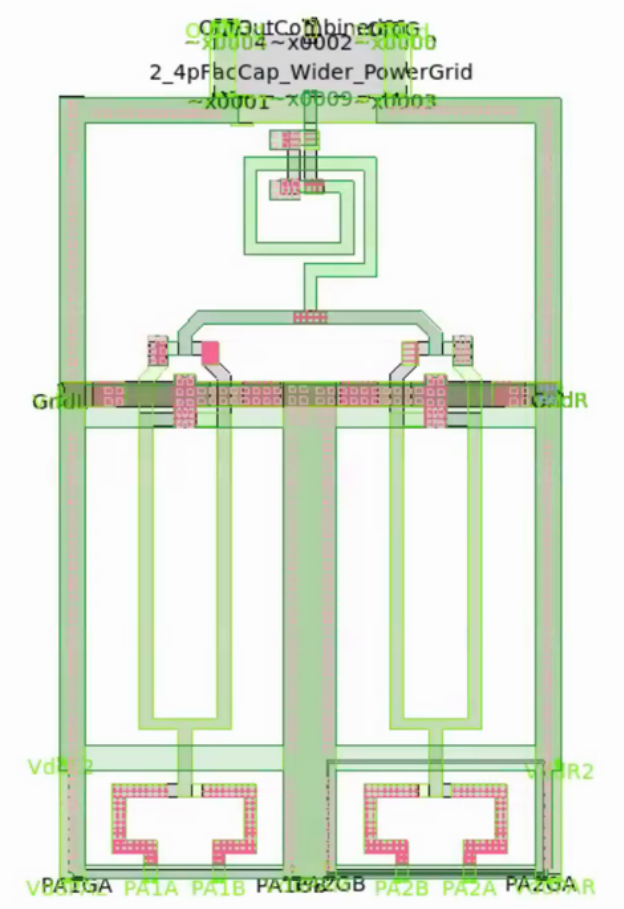

Figure 4.40: EMX view of the two-way PA output matching network. Both Guanellas are 25:100Ω. The top black box is 2.4pF AC coupling capacitor.

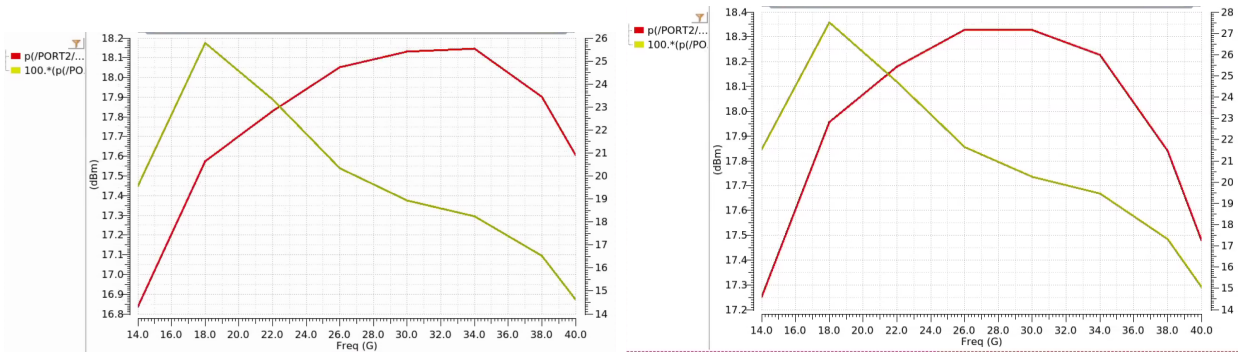

Figure 4.41: Performance of the output stage only with no spiral (left) and with a spiral (right). The yellow curve represents drain efficiency, red is output power.

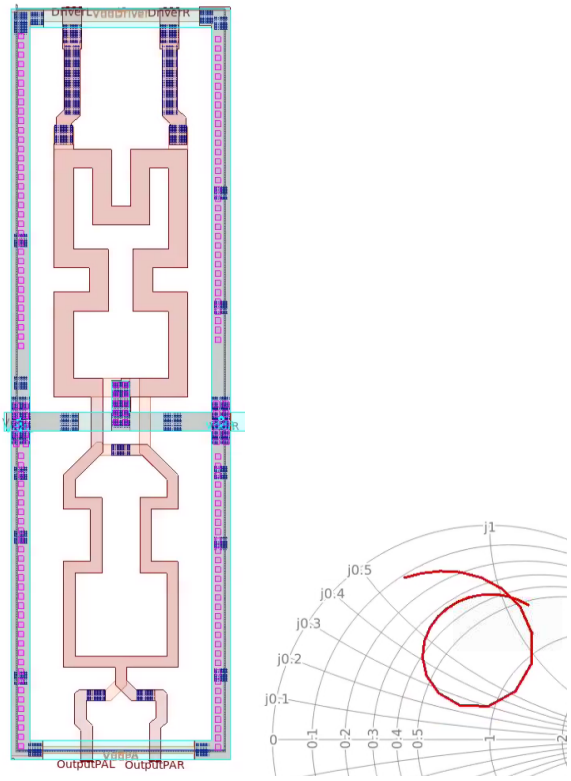

Figure 4.42: Interstage matching network of the 2-Way combined PA and the impedance presented to the driver. The dimensions are $115\mu m \times 400\mu m$ drawn.

achieves at best 15.9dBm with 24.8% efficiency. EVM results (64QAM) are shown in figure 4.49.

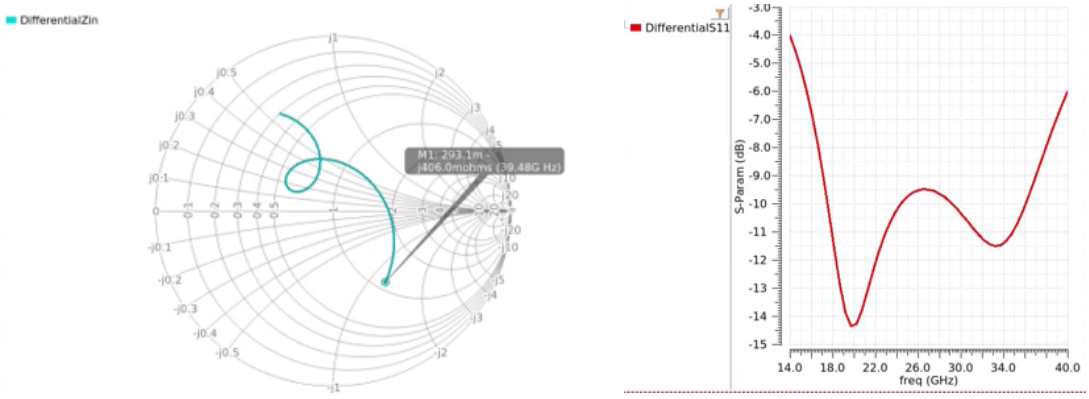

Figure 4.43: Simulated input matching of the two-way power combined PA

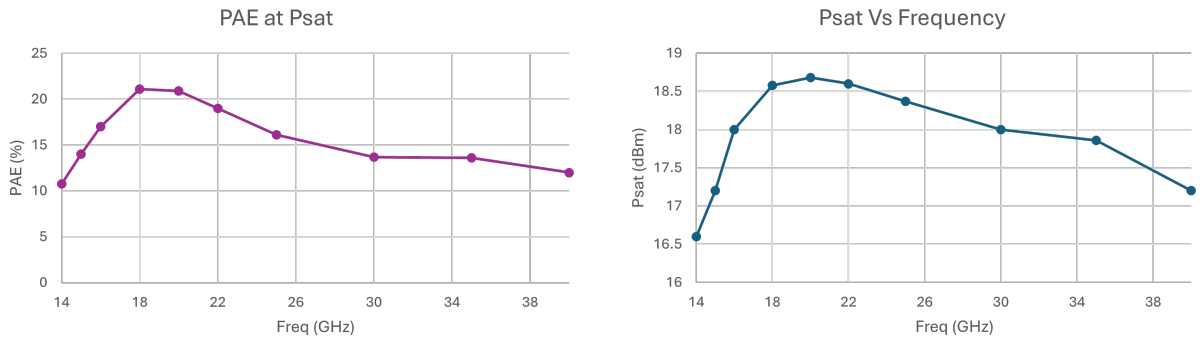

Figure 4.44: Simulated PAE and Psat of the Two-Way Power combined PA

Table 4.3: Comparison of the Designed PAs, Simulated

	VDD	Center Frequency (GHz)	Bandwidth (%)	Psat, Max (dBm)	Psat, Min (dBm)	PAE, Max (%)	PAE, Min (%)
No combining, (sim)	1V	28	85.7%	16.8	15	21%	13.9%
4-Way, sim	1V	25	80%	22	20.5	18%	11%
2-Way, sim	1V	22.5	113%	18.6	17	20.5%	11%

All Min/max values are over operating bandwidth. PAE is given at Psat.

4.7 Ancillary Design Considerations

In this section we cover a few additional passives and considerations regarding K/Ka and bulk-CMOS designs, which, while not novel, may still be of interest to the reader.

4.7.1 Reliability Guidelines: Electro-migration and Breakdown

While this is research, not a product, it is still important to meet the reliability guidelines set by the foundry to show the work is indeed applicable to real-life applications. For PAs, this actually sets a serious constraint on the design. For breakdown voltages, we ensure that no voltage between a node and the bulk exceeds twice the gate breakdown voltage[4]. Note we cannot simply tie the source to the bulk node as the resulting parasitic bulk diode capacitance is devastating at these higher frequencies. Electro-migration concerns place a limiting factor on the passives, especially the output stage, which has an extremely large ($>$ 100mA) DC current. This naturally sets a minimum allowed width for our metal traces,

[4]This is a rough guideline based on the author's and others' experiences

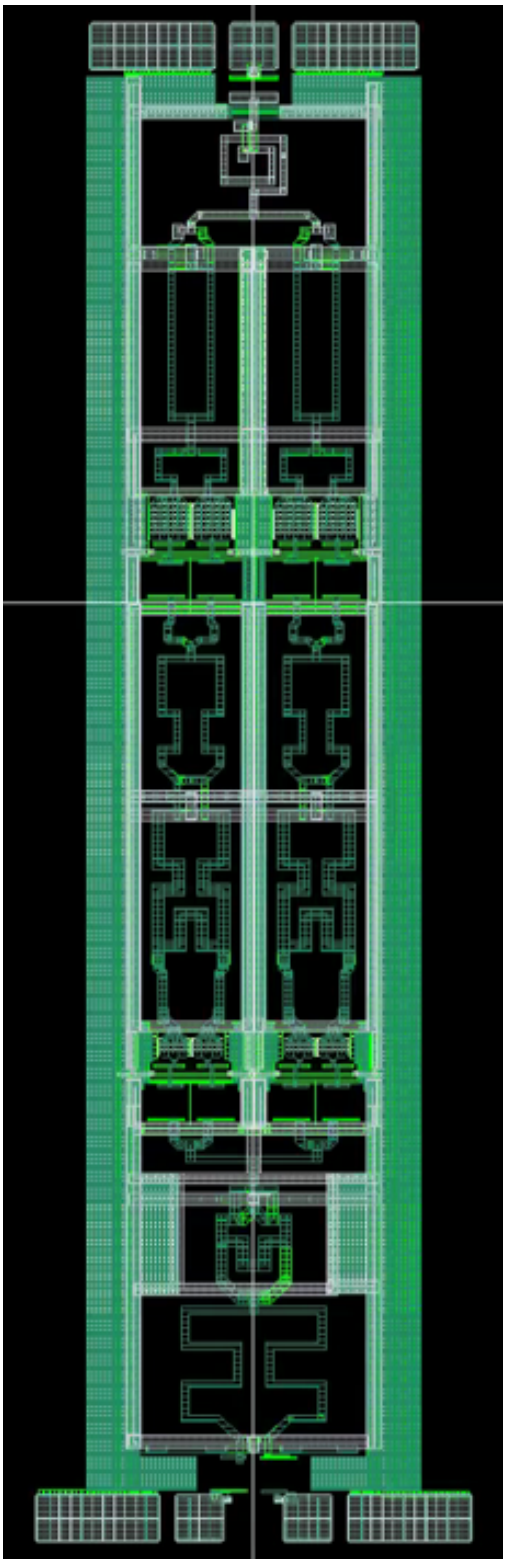

Figure 4.45: Layout of the full two-way power combined PA

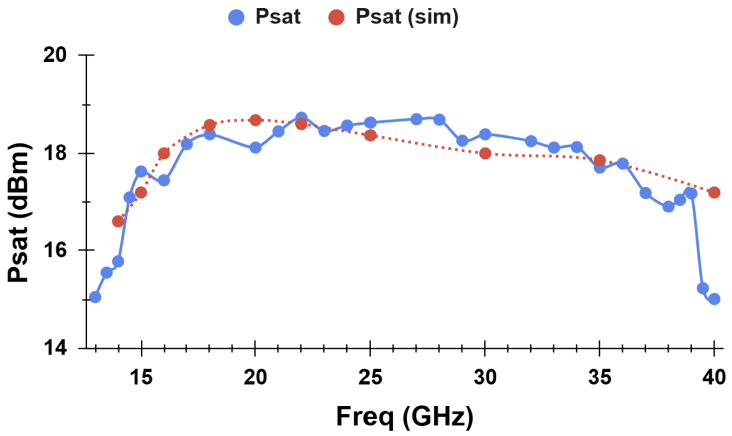

Figure 4.46: Measured Psat of the two-way power combined PA.

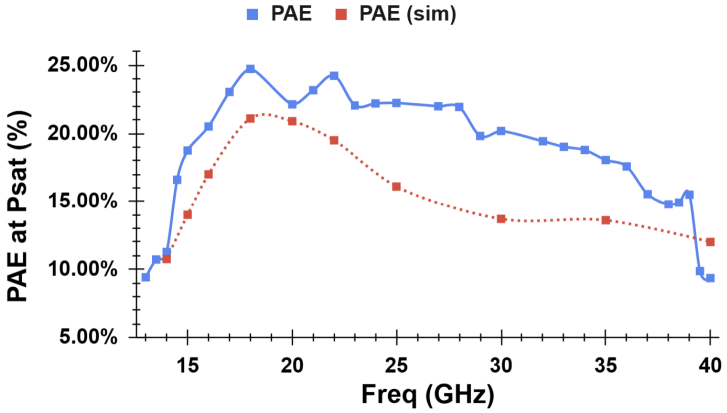

Figure 4.47: Measured PAE (at Psat) of the two-way power combined PA.

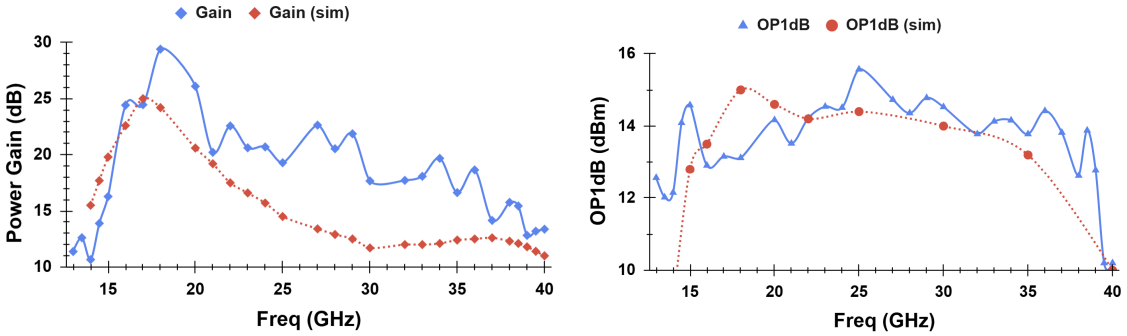

Figure 4.48: Measured power gain (-20dB back-off) and OP1dB of the two-way power combined PA. Simulated results are provided under nominal process conditions.

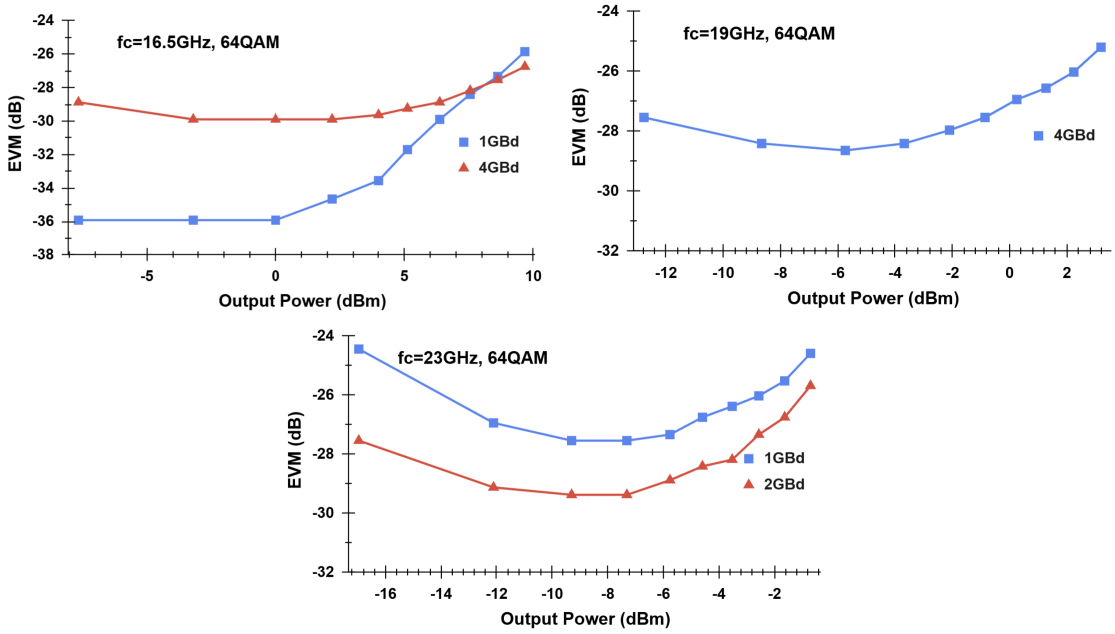

Figure 4.49: Measured EVM of the two-way power combined PA for various symbol rates and carrier frequencies.

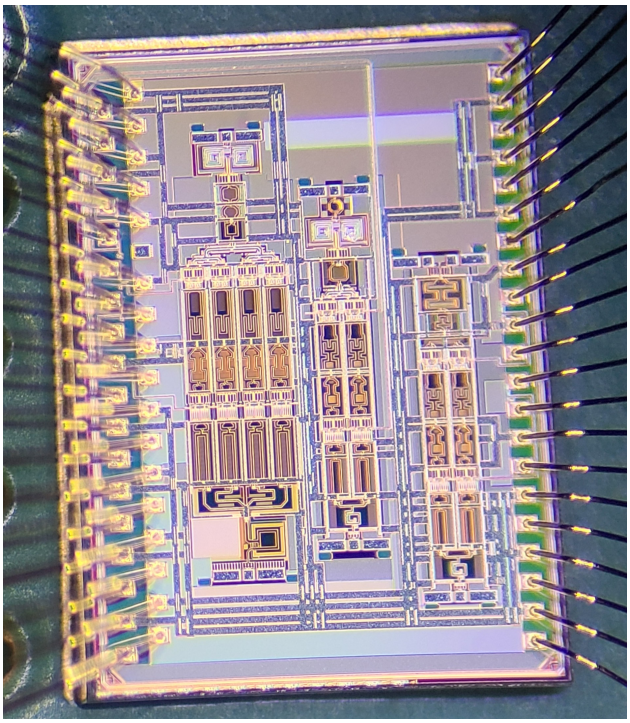

Figure 4.50: Die Photo of the Two-Way Combined PA chip.

and also entirely rules out the use of the topmost aluminum redistribution layer, as the aluminum layer current density limit is three times lower that of the lower copper layers. This is unfortunate as the topmost metal layer would have a lower parasitic capacitance to the substrate and thus a higher even mode impedance.

4.7.2 Resistors

Reliability rules make simulating and designing the termination resistors actually important. These resistors are used not only as a load for the driver, but are also necessary to provide low frequency stability for the PA (resistors are almost inevitable for an octave bandwidth PA; the idea here is that a resistor provides a lower shunt resistance to the gate, removing oscillations due to feedback to the gate at low frequencies). To achieve low (100Ω) resistance poly-silicon resistors that can handle the large AC currents at Psat requires many wide, parallel resistors to be used. The resulting resistor dimensions were huge: 32um x 11um, and therefore the additional capacitance added here was unfortunately no longer negligible.

4.7.3 MOM AC Coupling Capacitors

We know the component quality factor for a capacitor is $Q_c = 1/(\omega RC)$, where R is the parasitic series resistance of the capacitor. At low frequencies, this series resistance is of little concern. Indeed, most MOM capacitors provided CMOS foundries show little regard for this, and instead aim to maximize capacitance density. This is quite important for typical mixed-signal, switched capacitor circuits. As we reach the K/Ka band frequencies however, we cannot simply take the capacitor quality factor for granted anymore, and it becomes important to tweak the MOM-cap design to improve (lower) the series resistance. This is doubly important for the PA designs used here which require large MOM capacitance to maintain efficiency, especially on the output stage[5]. And a large capacitance of course further worsens the quality factor. With this in mind, we utilize a MOM design first shown by Nima Baniasadi [24]. While thin, densely packed fingers are needed to achieve decent capacitance density, the series resistance is considerable. Instead, the fingers are kept short and use all the offered thin metal layers (with vias connecting them) to create a pseudo-thick metal finger (or wall); the spacing is set by both the DRC rules and the maximum allowed electric field (even though it is an ac-coupling capacitor, there will still be a large voltage swing across the capacitor with respect to the body). The (considerably) thicker metal layers are incorporated into the design as well, with the top-most copper layer (M10) acting as a cover shield over the entire structure, and M9 feeding the fingers with a low resistance pathway, reducing the series resistance considerably. This is shown in figure 4.51.

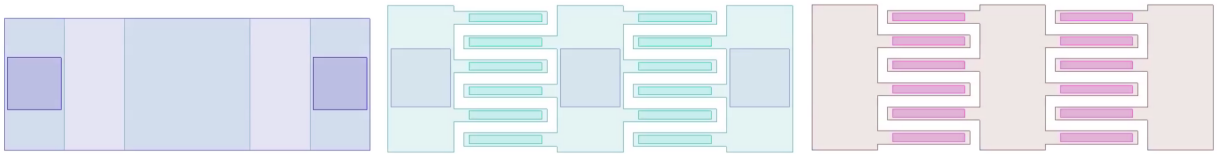

Figure 4.51: Layer by layer view of the K/Ka MOM cap. The leftmost image shows the top two metal layers M10 and M9, and via9. The middle picture shows the even, thin metal layers (m6, m4, m2); Via8, from the thick metals above to the thin metals, is shown here as well. The right image shows the odd thin metals (m7, m5, m3, m1). Note the capacitor fingers utilize all the thin-metal vias (Via1-Via7).

4.7.4 Actives: Transistor layout

It is well known for PA design, we need to minimize the parasitic resistance of the transistor at all nodes, including the gate. The typical double gated, multi-finger design is therefore used. For cascoded transistors, care must be taken to minimize the capacitance between the input

[5]Some quick math: at 20GHz, a 1pF capacitor has an impedance of roughly 8Ω. For a 50 ohm load then, this voltage divider reduces the voltage swing by 14%!

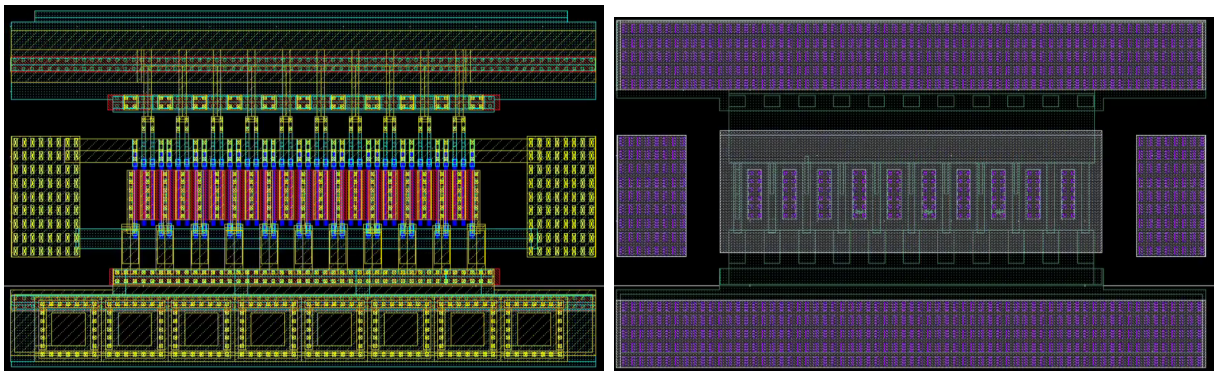

Figure 4.52: Left image shows the transistor gate (blue), doping (red), M1 (light blue), M2 (yellow), and their associated vias. The right image shows M3 (green), M4 (silver), and Via3 (purple).

transistor and casocde transistor (the middle node). Many designs utilize a shunt inductance here to resonate out this capacitance, but this is only practical for narrow-band designs (and certainly not for octave bandwidths). Therefore, since we can only hope to minimize the capacitance here, we chose to instead simply share the S/D between the cascode and input transistors, as commonly done in analog CMOS designs. A detailed look at the layout is shown in figure 4.52. The left-side image shows the bottom two metal layers along with the gate and doping layers, depicting the shared S/D layout and double gate contacts (the gate is fed from the right and left by M2 and M1 metals). The large DC and AC currents necessitate a low impedance ground plane and source connection; as a result, large ground traces (utilizing all available lower metal layers) are placed above and below the transistor cells and as close as possible to the source of the input transistor (this is connected by M2 and M3 to the ground). The right image depicts M3 and M4, which simply illustrates that the output of the cascode is taken directly above the transistor, using vias all the way to the top metal layers.

The overall differential pair is depicted in figure 4.53. The gate input is fed from the bottom, and the drain output is taken from the top, both by the topmost thick copper layer, M10. The ground plane is fed by the other thick copper layer, M9, horizontally (as well as all the thinner metals, not shown here). While utilizing the top metal layer for the ground increases the parasitic capacitance, the large (100mA DC) currents actually make reducing series resistance far more important to the overall PA performance. The structure has mirror symmetry around the center, befitting its differential nature. The cells for the positive and negative inputs are kept close together to reduce both the resistance to ground as well as any unwanted source inductance (often referred to in MMIC design as ‘lead inductance’).

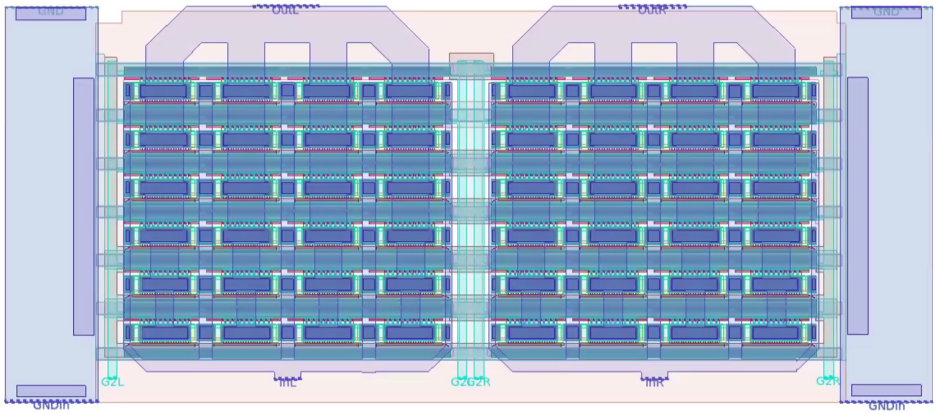

Figure 4.53: The output differential pair for the 4 Way and Single-slice PAs.

Chapter 5

VSWR Resilient Orthogonal Load Modulated Balanced Amplifier in Bulk CMOS

In this section we describe a prototype design of a Orthogonal Load Modulated Balanced Amplifier (OLMBA) in 28nm bulk digital CMOS, designed to offer VSWR resilience. This work was done prior to the Guanella transformer PAs, (the primary focus of this thesis), and unfortunately the performance of the active devices here needed to be improved significantly. This section is included however as it still relates to the overall topic of broadband K/Ka power amplifiers, and interest in load modulated balanced amplifiers has increased significantly recently [25]. Certainly incorporating the Guanella PAs into the OLMBA structure would be an interesting future work. Furthermore, for the OLMBA the author had used CPW implementations of passive structures, and we will be therefore be able to compare their performance with the broadside coupled implementations used later on in the Guanella PAs (the broadside coupled would perform better).

5.1 Overview

The concept of a load modulated balanced amplifier (LMBA) was first brought to attention by Cripps' paper in 2016 [26]. The LMBA schematic is shown in figure 5.1. The idea here is simple and elegant: by applying a control signal into the isolation port of a normal hybrid coupler, we can actively modulate the impedance seen at the input and coupled port. The impedance seen at ports 2 and 4 can be derived as follows [26]:

$$Z_4 = Z_2 = Z_o(1 + \frac{j2I_c e^{j\theta}}{I_b}) \tag{5.1}$$

where I_c is the control current, $I_4 = -jI_b$ and $I_2 = -I_b$, and θ is the phase difference between I_b and I_c. A major benefit of this structure is that hybrid couplers (when designed

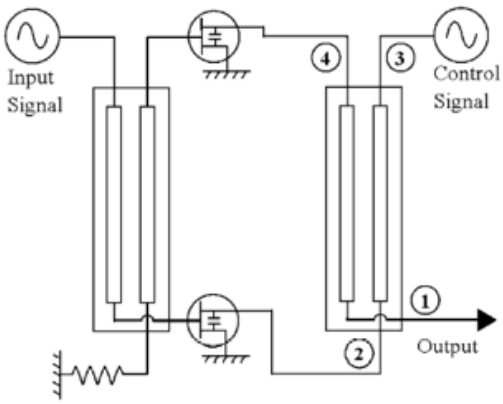

Fig. 1. Schematic representation of the LMBA.

Figure 5.1: Original schematic of the LMBA [26].

with backward-wave couplers) are inherently broadband (unlike a quarter wave line, which is traditionally used for active load modulation schemes). When the control signal is replaced with a separate PA, this turns into a broadband Doherty-like power amplifier, as demonstrated by numerous papers [25].

There is however a notable issue with the LMBA when used for cellular applications: it can be susceptible to instability when the output impedance (antenna) varies[1]. This is depicted in figure 5.2.

Essentially the presence of a control signal at the output hybrid leads to possible instability if the output of the hybrid is not perfectly matched.

5.1.1 Balanced Amplifiers

It's worth taking a look at the traditional balanced amplifier's performance over load impedance changes. First invented in 1965 [27], balanced amplifiers have been heavily used as broadband microwave amplifiers (not as integrated circuits), primarily for the military [28] in the late 20th century. They are most known for providing a broadband, low-VSWR input (as any reflections due to mismatches are directed into the terminated isolation port). When the output port is mismatched however, things are different; this scenario is depicted in figure 5.3. In this case, the impedance seen by each amplifier varies as follows [28]:

[1]For cellular applications, the antenna impedance has typically 2:1 VSWR variations during normal usage.

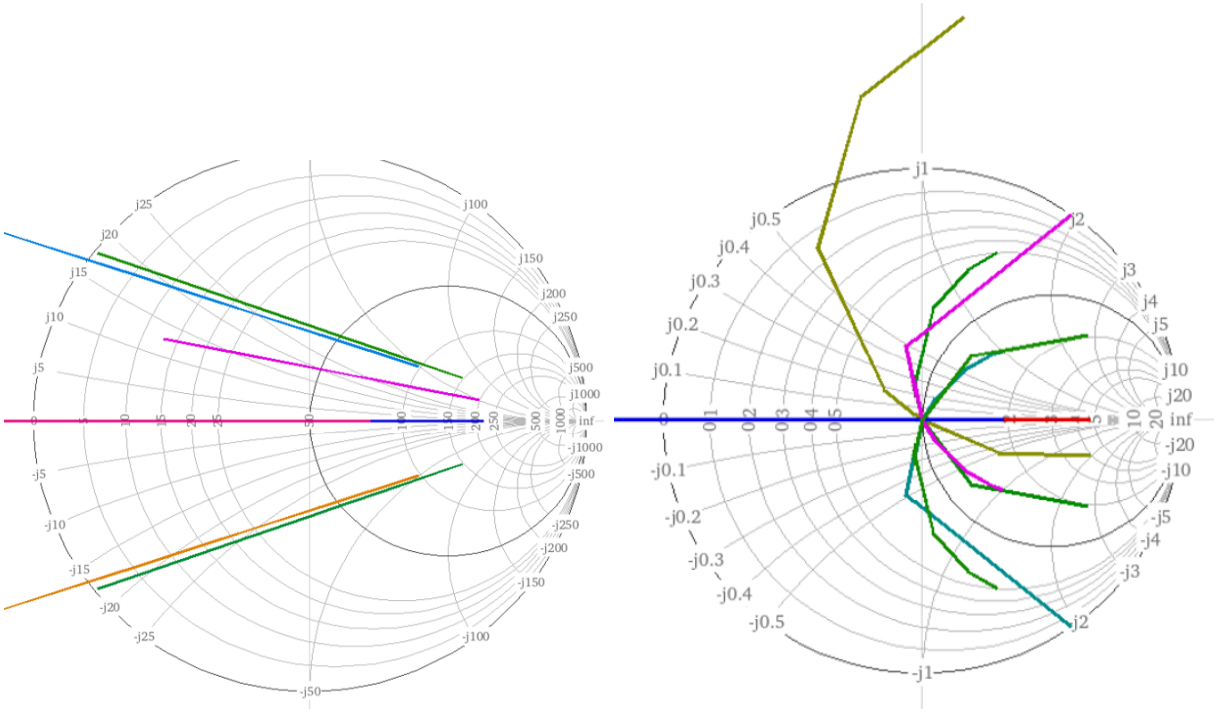

Figure 5.2: Impedance seen by the PA (left) and the control port (Right) when $I_C = 0.25I_B$. We are sweeping the phase θ for various values of Z_L across 2:1 VSWR. Note the possible instability issues for specific control signal and load impedance values.

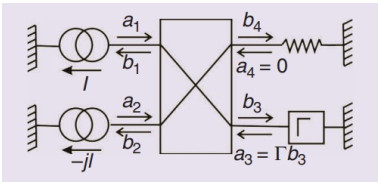

Figure 5.3: Balanced Amplifier under a mismatched load, from [28].

$$\rho_1 = \frac{a_1}{b_1} = \frac{\Gamma}{1+\Gamma} \tag{5.2}$$

$$\rho_2 = \frac{a_1}{b_1} = \frac{\Gamma}{1-\Gamma} \tag{5.3}$$

For the balanced amplifier, under the presence of a mismatched load, one PA sees an increased impedance magnitude, while the other sees a decreased impedance magnitude.

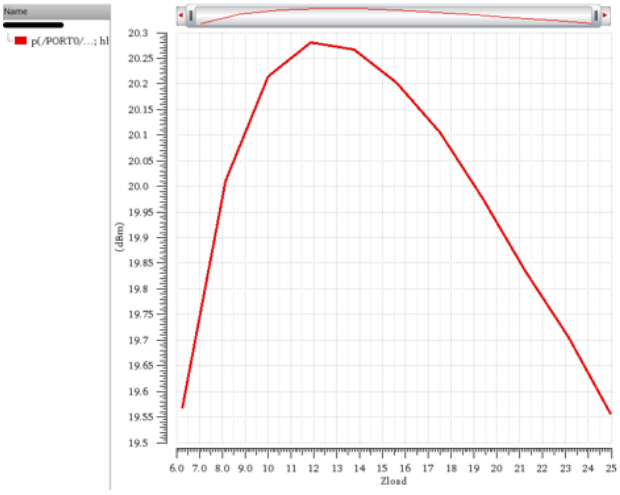

Figure 5.4: Here we see an example balanced amplifier provide 1dB of output power variation over 2:1 VSWR.

The net result here is that the overall system can be made to provide reasonable immunity towards load variations, as one PA can compensate (provide more output power as it sees a lower load impedance) for the other. An example simulation of this is shown in figure 5.4. Here we simulate (with an ideal hybrid of characteristic impedance 12.5Ω) a balanced amplifier over 2:1 VSWR, and see a fairly low 1dB of output power variation.

5.1.2 The Orthogonal Load Modulated Amplifier

The orthogonal LMBA (OLMBA) is a variant of the LMBA, first conceived in 2020 [29]. Shown in figure 5.5, the idea here is to supply the control/modulation power at the input hybrid, rather than the output hybrid. To still achieve load modulation, instead of terminating the output hybrid's isolation port, a purely reflective load (shown here with value jX) is instead used. The reflection at this port then mimics the control signal in the original LMBA concept. Note because the addition of the control signal at the input hybrid breaks the symmetry of the structure, the impedance seen by each PA is not as elegant as equation 5.1.

The original motivation for the OLMBA, provided in [29], was to reduce the required control signal power significantly (as notice now the control signal will be amplified as well as the desired signal by the PA/transistors). More interestingly however, we will see that because the output hybrid no longer has a control signal directly applied to it, we are able to choose the load impedance to provide the best performance for varying antenna load impedances. In [30] the authors show the impedance seen by the PAs are as follows (where

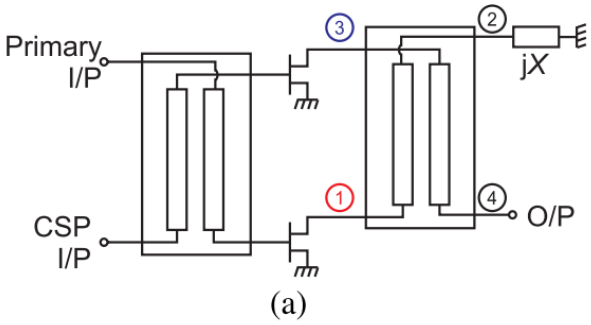

Figure 5.5: The OLMBA schematic, as first envisioned in [29].

α is the control signal power, relative to the PA power, Γ is the reflection coefficient):

$$Z_{PA1} = \frac{1 - \Gamma_L\Gamma_X}{1 + \Gamma_L\Gamma_X} - 2\frac{\Gamma_L + j\alpha\Gamma_X}{(1 + \Gamma_L\Gamma_X)(1 - j\alpha)} \tag{5.4}$$

$$Z_{PA2} = \frac{1 - \Gamma_L\Gamma_X}{1 + \Gamma_L\Gamma_X} - 2\frac{\alpha\Gamma_X + j\Gamma_L}{(1 + \Gamma_L\Gamma_X)(\alpha - j)} \tag{5.5}$$

From here we can see there is a rather complicated relationship between all the parameters. Some insight can be made if we assume $\alpha = 0$ and look at the reflection coefficients [30]:

$$\Gamma_{PA1} = \frac{-\Gamma_L(1 + \Gamma_X)}{1 - \Gamma_L} \tag{5.6}$$

$$\Gamma_{PA2} = \frac{\Gamma_L(1 - \Gamma_X)}{1 + \Gamma_L} \tag{5.7}$$

5.2 Idea

The authors in [30] would go on to show that there exists mathematically a possible value of α and Γ_X which brings each device to a single impedance point, for any Γ_L. What these values are however is not obvious, and would likely require some type of look-up table. Additionally, a continuously variable Γ_X would be extremely difficult to implement in practice on chip (the paper was for a board level design). What we propose instead is to simply alternate the termination jX between open and short, which is (relatively) easily done on chip with a load switch. Notice when $\Gamma_X = 1$ or $\Gamma_X = -1$ in equations 5.6 and 5.7, one of the PAs will see a perfect match ($\Gamma = 0$), while the other will see variation, fairly similar to that of a balanced amplifier. For instance, when $\Gamma_X = -1$:

$$\Gamma_{PA1} = 0;\ \Gamma_{PA2} = \frac{2\Gamma_L}{1 + \Gamma_L} \tag{5.8}$$

In this scenario (short circuited isolation port), we see that if the antenna impedance is smaller than the hybrid's characteristic impedance ($\Gamma_L < 0$), PA2 will see a smaller/reduced impedance than the hybrid's impedance. Similarly, for the open circuit scenario ($\Gamma_X = 1$), we can see:

$$\Gamma_{PA1} = \frac{-2\Gamma_L}{1+\Gamma_L};\ \Gamma_{PA2} = 0 \tag{5.9}$$

Here, when the antenna is larger than the hybrid's impedance, the other PA will see a reduced impedance. When compared to the balanced amplifier, this can potentially offer improved VSWR resilience. Recall in the balanced amplifier case, one PA sees its impedance decrease, and the other increase (by the same percentage). Here however, one PA will always see no change in impedance; the other PA will see a reduced impedance. With a switch we can always pick either open or short for the isolated port termination, whichever is more advantageous for the given antenna load. Using the same ideal simulation as seen for the balanced amplifier, we can observe the output power of the OLMBA with open/short conditions in figure 5.6. Notice now we can keep the output power approximately within 20.3-20.7dB over the 2:1 VSWR, by switching between open and short accordingly. Here the minimum output power actually occurs at perfect matching, because at this point both PAs will see the hybrid's characteristic impedance (whereas under mismatch we can always make one PA see a lower impedance and the other Z_o).

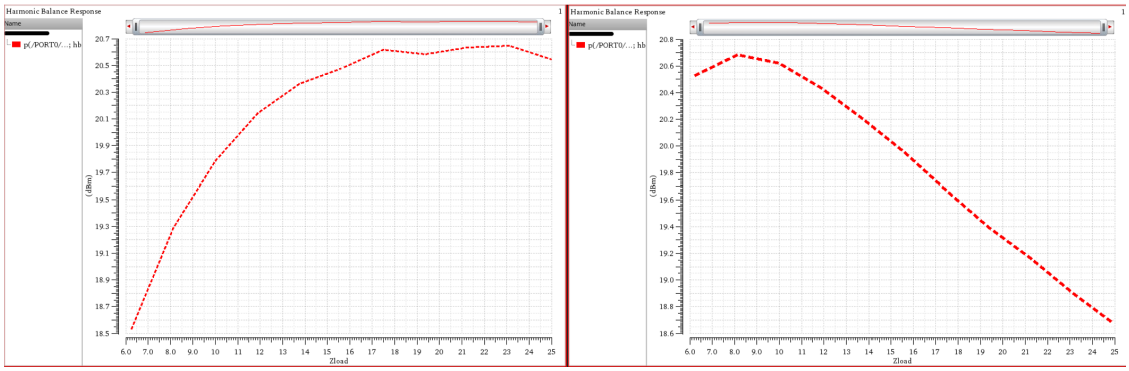

Figure 5.6: Left: OLMBA with $Z_X = 1$ (short) and right: with $Z_X = 1\text{k}\Omega$ (open)

5.3 Implementation Details

The overall schematic design is shown in figure 5.7. The design comprises of an input 50 ohm hybrid, followed by the PAs, a 12.5Ω hybrid, and a 4:1 Ruthroff transformer to the 50Ω output. Switch S1 provides the open/short condition. Note the entire design was single ended, done in 28nm bulk digital CMOS. The top-level layout is shown in figure 5.8.

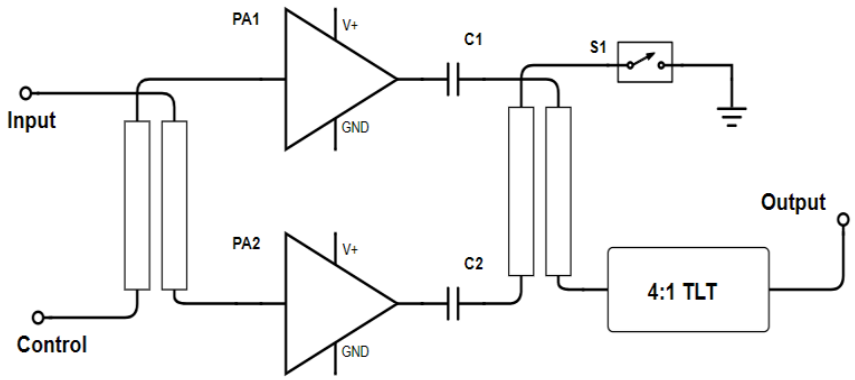

Figure 5.7: Overall schematic of the VSWR tolerant OLMBA idea.

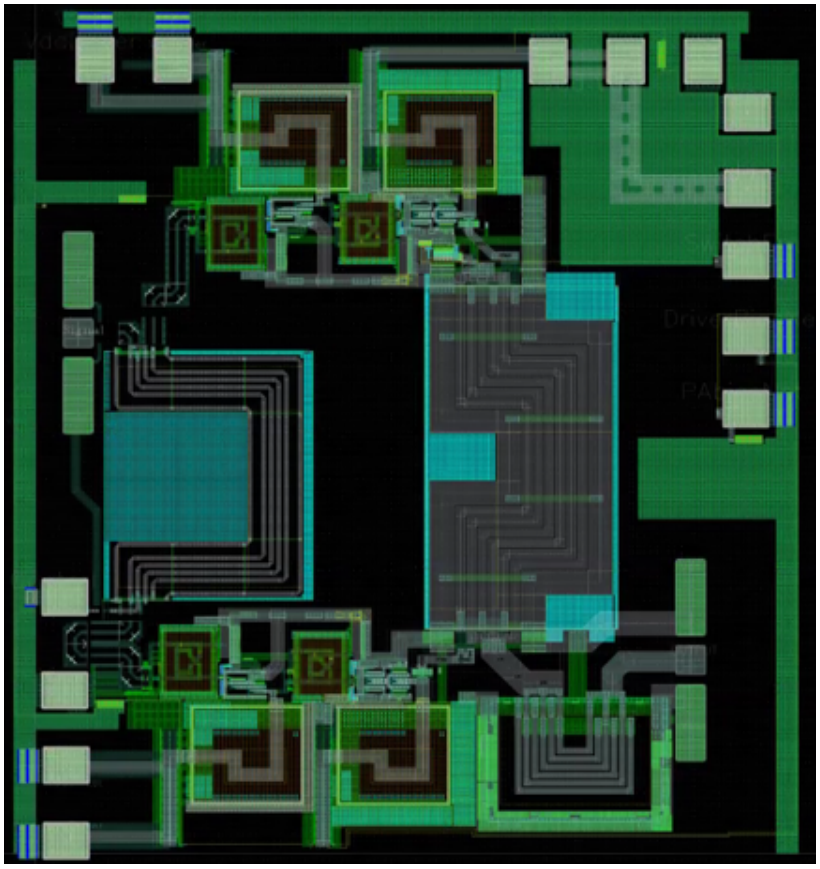

Figure 5.8: Top level layout of the OLMBA.

As mentioned in the beginning, the actives were poorly designed (ran out of time before the tape-out), and certainly the decision to make everything single-ended was a poor one in hindsight. It does however demonstrate the VSWR tolerance, with only 0.5dB of output power variation across 2:1 VSWR. The performance is specified in table 5.1.

Table 5.1: VSWR-Resilient Orthogonal LMBA Simulated Performance

Specification	Simulated Result
Vdd	1.4V
Psat across 2:1 VSWR	11.3 $\rightarrow$ 11.8 dBm
P1dB	6 dBm
PAE	5%
Minimum Power Gain across VSWR & Bandwidth	15 dB
Bandwidth (-3dB Power Gain)	27-42 GHz

5.3.1 Passives

It's worth taking a look at the passives in this design, as there is an interesting comparison to be made. In this tapeout, a more traditional, planar approach (akin to traditional planar microwave design) was taken. The hybrids and Ruthroff transformers all used co-planar waveguide (CPW), Lange-coupled lines. For planar structures, because the sidewalls (metal thickness) of the metals in the CMOS process are rather thin (even for 'thick' metals), multi-line (Lange) coupled lines are necessary to achieve the tight coupling necessary for -3dB couplers (see [14] for more details). Microstrip lines would not work as the even mode impedance would be far too low (as the ground plane would be right underneath, and the dielectric height in IC designs are typically quite small). CPW lines therefore allow us to design the return path accordingly to provide the required even mode for the passives.

The Lange-coupled Ruthroff 4:1 transformer is shown in figure 5.9, and its S11 (12.5Ω reference) is shown in figure 5.10. To maximize the coupling, a six-finger Lange coupler, with the minimum trace separation allowed by the DRC rules was used. The structure had an insertion loss of 0.8dB, a -10dB S11 matching bandwidth of 32 to 53GHz, and an overall size of 199x288um. If we compare this to the Ruthroff of figure 4.37, we can see the broadside coupled version is better across the board: the broadside version has lower insertion loss (0.4dB), smaller dimensions, and larger fractional bandwidth. This shows that the broadside lines offer much stronger coupling and higher even mode impedances. Additionally broadside designs are much easier to both simulate and design, showing that Lange-style designs should be restricted to situations where multi-layer designs are impractical (PCBs, thin-film alumina, etc).

The Lange hybrid coupler is shown in figure 5.11, and its magnitude response (S21 and S31) is shown in figure 5.12. Four fingers provided the necessary coupling for -3dB coupling, and the wide traces provide both the low characteristic impedance as well as the necessary width to handle electro-migration considerations. Note the total length of the structure is 750um ($\lambda/4$ at the center frequency).

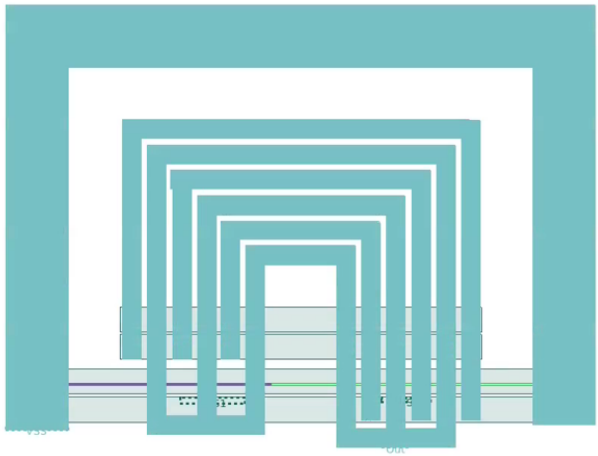

Figure 5.9: A CPW, Lange Coupled Ruthroff 4:1 Transformer.

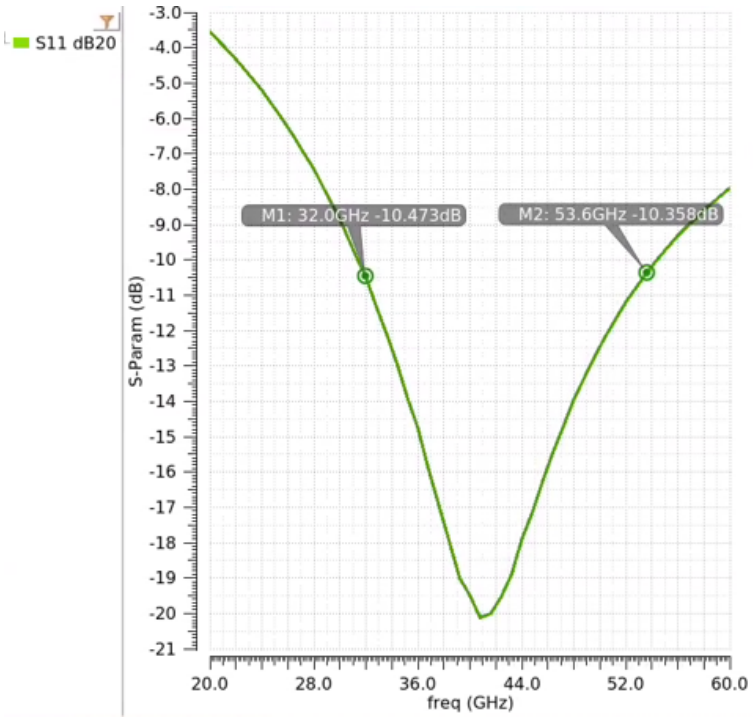

Figure 5.10: S11 of the Lange Ruthroff Transformer

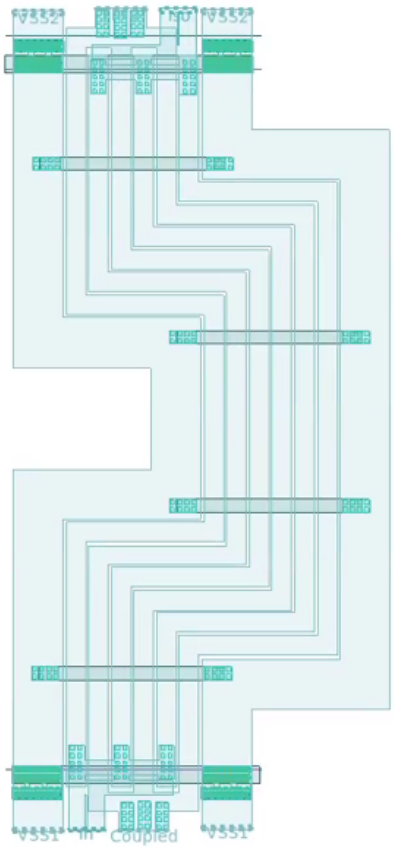

Figure 5.11: A 12.5Ω CPW Lange hybrid coupler.

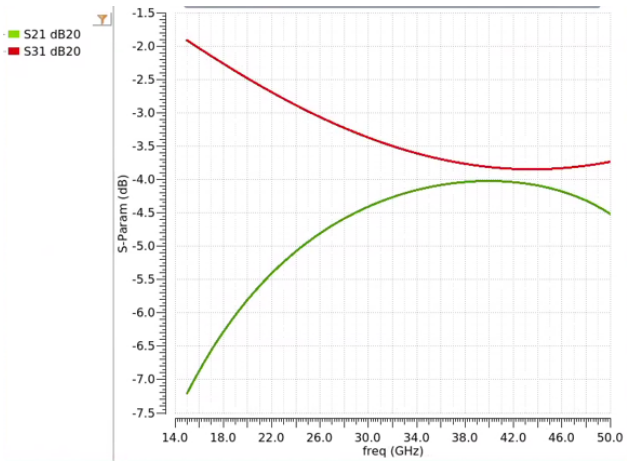

Figure 5.12: S21 and S31 Magnitudes for the CPW Lange hybrid coupler.

Chapter 6

Concluding Remarks

6.1 Summary

In this thesis we have covered the details of transmission line transformers, specifically Guanella transformers, and how they are applicable to modern CMOS power amplifier designs. In particular, their use as broadband output and interstage matching networks was highlighted. Practical tips were given on the design and simulation of these structures in the K, Ka bands, emphasizing the use of broadside coupled lines as the primary topology, and well-defined return paths for simulation. Then, we further demonstrated the Guanella transformer's usefulness in power combining networks, as replacements for traditional flux based series/parallel transformers. Four-way, two-way, and single slice power amplifiers were designed, taped-out, and measured, achieving octave-plus bandwidths, high power (for bulk CMOS), and reasonable efficiencies.

6.2 Future Directions

There are several future opportunities for improvement here. The novelty of this work is based on the passive networks, not so much the actives. These passives can still be used for more complex topologies, such as Doherty or polar architectures. The OLMBA for VSWR mitigation was also mentioned as a future work that could incorporate Guanella transformers. It may also be worthwhile to figure out if the package parasitics could potentially be incorporated into the transmission line network, creating a more optimized system/product.

Bibliography

[1] Qualcomm. *Global Update on 5G Spectrum*. https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/spectrum-for-4g-and-5g.pdf.

[2] Navy SBIR/STTR. "N254-C03: Catapult Challenge: Modular Electronic Warfare Counter-Unmanned Aerial System (C-UAS) Payload". In: *https://www.navysbir.com/n25_41/N254-C03.htm* (2024).

[3] et al. Hua Wang. "Power Amplifiers Performance Survey 2000-Present". In: *https://ideas.ethz.ch/research/surveys/pa-survey.html* ().

[4] Weisen Zeng et al. "25.2 A 19.7-to-43.8GHz Power Amplifier with Broadband Linearization Technique in 28nm Bulk CMOS". In: *2023 IEEE International Solid-State Circuits Conference (ISSCC)* (2023), pp. 372–374.

[5] Valdrin Qunaj and Patrick Reynaert. "A Ka-Band Doherty-Like LMBA for High-Speed Wireless Communication in 28-nm CMOS". In: *IEEE Journal of Solid-State Circuits* 56.12 (2021), pp. 3694–3703.

[6] Marco Vigilante and Patrick Reynaert. "A Wideband Class-AB Power Amplifier With 29–57-GHz AM–PM Compensation in 0.9-V 28-nm Bulk CMOS". In: *IEEE Journal of Solid-State Circuits* 53.5 (2018), pp. 1288–1301.

[7] Kaushik Dasgupta et al. "A 26 dBm 39 GHz Power Amplifier with 26.6% PAE for 5G Applications in 28nm bulk CMOS". In: *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2019), pp. 235–238.

[8] Mohamed M.R. Esmael, Mohamed A.Y. Abdalla, and Islam A. Eshrah. "A 19-43 GHz Linear Power Amplifier in 28nm Bulk CMOS for 5G Phased Array". In: *2019 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)* (2019), pp. 1–3.

[9] Jaehun Lee and Songcheol Hong. "A 24–30 GHz 31.7% Fractional Bandwidth Power Amplifier With an Adaptive Capacitance Linearizer". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 68.4 (2021), pp. 1163–1167.

[10] G Guanella. "New Method of Impedance Matching in Radio-Frequency Circuits". In: *The Brown Boveri Review* (Sept. 1944).

[11] C. L. Ruthroff. "Some Broad-Band Transformers". In: *Proceedings of the IRE* 47.8 (1959), pp. 1337–1342.

[12] Yoshitaka Niida et al. "A 0.6–2.1-GHz Wideband GaN High-Power Amplifier Using Transmission-Line-Transformer-Based Differential-Mode Combiner With Second-Harmonic Suppression". In: *IEEE Transactions on Microwave Theory and Techniques* 69.3 (2021), pp. 1675–1683.

[13] Wentao Zhou et al. "A Compact Multilayer-Stacked Guanella Transformer Based 50–680-MHz GaAs Power Amplifier". In: *IEEE Microwave and Wireless Technology Letters* 34.6 (2024), pp. 627–630.

[14] Rajesh Mongia, Inder Bahl, and Prakash Bhartia. *RF and Microwave Coupled-Line Circuits*. Artech House, 1999.

[15] I.J. Bahl. "Broadband and compact impedance transformers for microwave circuits". In: *IEEE Microwave Magazine* 7.4 (2006), pp. 56–62.

[16] Tim Canning, Jeff R. Powell, and Steve C. Cripps. "Optimal Design of Broadband Microwave Baluns Using Single-Layer Planar Circuit Technology". In: *IEEE Transactions on Microwave Theory and Techniques* 62.5 (2014), pp. 1183–1191.

[17] Steve C. Cripps. "Pushing and Pulling". In: *IEEE Microwave Magazine* (Oct. 2020).

[18] Nathan Marchand. "Transmission-Line CONVERSION". In: *Electronics* (Dec. 1944).

[19] Nathan Marchand. *Ultrahigh Frequency Transmission & Radiation*. Wiley, 1947.

[20] Tom Apel and June Zhao. "Pushpull Power Amplifiers for Handsets". In: ***https://k5tra.net/*** ().

[21] Kyu Hwan An et al. "Power-Combining Transformer Techniques for Fully-Integrated CMOS Power Amplifiers". In: *IEEE Journal of Solid-State Circuits* 43.5 (2008).

[22] Alan W. Davis and Krishna Agarwal. *Radio Frequency Circuit Design*. John Wiley & Sons, Inc, 2001.

[23] *1:4 Transmission Line Transformers: Coax Impedance versus Length*.

[24] Nima Baniasadi. "Millimeter-Wave Receiver and Package Design Close to the Device Activity Limits". PhD thesis. EECS Department, University of California, Berkeley, May 2023. URL: `http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-40.html`.

[25] Akram Sheikhi. "Historical Aspect of Load-Modulated Balanced Amplifiers". In: *IEEE Access* 12 (2024), pp. 7974–7986.

[26] Daniel J. Shepphard, Jeffrey Powell, and Steve C. Cripps. "An Efficient Broadband Reconfigurable Power Amplifier Using Active Load Modulation". In: *IEEE Microwave and Wireless Components Letters* 26.6 (2016), pp. 443–445.

[27] K. Kurokawa. "Design theory of balanced transistor amplifiers". In: *The Bell System Technical Journal* 44.8 (1965), pp. 1675–1698.

[28] Steve C. Cripps. "Coupling Factors [Microwave Bytes]". In: *IEEE Microwave Magazine* 22.3 (2021), pp. 12–87.

[29] David J. Collins et al. "The Orthogonal LMBA: A Novel RFPA Architecture With Broadband Reconfigurability". In: *IEEE Microwave and Wireless Components Letters* 30.9 (2020), pp. 888–891.

[30] Roberto Quaglia et al. "Mitigation of Load Mismatch Effects Using an Orthogonal Load Modulated Balanced Amplifier". In: *IEEE Transactions on Microwave Theory and Techniques* 70.6 (2022), pp. 3329–3341.