University of California
Santa Barbara

AlGaN/GaN High-Electron-Mobility-Transistors
Based
Flip-chip Integrated Broadband Power Amplifiers

A dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering
by

Jian Xu

Committee Members:
Professor Robert A. York, Chairperson
Professor Umesh K. Mishra
Professor Stephen I. Long
Doctor Yifeng Wu

December 2000
The dissertation of Jian Xu
is approved:

[Signatures]

[Signature]
Committee Chairperson

December 2000
Copyright by

Jian Xu
2000
Acknowledgments

This work would not have been possible but for the collective efforts of my advisory committee, my lab mates and the hard-working UCSB supporting staff team.

I'm deeply indebted to my graduate advisor, Professor Robert York, who is extremely hard working and dedicated to his work, for his guidance and encouragement during the past four years. His work ethics, pragmatic methodology have greatly influenced my approach to the professional world and surely will be with me for the years to come. I'm deeply grateful to Professor Umesh Mishra, Professor Steven Long and Doctor Y.F. Wu for their supervising of the thesis and being extremely supportive and inspirational throughout this endeavor.

My lab mates have made the work environment enjoyable and exciting. I thank you: Amit, Nick, Vicki, Paolo, Pengcheng, Bruce, Erich, Chinghui, Primit, Prashant, Rob, Peter, James, Rama, Lee, Gia, Can, Huili, Naiqiang, Likun, for giving me a long-lasting memory of many good moments in the hardest four years of my life.

Thanks to the efforts of Bob Hill, Mike Anzlowar, Jack Whaley and Gerry Robinson to keep the cleanroom operational 24 hours a day, 365 days a year. Without them, I might not need to spend that much time in the cleanroom.

Finally, I want to thank my family, especially my husband Xiaodong, for their faith in me, their sacrifice and their unconditional love and support. I'm afraid that I may not have anything to them in return. This work is for them and I hope I have made them proud.
### Vita

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 16, 1971</td>
<td>Born, Shanghai, China</td>
</tr>
<tr>
<td>June 1992</td>
<td>Bachelor of Electrical Engineering</td>
</tr>
<tr>
<td></td>
<td>Zhejiang University, Hangzhou, China</td>
</tr>
<tr>
<td>June 1995</td>
<td>Master of Electrical Engineering</td>
</tr>
<tr>
<td></td>
<td>Zhejiang University, Hangzhou, China</td>
</tr>
<tr>
<td>September 1995-</td>
<td>Research Assistant</td>
</tr>
<tr>
<td>September 1996</td>
<td>University of Nevada, Las Vegas</td>
</tr>
<tr>
<td>September 1996-</td>
<td>Research Assistant</td>
</tr>
<tr>
<td>October 2000</td>
<td>University of California, Santa Barbara</td>
</tr>
<tr>
<td>October 2000</td>
<td>Ph.D Electrical and Computer Engineering</td>
</tr>
<tr>
<td></td>
<td>University of California, Santa Barbara</td>
</tr>
</tbody>
</table>
Publications

Journal Publications:


Conference Papers:


Abstract

AlGaN/GaN High-Electron-Mobility-Transistors based Flip-chip Integrated Broadband Power Amplifiers

GaN HEMTs have enormous potential for realizing high-power solid-state amplifiers at microwave frequencies, due to their characteristics of wide band-gap materials such as high breakdown field, high electron saturation velocity and high operating temperature. Through flip-chip mounting the device onto a ceramic Aluminum Nitride substrate, which also hosts all the passive components in the GaN amplifiers, we have achieved record high single-device results of 4.4 W output power (CW) at 8 GHz with a 1mm-wide GaN HEMT on sapphire. Three generations of GaN broadband power amplifiers have been designed and fabricated. First, a modified traveling wave power amplifier with 1-8 GHz bandwidth and up to 4.5-Watt output power was demonstrated, which was also the first ever reported GaN amplifier. Higher efficiency can be achieved by eliminating the backward wave of the conventional TWA. But the unequal drive problem limited the efficiency to less than 15%. The second generation of GaN broadband power amplifier was a novel input LCR broadband matching structure. Record high 8.5-Watt output power, 20 % PAE at 8 GHz was achieved with a 3-dB-bandwidth of 3-10 GHz. Combining the advantages of the first and second circuits, the third generation GaN broadband power amplifier was a 2×2 matrix modified TWA. Optimized overall power performance was achieved with up to 7.5-Watt output power, close to 1 – 6 GHz bandwidth and up to 25 % efficiency.
Contents

Chapter 1 Introduction 1

Chapter 2 AlGaN/GaN HEMTs power device design and fabrication 7
  2.1 AlGaN/GaN HEMTs material system 7
  2.2 Device fabrication 10
  2.3 Device DC and RF characterization 14
  2.4 Large gate periphery device design 19
  2.5 Flip-chip mounting technique for thermal management 24
  2.6 Device small-signal and large-signal modeling 31
  2.7 Load-pull system 38

Chapter 3 Broadband matching design 45
  3.1 Conjugate match and load-line match 45
  3.2 Bandwidth limitation 47
  3.3 Bandwidth extend methods 49
  3.4 Broadband LC lowpass matching networks 54
  3.5 Broadband power combiner/divider design 58

Chapter 4 Power amplifier structures 71
  4.1 Challenges of broadband power amplifier design 71
  4.2 Power amplifier structures 75
  4.3 Distributed power amplifier (Traveling wave amplifier) 81

Chapter 5 GaN power amplifiers design and performance 93
  5.1 Modified traveling wave power amplifier 93
  5.2 GaAs modified traveling wave power amplifier 96
  5.3 GaN modified traveling wave power amplifier 112
  5.4 GaN LCR-matched power amplifier 122
  5.5 Modified 2×2 matrix traveling wave power amplifier 130
Chapter 6 Conclusions 137
  6.1 Summary 137
  6.2 Future work 140
Appendix A GaN HEMT process flow 145
Appendix B AlN-substrate based FC-IC fabrication 154
Chapter I INTRODUCTION

With the development of the wireless communications, telecommunications, data communications and aerospace systems, the demand for solid-state power amplifiers has been continually increasing over the last decade. The requirements include aspects of high power level, high efficiency, high linearity and high operating frequency and the relative importance of each of these features is application specific. One major trend is the continuous demand for more power at higher frequencies. Unfortunately, the existing technologies have struggled to sustain the higher demand. GaAs-based power devices have been very reliable workhorses at high frequencies especially microwave spectrum. However, their power performances have already been pushed close to the theoretically limit. In terms of power density, about 1 W/mm at 10 GHz would be the state of art performance for GaAs power pHEMTs. Under such circumstance, wide band-gap semiconductors with an order of magnitude or so higher breakdown voltage along with excellent thermal properties began to emerge. For many years, silicon carbide (SiC) has been widely touted as a potential candidate, but unfortunately it doesn’t appear to be a truly microwave technology in terms of its carrier mobility. With technological and cost advantages over competing SiC, Gallium Nitride (GaN) based family of semiconductor has quickly gained the center attention. In terms of electronic material properties, GaN possessed very attractive features such as a large band-gap (3.4 eV), high breakdown field (3 ×10^6 V/cm), the existence of modulation doped AlGaN/GaN structures with attendant high electron mobility (1500 cm^2V^{-1}s^{-1}) and extremely high peak (3 ×10^7 cm/s) and saturation electron velocity (2 ×10^7 cm/s). Also GaN has been successfully grown on SiC substrates, which have excellent thermal property. GaN epitaxial layers can also be grown on single-crystal sapphire (Al_2O_3) substrates. Sapphire has the advantage of lower cost and availability in larger wafer size than SiC, but
it is a poor thermal conductor. GaN HEMTs on SiC have achieved a power density of 6.9-9.2 W/mm, while GaN HEMTs on sapphire have demonstrated 6.6 W/mm, and the difference is believed to be due to the higher thermal conductivity of SiC (\(\sigma = 450\) W/mK) over sapphire (\(\sigma = 30\) W/mK). However, by paying close attention to thermal management, GaN-HEMT-on-Sapphire can still have competitive performance. In particular, adequate thermal management can be achieved through flip-chip bonding of the device onto a thermally-conductive, electrically-insulating substrate, such as ceramic Aluminum Nitride (AlN) (\(\sigma = 180\) W/mK), which also hosts the matching networks and combiner networks in the GaN amplifiers. Using this technique we have achieved record high single-device results of 4.4 W output power (continuous wave (CW)) at 8 GHz with a 1mm-wide GaN HEMT on sapphire.

Under the Office of Naval Research (ONR) funded MURI program, our main research goal of this work will focus on the development of multi-octave microwave bandwidth, high power, and high efficiency GaN amplifiers. Their main application would be the power sources for the phased-array radar that will be the key component of their proposed Advanced Multifunction RF Systems (AMRFS). The achievement of this research can also be readily transferred to commercial applications such as wireless communication, satellite communication and etc. However, in order to realize the objective there are several major challenges that we must face and be able to find corresponding solutions:

1) High impedance transformation ratio:
In order to achieve higher output power level, large periphery device would be required. Consequently, the lowered impedance levels of the devices would represent more difficulty in the design of the matching networks. For example,
the impedance levels for a GaN amplifier using a total of 4-mm-wide GaN HEMTs require transformation ratios of 7:1 for the output network and 50:1 for the input network. Impedance transformation is usually obtained with either band-pass networks such as λ/4 transformer or low-pass networks such as simple serious-L shunt-C matching circuit. However, all of these matching networks have limited bandwidth that is closely related to the ratio of the impedance transformation. Generally speaking, the ratio for each matching circuit should be less than 1.5 in order to achieve larger than 3:1 bandwidth. Multi-section of matching networks would be necessary to achieve the desired bandwidth, resulting in more complex and lossy matching networks.

2) Large signal modeling:
The evolutionary CAD technology advances that have been made in recent years have been greatly supported through the parallel development of more valid models for both active devices and passive circuit elements. In particular, the accurate large-signal performance predictions for microwave circuits depend very heavily on the accurate large-signal models for the active devices. Unfortunately, due to the early stage of GaN HEMTs technology, large-signal modeling is still an ongoing study field. We have to rely greatly on the load-line approach and the empirical yet very effective load-pull measurement for large-signal design.

3) Thermal management:
The efficiency of broadband power amplifiers tend to be fairly low, 20 -30 % power added efficiency would be the range for most of the amplifiers. With efficiency in this range, the disposal of heat becomes a serious problem, since too high a channel temperature will degrade performance and shorten the life of a semiconductor device. Careful thermal design of large-periphery GaN HEMTs
would be required, in order to achieve a balance between low thermal resistance and high RF gain at higher frequencies.

4) Poor input/output matching (large VSWR)
In general, the input and output matching of the broadband power amplifier tends to be poor. At the output of the power amplifier, the device is not conjugately matched; instead the optimum load has to be presented to the device for the maximum output power. Input broadband matching is extremely hard due to the high impedance transformation ratio. There are several ways to address the problem. First, we can employ a distributed network at the input to realize good broadband matching, but at the output, the distributed network is not a good choice because it would greatly reduce the efficiency. We can also design multistage amplifiers to achieve better input matching by distributing the high impedance transformation ratio among different stages, however the increased complexity of biasing circuitry design could overweigh its benefits. Finally, the balanced amplifier is a very good approach to achieve low VSWR at both input and output. However, due to the difficulty of CPW coupler design and the low efficiency of the structure, we will not consider this technique in this work. The output VSWR of all the amplifiers in this work will remain high since it’s not our main concern.

With the power density advantage of the GaN HEMTs technology and the novel approaches to the challenges of the broadband power amplifier design, we have successfully designed and fabricated three generations of GaN power amplifiers. First of all, a modified traveling wave power amplifier with 1-8 GHz bandwidth and up to 4.5-Watt output power was demonstrated, which was also the first ever reported GaN amplifier. Higher efficiency can be achieved by eliminating the backward wave of the conventional TWA. But the unequal drive problem limit
the efficiency to less than 15%. Second generation of GaN broadband power amplifier was a novel LCR-matched, high-efficient corporate power combing structure. Record high 8.5-Watt output power, 20% PAE at 8 GHz was achieved with a bandwidth of 3-10 GHz. Broader bandwidth can be achieved with the first circuit topology, and the second circuit tends to have higher efficiency. Therefore, a third generation of the GaN broadband power amplifier was a 2×2 matrix modified TWA which combines the advantages of the first and second circuits. Optimized power performance was achieved with up to 7.5-Watt output power, close to 1–6 GHz bandwidth and up to 25% efficiency. It proves to be the better circuit topology in terms of the overall performance of the bandwidth, power and efficiency. Compared with the commercial state-of-the-art GaAs broadband power amplifiers, which is best represented by the TriQuint product TGA9083-EEU (6.5-11.5 GHz bandwidth, 5 Watt output power and 40% PAE) and TGA9092-EPU (6-18 GHz bandwidth, 2.8 Watt output power and 25% PAE) with the 0.25μm pHEMT technology, GaN broadband power amplifiers have already demonstrated its power-bandwidth product advantage even with the existing immature 0.7μm HEMT technology. The relatively low PAE could be greatly improved with the progress of the GaN HEMTS material system in the near future.

This thesis includes six chapters. First chapter is the introduction of the thesis, it describes the background of the research and defines the objective of the thesis. Chapter 2 describes the design and fabrication of AlGaN/GaN HEMTs power device, from the basics of the material system, the characterization of small devices to the development of large gate-periphery devices, which are critical to the success of the GaN power amplifier design. Chapter 3 reviews the principle of the power amplifier design, and some basic bandwidth extending techniques
are also described. Chapter 4 discusses the major challenges of the broadband power amplifier design and some basic power amplifier structures: traveling wave amplifier is analyzed in details in terms of both the small-signal and large-signal characteristics. Chapter 5 presents the design of GaN broadband power amplifiers that include modified TWPA, LCR-matched PA and the 2×2 matrix TWPA, their respective performances are also included in this chapter. Chapter 6 concludes the thesis with the achievements summary and suggestions for the future work. Appendix A describes the GaN power device process flow. Appendix B summarizes the AlN substrate fabrication flow and addresses some common problems in the amplifier fabrication.
Chapter 2
AlGaN/GaN HEMTs Power Device Design and Fabrication

This Chapter describes the design and fabrication of AlGaN/GaN HEMTs power
device, from the basics of the material system, the characterization of small
devices to the development of large gate-periphery devices, which is critical to the
success of the GaN power amplifier design.

2.1 AlGaN/GaN HEMTs material system

The AlGaN/GaN material system possesses fundamental electronic properties that
make it an ideal candidate for high power microwave devices. First, as a wide
band-gap material ($E_g = 3.4$ eV), GaN has very high electric breakdown field (~2
MV/cm), as a result, GaN-based device can be biased at very high drain voltage
($V_{break-down} = 50 - 500$ V) depending on the application of the device, also it can be
operated at higher channel temperature (~ 300 °C). Second, it possesses high
saturation ($2 \times 10^7$ cm/s) electron velocity, which contributes to higher current
density since $I_{max} \propto qn_s v_s$ ($q = 1.6 \times 10^{-19}$ coulomb, $n_s$ is the sheet charge density, $v_s$
is the electron saturation velocity) and high operating frequency since $f_t \propto v_s/L_{eff}$.
Third, the existence of modulation-doped AlGaN/GaN structure leads to 1) high
sheet charge density ($n_s = 1 \times 10^{13}$/cm$^2$), which is the other reason for high $I_{max}$;
and 2) high electron mobility ($\mu = 1200 - 1500$ cm$^2$/V-s), which is largely
responsible for low on-resistance (low knee voltage) since channel resistance is
related to $1/(q\eta_s \mu E)$ at low electric field.

Consequently, AlGaN/GaN HEMTs can achieve very high break down voltage,
very high current density, and sustain very high channel operating temperature.
Furthermore, high operating frequency ($f_t$) and high drain power added efficiency (PAE) could be achieved. Fig. 2.1-1 illustrates the relationship of those characteristics as power microwave devices resulting from the intrinsic electronic properties of the AlGaN/GaN HEMTs structure.

![Diagram showing relationships between various properties of AlGaN/GaN HEMTs](image)

Fig. 2.1-1 Electronic properties of AlGaN/GaN HEMT structure

One important figure of merit to describe the high-frequency power performance of the device is the maximum RF power and $f_t$ product. According to Wemple’s model\(^5\), device gate-drain breakdown voltage can be approximated by

$$V_{df\beta} = \frac{\varepsilon L_{\text{eff}} E_{br}^2}{2qN_D a}$$

where $a$ is the active layer thickness, $N_D$ is the active layer donor density, $E_{br}$ is the break-down electrical field, $L_{\text{eff}}$ is the effective gate length ( according to Wemple, for $L_g < 0.5 \, \mu m$, $L_{\text{eff}} = L_g$, otherwise $L_{\text{eff}} = 0.5 \mu m$). Also, maximum current $I_{\text{max}}$ is given by

$$I_{\text{max}} = qN_D a v_f v_g$$
then the maximum linear output power $P_{\text{max}}$ can be approximated by

$$P_{\text{max}} = \frac{I_{\text{max}} V_{\text{dss}}}{8} = \frac{\varepsilon w g L_{\text{eff}} v_s E_{\text{br}}^2}{16}$$

Since approximately $f_t = v_s / (2\pi L_{\text{eff}})$, we can obtain the power $f_t$ product as following:

$$P_{\text{max}} f_t = \frac{\varepsilon w g v_s^2 E_{\text{br}}^2}{32\pi}$$

Although practically the power performance doesn’t scale proportionally to the $P_{\text{max}} f_t$ product (also referred to as “Johnson’s figure of merit”) due to other effects, it is nevertheless a clear indication of the advantage of GaN as a material for power devices over conventional semiconductors (Table 2.1).

<table>
<thead>
<tr>
<th>Material system</th>
<th>$v_s$ (cm/s)</th>
<th>$E_{\text{br}}$ (V/cm)</th>
<th>$P_{\text{max}} f_t$ (W-GHz/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silicon</td>
<td>$6 \times 10^6$</td>
<td>$3 \times 10^5$</td>
<td>$\sim 2$</td>
</tr>
<tr>
<td>GaAs</td>
<td>$10 \times 10^6$</td>
<td>$4 \times 10^5$</td>
<td>$\sim 9$</td>
</tr>
<tr>
<td>GaN</td>
<td>$20 \times 10^6$</td>
<td>$20 \times 10^5$</td>
<td>$\sim 400$</td>
</tr>
</tbody>
</table>

Table 2.1 Johnson’s figure of merit of Si, GaAs and GaN

The device layer structure used for this study is shown in Fig. 2.1-2. The substrate can be C-plane sapphire substrate or 4H-SIC substrate. Sapphire substrate is chosen here for two reasons: 1) Sapphire is much cheaper than SiC substrate, 2) Sapphire can be available in larger wafer size than SiC, and this could further reduce cost. Metal organic chemical vapor deposition (MOCVD) is used to grow
the sample. First, 200 Å GaN nucleation layer is grown on a sapphire substrate. This is followed by a 2-3 µm GaN channel unintentionally doped (UID) with a background doping density around $1 \times 10^{15} \text{cm}^{-3}$. AlGaN (Al composition is about 33 %) structure is consisted of a 30 Å UID AlGaN spacer, a 120 Å Si-doped heavily doped layer ($N_D = 5 \times 10^{18} \text{cm}^{-3}$) and a 50 Å UID cap.

![Fig. 2.1-2 Epi structure of the AlGaN/GaN HEMT](image)

Hall measurement reveals that the room temperature mobility is about 1200 – 1500 cm$^2$/V-s and sheet charge density of about $1 \times 10^{13} \text{cm}^{-2}$ could be achieved.

### 2.2 Device fabrication

Fig 2.2-1 shows the layout of a 100 – µm wide device. Four process steps are needed for the fabrication of these small devices: 1) Ohmic contact (Source/Drain Pattern), 2) Gate pad isolation, 3) Schottky contact (Gate Pattern), 4) Mesa isolation.
Fig. 2.2-1 Layout of a 100-µm wide small device

Fig. 2.2-2: TLM measurements for a typical ohmic contact of GaN HEMT. \(2R_c = 15.2 \, \Omega\) for 75 µm, \(R_c = 0.5 \, \Omega\)-mm)
1) Ohmic contact:
The first step is to place ohmic contacts, which are source and drain patterns. Al forms a natural ohmic contact to n-GaN. Multi-layer ohmic scheme of Ti/Al/Ni/Au (200/2000/400/500 Å) annealed in RTA (rapid thermal annealer) at 880 °C for 30 seconds could yield contact resistance $R_c$ ranging from 0.5 Ω-mm to 1.5 Ω-mm (Fig. 2.2-2). $R_c$ can be measured through TLM (transfer length method) pattern, which is comprised of more than three ohmic contacts with unequal spacing between contacts. The total resistance between any two contacts is: $R_T = 2R_c + \rho_s d/Z$. ($\rho_s$ is the sheet resistance: a measure of the resistivity averaged over the sample thickness, d is the contact spacing, Z is the contact width). From the plot of total resistance as a function of contact spacing, as shown in Fig. 2.2-2, we can obtain $\rho_s / Z$ (the slope of plot) and $2R_c$ (the intercept of the plot with the x-axis). The Ti layer can help the formation of n$^+$-GaN interface and its good attachment to the surface is desirable during metal deposition, Au layer was for preventing the oxidation of the Al and a better conductivity of the metal layers, and Ni layer was utilized for the reducing of the Au/Al mixing. With such ohmic contact scheme, good metal morphology can be achieved. RIE 5 Cl$_2$ dry etching was used to etch through the UID AlGaN cap before the metal deposition, but careful etching thickness control was required. If etching was so deep that the Si-doped AlGaN was mostly etched away, then there won’t be enough charges to be supplied to the 2-Deg electron gas; as a result, the current density will be greatly reduced. If etching wasn’t deep enough that the UID cap wasn’t etched through, then good ohmic contact couldn’t be achieved. We used very low etching power (Power = 60 W) and etched for about 14 – 16 seconds, the etching thickness was about 100 Å. The Cl$_2$ gas flow is 10 ml, the chamber pressure is 5 milliTorr, and it is good enough to start etching when the chamber pressure is below $2 \times 10^{-6}$ Torr.
2) Gate pad isolation and Schottky contact:

To reduce the gate pad parasitic capacitance, a SiO$_2$ layer was deposited underneath the gate metal. After the lithography of the SiO$_2$ pattern, RIE 5 dry etching was utilized to etch away about 1000 Å thick AlGaN/GaN layer, the thickness control was not as critical as the ohmic contact etching. The etching condition was similar to ohmic contact etching except that the etching power used was much higher (Power = 250 W), as a result, etching rate was much higher (about 1500 Å/min), and therefore about 45 second etching was needed. Then about 1000 Å thick SiO$_2$ was deposited just to fill up the etched trench to achieve a planar surface for the later gate metal deposition.

Au, Ni and Pt metals have Schottky barriers of about 0.8 ~ 1.1eV on n$^-$-GaN and are possible candidates for the gate metals. In this study, metal scheme of Ni/Au/Ni (200/5000/400Å) was utilized for the gate Schottky contact. Thick Au metal was deposited for the reduction of the gate metal resistance, this is especially important for the power device since larger current capacity is expected. Last layer of Ni was for the protection of the Au layer during the later mesa etching since long time dry Cl$_2$ etching with high power will damage Au metal. Schottky contact is the process step requiring high-resolution lithography; gate finger alignment and dimension are very critical to the performance of the device. Gate offset is employed, which shifts the gate closer to the source to reduce the source resistance $R_s$. The source-drain space was 2.5 μm; the gate length was 0.5 μm in the mask design and the fabricated gate turned out to be about 0.7 μm using the i-line Stepper lithography and $L_{gs} = 0.8$ μm, $L_{gd} = 1$ μm.

3) Mesa isolation:
RIE 5 dry etching was used again for the mesa isolation. It doesn’t need to etch down to the sapphire substrate as illustrated in Fig. 2.2-3, about 4500Å etching turned out to be good enough for the device isolation, which needed about 3 minutes etching using high etching power of 250 W. After the mesa isolation, we can carry on DC and RF characterization on the devices.

Fig. 2.2-3 Process steps of AlGaN/GaN HEMTs

2.3 Device DC and RF characterization

DC I-V curve of a 50-μm wide 0.7-μm long device was measured on the curve tracer. Typical DC characteristics (Fig. 2.3 -1) are following: $I_{\text{max}} = 1$ A/mm (at $V_{gs} = 1.5$ V), $V_{\text{break-down}} = 50$ V, $V_{knee} = 5 \sim 6$ V, $g_{m,\text{extrinsic}} = 200$ mS/mm, $V_{\text{pinchoff}} = -4 \sim -5$ V.
Fig. 2.3-1 Typical DC-IV curve for a 50-μm wide 0.7-μm long GaN HEMT

Fig. 2.3-2 Observed dispersion phenomena for GaN HEMTs. (Longer traces are “dispersed” IV curve, shorter traces are “un-dispersed” IV curve)

But the noticeable difference of I-V curve between DC and pulsed condition was observed, which is referred to as “dispersion” phenomena. Fig. 2.3-2 shows the output IV characteristics of a 50 μm wide device, under DC and pulsed (80 microsecond) gate drives. As can be seen, current is reduced and knee voltage is increased with pulsed operation. The pulsed drive was used to simulate AC
operation. With pulse operation, the device is first pinched off, and electrons are injected into the traps in the material on the drain side under this high field condition. When the channel is turned on by short pulse, the trapped electrons cannot respond in time. However, under DC condition, the gate bias is stepped gradually up from pinch off, giving enough time for the trapped electrons to respond and emit from the traps. Thus the maximum channel current can be obtained. Dispersion reduces the AC output power from what is predicted by DC-IV characteristics. The amount of dispersion is directly related to the trap density. The current drop can vary dramatically from 5-10 % to 40-50 %. The bulk traps can be reduced by improved growth and the surface trap density can be lowered by improved process technique. For this study, the dispersion of the devices is well controlled, mostly less than 20 %.

The simple equation to predict the maximum linear power level can be achieved for the device from the DC-IV curve as following (assuming class A operation):

\[ P_{\text{out}} = \frac{1}{8} I_{\text{max}} \times (V_{br} - V_{nee}) \]

With the numbers mentioned above, expected \( P_{\text{out}} \) should be about 5.6 W/mm. This can be further increased if the device is driven well into saturation, but the above calculation gives a fairly close indication of the intrinsic limit of the available output power level from the device. Because of the existence of the dispersion, real power level is less than the predicted value. Fig. 2.3-3 (a) shows the power measurement of a 100-μm wide device. ATN automatic load-pull system (with fundamental tuning only) was utilized to present the device with the optimum load state for maximum output power level. Scaled to one mm, a power density of 4.6 W/mm, PAE of 40 % with associated large signal gain of about 7 dB at 8 GHz were achieved. The device was biased at a drain voltage of 24 V and drain quiescent current of 200 mA/mm, which is class AB operation condition.
Fig. 2.3-3 (b) shows the power measurement when tuned for maximum PAE, the device was biased at 18 V, and $I_d = 200$ mA/mm, PAE of 52% with output power density of 3.6 W/mm at 8 GHz was achieved.

We also processed AlGaN/GaN HEMTs grown on SiC substrate. Even these samples were grown by MOCVD with un-optimized conditions, promising power performance was still demonstrated. When biased at $V_d = 30$ V, $I_d = 200$ mA/mm, the device achieved a power density of 6 W/mm and PAE of 30% with associated gain of 9 dB at 8 GHz. (Fig. 2.3-4). The improved power density is believed to be largely due to the better thermal property of the SiC over sapphire substrate, which could be beneficial to the quality of the grown epitaxy layer. Higher breakdown voltage comes from the less doping of AlGaN layer.

From these power measurements, it was noticed that the gain compression was relatively high when the highest power level was achieved, ranging from 5-7 dB. We believe traps within the material system are the main reason for high gain compression and it should be addressed as the material quality and process techniques are further improved.
Fig. 2.3–3 Power measurement for (a) maximum output power, $P_{\text{out}} = 4.6$ W/mm, PAE = 40 % ($V_d = 24$ V, $I_d = 200$ mA/mm, $f = 8$ GHz) (b) maximum PAE. $P_{\text{out}} = 3.6$ W/mm, PAE = 52 % ($V_d = 18$ V, $I_d = 200$ mA/mm, $f = 8$ GHz)

Fig. 2.3-4 Power measurement of AlGaN/GaN HEMT grown on SiC substrate ($W_g = 100$ μm, $L_g = 0.7$ μm, $f = 6$ GHZ, $V_d = 30$ V, class AB mode)
2.4 Large gate periphery device design

Fig 2.4-1 Layout of two one-mm multi-finger AlGaN/GaN HEMTs

After the evaluation of the unit device (100 μm wide), it's not simply an agglomeration of unit devices for a large gate periphery device. Major issues for the design of large devices are determination of the pattern layout, unit gate finger width, gate-to-gate spacing (gate pitch) and number of fingers per gate feed pad. Fig. 2.4-1 shows the layout of two one-mm wide devices. A typical inter-digital field effect transistor (FET) structure is employed. In the early design, air-bridges were employed to connect the source pads only at the extended portion of the pad. With this connection, probing was feasible to test the electrical characteristics such as DC-IV curve of the whole device. Since for large devices, heat dissipation became a serious issue, full scale testing with normal bias condition was impossible. As a result, we can only test some limited characteristics of the device such as pinch-off voltage, gate-leaking current, mainly for the defect-screening purpose. As the process technique was improved, this kind of source pad airbridges became unnecessary and can be eliminated. Instead, source pads connection was made on the AlN circuit board that also serves as a good heat sink. Thick Au bonding pads were deposited on both the source and drain pads as
well as on the corresponding source and drain electrodes on the AlN board. Then the device was flipped and bonded onto the AlN board using a flip-chip bonder.

1) Determination of unit gate finger width $W_{gu}$:

![Diagram](image)

*Fig 2.4-2 Transmission line model of a unit gate finger*

Gates are usually of 1μm or less in length; signals traveling along the gate line would have considerable conductive loss at microwave frequencies, also the signal leaking to the underneath active device further increase the loss. In order to have more uniform signal drive on the gate line, shorter unit gate width is preferred. But for a fixed total gate-periphery, more gate fingers are needed for shorter unit gate width. This will lead to wider lateral device size, consequently, more phase delay and signal loss are introduced for signals to reach each gate finger. Therefore, we need to compute the signal loss in the gate fingers and decide the allowable maximum gate width given a certain tolerable loss at the desired frequency range.

We can treat the gate finger as a distributed circuit system, with the model of unit section of the gate shown in Fig. 2.4-2. For gate length $L_g = 0.7\,\mu$m, gate resistance $r_g = 65\,\Omega/mm$, gate inductance $l_g = 36\,pH/mm$, gate-source resistance $r_{gs} = 1.2\,\Omega/mm$ and capacitance $C_{gs} = 2.5\,pF/mm$. Then the propagation constant of the artificial gate transmission line can be computed using the formula
\[ \gamma_{gate} = \alpha + j\beta = \sqrt{Z \times Y} = \frac{r_g + jwI_g}{r_{gs} + 1/jwc_{gs}} \]  

(2.4-1)

However, the gate resistance above is typically estimated from DC measurements, at high frequencies there is an increase in loss due to the skin effect. This can be approximated as follows (t is the thickness of the gate metal, \( \rho \) is the resistivity of the metal, \( \delta \) is the skin depth)

\[ r_g = r_{g-DC} \frac{t / \delta}{1 - \exp(-t / \delta)} \quad \text{where} \quad \delta = \sqrt{\rho / \pi f \mu_0} \]  

(2.4-2)

On the drain side, we can also treat the drain pad as distributed line with propagation constant

\[ \gamma_{drain} = \sqrt{\frac{r_d + jwI_d}{r_{ds} + 1/jwc_{ds}}} \quad \text{where} \quad r_d = 2.5 \Omega, r_{ds} = 150 \Omega, C_{ds} = 0.015 \mu F, l_d = 15 \mu H \]

Letting the signal voltage at the gate feed point be \( V_0 \), the signal voltage input along the gate finger can be expressed as a function of the gate position \( x \) as (assuming the gate is open-circuited at the other end):

\[ V(x) = \frac{V_0}{2} (e^{-\gamma_{gaw}x} + e^{-\gamma_{gaw}(2W_{gs} - x)}) \]  

(2.4-3)

Let us assume the unit device cell with unit gate finger width is comprised of the summation of unit voltage-controlled current source with the trans-conductance \( g_m = 300 \) mS/mm (similar to the concept of the traveling wave amplifier), then the output amplified current distribution on the drain line is

\[ I(x) = g_m \times \frac{V_0}{2} (e^{-\gamma_{gaw}x} + e^{-\gamma_{gaw}(2W_{gs} - x)}) \]  

(2.4-4)

If we neglect the loss on the drain line (since drain line typically is wide and thick) then the summation of the total drain current would be (the second term in the integral accounts for the distributed effect of the drain line):
\[ I_{out} = \int_{0}^{W_{g}} g_m \times \frac{V_0}{2} \times (e^{-\gamma x} + e^{-(2W_{g} - x)} \times \frac{1}{2} (e^{-\gamma_{d_{dr}} (W_{g} - x)} + e^{-\gamma_{d_{dr}} (W_{g} + x)})) \, dx \] (2.4-5)

and the ideal total drain current is \( I_{\text{ideal}} = g_m V_0 W_g \). The resulted reduction in the output power would be

\[ \frac{P_{\text{out}}}{P_{\text{ideal}}} = \left| \frac{I_{out}}{I_{\text{ideal}}} \right|^2 = \frac{1}{W_g} 10 \times \log \left( \int_{0}^{W_{g}} g_m \times \frac{V_0}{2} \times (e^{-\gamma_{d_{dr}} (W_{g} - x)} + e^{-\gamma_{d_{dr}} (2W_{g} - x)}) \times \frac{1}{2} (e^{-\gamma_{d_{dr}} (W_{g} - x)} + e^{-\gamma_{d_{dr}} (W_{g} + x)}) \, dx \right)^2 \] (2.4-6)

Fig. 2.4-3 shows the effects of unit gate finger width \( W_{gu} \) on the output power level. From the graph, we can tell that for the frequency range up to 10 GHz, \( W_{gu} \) should be less than 100 \( \mu \)m in order to keep the power gain loss below 2 dB. Dashed lines plotted the case when the distributed effect of the drain line is neglected.

Fig. 2.4-3 The power loss (compared with the ideal output power) vs. the unit gate finger width (solid line: with the drain line distributed effect, dashed line: without the drain line distributed effect.)

22
2) Determination of the gate-gate spacing $L_{gg}$ and number of fingers per cell n:
In general, the wider of the lateral span of a power FET, the lower is the gain. The reasons include: 1) Phase rotation from the gate feed pad to each gate finger. 2) Non-uniform operation from cell to cell due to the various in materials and process. 3) Non-uniform channel temperature. 4) Additional paths needed to connect between cells that will introduce parasitics as well as losses.

![Diagram](image)

Fig. 2.4-4 Simplified graph for a unit cell device (only gate electrode is shown)

The empirical rule for the maximum lateral device width is: When it's below $\lambda/16$, parallel operation is maintained throughout the active region and the phase rotation is not observable. When the width is between $\lambda/8$ and $\lambda/16$, there is some reduction in gain, but the device still operates efficiently. When the width is above $\lambda/8$, the device can no longer be considered as a lumped element, and external circuitry is required to control the phase rotation in power input and output. For AlGaN/GaN HEMTs, $\lambda/16$ is about 860 μm at 10 GHZ, which we chose as the maximum lateral width of the device.

As shown in Fig. 2.4-4, for a fixed total gate periphery per cell ($W_{gc}$) and unit gate finger width $W_{gu}$ ($n = W_{gc}/W_{gu}$), the number of fingers n is fixed, the lateral
device size, which is \((n-1) \times L_{gg}\), should be kept as narrow as possible, for the reasons just mentioned. As a result, small \(L_{gg}\) is preferred. Also small \(L_{gg}\) has the benefit of reducing the source and drain pad parasitic capacitance, since \(L_{gg}\) is very close to the source and drain pad size. But on the other hand, a large \(L_{gg}\) would reduce the thermal resistance of the device since the heat generated could be dissipated through wider area. In this study, we chose pitch size \(L_{gg}\) based on the following procedures: 1) Design 1mm devices with \(L_{gg}\) ranging from 20 to 50 \(\mu m\). 2) Bond devices onto test features on the AlN circuit board. 3) Measure the DC-IV curve of the devices. 4) Choose the smallest \(L_{gg}\) that can achieve \(I_{\text{max}}\) larger than 800mA (current density measured from a 100-\(\mu m\) device is 1000 mA/mm). With this approach, we decided to choose \(L_{gg}\) 45 \(\mu m\). Thermal simulation with this device layout pattern will be presented in the next section.

2.5 Flip-chip mounting technique for thermal management

![Diagram of flip-chip bonding process](image)

Fig. 2.5-1 The procedures of flip-chip bonding

In section 2.3, we reported the superior power performance of AlGaN/GaN HEMTs. However, the gate peripheries corresponding to these results are in the order of 100 \(\mu m\), the total amount of produced power is too small to be
significantly affected by the poor thermal properties of the sapphire substrate. At room temperature, the conductivity of the sapphire is very poor ($\kappa = 30 \text{ W/mK}$), compared with other materials, AlN: $\kappa = 180 \text{ W/mK}$, SiC: $\kappa = 330 \text{ W/mK}$, GaN: $\kappa \approx 200 \text{ W/mK}$. For GaAs (also a poor thermal conductor) based large-periphery power devices, heat removal is typically done through backside heat sink and wafer thinning. However, for GaN devices, both techniques are still premature due to the lack of proper etching methods. We circumvent the problem by developing the flip-chip mounting technology with AlN substrate as the heat sink as well as the circuit carrier. As a result, we are able to fabricate GaN HEMTs on sapphire at much lower cost, having competitive performance as GaN-grown-on-SiC without a compromise in thermal management. Also, this technique is scalable to larger devices (1mm, 2mm and 4mm have been tested) with proper design of the multi-finger layout pattern.

Fig. 2.5-1 describes the procedure of the flip-chip bonding. First, we fabricated about 5-µm gold bonding bumps on both the GaN chip and the AlN substrate; a flip-chip bonder was then used to align the GaN device to the AlN circuit board with coplanar waveguide (CPW) lines structure pre-fabricated to avoid backside grounding. Proper amount of pressure and heat were then applied to bond them together. Other advantages of the flip-chip mounting include: (1) The reduced parasitic inductance compared to the conventional wire bonding connection; this advantage will be further illustrated in the Chapter 5. The bonding wires shown in Fig. 2.5-2 could seriously degrade the circuit performance such as bandwidth and gain. (2) More cost-effective use of the GaN material with all the passive components built on in-expensive AlN circuit board; also AlN could provide a better substrate for the passive circuit components in terms of thermal property.
Fig. 2.5-2 Bonding wire connections in the hybrid circuit.

Thermal simulation was performed using commercial software provided by F-chart Company, to theoretically analyze the operating temperature. Two-dimensional finite element method is utilized. The partial difference equation governing heat flow is

\[
\frac{\partial}{\partial x} (\kappa \frac{\partial T}{\partial x}) + \frac{\partial}{\partial y} (\kappa \frac{\partial T}{\partial y}) + q - \rho c \frac{\partial T}{\partial t} = 0
\]

where \( \kappa \) is the thermal conductivity of the material, \( c \) is the material specific heat, \( q \) is the internal generation rate per unit volume and \( \rho \) is the material density. In the simulation, steady-state CW operation was assumed which led to \( \partial T/\partial t = 0 \). Fig. 2.5-3 (a) shows the simulated temperature of the AlGaN/GaN FET without flip-chip bonding, which could reach over 200 °C assuming generated heat density was 4W/mm. Fig. 2.5-3 (b) shows that with flip-chip bonding the temperature could be reduced to about 60 °C. Fig. 2.5-4 shows the dramatic improvement of the DC-IV characteristics of a 50-μm device after flip-chip bonding compared with the FET without flip-chip bonding.
Fig. 2.5-3 Thermal simulation of the AlGaN/GaN HEMT (a) without the flip-chip bonding and (b) with the flip-chip bonding.

Fig. 2.5-4: DC IV curve of a 50-μm wide device (a) before flip-chip bonding (b) after flip-chip bonding.
We have tested large-periphery GaN HEMTs of size 1mm, 2mm and 4mm respectively; 4mm device cannot be measured due to the equipment limitation. Fig. 2.5-5 shows the picture of a 1mm-wide device before and after the flip-chip bonding. Inter-digital pattern layout was chosen with gate and drain pads all connected respectively while source pads not connected. Since it is not trivial to flip-chip bond all the gold bumps firmly at the same time, careful design of device layout was needed such as reasonable bonding chip size, pad size and shape, pad alignment tolerance, etc, to facilitate the bonding process. The enlarged source and drain pads were used to reduce the thermal resistance of the device; the connection of the source pads was made on the AlN substrate to avoid the use of air-bridges. The trade-off is the increased pad parasitic capacitances, which must be taken into account during the circuit design.

Fig 2.5-5: Picture of a one-mm device (a) before flip-chip bonding, (b) after flip-chip bonding.

Fig. 2.5-6 is the measured power performance of the 1mm-wide device using ATN LP1 load-pull system, with only the fundamental tuning, total output power of 4.4 W and 35 % PAE at 8 GHz has been achieved. Fig. 2.5-7 is the measured
power performance of a 2mm-wide device, total output power of 6.3 W and 30 % PAE at 4 GHz has been achieved. However, typically gain compression was high, ranging from 4-7 dB, when the device reached its highest output power level. Also, the device will degrade to some extent after the RF test, which can be observed from the DC IV curve after the power measurement (Fig. 2.5-8), the $I_{\text{max}}$ was decreased from 800 mA to 700 mA, this was related to the immature material technology. This limited the large-signal gain and efficiency of the amplifiers. Nevertheless, in terms of realization of high power potential of GaN, our "baseline" GaN-grown-on-Sapphire power device process has proved to be more than adequate for the subsequent design and fabrication of the GaN power amplifiers.

Fig. 2.5-6: Power measurement of a one-mm device after flip-chip bonding. ($P_{\text{out}} = 4.4$ W, PAE = 35%, $f = 8$GHz, $V_{\text{ds}} = 25$ V, $I_{\text{ds}} = 200$mA, class AB mode)
Fig. 2.5-7: Power measurement of a two-mm device after flip-chip bonding. (Pout = 6.3 W, PAE = 32%, f = 4GHz, Vds = 25 V, Ids = 350 mA, class AB mode)

Fig. 2.5-8: DC IV measurement of a one-mm device after the power measurement
2.6 Device small-signal and large-signal modeling

1) Small-signal modeling:

From the circuit design point of view, the accurate models for the device are crucial for the success of the design. The small-signal model is based on the physics of the device that has general implication, while S-parameters are obtained from a specific device that could vary with the material as well as process. As a result, the small-signal model is preferred in the simulation over the measured S-parameters, while S-parameters obtained from multiple devices are utilized to extract and generalize the circuit components for the small-signal model. The methods are described in\textsuperscript{6}. The model shown in Fig. 2.6-1 includes 17 elements. The shaded area is the intrinsic FET; $R_g$, $R_s$ and $R_d$ are the resistive access parasitics of the terminals (including ohmic contact, terminal metal conductive and channel resistance); $L_g$, $L_s$ and $L_d$ are the series inductance of the terminals,
which are normally negligible for the frequency range of the concern (in the order of tens of pH); $C_{gs}$, $C_{gd}$ and $C_{ds}$ are parasitic capacitances introduced by the noticeable pad size of the terminals. (They have significantly increased values for the particular FET layout pattern in this study). The typical values of those components (device with $L_g = 0.7\mu m$, $W_g = 1 mm$) are following: $g_m = 320mS$, $C_{gs} = 2.5pF$, $C_{ds} = 0.015pF$, $C_{gd} = 0.01pF$, $R_{in} = 1\Omega$, $R_{ds} = 150\Omega$, $R_{gd} = 10\Omega$, $R_g = 0.75\Omega$, $R_d = 2.5\Omega$, $R_s = 2\Omega$, $C_{spad} = 0.15pF$, $C_{dpad} = 0.15pF$, $C_{dspa} = 0.25pF$ and $\tau = 7.8psec$. (All inductances are neglected)

Fig. 2.6-2: The simplified circuit model of (a) a FET, (b) the definitions of short circuit current gain (defines $f_0$) and (c) maximum available power gain (defines $f_{max}$).

32
2) $f_i$ and $f_{\text{max}}$:

$F_i$ and $f_{\text{max}}$ are two very useful figures of merit that can indicate the maximum achievable performance of the circuit using the existing devices, such as gain-bandwidth product, operating frequency limit and etc. They can be derived from the simplified circuit model shown in Fig.2.6-2 (a). $F_i$ is the frequency at which the short-circuit current gain is 1 ($10\times\log(|h_{21}|^2)=0$), it can be estimated using equation (usually $C_{gd}$ is negligible compared with $C_{gs}$, as in Fig. 2.6-2(a))

$$f_i \approx \frac{g_m}{2\pi(C_{gs} + C_{gd})}$$

We can also extract $f_i$ through the S-parameters measurement since we can convert S-parameters into H-parameters and then plot $h_{21}$ (in dB) vs. frequency using equation:

$$h_{21}(dB) = 20\log\left(\frac{-2S_{21}}{(1-S_{11})(1+S_{22}) + S_{12}S_{21}}\right)$$

From the extrapolation of the $h_{21}$ curve with the frequency axis, we can obtain $f_i$.

Normally, $f_i$ is a good indication of the maximum achievable gain-bandwidth for resistively terminated circuit, which represents most analog high-frequency amplifiers.

$F_{\text{max}}$ is the frequency at which Mason's unilateral power gain (UPG) is 1 ($10\times\log(|\text{UPG}|)=0$), it's obtained with the condition: 1) Both input and output of the device are conjugated matched, 2) The feedback of the device ($C_{gd}$ and $R_{gd}$) is cancelled out through careful matching. This frequency is the maximum possible frequency to achieve power amplification using the device. The simple way to calculate $f_{\text{max}}$ (for FET) is:

$$f_{\text{max}} \equiv \frac{f_i}{2\sqrt{(R_g + R_{gs} + R_s)/R_{ds} + 2\pi f_i R_g C_{gd}}}$$
and normally the term $2\pi f_t R_g C_{gs}$ can be neglected as the equation for $f_t$. We also can extract $f_{max}$ from S-parameters using the following express for UPG ($f_{max}$ is the frequency at which UPG (in dB) = 0):

$$UPG(dB) = 10\log\left[\left(\frac{1}{1-|S_{11}|^2}\right)|S_{21}|^2 \left(\frac{1}{1-|S_{22}|^2}\right)\right]$$

An impedance-matched amplifier could achieve the maximum available power gain over certain bandwidth through conjugate matching both input and output of the device; the achievable bandwidth is determined by the bandwidth of the matching networks. For the distributed amplifier in which device capacitances $C_{gs}$ are absorbed into the artificial gate line, the gain-bandwidth product could reach $f_{max}$ if proper matching conditions are satisfied. Fig. 2.6-3 shows the theoretical limit of the performance for the three kinds of circuits: resistively terminated, conjugate matched and distributed amplifiers.

From the above discussion, it can be seen that, $f_t$ is more related to the intrinsic (ideal) device parameters ($g_m$, $C_{gs}$) and $f_{max}$ is determined by both the intrinsic device ($f_t$) and extrinsic components (parasitic resistance $R_g$, $R_{gs}$, $R_s$). We can maximize $f_{max}$ by carefully fabricate the devices, minimizing those parasitic resistances. Fig. 2.6-4 shows the extracted $f_t$ and $f_{max}$ for 1mm-wide and 0.7μm-long AlGaN/GaN HEMTs used in this work, $f_t$ and $f_{max}$ are about 20 GHz and 38 GHz respectively. In the graph, we can see that UPG became MAG when it intercepted maximum stable gain (MSG = 10 log ($|S_{21}/S_{12}|$)) curve, MAG (in the stable region) and MSG (in the unstable region) defined a meaningful power gain boundary for the device.
Fig. 2.6-3: Gain-frequency constraints of (a) resistively terminated (b) conjugate matched (c) distributed amplifiers

Fig. 2.6-4: Extracted $f_t$ and $f_{\text{max}}$ from S-parameter measurement

2) Large-signal model:
For power amplifier design, it is desirable to have accurate large-signal model, we can then predict the maximum output power level and achievable power added efficiency (PAE) and other nonlinear phenomena over the whole operating
frequency range, using harmonic balance test bench of EEsof/HP Communications Design Suite. However, since the AlGaN/GaN material system is still immature, such as the phenomena of “dispersion” associated with traps in the system, the non-uniformity of the wafer, the degradation of the device after the DC and RF stress, etc, these complicate the process of developing accurate large-signal model for GaN HEMTs. Furthermore, the thermal property has to be modeled in order to realistically reflect the operating condition; itself is already a very challenge task. With those in mind, a crude Curtice-cubic model was developed which incorporates the small-signal model circuit parameters and the DC-IV characteristics. This model can give us a first-order estimate of the circuit power performance. The I-V characteristics are described by:

\[ I_d = (A_0 + A_1 V_1 + A_2 V_1^2 + A_3 V_1^3) \tanh \gamma V_{ds} \]

where

\[ V_1 = V_{gs} \times \left[ 1 + \beta (V_{OUTO} - |V_{ds}|) \right] \]  

(2.6-1)

Fig. 2.6-3 shows the large-signal model of a one-mm wide 0.7-μm long AlGaN/GaN HEMT. There are some approximations we made to extract and simplify the model: 1) Gate pinch-off doesn’t change with \( V_{ds} \), which leads to \( V_1 = V_{gs} \) in equation (2.6-1). 2) Trans-conductance \( g_m \) doesn’t vary with \( V_{gs} \), which leads to \( A_2 = 0 \) and \( A_3 = 0 \). 3) Intrinsic \( V_{gs} \) and \( V_{ds} \) are the gate and drain terminal applied voltages subtracting voltage drops on the extrinsic parasitic resistances \( R_s \) and \( R_d \), that is \( V_{gs} = V_{gs-terminal} - I_{ds} \times R_s \), and \( V_{ds} = V_{ds-terminal} - I_{ds} \times (R_s+R_d) \). 4) At saturation (\( V_{ds} > V_{knee} \)), \( \tanh(\gamma V_{ds}) = 1 \). 5) At the gate terminal pinch-off voltage \( V_{pinch-off} \), \( I_{ds} = 0 \). 6) At the gate terminal bias of \( V_{gs\text{-max}} \), \( I_{ds} = I_{max} \). 7) At the drain terminal bias of \( V_{knee} \), \( I_{ds} = x_f \times I_{max}(x_f = 0.99) \). Then we can have the following equations to calculate the parameters \( A_0 \), \( A_1 \), \( A_2 \), \( A_3 \) and \( \gamma \):
\[
A_0 = \frac{I_{\text{max}}}{1 + (V_{g_{\text{r-max}}} - I_{\text{max}} R_s) |V_{\text{pinch-off}}|}, \\
A_1 = \frac{A_0}{|V_{\text{pinch-off}}|}, \\
A_2 = 0, A_3 = 0, \\
\gamma = \tanh^{-1}\left(\frac{x_f I_{\text{max}}}{A_0 + A_1 (V_{g_{\text{r-max}}} - x_f I_{\text{max}} R_s)} \times \frac{1}{V_{\text{knee}} - x_f I_{\text{max}} (R_s + R_d)}\right)
\] (2.6-2)

Fig. 2.6-3: Curtice_3 large-signal model for AlGaN/GaN HEMTs

Fig. 2.6-4 is the DC I-V curve generated using the above model. Here, \(I_{\text{max}}\) has been modified to the realistic value of 800mA for a one-mm wide device instead of the 1000mA value extracted from the small device measurement. This was achieved by introducing an empirical relation of \(I_{\text{max}}\) with the scaling of the devices in the model. With this model, the limitation of the large-signal operation of the device was defined by the boundary of the DC I-V curve. This is a first-
order simulation because the approximations such as linear $g_m$, zero output conductance are simply not true. Also, without the modeling of the thermal affect and dispersion phenomena, this model can only be used as a guide for the large-signal design of the amplifier. As a result, intensive efforts were put into the load-pull measurement of the FET and the large-signal design relied heavily on the load-pull data.

![Graph](image)

Fig. 2.6-4: Modeled DC I-V curve using the Curtice-3 model of a one-mm AlGaN/GaN HEMT.

### 2.7 Load-pull system:

In this section, we will discuss how to achieve the maximum output power from the device. Cripps method is described briefly, then the load-pull method will be introduced and a detailed description of an ATN load-pull system used in our lab, its set-up, calibration and measurement procedure, would be followed.
1) Cripps method:

![Diagram showing output load line effects on the output power.](image)

Fig. 2.7-1: Output load line effects on the output power.

Fig. 2.7-1 illustrates the simple yet effective optimum load line (Cripps method) theory to achieve maximum output power. Assuming the DC-IV curve of a device is bounded by its maximum drain current $I_{\text{max}}$, knee voltage $V_{\text{knee}}$ and drain breakdown voltage $V_{\text{br}}$, and the device is biased in class A mode ($V_D = (V_{\text{br}} - V_{\text{knee}})/2$, $I_D = I_{\text{max}}/2$), then the optimum load for the maximum output power would be:

$$R_{\text{opt}} = \frac{V_{\text{br}} - V_{\text{knee}}}{I_{\text{max}}}$$

Since in that situation, the device could achieve the maximum current and voltage swing at the same time, and the power delivered to the load would be

$$P_{RF,\text{max}} = \frac{1}{8} I_{\text{max}} \times (V_{\text{br}} - V_{\text{knee}}) = \frac{1}{8} \left(\frac{V_{\text{br}} - V_{\text{knee}}}{R_{\text{opt}}}\right)^2 = \frac{1}{8} I_{\text{max}}^2 \times R_{\text{opt}}$$

39
If $R_L > R_{\text{opt}}$, then the voltage boundary would be the limiting factor and the power delivered to the load is given by

$$P_{RF} = \frac{1}{2} \frac{(V_{br} - V_{knee})^2}{R_L} = P_{RF,\text{max}} \times \frac{R_{\text{opt}}}{R_L} = P_{RF,\text{max}}$$

Similarly, for $R_L < R_{\text{opt}}$, with which the device is limited by the current boundary, the power delivered to the load is

$$P_{RF} = \frac{1}{8} I_{\text{max}}^2 R_L = P_{RF,\text{max}} \times \frac{R_L}{R_{\text{opt}}} < P_{RF,\text{max}}$$

Using this method for the GaN HEMT used in the work ($I_{\text{max}} = \sim 1A/mm$, $V_{br} = 50 V$, $V_{knee} = 5 V$), the predicted $R_{\text{opt}}$ should be about 45-50 $\Omega$.

2) Load pull method

Load-pull is an empirical approach to design power amplifier. In its basic form, an output tuner or its active equivalent is adjusted to maximize the output power; the tuner will record the load impedance that produces the maximum power. Variation of the output load can be passively carried out by means of a tuner (either mechanical or electronic) or actively carried out by injecting a power wave toward the transistor output and, thus forcing the desired wave ratio. The load variation can take place either at the fundamental frequency $f_0$ alone or at the $f_0$ and a number of its harmonics. The importance of the harmonic tuning (especially the second harmonic 2 $f_0$) could be well observed, especially for PAE. But for the broadband power amplifier, the concept of tune harmonics is unrealistic since the bandwidth desired here is at least an octave wide.
Fig. 2.7-2: ATN automatic load-pull system (a) front panel connection, (b) real panel connection.

With our system, a commercial automatic single-tone load-pull assembly (only have fundamental tuning) manufactured by ATN Microwave Company, two electronic tuners with bandwidth from 1.8 – 9 GHz are utilized. The purpose of the source tuner is mainly for the input match thus improving the power gain as well as the PAE; the effect of the source impedance on the optimum load is negligible. The electronic tuners are of limited use because of inherent losses, which impose limitations in reflection-coefficient magnitude ($|\Gamma_L|$). The maximum magnitude decreases with frequency, the number of elements and length of cables connected between the measurement plane and tuner. This is especially disadvantageous for on-wafer measurements and measurements of highly mismatched devices. As a result, carefully setup of the load-pull system is
very important, especially the connection of the output tuner, for the accuracy of the calibration as well as the measurement. The tuners should be placed as close as possible to the probe tips using hard line as short as possible and with minimal bends. Connector care plays a vital role in measurement accuracy. Cables connecting between equipments should use low loss cables (rigid cables are preferred).

Fig. 2.7-2 shows our load-pull system setup, it consists of the following components: ATN LP control modules (including three parts: signal control module (SCM), amplifier bypass module (ABM), electronic mainframe (EM)), two electronic tuners (electronic load module (ELM)), an 8722 network analyzer (NWA), a 4-8 GHz 50 Watt tube power amplifier (PA), a 438 HP power meter with power sensor of model 8425, a two-channel programmable power supply and a probe station with ACP-40 CPW probes. The EM is used as the load-pull system controller; it has control cables connected to SCM, ABM and two ELMs. The SCM module routes the signal from the network analyzer through attenuating, filtering or switching in the power amplifier paths, and finally provides the device with the desired input signal. The ABM is the module to connect the power amplifier. The power meter with the power sensor is used during the calibration process; it is connected to the output port of the output ELM module and will be removed during the measurement. The whole system is connected together through HP-IB cables; the EM module, NWA, the power meter and power supply all have HP-IB port connected to a window-operated PC HP-IB interface.

With the above system, we are able to perform power measurement as was mentioned in section 2.3 with source and load impedance information specified. The 1-mm 4.4 Watt results (at 8 GHz) were obtained with the source and load
impedance as following: $|\Gamma_S| = 0.71$, $\angle \Gamma_S = 150^\circ$, $|\Gamma_L| = 0.45$, $\angle \Gamma_L = 158^\circ$. Since the test feature includes long CPW lines for probing, after the de-embedding of the phase rotation, the optimum load impedance $R_{opt} = 35\,\Omega$, $C_{ds} = 0.35\,pF$ was obtained.
REFERENCE


Chapter 3
Broadband Matching Design

3.1 Conjugate match and load-line (power) match

In the previous chapter, we discussed the optimum load for achieving the maximum output power. Here we will compare the conjugate match and load-line (power) match,\(^1\) which is one of the principle differences between the small-signal and large-signal designs.

![Diagram of conjugate match and load-line match](image)

Fig. 3.1-1: Comparison of the conjugate match and load-line match.

We all understand that with the conjugate match, the maximum power gain can be achieved. As is shown in Fig. 3.1-1, when the load impedance (admittance) is conjugate matched to the generator impedance (admittance), the maximum power is delivered into the load from the generator. But this is true only with the assumption that the generator is an ideal generator, which can have infinite output voltage (current); or the operating condition is small signal, that the signal swing
is small enough to be well within the current and voltage limits of the generator. Since the transistors are not ideal current generators, which have maximum voltage limit (breakdown voltage) and maximum current limit ($I_{\text{max}}$), as are shown in Fig. 3.1-1, the load-line match is chosen for maximum output power rather than the conjugate match. Although under the small-signal drive, the output power is higher with the conjugate match than the load-line match, but the maximum achievable output power is significantly higher with the load-line match. Fig. 3.1-2 illustrated the improvement of the power performance of an amplifier with the load-line match compared with the conjugately matched design.

![Graph showing power performance with different matches](image)

**Fig. 3.1-2** Power performance of an amplifier with load-line match and conjugate match. (circle: conjugate match, square: load-line match)
3.2 Bandwidth limitation

Fig. 3.2-1: Network topologies used in the calculation of the gain-bandwidth limitations.

Now the broadband power amplifier design can be simplified to the broadband conjugate match at the input and broadband load-line match at the output of the device. Normally a matching network will produce the desired match at one frequency and matching degradation will occur at the other frequencies. Fano\(^2\) has derived a complete set of integrals that predict the gain-bandwidth restrictions for the lossless matching networks terminated in an arbitrary load impedance. For
the network shown in Fig. 3.2-1 (a), the best reflection coefficient \( \Gamma \) that can be achieved over a frequency range is restricted by the integral

\[
\int_0^\infty \ln \left| \frac{1}{\Gamma} \right| d\omega \leq \frac{\pi}{RC}
\]  

(3.2-1)

This equation shows that the product of the bandwidth \( \Delta \omega \) and \( \ln|1/\Gamma| \) cannot be greater than \( \pi/RC \). Therefore, if matching is required over a certain bandwidth, it can be obtained at the expense of the reduced matching (less power transfer) or higher \( |\Gamma| \).

![Graph](image)

**Fig. 3.2-2: Optimum values of \( |\Gamma| \).**

The ideal matching situation is obtained when \( |\Gamma| \) is constant in the desired bandwidth (from \( \omega_1 \) to \( \omega_2 \)) and equals to 1 outside that range, as illustrated in Fig. 3.2-2. It can be derived from equation (3.2-1) that

\[
|\Gamma| = e^{-\pi / (\omega_2 - \omega_1)RC} = e^{-\pi Q_2 / Q_1}
\]

where \( Q_1 = R/X \) for parallel network or \( X/R \) for series network as shown in Fig. 3.2-1, and \( Q_2 = \omega_b/(\omega_2 - \omega_1) \) (reciprocal of fractional bandwidth). Ideally this is the best \( |\Gamma| \) (lowest) that can be obtained in the band. For given bandwidth (\( Q_2 \)), lower \( Q_1 \) can produce lower \( |\Gamma| \). Therefore, load impedance with lower \( Q_1 \) can
achieve better broadband matching. For the matching of the transistor, Fig. 3.2-1 (a) represents the simplified output circuit model of the device and Fig. 3.2-1 (b) represents the simplified input circuit model of the device. As a result, smaller RC product for the output and larger RC product for the input of the device are preferred to achieve broadband matching. However, normally lower Qs at the input and output of the device result in lower maximum available gain. Therefore there is always a tradeoff between the broadband matching and available gain for the device.

3.3 Bandwidth extend methods

Most broadband design techniques include two categories: 1) interstage networks (including the input and output network), 2) feedback networks (or the combination of both methods). From a network viewpoint, broadband design involves 1) removing or canceling the poles nearest the jω axis, 2) changing the positions of the remaining poles to realize the desired transfer characteristic (filter characteristics such as Butterworth or Chebyshev, etc). Both 1 and 2 can be achieved by the employment of interstage networks or using feedback. Interstage networks can be either lossless (consist only L, C elements) or lossy (include R elements with the L and C elements). Some commonly used techniques for bandwidth extending are described briefly below:\(^3\).

1) Resistive Broadbanding:

Fig. 3.3-1 shows an example of a resistive interstage network. It is the simplest type of broadbanding, but it's also the most inefficient. It has lower gain, lower output power and degraded noise figure. As a result, it's seldom used in the simple form.
Fig. 3.3-1: Resistive interstage network (shunt resistance R).

Fig. 3.3-2: Device with shunt peaking inductor in series with the load $R_L$. 
2) Inductor Shunt Peaking:
At the output of the device, the parallel RC network (R is the load in parallel with
the $R_{ds}$, C is normally $C_{ds}$ if the load has negligible capacitance) will set the upper
3-dB bandwidth. Inductor shunt peaking, as shown in Fig. 3.3-2, can extend the
upper frequency of operation by designing proper poles and zeros of the response
function.

3) Inductor Series Peaking:
Series peaking is implemented by placing an inductor in series with the load
capacitance; normally the inductor splits the output and input capacitors of two
cascaded stages. Series peaking uses delay to increase the bandwidth by time
separating the charging of capacitors.

4) T-Coil Bridge:
Figure 3.3-3 shows the T-coil peaking technique. With the network of a controlled
coupling transformer with the common terminal connected to the load, the load
impedance can be matched to the desired value over a broadband. By proper choice of L₁ and L₂ and coupling M, the input impedance of the device (Zᵢ) can be matched to R (at the point 1) independent of frequency.

![Feedback Configuration Diagram](image)

(a) (b)

Fig. 3.3-4: Feedback configuration: a) shunt-shunt, b) series-series.

5) Feedback:
Feedback technique has the advantage of increasing bandwidth, stabilizing gain, establishing and controlling input/output impedances, reducing parametric sensitivities and reducing distortion. There are four basic feedback configurations: shunt-shunt, shunt-series, series-shunt and series-series. An example of the shunt-shunt case (shown in Fig. 3.3-4 (a)) is a common-source FET with a shunt feedback resistor R connected from the drain node back to the gate node. The gain, input and output impedances are reduced in the shunt-shunt configuration. An example of a series-series configuration (Fig. 3.3-4 (b)) would be a resistor in series with the source terminal of a common-source FET. In that case, the gain would be reduced, but the input and output impedances are
increased as a result of the feedback. Sometimes more than one feedback can be used to take full advantage of the benefits of the feedback.

6) Lossy Match:

Fig. 3.3-5 Lossy match networks at the input of the device (a) R-L shunt network, (b) R-C series network, (c) R-λ/4 line network, (d) R-L-C shunt network.
The lossy match is superior to pure resistive broadbording. Resistors are used in the matching networks to provide lossy gain compensation. One of the lossy match networks (R-L branch) at the input of the active device is shown in Fig. 3.3-5(a). At frequencies well below the upper edge of the band, the R-L behaves as a pure shunt resistance $R$. However, for frequencies approaching the upper band edge, the effective branch impedance increases because of the inductors' series impedance. There are also series-connected parallel R-C network (Fig. 3.3-5 b), shunt $R-\lambda/4$ line network (Fig. 3.3-5 c) and shunt R-L-C network (Fig. 3.3-5 d) that can achieve similar broadband matching at the input of the device. Lossy matching is normally avoided at the output for achieving best efficiency and output power.

3.4 Broadband LC lowpass matching networks

![Diagram of LC matching networks](image)

Fig. 3.4-1: The lowpass LC matching networks for: (a) $R_L > R_S$, (b) $R_L < R_S$. 

54
Fig. 3.4-2: (a) Transform from parallel network to series network \((X_p = 1/\omega C)\),
(b) The equivalent circuit of Fig. 3.4-1 (a) \((X_s = \omega L)\).

(1) Single-section lowpass LC network:

Single-section lossless lowpass LC matching network is one of the simplest matching networks.\(^4\) The main reason of choosing lowpass LC network over other LC networks is that we need to have a DC supply path. This network can realize perfect impedance transformation at one single frequency and have good match over certain frequency range. Fig. 3.4-1(a) shows the network when the load resistor \(R_L\) is larger than the source resistor \(R_S\) and Fig. 3.4-1 (b) shows the network when the load resistor \(R_L\) is smaller than the source resistor \(R_S\). Here we only illustrate the situation of real source and load impedance transformation, since we can always absorb the reactive parts into the LC network. We can calculate the values of \(L\) and \(C\) by transforming the parallel network into series network as shown in Fig. 3.4-2 (a). We use \(-X_p\) to denote the reactance of the capacitance, and \(-X_p^*\) and \(R_L^*\) as the transformed reactance and resistance, where
\[ X_p^* = \frac{X_p R_i^2}{X_p^2 + R_i^2} \]
\[ R_L^* = \frac{X_p^2 R_L}{X_p^2 + R_L^2} \]

For perfect match, it should be designed to achieve: \( X_S = X_p^* \), \( R_S = R_L^* \) in Fig. 3.4-2 (b), which is the equivalent circuit of the network shown in Fig. 3.4-1(a). Therefore, we can obtain the desired \( L \) and \( C \) values by defining the \( Q \) value of the network.

For \( R_L > R_S \),

\[ Q = \sqrt{\frac{R_L}{R_S}} - 1 \]
\[ X_S = \omega L = R_S \times Q, \ X_p = 1/ \omega C = R_L/Q. \]

For \( R_L < R_S \),

\[ Q = \sqrt{\frac{R_S}{R_L}} - 1 \]
\[ X_p = 1/ \omega C = R_S/Q, \ X_S = \omega L = R_L \times Q. \]

The bandwidth of the LC network is determined by the \( Q \) value of the circuit and is inversely proportional to \( Q \). As we can see from those equations, given the source and load impedance \( R_S \) and \( R_L \), \( Q \) is fixed for the simple single-section lowpass LC network. The larger is the ratio of impedance transformation \( R_L/R_S \) (or \( R_S/R_L \)), the higher is the \( Q \) value of the network. As a result, we don’t have much control of the bandwidth of those simple single-section LC networks for given \( R_S \) and \( R_L \). For broadband matching network design, we need to improve the bandwidth characteristic of those LC networks.
(2) Multi-section (low Q) lowpass LC networks:

Fig. 3.4-3: Multi-section low Q LC matching networks ($R_L > R_S$).

Since the bandwidth of the LC network is inversely proportional to the Q value, we have to design low Q LC networks to achieve the broadband matching. Naturally, we can employ multi-section LC networks, each has lower impedance transformation ratio, as shown in Fig. 3.4-3 (assuming $R_L > R_S$). The intermediate impedances are decided by distributing the impedance transformation ratio equally among $N$ sections, that is $R_1/R_S = R_2/R_1 = R_3/R_2 = \ldots = R_L/R_N = (R_L/R_S)^{1/N}$. As a result, the Q of the network will be reduced from the single-section value of $Q_1 = \sqrt{\frac{R_L}{R_S}} - 1$ to $Q_N = \sqrt{\left(\frac{R_L}{R_S}\right)^{\frac{1}{N}} - 1}$. With the knowledge of the intermediate impedances, the L and C values of each section can be obtained using the same formula as the design of single-section LC network. In this study, we use high impedance line to realize the inductances and metal-insulator-metal (MIM) to realize the capacitances. Since high impedance CPW lines normally have narrow signal lines, they tend to be lossy especially when the L values become relatively large (longer CPW lines).
3.5 Broadband power combiner/divider design

In most situations, power combining of signals will be needed to produce large amount of power for power amplifiers. In this section, Wilkinson power combiner is introduced, and then based on it, N-way multi-section power combiner and corporate power combiner will be discussed for broadband power combining.

1) The Wilkinson power combiner/divider.

A Wilkinson power combiner/divider offers broad bandwidth and equal phase characteristics at each of its output ports.\(^5\) Fig. 3.5-1 shows the schematic diagram of a Wilkinson combiner/divider. Each of the quarter-wave lines has the characteristic impedance of \(\sqrt{2}Z_0\), and the output (port 1 or 2) is terminated by a resistor of \(2Z_0\Omega\) (\(Z_0\) is the system impedance).

![Schematic diagram of a Wilkinson combiner/divider](image)

Fig. 3.5-1: Single-section Wilkinson combiner/divider.
Fig. 3.5-2. (a) The Wilkinson power combiner/divider in symmetric form. (b) Bisection of the circuit (a) for even-mode excitation. (c) Bisection of the circuit (a) for odd-mode excitation.

We can employ even-odd mode theory to analyze the Wilkinson power combiner/divider. Fig. 3.5-2 (a) shows the Wilkinson power combiner in symmetric transmission line form with excitation at port one and two. For the even mode excitation, we assume port one and two have the equal signal drives and there is no current flow through the R/2 resistors or the short circuit between the transmission lines at port three. Thus we can bisect the network of Fig 3.5-2 (a) with open-circuits at these points, to obtain the network of Fig. 3.5-2 (b) (the grounded side of the λ/4 line is not shown). Then looking into port one, we see an impedance of \( Z_{inc} = (\sqrt{2}Z_0)^2 / (2Z_0) = Z_0 \), because of the impedance transformation of the λ/4 line. Thus port one is matched as well as the symmetric port two and all the power will be delivered to the load connected at port three. In this case the resistor R is superfluous. For the odd mode excitation, the drives at
port one and two are opposite and there is a voltage null along the middle of the circuit in Fig. 3.5-2 (a). Thus we can bisect this circuit by grounding it at two points on its mid-plane, to give the network of Fig. 3.5-2 (c). Looking into port 1, we see an impedance of $R/2$, since the parallel-connected transmission line is $\lambda/4$ long and shorted at port 3, and so looks like an open circuit at port 1. Thus, port 1 will be matched since $Z_{in} = R/2 = Z_0$. In this mode of excitation, all power is delivered to the R/2 resistors, with none going to port 3. The typical frequency response of the single-section Wilkinson combiner is shown in Fig. 3.5-3. Using output match less than $-15$ dB (or VSWR ratio less than 1.43) as the criteria to evaluate the bandwidth, more than one octave bandwidth can be achieved.

Fig. 3.5-3: Frequency response of Wilkinson combiner/divider.
Fig. 3.5-4: Comparison of the bandwidth of the quarter-wave transformer with different transformation ratio.

2) N-way multi-section Wilkinson combiner.

We can incorporate the power combining and impedance transformation property by deploying the multi-section Wilkinson combiner/divider (N-way Wilkinson combiner), also the bandwidth can be further extended. Since the bandwidth of the \( \lambda/4 \) transformer is decided by the ratio of the impedance transformation, which is shown in Fig. 3.5-4, the transformation ratio of each \( \lambda/4 \) line should be kept as low as possible. The single-section Wilkinson combiner/divider (shown in Fig. 3.5-1) employs \( \lambda/4 \) line with the impedance transformation ratio of two (from 50\( \Omega \) to 100\( \Omega \)), so its bandwidth in terms of \( S_{33} \) is determined by the \( S_{11} \) curve with ratio 2:1 in Fig. 3.5-4. By employing two sections of \( \lambda/4 \) lines, each line has the impedance transform ratio of \( \sqrt{2} \), a two-section Wilkinson
combiner/divider is shown in Fig. 3.5-5. From the simulation shown in Fig. 3.5-6, we can observe that the bandwidth of the two-section Wilkinson combiner/divider is greatly improved (4GHz -15GHz), as compared with the previous one octave bandwidth (6GHz -12 GHz).

![Diagram](image)

Fig. 3.5-5: Two-section Wilkinson combiner/divider. \( Z_1 = 59\Omega, Z_2 = 84\Omega, R_1 = 94\Omega, R_2 = 118\Omega \), the system impedance \( Z_0 = 50\Omega \)

![Graph](image)

Fig. 3.5-6: Frequency response of the two-section Wilkinson combiner/divider.
3) Binomial (maximally flat) and Chebyshev (equal-ripple) multi-section matching transformers.

There are two common types of multi-section λ/4 transformers: Binomial and Chebyshev transformers. The pass-band response of a binomial transformer is optimum in the sense that the response is as flat as possible near the design frequency. Thus, it is also known as maximally flat transformer. The Chebyshev transformer optimizes bandwidth at the expense of pass-band ripple. If such a pass-band characteristic can be tolerated, the bandwidth of the Chebyshev transformer will be substantially better than that of the binomial transformer, for a given number of sections.

![Diagram of a multi-section λ/4 transformer](image)

Fig. 3.5-7: Maximally flat multi-section λ/4 transformer design. (Z_{in1}, Z_{in2}, Z_{in3} are the intermediate impedances, \(\Gamma_n\) (n=1,2,3,4) are partial reflection coefficients)

Fig. 3.5-7 shows an example of the design of a four-section maximally flat transformer. Suppose we intend to transform impedance from \(Z_L\) to \(Z_S\), then the
intermediate impedances would be the geometric mean of the corresponding source and load impedances, that is: \( Z_{in1} = \sqrt{Z_S Z_L} \), \( Z_{in2} = \sqrt{Z_{in1} Z_L} \), \( Z_{in3} = \sqrt{Z_S Z_{in1}} \). Next the characteristic impedances of each sections can be easily calculated as: \( Z_1 = \sqrt{Z_{in3} Z_S} \), \( Z_2 = \sqrt{Z_{in3} Z_{in1}} \), \( Z_3 = \sqrt{Z_{in1} Z_{in2}} \), \( Z_4 = \sqrt{Z_L Z_{in3}} \). Or we can obtain partial reflection coefficients shown in Fig. 3.5-6 by using formula:

\[
\Gamma_n = \frac{A C_n^N}{\frac{Z_L - Z_s}{Z_L + Z_s}}
\]

where \( A = 2^{-N} \left| \frac{Z_L - Z_s}{Z_L + Z_s} \right| \)

and then the characteristic impedances of each sections \( Z_n \) (\( n = 1, 2, \ldots, N \)) are be easily calculated using the following equations:

\[
\Gamma_0 = \frac{Z_1 - Z_S}{Z_1 + Z_S}
\]

\[
\Gamma_n = \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} \quad (n = 1, 2, \ldots, N - 1)
\]

\[
\Gamma_N = \frac{Z_L - Z_N}{Z_L + Z_N}
\]

The fractional bandwidth of the binomial transformer can be evaluated as follows:

\[
\frac{\Delta f}{f_0} = \frac{2(f_0 - f_m)}{f_0} = 2 - \frac{4}{\pi} \cos^{-1} \left[ \frac{1}{2} \left( \frac{\Gamma_m}{A} \right)^{1/N} \right]
\]

\( A = 2^{-N} \left| \frac{Z_L - Z_s}{Z_L + Z_s} \right| \)

65
where \( f_m \) is the lower cut-off frequency of the pass-band, \( N \) is the number of sections, \( \Gamma_m \) is the maximum allowed reflection coefficient, \( Z_L \) and \( Z_S \) are the load and source impedance respectively.

The partial reflection coefficients of Chebyshev multi-section transformer \( \Gamma_n \) \((n = 1,2, \ldots , N)\) are achieved by making overall reflection coefficient \( \Gamma(\theta) \) \((\theta = \beta l, \quad \theta = \pi / 2\) at center frequency \( f_0 \), \( \theta = \theta_m \) at lower cutoff frequency \( f_m \) where reflection coefficient \( \Gamma = \Gamma_m \) proportional to \( n \)th order Chebyshev polynomial \( T_n(\sec \theta_m) \), that is:

\[
\Gamma(\theta) = 2e^{-jN\theta}[\Gamma_0 \cos N\theta + \Gamma_1 \cos(N-2)\theta + \ldots + \Gamma_n \cos(N-2n)\theta + \ldots ]
\]

\[
= Ae^{-jN\theta} \Gamma_n(\sec \theta_m \cos \theta)
\]

where

\[
A = \Gamma_m
\]

\[
\sec \theta_m = \cosh[\frac{1}{N} \cosh^{-1}\left(\frac{1}{\Gamma_m} \frac{Z_L - Z_S}{Z_L + Z_S}\right)]
\]

and the fractional bandwidth can be calculated from the following equation once \( \theta_m \) is known:

\[
\frac{\Delta f}{f_0} = 2 - \frac{4\theta_m}{\pi}
\]
Fig. 3.5-8: (a) Four-section 4-way combiner with high impedance $\lambda/4$ lines, (b) Four-section corporate combiner.
4) Corporate power combiner.

In the previous N-way multi-section combiner (N parallel quarter wave networks) design, high impedance \( \lambda/4 \) lines are needed which is a big disadvantage of the design. As a result, the corporate combiner/divider is employed to accommodate the problem. Fig. 3.5-8 (b) illustrates a four-section corporate combiner/divider design with impedance transformation from 100\( \Omega \) to 200\( \Omega \), compared with the N-way combiner design shown in Fig. 3.5-8 (a), high impedance \( \lambda/4 \) lines are avoided. In both N-way combiner and corporate combiner design, the termination resistors are eliminated because of the difficulty in the layout of those resistors between adjacent CPW signal lines. But this change should not affect the performance of the combiner if all the combined signals have the same signal (amplitude and phase) drive. In other words, assuming the even mode excitation (as shown in the Wilkinson power combiner) for the combiner inputs, we can safely ignore the termination resistors since those resistors are intended to offset the difference of the combined signals. We can simulate the frequency response of the combiner using Essof/HP harmonic balance test bench shown in Fig. 3.5-9, each combined branch is terminated with the desired transformed resistance and driven with the same signal source, the power delivered onto the load resistor is then measured. Fig. 3.5-10 shows the simulation results of the corporate combiner, we can achieve more than 80% combining efficiency. One of the disadvantages of the corporate combiner is that it is not area-efficient, but since we use inexpensive thermally-conductive, electrically-insulating AlN substrate, it's not a big concern. But on the other hand, the conductive loss of the corporate power combiner/divider due to the lengthy transmission lines could be costly, especially at the output of the devices, since the loss will directly reduce the output power from the power devices.
Fig. 3.5-9: Test bench for the frequency response of the combiner/divider.

Fig. 3.5-10: Four-section corporate combiner simulation results. (Transform from 100Ω to 200Ω)
REFERENCE

Chapter 4
Power Amplifier Structures

4.1 Challenges of broadband power amplifier design

For given device technology, broadband amplifier design is constrained by the gain-bandwidth product, impedance matching limitation and other requirements, and various broad-banding techniques are attempted and utilized in order to achieve the desired broadband performance. Broadband power amplifier design faces even more challenges due to the additional notorious issues such as nonlinear effect, thermal management, device reliability and so forth. In this section, we will focus on four major design considerations for broadband power amplifier, the respective approaches and techniques that could address those problems.

1) High impedance transformation ratio:
In order to achieve large amount of output power from the amplifier, large size devices have to be utilized, as a result, the input and output impedance of the device would normally be too low to be matched to the 50 Ω source and load impedance. In other word, the input and output impedance transformation ratio would be too high to realize broadband matching since the ratio will have a great effect on the achievable bandwidth.

For example, for a 4-mm wide GaN HEMT, the input resistance is about 1 Ω and the desired output resistance (optimum resistance for maximum output power) is about 8 Ω. Therefore, we have to realize 50:1 impedance transformation at the input and 50:8 at the output. In order to achieve broadband matching (at least 3:1
bandwidth in our work), we can employ multi-section LC matching networks and corporate power combiner/divider (consisted of multi-section \(\lambda/4\) transformer) or the combination of both. As we discussed in section 3.5, we have to limit the impedance transform ratio of each section of \(\lambda/4\) transformer to less than 1.5 to be able to achieve bandwidth great than 3:1, and the same criteria will apply to the LC network design as was discussed in section 3.4. Consequently, it’s necessary to employ more than two sections of both \(\lambda/4\) transformer and LC network, which could create too lengthy transmission lines and have significant conductive loss.

Another issue of the impedance matching is parasitic absorption, both at the input and output of the device (\(C_{gs}\) and effective \(C_{ds}\)). The parasitic absorption problem is measured in Q (related to RC product), which was discussed in section 3.2, the achievable bandwidth for matching an arbitrary load using lossless matching network is decided by the Q value according to Fano’s rule. Hence, theoretically matching (from a parasitic absorption point of view) is the same for device with different size since Q doesn’t scale with gate-width. But as we just mentioned, due to the practical issue of matching very low impedance to 50 \(\Omega\) for large periphery FETs, that is: since too many sections of LC network and \(\lambda/4\) transformer will be needed, the whole matching circuitry will become too lossy to be functional, it’s practically much more difficulty to match large periphery FETs than smaller FETs.

2) Immature large-signal modeling:
In order to predict the large-signal performance of the amplifier, accurate large-signal model is required. We have developed a curtice-3 model that can produce the DC-IV characteristic of the devices with first-order accuracy by assuming linear \(g_m\), zero output conductance as discussed in section 2.6. But due to the
complex thermal distribution issue and dispersion phenomena of the device are not included in this model, it can only be used as a guide for circuit simulation. Even though, at the end of this work, we have obtained reasonably good agreement between simulation and measurement result by further modifying the large-signal model, large-signal design is relied heavily on the load-pull data.

Cripps method (Optimum load-line method) is a simple but effective approach to achieve maximum output power; it is very effective for the GaAs power amplifier design. However, due to the immaturity of GaN material system, usually there is considerable disagreement between DC and RF IV characteristic, Cripps method could produce erroneous optimum load impedance with the use of maximum current $I_{\text{max}}$ and knee voltage $V_{\text{knee}}$ obtained from DC-IV curve. For example, using Cripps method, the optimum load of GaN HEMTs is 45-50 $\Omega$-mm, but actually, it should be 32 $\Omega$-mm from the load-pull measurement. Loadpull method is an empirical approach; it requires complex experiment setup and device test feature before you can proceed to amplifier design. But it can produce very accurate optimum load impedance, including both resistance and reactance part. As a result, in the amplifier design, we will first design passive matching networks that will present the devices with the optimum load impedance obtained from the load-pull measurement at both input and output, and then use curtice-3 large-signal model for the device to simulate the large-signal performance of the amplifier.

3) Thermal Management:
For power amplifier, large amount of power will be generated onto the load and meanwhile even more power needs to be dissipated from the circuit down to the heat sink since normally PAE is relatively low (~20 -30 %) for broadband power amplifier. Therefore, it is very important to have good thermal management in order to keep the devices working at relatively low temperature. Since, in our
work, GaN HEMTs are grown on sapphire substrate, which is a very poor thermal conductor, it’s even more critical to be able to dissipate the heat out of the active region of devices. We solve the thermal issue by flip-chip bonding the devices onto the AlN circuit board, which is described in details in section 2.5. Even with the flip-chip technique, it’s still very critical to design the pattern of the device to have a low thermal resistance. Gate-to-gate spacing (gate pitch) is the most important parameter to determine the thermal resistance. In general, the wider is the gate pitch size; the lower is the thermal resistance. Also, we can achieve good thermal management by utilizing GaN HEMTs that are grown on excellent thermal conductor: SiC substrate. Pulsed operation can sometimes be used to obtain a peak output power that is greater than the CW power output.

4) Poor input/output matching (large VSWR)
As mentioned in the first consideration, the input impedance transformation ratio can be as high as 50:1, which makes it difficult to achieve a good matching. Also, sometimes we could intentionally introduce mismatch at the input to compensate for the 6dB/octave gain roll-off characteristic of the device, therefore achieving gain flatness over the whole bandwidth. In general, the input matching of the broadband power amplifier tends to be poor. At the output of the power amplifier, the device is not conjugately matched; instead the optimum load has to be presented to the device for the maximum output power. Therefore the output mismatch normally would be significant as well. There are several ways to address the problem. First, we can employ a distributed network at the input to realize good broadband matching, and be able to achieve gain flatness at the meantime. But at the output, the distributed network is not a good choice because it would greatly reduce the efficiency and output power due to the backward wave traveling to the dummy drain-line load. Other broadband matching networks can be designed at both the input and output, but the bandwidth is usually not as broad.
as the distributed approach and they tend to be lossy. Also we can design multi-stage amplifiers to achieve better input matching by distributing the high impedance transformation ratio among different stages. Finally, balanced amplifier is a very good approach to achieve low VSWR at both input and output. The output VSWR of all the amplifiers in this work will remain high since it's not our main concern.

4.2 Power amplifier structures

The simplest power amplifier would contain single gain stage, using a single transistor. However, any practical power amplifier will be a subassembly consisting of several driver and gain stages, and the final power stage itself probably will use some form of power combining network. Due to the specific requirements such as linearity, efficiency and VSWR, we have to consider some other conventional yet important techniques such as push-pull and balanced amplifiers. Also, conventional distributed amplifier, which has excellent broadband characteristics, will be discussed in details in terms of both the small-signal and large-signal analysis.

1) Multi-stage power amplifier:
It's always necessary to have a driver to provide enough drive power for the last stage that contains the largest size of power device. Sometimes, additional gain stage will be required in front of the driver to provide enough overall gain for commercial system as shown in Fig. 4.2-1 (a). Even though both gain and driver stage design normally only involve linear (small signal) amplifier, there are some issues that need to be treated carefully, such as inter-stage match, biasing circuitry and stability consideration. As shown in Fig. 4.2-1 (b), for a single-ended driver-PA configuration, the inter-stage matching networks needs to match between two
complex impedances over a broadband, possess certain gain compensation slope (counterpart of 6dB/octave gain roll-off slope of device) to realize gain flatness, and provide both output and input bias insertion networks for the driver and last stage PA devices. Those tasks could require very cautious design and intricate techniques that further complex the power amplifier design process. Therefore, in this study, we will only focus on the design of the last stage of the power amplifier and try to fully demonstrate the advantage of GaN HEMTs as microwave power device without the distraction of other complex circuit design issues.

Fig. 4.2-1: (a) Multi-stage power amplifier structure. (b) Inter-stage matching network design for single-ended driver-PA configuration.
2) Push-pull power amplifier:

![Diagram of push-pull power amplifier](image)

Fig. 4.2-2: Simple integrated circuit Class B push-pull amplifier.

Class B push-pull amplifier has been an old technique invented in the early days of tubes. It can preserve the high-efficiency characteristic of class B operation while maintaining broadband linear operation. The conventional push-pull amplifier is shown in Fig. 4.2-2, two active devices (nnp and pnp bipolar transistors) are used to deliver the power instead of one, and each device conducts for alternate half cycles in class B mode. Theoretically, this amplifier can have high-efficiency of class B operation without the distortion of the sinusoidal input signal. For RF and microwave application, we normally have push-pull amplifier in the form as shown in Fig. 4.2-3. Two identical devices are employed instead of complimentary pair due to the unavailability of decent p-type devices. The two devices will be driven differently and then two output signals will be combined in anti-phase using some sort of baluns such as center-tapped transformer shown in the figure. But we have to point out that: for broadband amplifiers that have an octave or more bandwidth, we cannot filter out second harmonics such would be
the case in narrower band class-B amplifiers. This means the efficiency will not be as high as 78%, but there are still considerate benefits from making amplifiers this way such as improved linearity and much higher efficiency than class-A operation. Unfortunately, as the operating frequency is pushed into over giga-hertz region, broadband balun design becomes an increasingly intimidating task. Especially for the CPW environment, it is even more challenging than the microstrip line approach.

![Conventional RF push-pull amplifier with center-tapped transformer as balun.](image)

**Fig. 4.2-3:** Conventional RF push-pull amplifier with center-tapped transformer as balun.

3) Balanced power amplifier:
Typical balanced amplifiers are shown in Fig. 4.2-4, two identical amplifiers are fed from an input power splitter, which produces two signals in phase quadrature,
the outputs being recombined using a similar power combiner as the power splitter but connected in reverse. The principle advantage of this balanced configuration is that the mismatches at both the input and output will be cancelled out at the termination dummy loads. Theoretically, the match seen at the input and output ports will be essentially perfect if the two amplifiers are identical and the couplers have ideal amplitude and quadrature phase response. Not surprisingly, the broadband quadrature 3dB coupler design turns out to be quite a challenging. It's very difficult to achieve two output ports with identical -3dB amplitude and quadrature phase signals over broad bandwidth, especially at higher frequency region. Similar to the balun design for push-pull amplifier, CPW line approach is more difficulty than micro-strip line design due to the difficulty with the layout and unavailability of accurate CPW line models. Meanwhile, improved input and output matching for the balanced approach are realized at the expense of doubled circuit components and DC power consumption, which is a disadvantage.

Fig. 4.2-4: Configuration of balanced amplifier.
Fig. 4.2-5: (a) Configuration of conventional traveling wave amplifier (TWA). (b) The equivalent circuit model for the TWA using the simplified small-signal model for a FET.
4.3 Distributed power amplifier (Traveling wave amplifier).

As was mentioned in section 2.6, distributed amplifiers can achieve the maximum possible gain-bandwidth product ($f_{\text{max}}$) for a given device technology. Also, it's possible to achieve flat gain over the frequency range from DC to certain high cut-off frequency with excellent input and output match using this kind of circuit structure. Therefore, distributed amplifiers have been established as a reliable candidate for many broadband applications such as broadband receivers, satellite communication systems, pulsed radars, test instruments etc. In the 1980s, hybrid and monolithic GaAs FET-based distributed amplifiers were successfully demonstrated, which marked an important point in the evolution of the distributed amplifier.\(^1\)\(^2\)\(^3\) During that period, Ayasli and coworkers at Raytheon introduced "capacitive-division concept"\(^4\) in their GaAs traveling wave power amplifier (TWPA), which further extended the application of distributed amplifiers into the field of microwave power amplifiers. However, there are distinctive disadvantages for distributed amplifiers as power amplifiers, poor efficiency is one major drawback. As a result, in this work, we modified the conventional traveling wave amplifier (TWA) and developed a more suitable topology for broadband power amplification. We will discuss the modified TWPA in details in section 5.1.

Fig. 4.2-5 (a) and (b) shows the configuration of a conventional TWA and its equivalent circuit by using the simplified small-signal model (shown in Fig. 4.2-5 (a)) for a FET respectively. Since the inductors are always realized by using very high impedance transmission lines, sometimes in the figures of TWA the transmission lines will replace the inductors. Assuming for the moment that $R_{\text{in}} = 0$ and $R_{\text{ds}} = \infty$, then each gate-source capacitance $C_{\text{gs}}$ is symmetrically embedded
between a pair of inductors \( I_{\text{gate}}/2 \) and forms an artificial transmission line with characteristic impedance
\[
Z_i^g = \frac{Z_0^g}{\sqrt{1 - \left(\frac{\omega}{\omega_c^g}\right)^2}} \quad (4.2-1)
\]
where
\[
Z_0^g = \sqrt{\frac{L_{\text{gate}}}{C_{gs}}} \quad \text{and} \quad \omega_c^g = \frac{2}{\sqrt{L_{\text{gate}} \times C_{gs}}} \quad (4.2-2)
\]

As a result, the artificial gate transmission line will have characteristic impedance \( Z_0^g \) if the signal frequency is well below the cutoff frequency \( \omega_c^g \) (we call it Bragg frequency). If the gate transmission line is terminated in source and load impedance \( Z_0^g \), then the line will be perfectly matched if the frequency is well below Bragg frequency.

1) Small-signal analysis
The power gain of the TWA, \( |S_{21}|^2 \), is by definition the ratio of the power delivered to the load to the power available from the source. Assume \( v_s \) is the source voltage, \( i_0 \) is the current flowing through the load, \( Z_0^g \) and \( Z_0^d \) are the characteristic impedance of the artificial gate and drain lines respectively, then
\[
|S_{21}|^2 = \left| \frac{i_0^2 Z_0^d}{v_s^2 / 4Z_0^g} \right| \quad (4.2-3)
\]
where \( i_0 \) can be derived from the following equations
\[
i_0 = i_n^g + i_{n-1}^g \exp(-j\theta_d) + \ldots + i_1^g \exp[-j(n-1)\theta_d]
\]
\[
i_k^g = (g_m v_{gs}^k) / 2, \quad \text{for} \quad k = 1 \rightarrow n
\]
\[
v_{gs}^k = \frac{v_s}{2} \exp[-j(k - \frac{1}{2})\theta_d]
\]
where \( i_k^R \) is the amount of RF drain current emanating from the \( k \)th FET flowing to the right, \( \theta_g \) and \( \theta_d \) are the gate and drain line phase change per section, \( \theta_g = \omega(L_{\text{gate}} C_{gs})^{1/2} = 2\omega/\omega_c^g \), \( \theta_d = 2\omega/\omega_c^d \). After substitution and summation, we can obtain

\[
|i_0| = \left| \frac{g_m v_f}{4} \left( 1 + \exp[-j(\theta_d - \theta_g)] + \ldots + \exp[-j(n-1)(\theta_d - \theta_g)] \right) \right|
\]

\[
= \frac{g_m v_f}{4} \left| \frac{\sin \frac{n}{2}(\theta_d - \theta_g)}{\sin \frac{1}{2}(\theta_d - \theta_g)} \right|
\]

Finally, we can obtain the gain of the TWA given by

\[
|S_{21}|^2 = \frac{g_m^2 Z_0^2 Z_d^2}{4} \left| \frac{\sin \frac{n}{2}(\theta_d - \theta_g)}{\sin \frac{1}{2}(\theta_d - \theta_g)} \right|^2 \tag{4.2-4}
\]

Normally the following conditions are imposed: \( Z_0^g = Z_0^d \), \( \theta_g = \theta_d = \theta \) and therefore \( \omega_c^g = \omega_c^d = \omega_c \), \( \theta = 2\omega/\omega_c \), in which case we must have \( L_{\text{gate}} = L_{\text{drain}} \), \( C_{gs} = C_{ds} = (C_{ds} C_d)/(C_{ds} + C_d) \). Then the power gain is simplified to

\[
|S_{21}|^2 = \frac{n^2 g_m^2 Z_0^2}{4} \tag{4.2-5}
\]

If we consider the complete frequency range (0 - \( \omega_c \)), then the gain has to be modified as following (the derivation is skipped)

\[
|S_{21}|^2 = \frac{1}{4} \frac{n^2 g_m^2 Z_0^2}{1 - \left( \frac{\omega}{\omega_c} \right)^2} \tag{4.2-6}
\]

Thus the gain actually increases monotonically with increasing frequency, but the gain is within 1 dB of its dc value up to \( \omega = 0.45 \omega_c \).
The reverse gain of the TWA, that is, the gain referred to the power dissipated in the drain-line dummy load, can be calculated in a similar manner to that used to calculate the forward gain. The result is

\[
|S_{21}^R|^2 = \frac{1}{4} \left( \frac{g_m Z_0}{\sin \theta} \right)^2 \left( \frac{n \theta}{\sin \theta} \right)^2 \tag{4.2-7}
\]

where \( \theta = \theta_e = \theta_d = 2\omega / \omega_c \). It's not difficult to find out that the reverse gain is almost the same at very low frequency (\( \sim \) DC, \( \theta = 0 \)) and at very high frequency (\( \sim \omega_c \), both \( |S_{21}|^2 \) and \( |S_{21}^R|^2 \rightarrow \infty \)) as the forward gain, and more than 10 dB lower than the forward gain in the middle region (\( \sim 0.5\omega_c \)).

So far we only consider the ideal situation of both artificial gate and drain lines being lossless by assuming \( R_{in} = 0 \), \( R_{ds} = \infty \) and neglect the conductive loss of the transmission lines. In practical, the loading effect of both gate and drain lines is rather significant, especially the gate-line loss. The \( R_{in} \) causes attenuation of the input voltage wave as it travels down the gate transmission line such that the ac gate-source voltage of the right-hand end FETs in Fig. 4.2-5 (a) is smaller than that of the left-hand ones. The effect is more pronounced as the frequency increases since the attenuation constant \( \alpha_g \) is proportional to the frequency. We can alleviate the unequal drive problem by introducing “capacitive division” which will be discussed in section 5.1. Unfortunately a perfect compensation can only be realized at one frequency, since the gate-line loss is frequency dependent.

The attenuation constant \( \alpha_g \) can be derived from the equation for the propagation constant \( \gamma \) of the periodic equivalent circuit structure for the transmission line shown in Fig. 4.2-6.
\[
\gamma = \alpha + j\beta = \sqrt{Z \times Y} = \sqrt{\frac{1}{j\omega L \times R + \frac{1}{j\omega C}}}
\]

Since \( R \ll 1/\omega C \) for the typical FETs (\( C \approx 1 \text{ pF}, R \approx 1 \Omega \)) at microwave frequencies, and \( Z_0 = (L/C)^{1/2} \), it can be further simplified to

\[
\gamma = j\omega \sqrt{LC} (1 - \frac{1}{2} j\omega RC)
\]

\[
= \frac{1}{2} \omega^2 RC^2 Z_0 + j\omega \sqrt{LC}
\]

Therefore, we can obtain that

\[
\alpha_g = \frac{1}{2} \omega^2 R_{in} C^2_{gs} Z_0 \quad (4.2-8)
\]

Similarly, we can obtain the attenuation constant of the drain-line, that is

\[
\alpha_d = \frac{1}{2} Z_0 / R_{ds} \quad (4.2-9)
\]

We have to point out that since \( L, C \) and \( R \) in Fig. 4.2-6 are the values for each section (not unit length values), \( \alpha_g \) and \( \alpha_d \) given by equation (4.2-8) and (4.2-9) are actually the losses per section instead of losses per unit length.

![Diagram](Image)

Fig. 4.2-6: Periodic equivalent circuit structure of the artificial gate line (neglecting the signal conductive loss).
From the equation for $|S_{21}|^2$ (4.2-5), it seems that if we increase the number of the devices $n$ while keep each device size small enough, then we can achieve higher gain over the desired bandwidth, in the other words, the gain-bandwidth product can be increased without limit as long as we increases the number of devices. In fact, the number of the devices is limited by the loss of the gate and drain lines. With the impose of the following condition

$$n\alpha_g \times n\alpha_d \leq \left(\frac{1}{2}\right)^2$$  \hspace{1cm} (4.2-10)

which defines the upper-limit of the number of the devices that can be employed in a TWA. Using equations (4.2-5), (4.2-8), (4.2-9) and (4.2-10), the gain-bandwidth product can be derived as

$$|S_{21}| \times BW = \frac{1}{2} n g_m Z_0 \frac{\omega}{2\pi} \leq \frac{1}{2} \left(\frac{1}{2\sqrt{\alpha_g \alpha_d}}\right) g_m Z_0 \frac{\omega}{2\pi}$$

$$= \frac{1}{2} g_m \sqrt{\frac{R_{ds}}{2\pi C_{gs}}} \sqrt{\frac{R_{ds}}{R_{in}}}$$

$$= \frac{1}{2} f_t \sqrt{\frac{R_{ds}}{R_{in}}}$$

$$= f_{\text{max}}$$  \hspace{1cm} (4.2-11)

which shows that TWA can have gain-bandwidth product much higher than $f_t$, which normally defines gain-bandwidth product for the most analog high-frequency amplifiers.

2) Large-signal analysis

We must first calculate the drain-line voltage distribution, in order to characterize the large-signal properties of the TWA$^6$. For simplicity, assume the ideal case again, that is both gate and drain artificial lines are lossless. As is shown in Fig. 4.2-4 (b), each FET will inject a current of $g_m v_{gs}^k$ ($k = 1 \rightarrow n$) into the drain line,
half of which travel to the left and half to the right since the device sees the same impedance $Z_0$ in both directions. Then the voltage at the drain of the kth FET $v_{ds}^k$ is readily calculated by the following equation

$$v_{ds}^k = \frac{Z_0}{2} \{ i_1^R \exp[-j(k-1)\theta] + ... + i_{k-1}^R \exp(-j\theta) + i_1^L \exp(-j\theta) + ... + i_n^L \exp[-j(n-k)\theta]\}$$

where $i_k^R = i_k^R = g_m v_{gs}^k/2$ and $v_{gs}^k = v_s \exp[-j(k-1/2)\theta]$ ($k = 1 \rightarrow n$). We can further simplify the expression for $v_{ds}^k$, after some lengthly yet straightforward manipulations, to the following equation

$$|v_{ds}^k| = \frac{Z_0}{2} g_m v_s \frac{\sqrt{k^2 + \frac{2k \cos(n-k+1)\theta}{\sin \theta} + \frac{\sin^2(n-k)\theta}{\sin^2 \theta}}}{n/2}$$

It's not difficult to figure out that when $n = k$, $|v_{ds}^k| = Z_0 g_m v_s n/2$ and at all frequencies

$$|v_{ds}^k| \geq |v_{ds}^{k-1}|, \quad \text{for} \quad k = 2 \rightarrow n$$

Therefore, we can draw the conclusion that the largest drain-source voltage appears across the last FET. Then the largest possible linear output power from the TWA (in a $Z_0$ system) under class A conditions should be

$$P_{RF \max} = \frac{(V_{dgB} - |V_{pinch}| - V_{knee})^2}{8Z_0} \quad (4.2-12)$$

$$\quad \text{if} \quad Z_0 \geq R_{spe} = \frac{V_{dgB} - |V_{pinch}| - V_{knee}}{(nI_{max}/2)} \quad (4.2-13)$$

where $V_{dgB}$ is the drain-gate breakdown voltage, $V_{pinch}$ is the device pinch-off voltage, $V_{knee}$ is the device knee voltage and $I_{max}$ is the maximum drain current from one single FET. The best design would be: each single FET has the right gate-width therefore right $I_{max}$ that makes the optimum load defined in equation (4.2-13) exactly the same as the system impedance $Z_0$. The implication with the condition (4.2-13) $nI_{max}/2 \geq (V_{dgB} - |V_{pinch}| - V_{knee})/Z_0$ is that: the signal swing at the
last FET shouldn’t be limited by the current boundary, in other words, the FETs should be able to provide enough current \((nI_{\text{max}}/2)\) into the load to support the maximum voltage swing allowed by the last FET. For a typical GaAs FET \(V_{\text{dgB}} = 20 \text{ V}, V_{\text{pinch}} = -2.5 \text{ V}, V_{\text{knee}} = 1 \text{ V}\), then \(P_{\text{RFmax}} = 0.7 \text{ W}\) if \(Z_0 = 50\Omega\). For a typical GaN HEMT, \(V_{\text{dgB}} = 50 \text{ V}, V_{\text{pinch}} = -5 \text{ V}, V_{\text{knee}} = 5 \text{ V}\), then \(P_{\text{RFmax}} = 4 \text{ W}\). The only way to circumvent this limitation is to use a very low value for \(Z_0\), which will result in a gain reduction and also will need a broadband impedance transformer at the output.

To achieve the maximum output power level, the input signal drive should be able to swing between \(V_{\text{gs}} = 0\) and \(V_{\text{gs}} = -|V_{\text{pinch}}|\), which means that the RF input power delivered to the amplifier (in a \(Z_0\) system) should be

\[
P_{\text{in}} = \frac{V_{\text{pinch}}^2}{8Z_0}
\]

Therefore, the large signal gain should be

\[
|S_{21}|^2 = \frac{P_{\text{RFmax}}}{P_{\text{in}}} = \left(\frac{V_{\text{dgB}} - |V_{\text{pinch}}| - V_{\text{knee}}}{V_{\text{pinch}}}\right)^2
\]

That is, the gain of the TWA designed for maximum power output and efficiency is determined only by the DC parameters of the FET. The large-signal gain will only be identical to the small-signal gain if the FET has a linear \(g_m\). For a given device technology, the DC parameters are fixed, as a result, the maximum output power is pre-determined no matter what size of devices being employed. Also, since only half of the current emanated from each device can contribute to the output power, the maximum linear power output, hence, efficiency of the TWA can only be half that achievable from lossless combining of the individual FETs. Therefore, theoretically the maximum achievable efficiency for a TWA under
Class A condition would only be 25% due to the wave traveling to the left, which is sometimes referred as the backward wave.

![Diagram of tapered output traveling wave power amplifier](image)

**Figure 4.2-7** Tapered output traveling wave power amplifier

We cannot increase the maximum available output power level for a given device technology and system impedance level $Z_o$, but we can increase the efficiency by requiring fewer FETs (less total device size) and, therefore, less DC power dissipation to achieve the same current swing (the same output power level at the load). Tapering of the drain-line has been used to force more of the injected current from the devices to travel to the right than to the left. In the topology shown in Fig. 4.2-7, the device output current is not equally split, by using sections of transmission lines with varying characteristic impedances, therefore most of the injected current is directed to the load. However, this configuration has two major disadvantages: first of all it will result in unrealistic transmission line impedances since the drain-line impedances ranging from $Z_1$ to $Z_1/N$ ($N$ is the number of devices). If $N = 4$ and $Z_1/4 = 50 \Omega$ as shown in Fig. 4.2-7, then the required impedances would be ranging from 200 $\Omega$ to 50$\Omega$, which cannot be practically implemented. Second, this technology requires a step discontinuity in the drain line characteristic impedance to occur at each FET, and hence it tends to
result in degraded output VSWR over wide bandwidths. Based on the tapered drain-line approach, we developed a new TWA topology, which circumvents the previous problems yet still be able to achieve much high drain efficiency than the conventional TWA. We will discuss the modified TWA in details in section 5.1.
REFERENCE


Chapter 5
GaN Power Amplifiers Design and Performance

5.1 Modified traveling wave power amplifier (TWPA).

Traditional distributed amplification techniques have allowed for the flat gain and wide bandwidth at the expense of low efficiency. The decreased efficiency is primarily due to the existence of the dummy load termination of the output artificial transmission line. Backward wave propagation results in no more than a 50% overall output network combining efficiency with class-A operation. In order to improve the efficiency of the power amplifier, we developed a modified TWPA topology; the schematic is shown in Fig. 5.1-1. It's not difficult to figure out that this topology is similar to the tapered drain-line TWA shown in Fig. 4.2-7 if all the drain delay lines are stacked up and connected. It employs the conventional TWA's artificial input gate transmission line to realize the input broadband matching condition. The artificial drain line is replaced by delay lines for phase equalization followed by a broadband corporate combiner. The corporate combiner was designed to present each device with the optimum load for maximum power. Compared with the conventional TWA, this modified TWA topology eliminates the backward wave, thus theoretically increasing the output efficiency by a factor of two. Furthermore, the corporate combiner structure can avoid the problems associated with realizing very high impedance lines. These advantages make the modified TWA topology an excellent candidate for broadband high-efficiency power amplifiers. The circuit is not as area-efficient as a MMIC, but in this work all of the transmission lines will be fabricated on a low-cost AlN carrier.
Fig. 5.1-1: Schematic of the modified TWPA.

Fig. 5.1-2: Schematic of capacitive-division modified TWPA.
In order to achieve high output power, large gate periphery devices are needed which introduce large input capacitances \( C_{gs} \). Capacitive division\(^2\) was then employed to increase the bandwidth, as was shown in Fig. 5.1-2. Since the Bragg frequency \( f_{brag} \) of an artificial transmission line is inversely proportional to the square root of the gate-source capacitance, we have to reduce the effective input gate capacitance by introducing capacitive-division in order to realize the desired bandwidth. Normally, we have to keep the highest operating frequency less than at least half of the Bragg frequency. For given \( f_{brag} \) and \( C_{gs} \), we can use the subsequent equations to calculate parameters \( L_{gate}, C_{in} \).

\[
Z_0 = \sqrt{\frac{L_{gate}}{C_{gate}}} = 50\Omega
\]

\[
f_{brag} = \frac{1}{\pi \sqrt{L_{gate} \times C_{gate}}}
\]

\[
C_{gate} = \frac{C_{in}C_{gs}}{C_{in} + C_{gs}}
\]

Since the series capacitor \( C_{in} \) and gate-source capacitance \( C_{gs} \) combination forms a potential divider, it will cause a substantial loss of gain depending on the ratio of the \( C_{in}/(C_{in} + C_{gs}) \), the lower the ratio, the more the gain is reduced. But since the device gate-width and therefore the total output power can be substantially increased while maintaining the same bandwidth, this technique is still very beneficial for the power amplifier application. Also, to obtain equal input drive for each device, the capacitances can be varied \( (C_{in4} > C_{in3} > C_{in2} > C_{in1}, \text{ in Fig. 5.1-2}) \) to compensate for the loss along the input gate-line. Since such a loss is frequency dependent (gate attenuation \( \alpha_g = \omega_c^2 R_{in} C_{gs}^2 Z_0 \)), a perfect compensation can only be realized at one frequency, which normally is chosen at the mid-band in our design.
5.2 GaAs modified TWPA:

1) GaAs 1.2mm-wide discrete power device.

The first modified TWPA we have fabricated is based on GaAs commercial FETs (Fig. 5.2-1) that are manufactured by TriQuint Semiconductor Company. The TriQuint TGF4230-EEU is a single gate 1.2mm-wide, 0.5μm-long discrete GaAs Hetero-structure Field Effect Transistor (HFET) designed for high-efficiency power application up to 12 GHz in class A and class AB operation. It is readily assembled using automatic equipment, the bond pads and backside metalization are gold plated for compatibility with eutectic alloy attach methods as well as thermo-compression and thermo-sonic wire-bonding processes. The device chip size is: 0.572 x 0.699 x 0.102 mm (0.023 x 0.028 x 0.004 in.).

![Fig. 5.2-1: Picture of TriQuint 1.2mm-wide 0.5 μm-long GaAs HFET (product number TGF4230-EEU).](image)

We have the device small-signal model as shown in Fig. 5.2-2, the parameters are:

\[
R_g = 0.43 \, \Omega, \quad R_i = 1.21 \, \Omega, \quad R_s = 0.4 \, \Omega, \quad R_{ds} = 96 \, \Omega, \quad R_d = 0.66 \, \Omega, \quad C_{gs} = 1.21 \, \text{pF}, \quad C_{dg} = 0.01 \, \text{pF}, \quad C_{ds} = 0.25 \, \text{pF}, \quad L_g = 0.04 \, \text{nH}, \quad L_d = 0.02 \, \text{nH}, \quad L_s = 0.01 \, \text{nH}, \quad g_m = 0.2 \, \text{S}, \quad \tau = 5.49 \, \mu\text{sec}.
\]

The device DC and RF performance are shown in Fig. 5.2-3. The
nominal power performance of the device is: $P_{\text{out}} = 28$ dBm, PAE = 55% at 8.5 GHz with the biasing point of $V_{ds} = 8$ V, $I_{ds} = 50$ mA. We also know that the drain saturation current $I_{DSS} = -0.3$ A, the gate pinch-off voltage $V_{\text{pinch}} = -1.9$ V, the knee voltage $V_{\text{knee}} = -1.5$ V and the gate-drain breakdown voltage $V_{br} = -22$ V. Due to the unavailability of the large-signal model for this particular device, we have to use the Cripps method discussed in section 2.7 to calculate the optimum load for achieving the maximum output power. Since we will normally bias the device at 8 – 9 V for stability concern, the optimum load should be: $R_{\text{opt}} = 2(V_{DS} - V_{\text{knee}}) / I_{DSS} = 2(9-1.5) / 0.3 = 50$ Ω.

Fig. 5.2 -2: Small-signal model for the 1.2mm wide GaAs PHEMT.
(a) $V_D = 0.0 \text{ to } -2.25$
$0.25 \text{ V steps}$
$T_a = 25 \degree C$

(b) $f = 8.5 \text{ GHz}$
$V_D = 8.0 \text{ V}$
$I_D = 50 \text{ mA}$
$T_a = 25 \degree C$
Fig. 5.2-3: (a) DC IV curve of the 1.2mm wide GaAs HFET. (b) Output power sweep at 8.5 GHz. (c) Power added efficiency vs. input power at 8.5 GHz. (d) Gain vs. input power at 8.5 GHz.
2) GaAs modified TWPA design:

Based on the small signal model and calculated optimum load $R_{opt}$ for the GaAs HFETs, we were able to design the circuit as shown in Fig. 5.2 - 4. As you can see from the schematic, four cells of devices were employed, number 4 was chosen for optimum power combining efficiency and to avoid extremely high-impedance transmission lines.

We first set our design goals as followings: Brag frequency of the artificial gate line $f_{brag}$ would be about 20 GHz, the impedance of the artificial gate line $Z_0$ would be about 50 $\Omega$ and the impedance of the gate transmission line itself $Z_{gate}$ is 100 $\Omega$ that will be used to implement the inductors. Then based on the three equations discussed earlier in this section, that is:

$$Z_0 = \sqrt{\frac{L_{gate}}{C_{gate}}}, f_{brag} = \frac{1}{\pi \sqrt{L_{gate} \times C_{gate}}}, C_{gate} = \frac{C_{in} \times C_{gs}}{C_{in} + C_{gs}},$$

we can derive the values for $L_{gate}$, $C_{gate}$ and $C_{in}$. Then we can obtain the physical length of the transmission lines based on the following equations for calculating the inductance of transmission lines:

$$L = Z_0 \times \tau = Z_0 \times \frac{1}{f} \times \frac{l}{\lambda} = Z_0 \times \frac{l \times \sqrt{\varepsilon_{eff}}}{f \times \lambda_0} = Z_0 \times \frac{l \times \sqrt{\varepsilon_{eff}}}{\nu_o} = Z_0 \times \frac{l \times (\varepsilon_r + 1) / 2}{\nu_o}$$

where $Z_0$ is the impedance of the transmission line itself, $\tau$ is the signal transmit time, $l$ is the physical length of the transmission line, $\lambda$ is the wavelength of the signal, $\lambda_0$ is the wavelength of the signal in the vacuum, $\nu_0$ is the wave propagation velocity in the vacuum that is $3 \times 10^{11}$ mm/s, $\varepsilon_r$ is the relative dielectric constant of the medium and $\varepsilon_{eff}$ is the effective relative dielectric constant of the medium. The relative dielectric constant of AlN substrate is 8.5. It turned out to be: with the unit gate line length of 2mm (1mm at the start and end
of the gate line) and $C_{in}$ of 0.4 pF that requires gate series capacitor value of 0.6 pF, we can achieve Bragg frequency $f_{bragg} = 19$ GHz and the artificial gate line impedance $Z_0 = 42$ $\Omega$. Also, in the design, the gate and drain transmission line has the following parameters: $W_{gate} = 0.025$ mm, $G_{gate} = 0.1$ mm, $W_{drain} = 0.12$ mm, $G_{drain} = 0.08$ mm and the gate bias resistor $R_{bias} = 1000$ $\Omega$.

![Diagram of GaAs modified TWPA](image)

**Fig. 5.2-4: Schematic of the GaAs modified TWPA.**

Second, we decided to vary the capacitances along the gate line to achieve better equal drive for each device, and then drain delay lines with different lengths at the output of the devices were needed to achieve phase equalization of the signals. The impedance of those lines was 50 $\Omega$, that was the optimum load of the GaAs HFETs as we discussed earlier. CAD software essof/HP was then used to optimize the component values, $C_1 = 0.6$ pF, $C_2 = 0.8$ pF, $C_3 = 1.2$ pF, $C_4 = 1.5$ pF.
were the new capacitance values for those gate capacitors, and \( L_{\text{drain1}} = 21.5\text{mm} \), \( L_{\text{drain2}} = 17\text{ mm} \), \( L_{\text{drain3}} = 11\text{ mm} \), \( L_{\text{drain4}} = 4.2\text{ mm} \) were the final values for the lengths of the drain delay lines. Fig. 5.2-5 shows the amplitude of the signals at the input test points (shown in Fig. 5.2-4) and Fig. 5.2-6 (a) and (b) shows the phases of the signals at the input and output test points (shown in Fig. 5.2-4) respectively. Those graphs illustrated that the signal amplitude at the input of the four devices have been equalized to some extent and the phase of the output signals after the drain delay lines (before the corporate combiner) have almost been totally equalized. But we have to point out that the equalization of the input drive amplitude can only be realized at one frequency since the loss along the gate line is frequency dependent, in this design, we chose the mid-band of the desired bandwidth, which is about 5 GHz. Unfortunately, this unequal drive problem could have serious effect for the efficiency and stability of the GaN power amplifiers, which we will discuss in details later in this chapter.

![Graph](image)

**Fig. 5.2-5:** The amplitude of signals (in V) at the input of the four devices is equalized at 5 GHz only (at the input test points shown in Fig. 5.2-4).
Fig. 5.2-6: Phase of the four signals (in degrees) (a) before (at the input test points shown in Fig. 5.2-4) and (b) after the phase equalization of the drain delay lines (at the output test points shown in Fig. 5.2-4).
Next the corporate power combiner was designed, which would provide both the power combining and impedance transformation functions. Since the optimum load for each 1.2-mm GaAs HFET was 50 \( \Omega \) and we needed to combine four branches of signals, the impedance transformation was actually from 50 \( \Omega \) to 200 \( \Omega \). In order to achieve the bandwidth of 1-8 GHz, five sections of \( \lambda/4 \) transformers were used. Following the design method discussed in Section 3.5 and after the optimization of essof/HP software, the impedance values of five sections were: \( Z_1 = 63 \Omega, Z_2 = 80 \Omega, Z_3 = 52 \Omega \) (actually 104 \( \Omega \)), \( Z_4 = 63 \Omega \) (actually 126\( \Omega \)), \( Z_5 = 78 \Omega \) (actually 156 \( \Omega \)). The schematic of the combiner was shown in Fig. 5.2-7 and the frequency response was shown in Fig. 5.2-8. It
illustrated that the loss of the combiner was less than 1 dB, which corresponded to 80% combining efficiency, from 1.5 GHz to 15 GHz. The ripple was less than 0.2 dBm for the whole bandwidth. This result was obtained under the assumption that the combined signals had the same magnitude and phase. As was mentioned in section 3.5, those conditions are also required to be satisfied before we can safely eliminate the Wilkinson termination resistors in the combiner design.

Fig. 5.2-9 shows the simulation of the small-signal performance of the amplifier. $S_{21}$ is about 8 dB with the 3dB bandwidth of 0.5 – 10 GHz, both $S_{11}$ and $S_{22}$ are less than $-10$ dB for most of the bandwidth and $S_{12}$ is always less than $-20$ dB. Large-signal simulation is not available due to the lack of the large-signal model, but we predict the achievable output power would be about 33 dBm (2 Watts) with the PAE of about 35%, based on the power performance of the GaAs HFETs and the combining efficiency of the corporate combiner.

![Graph](image.png)

**Fig. 5.2-8:** Frequency response of the corporate power combiner.
Fig. 5.2-9: Simulation of small-signal performance.

3) Hybrid circuit implementation:

Fig. 5.2-10: Layout of the CPW layer for GaAs hybrid modified TWPA.

The GaAs amplifier was first implemented with the conventional hybrid circuit approach. The hybrid circuits were constructed on a 0.25mm thick AlN ($\varepsilon_r = 8.5$) substrate with a 2.5-μm-thick top gold layer. The CPW layout was shown in Fig.
5.2-10, the circuit dimension was 29 mm x 20 mm; air-bridges and device connections were made with 1 mil bond wire. The surface mount components were fastened with conducting Epoxy, which required 175°C for solidification. The gate termination resistor was 50 Ω and the gate bias resistors were 1000 Ω. The gate series capacitors were surface mount capacitors for the applications up to 20 GHz; their values were within 10% of the designed values.

Fig. 5.2-11: Picture of part of the hybrid GaAs modified TWPA and the close-up look of the active device area.

Fig. 5.2-11 shows the picture of the input gate-line and gate biasing part of the finished amplifier and the close-up look of the active device area. Fig. 5.2-12 and Fig. 5.2-13 show the measured small-signal and large-signal performance of the amplifier respectively. We can see that the small-signal gain varied from 6 dB to 9 dB with the bandwidth of 1 - 5 GHz and the output power ranged from 1 W to 2 W within the bandwidth with PAE of about 20 - 30 %. The major reason for the reduced bandwidth was the parasitics introduced by the bonding wire connections, especially the wires from the gate-line to the device gate electrode.
The circuit simulation had proved this conclusion by intentionally introducing the inductors to model those bonding wires. In order to achieve broader bandwidth and have better control of the circuit performance, it shows that the flip-chip integrated circuit implementation has to be developed.

![Graph showing gain vs frequency](image)

Fig. 5.2-12: Small-signal performance of the GaAs hybrid modified TWPA.

![Graph showing power output and efficiency vs frequency](image)

Fig. 5.2-13: Power performance of the GaAs hybrid modified TWPA.
4) Flip-chip integrated circuit (IC) implementation:

We developed the flip-chip integrated circuit (FC-IC) process on AlN circuit board, in which the active-device was flipped and bonded onto the board using the flip-chip bonder, resistors were fabricated by growing NiCr onto the board, capacitors were metal-insulator-metal (MIM) capacitors with the SiN as the insulating layer, and air-bridges were made by evaporating thick Au bond pads onto the board. The detailed process will be described in Appendix B. Fig. 5.2-14(a) was the layout of the GaAs FC-IC; the circuit size was 29mm × 19mm. Fig. 5.2-14(b) shows the detailed layout of the device gate region, which clearly illustrated the arrangement of the gate series capacitor as well as the gate biasing. Au bond pads would then be connected to the corresponded gate, drain and source electrodes of the flipped GaAs HFETs.
Fig. 5.2-14: (a) Layout of the GaAs modified TWPA. (b) Layout of the device gate area.

3) GaAs modified TWPA performance:

Fig. 5.2-15 shows the picture of the input part of the finished amplifier. As we can see, the GaAs devices have been flip-chip bonded onto the circuit board. The backside via holes of the devices can be seen from the picture. Fig. 5.2-16 and Fig. 5.2-17 show the measured small-signal and large-signal performance of the amplifier respectively. We have obtained $S_{21}$ of about 7 dB with 3-dB bandwidth of 1-8 GHz, $S_{11}$ and $S_{22}$ were less than $-10$ dB for most of the bandwidth and $S_{12}$ was less than $-25$ dB. But the output power was less than we expected, it was between 28 dBm (0.6W) and 31 dBm (1.3W). We believe that the main reason is that the poor thermal management of the device, since it was designed to be attached to the circuit board facing up instead of being flip-chipped. The device
couldn't achieve its maximum output power when its operating temperature is too high.

Fig. 5.2-15: Picture of the finished GaAs TWPA (only input part of the whole circuit).

Fig. 5.2-16: Small-signal performance of the GaAs TWPA.
Fig. 5.2-17: Large-signal performance of the GaAs TWPA.

From the design and fabrication of the GaAs modified TWPA, we have demonstrated that the modified TWPA can achieve higher efficiency than the conventional TWA over a broad bandwidth, also in order to improve the bandwidth the FC-IC instead of the hybrid circuit has to be implemented. Next, we are going to incorporate GaN HEMTs technology to further investigate the modified TWPA topology as a broadband power amplifier.

5.3 GaN modified TWPA

In this amplifier, we utilized four 0.75-mm wide, 0.7-μm long GaN HEMTs since we found out that it’s difficult to achieve the bandwidth of 1 – 8 GHz we desired using larger size devices.

1) 0.75-mm wide 0.7 μm long GaN HEMTs:
Fig. 5.3-1 shows the layout of the 0.75- mm wide 0.7- μm long device. It had typical inter-digital multi-finger structure. The actual gate length in the mask
design was 0.5 μm, which would produce about 0.7-μm-long gate finger using the Stepper lithography. The gate-to-source spacing \( L_{\text{GS}} \) was 0.7 μm and the gate-to-drain spacing \( L_{\text{GD}} \) was 1 μm. The gate-to-gate spacing was about 47 μm and the unit gate finger length was 75μm. On the expanded gate, drain and source electrodes, thick gold bonding pads would be deposited using E-beam evaporation. The connection of the source electrodes would be made through the corresponded source bonding pads on the circuit board.

![Bonding Pads Diagram](image)

**Fig. 5.3 - 1: Layout of the 0.75- mm wide 0.7- μm long device. \( L_{\text{GS}} = 0.7 \, \mu \text{m}, L_{\text{GD}} = 1 \, \mu \text{m}, L_g = 0.5 \, \mu \text{m} \) in the layout**

The small-signal model of the GaN HEMTs had the typical structure for a FET, as was shown in Fig. 5.3-2. The parameters were: \( g_m = 220 \, \text{mS}, C_{gs} = 2 \, \text{pF}, C_{ds} = 0.01 \, \text{pF}, C_{gd} = 0.01 \, \text{pF}, R_{in} = 1.3 \, \Omega, R_{ds} = 200 \, \Omega, R_{gd} = 14 \, \Omega, R_g = 1 \, \Omega, R_d = 3.4 \, \Omega, R_s = 2.3 \, \Omega, C_{gspad} = 0.1 \, \text{pF}, C_{gdpad} = 0.1 \, \text{pF}, C_{dspad} = 0.25 \, \text{pF} \). Most of the circuit components were scalable from the parameters of the 1mm device except the pad parasitic capacitances \( C_{gspad}, C_{gdpad} \) and \( C_{dspad} \) since the layout of those
expanded electrode pads for the 0.75mm device was very similar to that of the 1mm device.

![Small-signal model for the 0.75 mm-wide 0.7 μm–long GaN HEMT.](image)

**Fig. 5.3-2**: Small-signal model for the 0.75 mm-wide 0.7 μm–long GaN HEMT.

2) GaN modified TWPA design and simulation:

The Schematic of the modified GaN TWPA is shown in Fig. 5.3-3, it was very similar to that of the modified GaAs TWPA. Gate series capacitance were: $C_1 = 0.65 \, \text{pF}$, $C_2 = 0.9 \, \text{pF}$, $C_3 = 1.15 \, \text{pF}$, $C_4 = 1 \, \text{pF}$, the physical gate-line length of each section $L_{\text{gate}} = 1.65 \, \text{mm}$, the length of drain delay lines were: $L_{\text{drain1}} = 16.5 \, \text{mm}$, $L_{\text{drain2}} = 12.5 \, \text{mm}$, $L_{\text{drain3}} = 8.25 \, \text{mm}$, $L_{\text{drain4}} = 2.1 \, \text{mm}$, the gate transmission line impedance $Z_g = 100 \, \Omega$ and its width $W_{\text{gate}} = 0.028 \, \text{mm}$, gap $G_{\text{gate}} = 0.135 \, \text{mm}$, the drain delay line impedance $Z_d = 50 \, \Omega$ and its width $W_{\text{drain}} = 0.12 \, \text{mm}$, gap $G_{\text{drain}} = 0.06 \, \text{mm}$, the gate termination resistor $R_{\text{gate}} = 50 \, \Omega$ and the gate biasing resistor $R_{\text{bias}} = 1000 \, \Omega$. 
The circuit was designed to operate over a 1-9 GHz frequency range, with a flat transducer gain ($S_{21}$) of 8 dB, based on HPEEs of simulation using small signal model shown in Fig. 5.3-2. Since GaN based HEMT technology is still at a relatively immature stage, an accurate large signal model was unavailable at the time of design. DC-IV characteristics were employed to predict the optimum load for maxim output power of the devices used. The optimum load seen by the device was chosen to maximize the product of current and voltage swing available from the device. GaN HEMTs with 0.75 mm gate periphery were utilized which had a 50 $\Omega$ optimum load with the assumption that $I_{\text{max}} = 600$mA, $V_{\text{knee}} = 5$ V and a bias point of $V_d = 20$ V. The corporate power combiner, comprised of 5 sections of quarter-wavelength CPW transmission lines, was designed to transform $N \times Z_0$ (N is the number of the devices, $Z_0 (= 50 \, \Omega)$ was the amplifier load) to the optimum load to be seen by the device (50 $\Omega$). The number $N = 4$ was chosen for optimum power combining efficiency and to avoid extremely
high-impedance transmission lines. The CPW transmission lines were meandered to reduce the circuit size.

In the corporate combiner, the impedances of each λ/4 transformer were: \( Z_1 = 68 \ \Omega \), \( Z_2 = 82 \ \Omega \), \( Z_3 = 55 \ \Omega \), \( Z_4 = 67 \ \Omega \), \( Z_5 = 88 \ \Omega \), with their respective CPW line width and gap size as: \( w_1 = 0.12 \ \text{mm} \), \( g_1 = 0.145 \ \text{mm} \), \( w_2 = 0.05 \ \text{mm} \), \( g_2 = 0.11 \ \text{mm} \), \( w_3 = 0.146 \ \text{mm} \), \( g_3 = 0.08 \ \text{mm} \), \( w_4 = 0.094 \ \text{mm} \), \( g_4 = 0.1 \ \text{mm} \), \( w_5 = 0.08 \ \text{mm} \), \( g_5 = 0.25 \ \text{mm} \). The frequency response of the combiner is shown in Fig. 5.3-4 by using the harmonic balance test bench shown in Fig. 3.5-9. We can see that the combining loss was less than 1 dB (combing efficiency of 80%) over the frequency range from 1-13 GHz.

![Graph](image)

Fig. 5.3-4: Frequency response of the corporate power combiner for GaN modified TWPA.

The layout of the GaN modified TWPA is shown in Fig. 5.3-5. The drain delay lines and the λ/4 transformers of the corporate combiner were meandered to reduce the circuit size. Fig. 5.3-6 shows the simulation of \( S_{21} \) and \( S_{21} \), over the bandwidth of 1-9 GHz, \( S_{21} \) was about 8 dB and \( S_{12} \) was less than –20 dB. Fig.
5.3-7 shows the simulation of $S_{11}$ and $S_{22}$, they were less than $-10$ dB for most of the bandwidth.

Fig. 5.3 - 5: Layout of the GaN modified TWPA.

Fig. 5.3- 6: Simulation of $S_{21}$ and $S_{12}$ of the GaN modified TWPA.
Fig. 5.3- 7: Simulation of $S_{11}$ and $S_{22}$ of the GaN modified TWPA.

Fig. 5.3- 8: The effect of the variation of the gate artificial line termination resistor $R_g$ on $S_{21}$.

We found out the gain termination resistor $R_g$ would affect $S_{21}$ at very low frequency, as was shown in Fig. 5.3-8, $S_{21}$ with $R_g$ of value 35 $\Omega$, 50 $\Omega$ and 70 $\Omega$ were quite different at frequencies from DC – 2 GHz. This simulation result is
useful since it's not uncommon that the resistance value can have 20% - 30% variations during the process. Actually, on the finished amplifier, the gate termination resistor happened to be 35 \( \Omega \) instead of the designed 50 \( \Omega \).

3) GaN modified TWPA performance:

Fig. 5.3-9 shows the finished amplifier and a close-up view of the flip-chip bonded GaN HEMT. The circuit was fabricated on a 10-mil thick polished AlN substrate (\( \varepsilon_r = 8.5 \)). The board dimensions were 17mm \( \times \) 12mm. The gate bias and termination resistors were made of NiCr, the gate capacitors were MIM capacitors with 0.3 \( \mu \)m thick Si\(_3\)N\(_4\) as the insulating dielectric medium. Air-bridges were used for the CPW grounding and crossover connections. Three-\( \mu \)m-thick gold bumps were evaporated for bonding the flipped GaN HEMTs.

![Fig. 5.3-9: (a) Picture of the GaN modified TWPA (part of the whole circuit), (b) The close-up look of the flipped GaN HEMT.](image)

Stepper lithography was used to fabricate the GaN HEMTs with a 0.75 \( \mu \)m gate-length and 750 \( \mu \)m total gate-width, each gate finger length was 75 \( \mu \)m, and gate-to-gate spacing was about 50 \( \mu \)m. Two-\( \mu \)m-thick gold bumps were evaporated on
gate, source and drain electrodes for flip-chip bonding. The detailed process of the device was described in Appendix A. The wafer was subsequently diced into 1.2mm × 1mm discrete devices. Then a flip-chip bonder was used to align and bond 4 devices onto the AlN circuit board with all the passive circuit components pre-fabricated.

Fig. 5.3-10: Small-signal performance of the GaN modified TWPA.

Fig. 5.3-10 shows the measured S-parameters of the amplifier with the simulated $S_{21}$ for comparison. The gain was around 7 dB with a 3-dB bandwidth of 1-8 GHz. The input and output return losses ($S_{11}$ and $S_{22}$) were less than 15 dB. Fig. 5.3-11 shows the output power vs. frequency, the amplifier was biased in class AB mode with a drain voltage of 18 V; output power was measured at about 4 dB gain compression point. The mid-band output power was 3.6 W. Fig. 5.3-12 was the power measurement at 4 GHz, biased at 22 V. Output power was increased to 4.5 W. Due to the relatively immature device technology, such as poor uniformity, the circuit could not be reliably biased above 20 V, hence the power performance over the entire bandwidth at 22 V drain bias was not available.
Fig. 5.3 11: Large signal performance of the GaN modified TWPA.

Fig. 5.3-12: Power sweep at 4 GHz. (V_{ds} = 22 V, I_{ds} = 200 mA)

Because of the power limitation of the driver, measurements at 1 and 2 GHz were not performed. The power added efficiency (PAE) was less than 16%, which was about half of the expected value. The reduction of PAE is believed to be caused mainly by the non-uniformity and the un-equal drive of the 4 devices, since the
combiner was designed assuming four output signals with equal amplitude and phase. The DC resistive loss in signal lines and RF loss at corners of the meandering CPW lines may also contribute to the lowered PAE especially at higher frequencies.

5.4 GaN LCR-matched power amplifier.

1) GaN LCR-matched power amplifier design

![Diagram](image)

Fig. 5.4-1: Schematic of the 3-10 GHz LCR-matched GaN PA.

In this second generation GaN amplifier design, we intend to circumvent the uneven drive problem of modified TWPA by employing novel LCR-matched broadband power amplifier. The circuit schematic is shown in Fig. 5.4-1. A multistage corporate power divider and combiner were employed at the input and output, respectively, to distribute and collect power from these devices. The
combiner networks used multiple-stage quarter-wave transformers to realize the desired bandwidth and impedance transformation ratios (the desired impedance levels are naturally different at input and output). Input and output LC match networks were then utilized to reactively compensate for $C_{gs}$ and $C_{ds}$, and provide additional impedance transformation. A lossy LCR network (similar to what was reported in\textsuperscript{5} for a single-device amplifier) was used at the input of the device to achieve the desired gain flatness over the designed bandwidth.

Since the overall efficiency of the amplifier is given by

$$\eta = \frac{P_{\text{out}} - P_{\text{in}}}{P_{\text{DC}}} = \frac{P_{\text{out},\text{dev}} \cdot \eta_{\text{out}} - P_{\text{out},\text{dev}} / (\eta_{\text{in}} \cdot G)}{P_{\text{DC}}} = \frac{P_{\text{out},\text{dev}}}{P_{\text{DC}}} \left( \eta_{\text{out}} - \frac{1}{\eta_{\text{in}} \cdot G} \right)$$

where $\eta_{\text{in}}$ and $\eta_{\text{out}}$ are the combining efficiency of the input corporate power divider and output combiner respectively, $G$ is the gain of the device, $P_{\text{out},\text{dev}}$ is the output power right at the output of the device. This equation shows that the efficiency is directly proportional to $\eta_{\text{out}}$ if $G$ is large enough to offset the loss of the input matching networks. We used two sections of quarter-wave transmission lines at the output to realize impedance transformation while keeping the conduction loss as low as possible; at the input, four sections of quarter-wave transmission lines need to be employed to be able to realize the high-ratio impedance transformation. The CPW transmission lines were meandered to reduce the circuit size. As was shown in Fig. 5.4-2, the frequency response of the input and output corporate divider/combiner was plotted.
Fig. 5.4-2: (a) Frequency response of the input corporate power divider, (b) Frequency response of the output corporate power combiner.

Multi-section lowpass LC networks were used at both the input and output of the device to absorb the reactance and provide additional impedance transformation as well (Fig. 5.4-1). We can extend the bandwidth by transforming resistance step by step, using virtual R as intervals for calculating values of L and C of each LC section. At the input of each device, a lossy LCR matching network was employed to provide the gain compensation, thus achieve the gain flatness over
the whole bandwidth, also \( R \) served as a stabilizing resistor. Fig. 5.4-3 shows that LCR matching network was aimed to eliminate gain peaks at low band and has little effect at high band. Its frequency response \( (S_{21}) \) was plotted with parameters values as: \( L = 0.9 \) nH, \( C = 0.65 \) pF, \( R = 30 \) \( \Omega \).

Fig. 5.4-3: (a) LCR matching network structure, (b) \( S_{21} \) of the LCR network.

If we arrange the LCR network and the gate-source capacitance \( C_{gs} \) of the device as shown in Fig. 5.4-4, it turns out that the combination of the multi-section LC networks, LCR network and the \( C_{gs} \) is actually the equivalent of an transmission
line. This finding shows that those bandwidth-extending techniques have some similarity with the distributed amplifier design that could achieve the highest possible gain-bandwidth product. Since the values of L can C are not the same, they are not the real transmission line structure. However this similarity still can give us some insights about the broadband network design.

![Graph showing Simulation of the small-signal performance of the GaN LCR-matched PA.](image)

Fig. 5.4-5: Simulation of the small-signal performance of the GaN LCR-matched PA.

The circuit was designed to operate over 3-11.5 GHz frequency range using HPEssof simulation (as shown in Fig. 5.4-5), with small-signal gain $S_{21}$ of 7-10 dB, input match $S_{11}$ of $-10$ dB and output match $S_{22}$ of $-6$ dB. The GaN HEMTs used in this work were fabricated using Stepper lithography with a 0.7 μm gate-length, each gate finger length is 75 μm, and gate-to-gate spacing is about 50 μm. Due to the relatively low $f_{\text{max}}$ and high $C_{gs}$ of these devices, it proved more challenging to achieve the desired low input reflection throughout the bandwidth with this reactively-matched circuit topology as compared with the earlier TWPA design.
2) GaN LCR-matched PA measurement results

Fig. 5.4-6 shows the photo of the finished amplifier, the circuit dimension was 12mm × 8mm. Due to difficulties in flip-chip mounting a large-area HEMTs device, our design used two separate chips (each had two adjacent one-mm devices) for a total of 4mm of gate periphery. Fabricated on a 10-mil thick polished AlN substrate (εt=8.5), the circuit components included NiCr resistors, MIM capacitors using SiN as the insulating layer, air bridges and two flipped 2mm-wide GaN HEMTs chips. All the device and ceramic circuit processing was carried out at UCSB. Fig. 5.4-7 shows the small-signal performance of the circuit. The measured $S_{21}$ was about 7 dB with 3dB-bandwidth of 3-10 GHz. We expect to extend the bandwidth to above 11 GHz by shrinking the gate length to 0.6 μm. The measured $S_{11}$ was less than −8 dB, and $S_{22}$ was less than −5 dB over the band which was expected since output is not conjugated-matched; both were close to the simulation mentioned earlier.

Power measurement over the whole bandwidth is shown in Fig. 5.4-8, the highest power of 8 W (CW) was obtained at 9.5 GHz, and the lowest power of 4.5 W (CW) at 4.5 GHz were measured. The circuit was biased at 24 V, $I_d$ =500mA (about 15 % of $I_{max}$), in class AB mode. PAE ranged from 5% at 4.5 GHz to about 20% at 9.5 GHz. Fig. 5.4-9 shows the power sweep at 8 GHz, the output power of 8.5 W with PAE of about 20% was obtained; the associated gain was only 3 dB. Although the PAE has been improved by about 5% - 10% compared with the previous TWPA, it was still about half of the expected value. One of the reasons was that the large-signal gain was only about 3 ~ 4 dB since the gain compression was almost 4 dB when devices were driven into saturation, which causes much lower PAE than the drain efficiency. Also, non-uniformity of each
device due to the non-uniformity of the wafer, the DC resistive loss in signal lines and RF loss at corners of the meandering CPW lines may also contribute to the lowered PAE.

Fig. 5.4-6: Photo of the finished GaN LCR-matched PA.

Fig. 5.4-7: Small-signal performance of the GaN LCR-matched PA.
Fig. 5.4-8. Power measurement from 4.5 GHz to 10 GHz. (V_{ds} = 24 V, I_{ds} = 500 mA)

Fig. 5.4-9. Power sweep at 8 GHz. (V_{ds} = 16 V, I_{ds} = 500 mA)
5.5 Modified 2×2 matrix TWPA.

Fig. 5.5-1: Schematic of the 2×2 modified TWPA.

Based on the previous work, a 2×2 matrix modified traveling wave amplifier combined with broadband power divider/combiner was designed. Fig. 5.5-1 is the schematic of the circuit. A Wilkinson power divider (without isolation resistor) and two sections of LC networks were utilized for the broadband input matching circuitry, as a result, the input power was equally divided into two artificial gate lines (each had a characteristic impedance $Z_0$ of 15 Ω), each employing two 1-mm devices. The artificial drain line was replaced by drain delay lines to equalize the phase of output signal of the devices, then three sections of LC networks and two sections of Wilkinson power combiner (corporate power combiner) were employed to achieve a broadband impedance
transformation and power combining. This $2 \times 2$ matrix system is preferred over the previous $1 \times 4$ system (single gate line feeding four 1-mm devices) for the following reason: the attenuation on the gate line due to the period loading of the device is approximately $R_g \omega^2 C_{gs} Z_0$ ($R_g$ is the input resistance of the device, $C_{gs}$ is the effective input capacitance of the device, $Z_0$ is the characteristic impedance of the artificial gate line ($Z_0 = 15 \Omega$)), this loss causes the devices to be driven unevenly, which can seriously degrade the power performance of the devices since some devices will be overdriven in order to drive all the devices into saturation. Due to the immaturity of the GaN device technology, overdrive tends to have non-recoverable damage to the device performance. Also unequal drive will greatly reduce the combining efficiency of the corporate combiner. Capacitive division $^{[4]}$ can alleviate the problem to some extent but only at one frequency since this loss is frequency dependent. With only 2 devices loaded on each gate line for $2 \times 2$ system, and $Z_0$ was chosen as 15 $\Omega$ instead of 50 $\Omega$, the uneven drive problem was less severe. Another design concern is the broadband matching network. In order to achieve a 6:1 bandwidth for impedance transform ratio of six, multi-section LC networks and multi-section Wilkinson combiners were employed at the input and output. Though LC matching networks tend to be lossy due to the use of inductors, realized by high impedance transmission lines ($Z_0 = 100 \Omega$), the LC matching network is more compact than the Wilkinson divider/combiner, as a result, proper combination of LC networks and Wilkinson divider/combiner was chosen for optimum performance.
Fig. 5.5-2 (a) Simulated small-signal performance. (b) Simulated large-signal performance.

Fig. 5.5-2 (a) and Fig. 5.5- (b) show the simulation results of the small-signal and large-signal performance of the amplifier respectively. A modified curtice3 model, which was described in details in Chapter 2 was utilized for the simulation. With this model, the limitation of the large-signal operation of the device was defined by the boundary of the DC I-V curve. This is a first-order simulation because the approximations such as linear $g_m$, zero output conductance are assumed. Also, without the modeling of the thermal affect and dispersion phenomena, this model can only be used as a guide for the large-signal design of the amplifier. As a result, load-pull data was used to design the input and output broadband matching networks for the GaN HEMTs before the power performance simulation with the Curtice-3 model. Small-signal gain of 11 dB with a 3-dB bandwidth of 1-6 GHz has been obtained; large-signal simulation shows that output power up to 8 W/39dBm with PAE of 15-20% can be achieved.
Fig. 5.5-3: Picture of the finished amplifier. (the flipped GaN HEMTS not included)

The picture of the finished amplifier was shown in Fig. 5.5-3 (the flipped devices were not included), the circuit size was 10 mm × 8mm and it was fabricated on a 20-mil thick AlN ($\varepsilon_r = 8.5$) board. With all the passive circuit components such as resistors, capacitors, air-bridges built onto the AlN board and the GaN HEMTs flip-chip bonded, the amplifier was actually a flip-chip integrated circuit (FC-IC) that can achieve comparable performance as microwave and millimeter wave integrated circuit (MMIC). The measured small-signal and large-signal performances were shown in Fig. 5.5-4 (a) and Fig. 5.5-4 (b) respectively. The gain $S_{21}$ varied from 9 dB –12 dB with a 3dB bandwidth of 1.8 –6.8 GHz, and the input match $S_{11}$ was only –4 dB at mid-band. We believe that the gain ripple and poor input match were mainly due to the variation of the capacitance during the circuit process, which can be easily improved by better process control. The output match $S_{22}$ was from –4 dB to –16 dB over the whole bandwidth, which was normal for the power amplifier since the output of the device was matched to the optimum load instead of its conjugate load. The isolation $S_{12}$ was less than –20dB
for most of the bandwidth. The power performance was obtained under the drain bias voltage $V_{ds} = 18$ V and the drain bias current $I_{ds} = 400$ mA, the output power varied from 5 W to 7.5 W and the PAE was from 18% - 25% over the frequency range of 2.5 - 6 GHz. Compared with the previous GaN broadband power amplifier, the PAE was much improved and much more uniform over the whole bandwidth.

![Image](a)

![Image](b)

Fig. 5.5-4 (a) Measured small-signal performance of the amplifier, ($V_{ds} = 10$ V, $I_{ds} = 400$ mA) (b) Measured large-signal performance of the amplifier, ($V_{ds} = 18$ V, $I_{ds} = 400$ mA)
REFERENCE


Chapter 6 Conclusions

6.1 Achievements

We have demonstrated 4.6 Watt/mm power density at 8 GHz with GaN HEMTs on sapphire substrate. The tested device was 100μm wide and the gate length was 0.7 μm. Short circuit current gain cutoff frequency \( f_t \) was 20 GHz and power gain cutoff frequency \( f_{\text{max}} \) was 38 GHz. This power density is the highest for un-passivated GaN HEMTs grown on a sapphire substrate. We have also demonstrated 6 Watt/mm power density at 6 GHz with GaN HEMTs grown on SiC substrate. The improvement of the power density is believed to be mainly due to the higher thermal conductivity of SiC over sapphire. In this work, our efforts were focused on the development of GaN HMETs on sapphire substrate for their lower cost and availability. By paying close attention to the thermal management, GaN-HEMT-on-Sapphire can still have competitive performance. In particular, adequate thermal management can be achieved through flip-chip bonding of the device onto a thermally-conductive, electrically-insulating substrate, such as ceramic Aluminum Nitride, which also hosts all the passive circuit components in the later power amplifier fabrication. Using this technique, we have achieved record high single-device results of 4.4 W output power (CW) and 35% PAE at 8 GHz out of a one-mm-wide GaN HEMT on sapphire substrate. We also demonstrated 6.4 W output power (CW) and 30% PAE at 6 GHz using a two-mm-wide discreet GaN HEMTs. Step lithography was utilized to fabricate GaN HEMTs with gate length of 0.7 μm. The process includes ohmic contacts, gate-pad SiO₂ isolation, Schottky contacts, mesa isolation and bonding pads. The wafer will be diced up into discreet large-periphery devices, which will then be flip-chip mounted onto the pre-fabricated circuit board.
Based on the large-gate periphery GaN power device technology, we developed flip-chip integrated circuit (FC-IC) on AlN substrate process for the realization of GaN power amplifiers. The process was able to realize thick-metal deposition that is crucial for the very high power amplifiers. This FC-IC technology can achieve comparable performance as microwave and millimeter wave integrated circuit (MMIC) by minimizing the parasitics associated with the bonding wires in the conventional hybrid circuits. Since backside processing is not available for AlN substrate, all the amplifiers designed employ coplanar wave-guide (CPW) transmission line system. Through a total of seven steps (six masks layers): 1) Si protection layer 2) Resistor 3) CPW 4) Capacitors 5) SiN etch 6) Bonding pads 7) Air-bridges, all the passive circuit components such as CPW lines, NiCr resistors, MIM capacitors (SiN as the insulating medium) and air-bridges were fabricated onto the AlN substrate.

The first GaN-based amplifier was successful designed and fabricated. A modified traveling wave power amplifier topology was introduced, in order to achieve improved efficiency as well as output power over a wide bandwidth. It was realized by replacing the artificial drain line with the drain delay lines (which were to equalize the phases of the combined signals) and a more efficient corporate power combiner. By using four 750-μm GaN HEMTs, the amplifier had about 7dB small-signal gain with 1-8 GHz 3-dB bandwidth and very low return loss at both input and output. An output power of 3.6 W when biased at 18 V and 4.5 W when biased at 22 V was obtained at mid-band (4 GHz). This is the highest output power for a TWPA using solid-state field-effect-transistors (FETs). However, the frequency-dependent loss on the input artificial transmission line made it difficult to uniformly drive each device over the entire operating bandwidth, even using non-uniform capacitive division along the line. Such
uneven input drive will reduce the efficiency and reliability of the circuit, especially at higher frequency of the band. Combined with the non-uniformity of the devices themselves, the PAE of the amplifier was only 5-15 % over the whole bandwidth, about a half of what we expected.

In the second generation of GaN broadband power amplifier, we circumvent the unequal drive problem of TWPA by employing novel LCR matching networks, multi-section LC networks and multi-section quarter-wave transmission lines to realize the input broadband matching. The broadband corporate power combiner was also employed at the output. However it proved more challenging to achieve the desired low input reflection throughout the bandwidth with this reactively-matched circuit topology as compared with the earlier TWPA design. The circuit had about 7dB small-signal gain with 3-dB bandwidth of 3-10 GHz, the input return loss was about –8 dB and output return loss was about –5 dB. At 8 GHz, the circuit was able to generate about 8.5 W output power with 20% PAE using total of 4mm-wide devices that corresponds to more than 2 W/mm output power density. Although lower than 4.4 W/mm results using 1mm-wide devices, which is due to the increased complexity in thermal management and matching network as the device size increases, this power density is nonetheless about twice as high as what GaAs-based counterparts with the same size devices can achieve.

The third generation GaN amplifier was designed to achieve a balanced performance between the bandwidth (1- 6 GHz) and efficiency (~ 20%). A modified traveling wave amplifier combined with the broadband power divider/combiner was our solution. With the 2 × 2 input artificial gate feeding line topology, the unequal drive problem associated with the previous GaN modified TWPA has been greatly reduced. The much-improved PAE ranging from 18 % to 24 % with the output power up to 7.5 W over the bandwidth of 1.8 GHz – 6.8 GHz
has been achieved. We believe that the reduced bandwidth was mainly due to the variation of the capacitance during the circuit process, which can be improved by better process control.

6.2 Future work

1) Optimization of GaN HEMTs:

In terms of power density, GaN HEMTs, grown on sapphire substrate, have demonstrated at least 4 times that of their GaAs-based counterpart with small gate-periphery devices. By introducing the flip-chip bonding technique, large-periphery GaN HMETs-on-sapphire have also achieved 2 to 4 times as high as what their GaAs-based counterparts can achieve. However, the drain efficiency was still relatively low, normally below 50% for small devices and below 40% for large-periphery devices. The main reason was due to the immaturity of the GaN HEMTs material system. RF dispersion caused by traps within the GaN epi structure would have serious effects on the device RF performance, such as reduced maximum drain saturation current, increased knee voltage and high gain compression when the device was biased up to certain drain voltage. Normally, the gain compression was about 7 dB, sometimes even up to 10 dB, when the output power reached the saturation level. This further reduced the power added efficiency (PAE), which is closely related with the large-signal gain. The highest PAE of all the GaN broadband power amplifiers we've fabricated was about 25 %, therefore the combining efficiency of the circuit was over 70% since the PAE of each discrete device was only 35%. The 70% combining efficiency was already very close to the simulated 80% value. In order to further improve the PAE of the GaN broadband power amplifier, the most effective way would be to improve the PAE of the GaN HEMTs. By optimizing the growth conditions and
device process techniques, we can reduce the dispersion and therefore improve the PAE as well as the output power of the GaN HEMTs. Various experiments such as low temperature growth, passivation study etc. are being conducted and progress has been made.

The uniformity of the wafer is very important for the realization of GaN amplifiers. So far, the usable area is normally less than 50%, sometimes only 30-40% of the whole wafer. This has been a serious problem for the fabrication of large-periphery devices (especially when the device size is larger than two mm) and amplifiers that need to bond more than one chip. In general, the yield of GaN amplifiers would be seriously reduced. New MOCVD machine with new chamber structure has been installed and the uniformity would be greatly improved in the near future.

Deep submicron gate process is needed to achieve high $f_t$ and $f_{max}$, and therefore broad bandwidth and higher gain. E-beam writer can obtain gate length of less than 0.25 μm, which should be able to achieve $f_t$ and $f_{max}$ above 40 GHz and 80 GHz respectively. Then we are able to produce GaN power amplifiers with up to 20 GHz operating frequency range.

We found out that it’s extremely hard to flip-chip bond devices large than 4mm due to the limitation of the flip-chip bonder. We cannot achieve good contacts at the same time with too many bonding pads during the bonding period. Reducing the size of the bonding pads can improve the bonding condition, but at the expense of the increased thermal resistance of the device. For GaN HEMTs grown on SiC substrate, we can reduce the gate pitch size by shrinking the source pads without jeopardizing the thermal situation of the devices.
2) Incorporate GaN wide-band amplifiers into wide-band efficient spatially combined modules

In the previous GaN amplifiers, broadband corporate power combiners were employed to efficiently combine output powers from separate devices. However, as the number of stages of Wilkinson combiner increases, the combining efficiency will decrease significantly. But for the spatial power combining, theoretically the efficiency will be independent of the number of the amplifiers. Normally, as the combined stage is greater than four, the spatially power combining would have higher efficiency than the corporate power combining. Also, the thermal property of the spatial power combiner could be very well managed by the proper test fixture design. Furthermore, the spatial power combiner could achieve broadband characteristic with the optimized passive structure design. Hence, to be able to produce broadband power sources with up to 100-Watt output power and up to 10-GHz operating frequency, the spatial power combined module incorporating GaN broadband power amplifiers seems to be very promising approach.

Fig. 5.1: (a) Picture of the spatial power combined module (including 8 cards). (b) Picture of one card (including 4 GaAs MMICs).
Nick Cheng et al. have developed an X-band spatial power combined module using standard wave-guide structure and tapered antenna array to combine GaAs MMIC power amplifiers. Fig. 5.1 (a) is the picture of the spatial power combined module. It accommodated a total of 8 cards, and each card (shown in Fig. 5.1 (b)) hosted 4 GaAs MMIC power amplifiers. This spatial combiner has demonstrated up to 150 W output power at 8 GHz, which corresponds to 70% combining efficiency. Naturally, our next step would be to incorporate the GaN power amplifiers into a spatial combiner module. Based on the previous work, we should be able to produce GaN amplifiers with about 5-W output power, 30% PAE over a bandwidth of 1-8 GHz, using a total of 2mm-wide GaN HEMTs. With the improvement of GaN high power HEMTs technology, GaN spatial power combined module could be a promising candidate for replacing traveling wave tube amplifier (TWTA) as broadband high power sources in the near future.
Appendix A
GaN HEMT process flow

1. Source Drain Ohmic Contacts (mask layer 1, dark field)

A Solvent Cleaning

1. Check the resistivity of D.I. water. It should be > 17MΩ.
2. ACE 3 minutes, ultrasonic 1 minute.
3. ISO 3 minutes, ultrasonic 1 minute.
4. Running DI 3 minutes.
5. Blow dry with N₂.
6. Dehydration bake, 120 °C, 10 minutes in petri dish without cover.

B Photoresist Application and Exposure
Safety Note: The photoresist vapor is extremely harmful. Never breathe if you put your head under the hood.

1. Cool down after dehydration, 5 minutes.
2. Put wafer on spinner chuck with vacuum on, blow with N₂.
3. Apply SPR 950-0.8 with syringe and filter to cover whole wafer.
4. Spin at 2 krpm for 30 seconds.
5. Soft bake, 90 °C for 1 minute on hot plate.
6. Apply CEM 365 with syringe and filter to cover whole wafer.
7. Spin at 3 krpm for 30 seconds.
8. Expose for 1.8 seconds. (Focus should be set according to condition of the Stepper.)

C Development

1. Post bake, 100 °C, 2 minutes on hot plate.
2. Rinse in running DI water for 30 seconds.
3. Develop in MF–701 for 2 minutes and 20 seconds.

D Oxygen Plasma Descum of Photoresist
1. Chamber pressure = 300 mtorr of O₂.
2. Power = 100 W at low frequency.
3. Run for 30 seconds.

E RIE Etch of GaN Cap Layer
1. Place GaN wafer on thick Si wafer and load into RIE #5 chamber.
2. Etch conditions: Cl₂ flow rate = 10 sccm, Chamber pressure = 5 mtorr, Power = 60 W.
3. Etch for 14 seconds. (etch about 100 Å AlGaN layer)

F Surface Preparation
1. Mix a dilute solution of HCl: H₂O :: 1:10.
2. Dip in dilute HCl for 15 seconds.
3. Rinse in D.I. water for 1 minute.
4. Blow dry with N₂.

G Evaporation
1. Place wafer on E-beam mount and level it.
2. Pump down to below 1.5 × 10⁻⁶ torr.
3. Deposit metal Ti/Al/Ni/Au of 200/2000/450/550 Å. (spitting needs to be prevented)

H Liftoff
1. Soak wafer in ACE long enough till metal comes loose.
2. Squirt with ACE squirt bottle to speed up liftoff.
3. Rinse with ISO.
4. Rinse in running D.I. water for 1 minute.
5. Blow dry with N₂.
I Annealing

1. Use thermocouple to monitor temperature.
2. Test run program GN880 couple of times to check the stability of the RTA. Place wafer on Si substrate and load them into the chamber slowly. Be cautious that wafer might slide on Si substrate during annealing. (Anneal one sample at a time is recommend.)
3. Anneal at 880 °C for 20 seconds.

2 SiO Isolation of Gate Pad (mask layer 2, light field)

A Solvent Cleaning

1. Check the resistivity of D.I. water. It should be > 17MΩ.
2. ACE 3 minutes, ultrasonic 1 minute.
3. ISO 3 minutes, ultrasonic 1 minute.
4. Running DI 3 minutes.
5. Blow dry with N₂.
6. Dehydration bake, 120 °C, 10 minutes in petri dish without cover.

B Photoresist Application and Exposure

Safety Note: The photoresist vapor is extremely harmful. Never breathe if you put your head under the hood.

1. Cool down after dehydration, 5 minutes.
2. Put wafer on spinner chuck with vacuum on, blow with N₂.
3. Apply AZ5214 with syringe and filter to cover whole wafer.
4. Spin at 4 krpm for 30 seconds.
5. Soft bake, 95 °C for 1 minute on hot plate.
6. Expose for 0.45 seconds. (Focus should be set according to the condition of the Stepper.)

C Development

1. Post bake, 110 °C, 1 minutes on hot plate.
2. Flood exposure for 1 minutes.
3. Rinse in running DI water for 30 seconds.
4. Develop in AZ 400K (diluted with water 1:5.5) for 35-45 seconds. (Check under microscope if longer development is needed).
5. Rinse in running D.I. water for 30 seconds.

D Oxygen Plasma Descum of Photoresist
1. Chamber pressure = 300 mtorr of O₂.
2. Power = 100 W at low frequency.
3. Run for 30 seconds.

E RIE Etch of GaN Cap Layer
1. Place GaN wafer on thick Si wafer and load into RIE #5 chamber.
2. Etch conditions: Cl₂ flow rate = 10 sccm, Chamber pressure = 5 mtorr, Power = 220 W.
3. Etch for 45 seconds. (etch about 1000 Å AlGaN & GaN layer)

F Surface Preparation
1. Mix a dilute solution of HCl: H₂O :: 1:10.
2. Dip in dilute HCl for 15 seconds.
3. Rinse in D.I. water for 1 minute.
4. Blow dry with N₂.

G Evaporation
1. Place wafer on E-beam mount and level it.
2. Pump down to below 1.5 × 10⁻⁶ torr.
3. Deposit SiO₂ of 1000 Å.

H Lift-off
1. Soak wafer in ACE for 3 minutes, ultrasonic for 30 seconds.
2. Rinse with running ISO.
3. Rinse in running D.I. water for 1 minute.
4. Blow dry with N₂.
3 Schottky Contacts (mask layer 3, dark field)

A Solvent Cleaning
1. Check the resistivity of D.I. water. It should be > 17MΩ.
2. ACE 3 minutes, ultrasonic 1 minute.
3. ISO 3 minutes, ultrasonic 1 minute.
4. Running DI 3 minutes.
5. Blow dry with N₂.
6. Dehydration bake, 120 °C, 10 minutes in petri dish without cover.

B Photoresist Application and Exposure
Safety Note: The photoresist vapor is extremely harmful. Never breathe if you put your head under the hood.
1. Cool down after dehydration, 5 minutes.
2. Put wafer on spinner chuck with vacuum on, blow with N₂.
3. Apply SPR 950-0.8 with syringe and filter to cover whole wafer.
4. Spin at 2 krpm for 30 seconds.
5. Soft bake, 90 °C for 1 minute on hot plate.
6. Apply CEM 365 with syringe and filter to cover whole wafer.
7. Spin at 3 rpm for 30 seconds.
8. Expose for 3.0 – 3.8 seconds. (Exposure time and focus should be set according to condition of the Stepper.)

C Development
1. Post bake, 100 °C, 2 minutes on hot plate.
2. Rinse in running DI water for 30 seconds.
3. Develop in MF –701 for 2 minutes and 20 seconds.
D Oxygen Plasma Descum of Photoresist
1. Chamber pressure = 300 mtorr of O₂.
2. Power = 100 W at low frequency.
3. Run for 15 seconds.

E Surface Preparation
1. Mix a dilute solution of HCl: H₂O :: 1:10.
2. Dip in dilute HCl for 15 seconds.
3. Rinse in D.I. water for 1 minute.
4. Blow dry with N₂.

F Evaporation
1. Place wafer on E-beam mount and level it.
2. Pump down to below 1 × 10⁻⁶ torr.
3. Deposit Ni/Au/Ni of 200/4000/400 Å.

G Lift-off
1. Soak wafer in ACE until metal comes loose, ultrasonic for 30 seconds.
2. Squirt with ACE squirt bottle to speed up liftoff.
3. Rinse with running ISO.
4. Rinse in running D.I. water for 1 minute.
5. Blow dry with N₂.

4 Meas Isolation (mask layer 4, dark field)

A Solvent Cleaning
1. Check the resistivity of D.I. water. It should be > 17MΩ.
2. ACE 3 minutes, ultrasonic 1 minute.
3. ISO 3 minutes, ultrasonic 1 minute.
4. Running DI 3 minutes.
5. Blow dry with N₂.
6. Dehydration bake, 120 °C, 10 minutes in petri dish without cover.

B Photoresist Application and Exposure

Safety Note: The photoresist vapor is extremely harmful. Never breathe if you put your head under the hood.

1. Cool down after dehydration, 5 minutes.
2. Put wafer on spinner chuck with vacuum on, blow with N₂.
3. Apply SPR 950-0.8 with syringe and filter to cover whole wafer.
4. Spin at 2 krpm for 30 seconds.
5. Soft bake, 90 °C for 1 minute on hot plate.
6. Expose for 1.8 seconds. (Focus should be set according to condition of the Stepper.)

C Development

1. Develop in MF –701 for 1.5 minutes.
2. Rinse in running D.I. water for 30 seconds.

E RIE Etch of GaN Cap Layer

1. Place GaN wafer on thick Si wafer and load into RIE #5 chamber.
2. Etch conditions: Cl₂ flow rate = 10 sccm, Chamber pressure = 5 mtorr, Power = 220 W.
3. Etch for 2 minutes and 45 seconds. (etch about 4000 Å AlGaN & GaN layer)

F Photoresist strip

1. ACE 3 minutes, ultrasonic 1 minute.
2. ISO 3 minutes, ultrasonic 1 minute.
3. Running DI water 3 minutes.
4. Blow dry with N₂.

5 Bonding Pad Metalization (mask layer 5, light field)
A Solvent Cleaning

1. Check the resistivity of D.I. water. It should be > 17MΩ.
2. ACE 3 minutes, ultrasonic 1 minute.
3. ISO 3 minutes, ultrasonic 1 minute.
4. Running DI 3 minutes.
5. Blow dry with N₂.
6. Dehydration bake, 120 °C, 10 minutes in petri dish without cover.

B Photoresist Application and Exposure

Safety Note: The photoresist vapor is extremely harmful. Never breathe if you put your head under the hood.

1. Cool down after dehydration, 5 minutes.
2. Put wafer on spinner chuck with vacuum on, blow with N₂.
3. Apply AZ5214 with syringe and filter to cover whole wafer.
4. Spin at 2 krpm for 30 seconds.
5. Soft bake, 95 °C for 1 minute on hot plate.
6. Apply second layer of AZ5214 with syringe and filter to cover whole wafer (Must be fast otherwise Photoresists will mix).
7. Spin at 2 krpm for 30 seconds.
8. Soft bake, 95 °C for 1 minute on hot plate.
9. Expose for 0.48 seconds. (Focus should be set according to the condition of the Stepper.)

C Development

1. Post bake 110 °C, 1.5 minutes on hot plate.
2. Flood exposure for 3 minutes of longer if needed.
3. Develop in AZ 400K (diluted with water 1:5.5) for 45-60 seconds.(Check under microscope if longer development is needed).
D Oxygen Plasma Descum of Photoresist

1. Chamber pressure = 300 mtorr of O₂.
2. Power = 100 W at low frequency.
3. Run for 30 seconds.

E Surface Preparation

1. Mix a dilute solution of HCl: H₂O :: 1:10.
2. Dip in dilute HCl for 15 seconds.
3. Rinse in D.I. water for 1 minute.
4. Blow dry with N₂.

F Evaporation

1. Place wafer on E-beam mount, lower boom to about 18 cm from the source. (The actual rate is about 3 times of the reading)
2. Pump down to below 1.5 × 10⁻⁶ torr.
3. Deposit Ti/Au of 100/15000 Å. (It is very critical to control the rate to prevent spitting)

G Liftoff

1. Soak wafer in ACE for long enough till metals come loose, ultrasonic for 30 seconds.
2. Rinse with running ISO.
3. Rinse in running D.I. water for 1 minute.
4. Blow dry with N₂.
Appendix B

AlN-substrate based FC-IC fabrication

In order to achieve a superior circuit performance especially at microwave frequencies, we developed the AlN-substrate based flip-chip integrated circuit (FC-IC) process for realizing GaN broadband power amplifiers. Since GaN HEMTs are grown on Sapphire substrate, which is a very poor thermal conductor, we choose to flip-chip mount the devices onto the AlN substrate. AlN substrate is an insulator with good thermal conductivity; it not only serves as a good heat sink for GaN HEMTs, but also hosts all the passive circuit components such as transmission lines, resistors, capacitors and air-bridges. Since backside processing is not available for AlN substrate, all the amplifiers designed employ coplanar wave-guide (CPW) transmission line system. Stepper lithography is utilized to achieve better uniformity and more controlled process, especially for thick metal layer lithography that is required for power amplifiers. The maximum exposure aperture is only 20 mm × 20 mm and there is distortion at the corner, so the real usable area is less than 18mm × 18 mm, which sometimes can only accommodate 2-3 GaN amplifiers with size around 10mm × 10mm. Therefore, the GaN amplifiers must be made more compact for more efficient use of the Stepper lithography. Furthermore, with the use of 2” × 2” AlN substrate, we can only have 4 cells on each substrate, which will produce about 8-12 amplifiers at a time. This number of amplifiers usually is not enough for achieving enough good circuits with flipped GaN devices, considering the subsequent relatively complex device flip-chip bonding process. Normally, more than one round of process is needed to produce enough AlN circuits for the last step of the circuit fabrication: device flip-chip bonding. There are a total of seven steps for the whole AlN
substrate process, very cautious and skillful process is required to improve the yield of the fabrication.

5.1 AlN process realization flow

1) Si protection layer.
Before any process, a thin layer of Si (800 –1000 Å) has to be deposited onto the blanket AlN substrate. The Si layer can protect the AlN from the attack of the developer AZ400k. Meanwhile it can help to achieve better lithography, since it will prevent the light to penetrate through and then reflect back from the semi-transparent AlN substrate, therefore better defined pattern can be obtained. The Si layer will be etched away after the lithography using CF₄ gas. If there is any residue of the Si after etching, it will cause the top layers to peel off in the later process. Therefore, it’s crucial to achieve clean Si etching. We normally need to over etch for about one minute. Also it’s very important to have a clean chamber. First, we have to clean the chamber with ISO and clean wipes (sometimes sandpapers if necessary), and then use 10 minutes 300 W oxygen clean before start etching Si.

2) Resistors
NiCr resistors are designed to achieve 50 Ω/ sheet resistivity. However, it’s difficult to determine the thickness of NiCr for two reasons: first, NiCr source could be oxidized over time and its resistivity could change, as a result, the different source tends to have different resistivity. Second, since the thickness of NiCr is only about 500 Å, which is comparable with the roughness of the AlN substrate, it is crucial to have strict thickness control. Hence, it is recommended to use new NiCr source and check the thickness reading of the E-beam evaporator before the real process starts. According to our experience, normally 525 Å NiCr
can achieve 50 Ω/ sheet resistivity. Still, it’s not uncommon that there is up to 20% variance of the resistance.

3) CPW
This layer defines the pattern of CPW transmission lines. E-beam evaporator is utilized to deposit Ti/Au/Ti of thickness 200/30000/100 Å. Ti is used to achieve better adhesion for Au to the substrate and for the later layers especially SiN to the Au layer. It’s very beneficial to keep the Au thickness to at least 3 µm, especially at the output section, where large DC current up to 2 amperes could be flowing through. In order to define thick photo-resist pattern, double layers of EIR 5214 with lowered spin speed (~ 3k/sec) need to be applied to the substrate to achieve at least 4-µm thick photo-resist layer. During the deposition, the sample holder needs to be lowered to the height of about 17 mm above the base platform, in order to increase the deposition ratio three times of the reading. The evaporation has to be done slowly (ratio should be less than 15 Å/sec) to avoid spitting, which could cause short circuit for the later layers especially capacitors and air-bridges.

4) Capacitors
The capacitors are all shunt capacitors, so their bottom planes are on the ground of CPW layer. SiN is utilized as the insulating medium; its capacitance per unit area is 0.2 fF/ µm² for 3000 Å thick SiN under the normal plasma enhanced chemical vapor deposition (PECVD) conditions. After the deposition of a blanket of SiN on top of the CPW layer, capacitor top plates will then be defined by Stepper lithography. Again, a thick Au metal of 30000 Å needs be deposited to reduce the DC contact resistance of the capacitor plates. Sometimes 6000 Å thick SiN is deposited to achieve 0.1fF/ µm² capacitance, it’s recommended to perform
the deposition 3000 Å at a time. In between, ISO clean can reduce the chance of through pinholes in SiN, also the uniformity of the SiN layer can be improved by rotating the samples.

5) Capacitors etch
This step is aimed to etch away SiN wherever unnecessary. The top plates of capacitors and NiCr resistors need to be protected by SiN, also underneath the air-bridges, SiN can be left as an insulating medium for preventing the air-bridges from short-circuited. This SiN etching step can be eliminated, but cautious process will be needed in the later steps of process since SiN could cause lithography difficulty. Similar to Si etching, CF₄ gas is used to etch SiN. The same clean procedure is required to be able to achieve clean etching.

6) Bonding pads
Since bonding pads in this work are required to be at least 6 μm up to 8 μm thick, relative complex bi-layer resist process has to be developed. We first use two layer of PMGI-SF15, which can obtain at least 5-μm thickness, and then EIR 5214 will be utilized to define the pattern as well as providing more thickness (at least 2 μm by spinning at lower speed 3k/sec). After the Stepper lithography and development of EIR 5214 resist, the PMGI layer will be exposed. Subsequent deep-UV exposure and development in SAL-101 developer can define the PMGI layers. The thick Au deposition has to be done twice, 3 - 4 μm each time. Extra caution is required to reduce Au spitting, since the flat bonding pad surface is very important for the success of device bonding.

7) Airbridges
This is the last yet most complex step of the whole AlN substrate process. Total of two masks would be required. First mask is used to define the pattern of posts, which open up the metal areas for the air-bridges contacts. PMGI-SF15 is the photo-resist used for the post layer. The post mask is dark field because the positive photo-resist is preferred to define PMGI. Positive photo-resist AZ P4210 or AZ P4330 need to be utilized to protect the PMGI from the exposure during the Stepper lithography, since the exposed PMGI will be attacked by acetone that has to be used to strip off the positive resist after the pattern of the PMGI. Then re-flow of PMGI at high temperature (250°C) for 10 minutes is required to smooth the edge of the posts for better air-bridge connections. Then, the second mask that defines the air-bridge span will be utilized. Two layers of EIR-5214 need to be applied to achieve at least 3-μm-thick photo-resist layer. Subsequently, E-beam evaporation that is similar to the previous thick Au deposition will be done to obtain at least 2 μm thick Au air-bridges. It can improve the reliability of the air-bridges to leave the PMGI underneath the bridges. But the drawback is that in the subsequent device flip-chip bonding procedure, the bonding temperature cannot be too high, otherwise PMGI will be re-flowed. However, the optimized bonding condition will require 300 °C high temperatures. As a result, it’s recommended to remove the PMGI underneath the air-bridges and then be able to bond the devices at 300 °C.

5.2 Other common fabrication problems

During the AlN substrate and GaN HEMTs process development, we encountered numerous problems that we didn’t expect. We hope to provide valuable information for the later graduate students by summarizing some of the common problems.
1) Dark-field mask for resistors layer
Since we need to rely on the alignment mark on the NiCr resistor layer for the Stepper lithography alignment, we recommend using dark-field mask for this layer. The standard Stepper alignment mark has about 1μm size pattern and we can easily achieve 1μm feature with SPR-950 photo-resist. If light-field mask was used, it’s very difficult to pattern 1μm feature with EIR-5214 photo-resist.

2) Larger alignment tolerance for posts layer
After the pattern of posts layer, reflow of PMGI photo would be required to obtain flat edges. Normally the size of the posts opening will be about 1μm - 2μm wider. Therefore, in the layout design, the posts pattern should be at least 1μm more margin of alignment tolerance to the other layers.

3) Avoid of small closed-loop layout pattern
We found out that for small closed-loop layout pattern, there will be lift-off problem. So patterns such as circle, ring or island should be avoided or be made as open as possible. For example, the pattern shown in Fig. 5.1 (a) will have lift-off problem and it can be changed to the pattern in Fig. 5.1 (b) that doesn’t have lift-off problem.

![Fig. 5.1(a) Pattern that will have lift-off problem, (b) Pattern that will not have lift-off problem.](image)
4) Exposed PR4220/4330 strip-off
During the process of posts step, positive photo-resist PR 4220/4330 will be used
to pattern PMGI, afterwards we will use acetone to strip-off these positive photo-
resists. There will be problems to cleanly strip off those exposed chemically
changed photo-resists. One of the solutions is to use Oxygen etching for 2-3
minutes before acetone cleaning.

5) Double layer photo-resist mixing
Double layer photo-resist will always be needed to pattern thick photo-resist
profiles. Sometimes there is problem of mixing of the layers that prevents us
from achieving thick photo-resist layer. By shortening the time of apply of the
second layer before spinning, we can greatly reduce the mixing problem.

6) Length of air-bridges
The reliability of the air-bridges (without PMGI support) is closely related to the
length of the bridges. The empirical length for best reliability is less than 50 μm
and the upper limit should be always less than 150 μm. You can always shorten
the air-bridges by pull the ground closer for a very short distance, or if the signal
line is too wide, you can make the air-bridges as the connection of the signal lines
to cross the ground plane.

7) Size of the diced-up chip
If the size of each diced-up chip is too small, the chips could pop up during the
dicing because the wax couldn’t hold them anymore. We found out that if the
GaN chip is smaller than 1mm × 1mm, it’s very difficult to hold them using the
crystalbond-509 mounting wax. There is wax that can provide stronger adhesion,
but it’s harder to strip the wax off after the dicing. We can improve the adhesion of the chips by employing multi-cuts with slower speed. Unfortunately, for GaN wafer we haven’t found a reliable solution to achieve smaller than 1mm × 1mm dice. Therefore, we need to design chips with larger than 1mm × 1mm size.

8) Overheat of protecting photo-resist layer for dicing
When we dice the sample, a blanket photo-resist layer needs to be applied to protect the patterns on the sample from the dicing dust. Since we need to apply about 200°C temperature to wax the sample onto the Si substrate before the dicing, we must avoid overheating the protection photo-resist layer. We need to apply the photo-resist layer after the mounting or use PMGI as protection photo-resist. Similarly, after the dicing you should be very careful not to overheat the photo-resist if you want to take off the chips right away by heating up the Si substrate to about 200°C. The best way to take off the diced chips would be to leave the sample and the Si substrate in the acetone for overnight.

9) Sample holder size for the flip-chip bonder
When you need to bond more than one chip onto the circuit board, there is requirement for the size of the sample holder depending on the distance of the bonded chips and the size of the chips. For example, we need to bond 4 GaN chips onto the AlN circuit board in the fabrication of GaN modified TWA, these chips are 2mm apart and they are about 1.2mm × 1.5mm in dimension. It turned out that we have to make a custom sample holder with about 2mm × 2mm contact area, since the existing sample holders are too big that they will touch the previous bonded chips when bonding the subsequently chips.