BASEBAND LINEARIZATION SCHEMES FOR HIGH EFFICIENCY POWER AMPLIFIERS

A thesis submitted to Cardiff University
In candidature for the degree of
Doctor of Philosophy

By

Guofeng Wang, MSc

Centre for High Frequency Engineering
Cardiff School of Engineering
Cardiff University
United Kingdom

June 2019

Abstract

High efficiency and high linearity microwave power amplifiers (PAs) are a critical element in modern wireless applications. Over recent years, modern communications systems and the complex modulated signals they use have presented significant challenges in terms of maintaining acceptable efficiency and achieving the high degrees of linearity required in microwave radio frequency power amplifier (RFPA) designs. The next 'big' challenge is the deployment of the fifth-generation (5G) mobile network, which is scheduled for commercial launch in 2020. Although the specification for 5G is not completely known at this point, the expectations in terms of what 5G will bring most certainly are; including 1000x more capacity, less than 1ms latency and 100x network energy efficiency. New 5G systems will need to provide higher spectral efficiency, wide and fragmented signal spectra and dynamic spectrum access (DSA). As a result, the waveforms used in 5G systems will be characterised by high peak to average power ratio (PAPR) and high bandwidth, especially for high data rate applications, which brings additional challenges in terms of achieving system efficiency and linearity.

Digital Predistortion (DPD) has been widely and very successfully applied in modern communication systems to linearize PAs and meet system requirements. However, as the signal bandwidth widens and carrier aggregation becomes commonplace in 5G system, higher complexity DPD algorithms and an

increased number of associated parameters will be required. This will inevitably result in a more complex DPD systems with higher power consumption and overall, lower system efficiency. This is especially problematic when systems advance into massive multiple-input, multiple-output (MIMO) scenarios, where the distributed systems are smaller in size and massive in number.

The research work in this thesis starts by analysing the different nonlinear distortion mechanisms present in the typical microwave power transistor devices that would be deployed in an RFPA within a 5G system. A tunable analytical device model is established to investigate the individual contributions of key nonlinear elements in the device. A number of important observations, such as “sweet-spots”, sideband asymmetry and drive dependent optimum baseband termination have been discovered and analysed in detail. Using the developed analytical model, a linearity optimization strategy in circuit design has been discussed and applied to a commercially available and widely used nonlinear device model CGH60015D from Cree (now Wolfspeed). For the first time, a systematic study of all main nonlinear components has been done and the interaction between these components has been discussed.

In the second part of the thesis, a pair of novel system-level envelope domain linearization techniques are presented and analysed. They are applied at the input node and output node of the power amplifier, respectively. The envelop linearization techniques have been demonstrated with both the analytical model, developed in this thesis, and the nonlinear device model CGH60015D. The

advantages of envelope linearization has been discussed as well as the challenges such an approach presents.

The *Linearizability* of a system, both in terms of circuit design and linearization techniques are discussed. In fact, linearity and linearizability of power amplifiers forms the central thread that runs through this thesis together with linearity, which provides guidance for a top-to-bottom level PA linearization strategy.

Acknowledgments

Firstly, I would like to thank my family for all the motivational and financial supports during my PhD journey. My deepest gratitude goes to my parents without whom there will be nothing possible in my life. I am very thankful to my wife, Han Tong, who has been very supportive all the time and taking care of this family. This work wouldn't be possible even from the starting without their supports and sacrifices.

I would like to express gratitude to my amazing supervisors; Prof. Paul Tasker and Dr. Jonathan Lees, without whom this work will never be possible to be finished. Prof. Paul Tasker has provided lots of insights and intellectual advises to this work. During the past 4 years in Cardiff, there are uncountable number of meetings in which I easily forgot the time while enjoying listening to and discussing with him. He is unbelievable energetic and such enthusiastic in his work as well as his life, who is obviously setting a great positive image for all the students and staff in this group. Dr. Jonathan Lees is one of the best supervisors I have ever seen. He has gone far beyond the call of duty to provide sufficient supports, advises and time. I can't remember how many times I received his immediate email response at late night. Moreover, he is always caring about his students, for their mind and health well-beings. Those kindly and encouraging chats will be always remembered.

I would like to thank all the members of Centre for High Frequency Engineering (CHFE) in Cardiff University. Special thanks go to my colleagues, Zulhazmi A. Mokhti, David Loescher, Elango Nagasundaram, Azam Al Rawachy, Sattam Alsahali, Alexander Alt and Peng Chen for their friendship, inspiration and fruitful discussions. Thanks to Dr. Roberto Quaglia and Dr. James Bell, for the inspiring discussion on DPD and analytical models.

Finally, a special thanks to my research office staffs, Jeanette Whyte, Chris Lee and Aderyn Reid, as I received so much help from them in the past 4 years. They are all lovely people and always make me feel supported with peace of mind.

Thank you all for your continuous supports throughout my PhD.

List of Acronyms

1D	One-dimension
2D	Two-dimension
4G	Fourth generation communication systems
5G	Fifth generation communication systems
ACLR	Adjacent channel leakage power ratio
ACPR	Adjacent channel power ratio
ADC	Analog to digital converters
AM	Amplitude modulation
ANN	Artificial neural network
BTS	Base transceiver station
CA	Carrier aggregation
C_{ds}	Drain to source capacitance
C_{gs}	Gate to source capacitance
C_{gd}	Gate to drain capacitance
D2D	Device to device
DAC	Digital to analogue converters
DAS	Distributed antenna systems
DC	Direct current
DDR	Dynamic deviation reduction

DPA	Doherty power amplifier
DPD	Digital predistortion
DSA	Dynamic spectrum access
EER	Envelope elimination and restoration
eMBB	Enhanced mobile broadband
ET	Envelope tracking
ETPA	Envelope tracking power amplifier
EVM	Error vector magnitude
FET	Field-effect transistor
FIR	Finite impulse response
GaAs	Gallium arsenide
GaN	Gallium Nitride
GMP	Generalized memory polynomial
HB	Harmonic balance
IEL	Input envelope linearization
IFB	Indirect feedback
IM	Intermodulation
IM3	3rd order intermodulation
IM5	5th order intermodulation
IMD	Intermodulation distortion
IMD3	3rd order intermodulation distortion
IoT	Internet of things

ITU	International telecommunication union
LDMOS	Laterally diffused metal oxide semiconductor
LES	Least squares estimation
LINC	Linear amplification using nonlinear components
LIP	Linear in parameters
LO	Local oscillator
LTE	Long term evolution
LTE-A	Long term evolution advanced
LUT	Lookup table
M2M	Machine to machine
MIMO	Multiple input multiple output
mMTC	Massive machine-type communications
MP	Memory polynomial
OBO	Output back-off
OEL	Output envelope linearization
OPEX	Operational expenditure
PA	Power amplifier
PAE	Power added efficiency
PAPR	Peak to average power ratio
PLL	Phase lock loop
PM	Phase modulation
RF	Radio frequency

RFPA	Radio frequency power amplifier
SDD	Symbolically defined device
SiC	Silicon carbide
Tanh	Hyperbolic tangent
uRLLC	Low latency communications
VCO	Voltage controlled oscillator
VCVS	Voltage controlled voltage source
VR	virtual reality
VS	Volterra series
WCDMA	Wideband Code Division Multiple Access

List of Tables

List of Figures

Table of Contents

Chapter 1 Introduction

1.1 Next generation – 5G communication

Over recent years, modern communication systems have needed to develop rapidly to be able to accommodate the demanding requirements from various communication applications. These have ranged from simple text message delivery to real-time video streaming to the future requirement of virtual reality (VR) communication. Generally, the requirement involves transmitting high volumes of data to support increasing numbers of the users, with increasing speed and reducing latency. The next, fifth generation (5G) of mobile networks are scheduled for commercial launch in 2020 [1]. These networks will principally be designed for greater

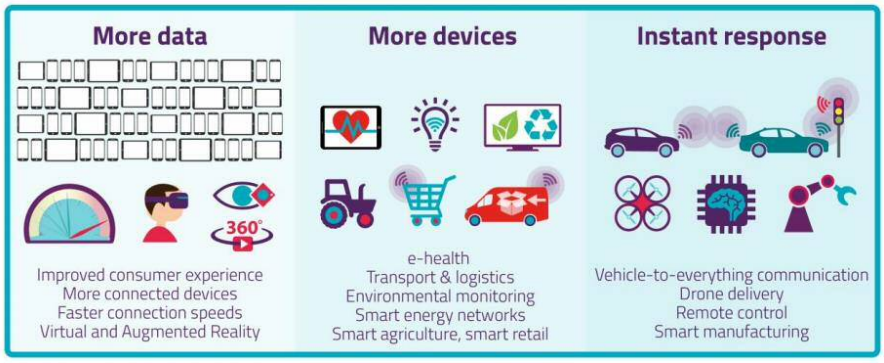


Figure 1.1 Expected 5G applications, copy from reference [2]

diversity of data services, such as the internet of things (IoT), machine to machine (M2M) communications and device to device (D2D) communications rather than

simple website browsing and video streaming that have been the main drivers in previous generations. The expected uses, as shown in Figure 1.1, range from "more data" with more connected devices, faster connection speeds, virtual and

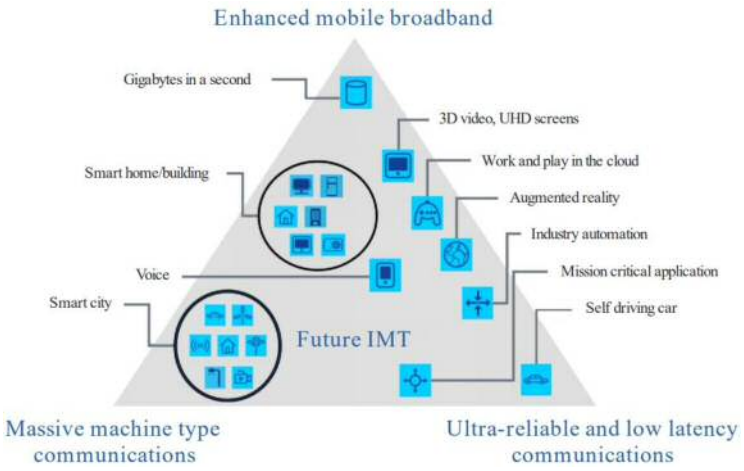


Figure 1.2 5G applications categorized by IMT-2020, copy from reference [1]

augmented reality to "more devices" with automation e-health, transportation and logistics and smart industry to less latency and "instant response", which enables, for example, vehicle to everything, drone delivery and other disruptive technology concepts [2]. These applications are proposed by the international telecommunication union (ITU) and categorized into three categories: enhanced mobile broadband (eMBB), ultra-reliable and low-latency communications (uRLLC), and massive machine-type communications (mMTC), Figure 1.2.

5G networks are expected to accomplish aggressive technical specifications compared to its earlier fourth generation (4G) systems [3]: for instance, less than 1ms latency compared to 50ms for 4G, 100x network energy efficiency, 20Gbps peak data rates compared to 20Mbps average for 4G, and 10Mps/m^2 area traffic capacity compared to 7Mps/m^2 for 4G. Although 4G has only been around

from 2013 when Long Term Evolution Advanced (LTE-A) was released, the current technology and infrastructure for 4G LTE networks rapidly matured with the support from communication industry and research groups. However, in comparison, significant work remains to bridge the technical gap to 5G, if the established roll-out timescale is to be achieved. Emerging technologies such as millimetre wave

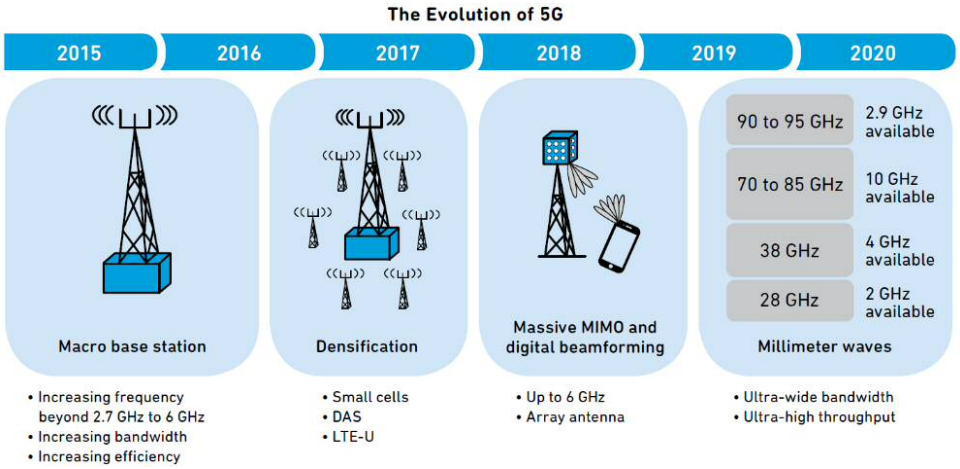


Figure 1.3 The evolution of 5G, copy from reference [4]

(mmWave), massive multiple-input multiple-output (MIMO) antenna arrays, and advanced beamforming have become promising candidates in 5G implementation. Typical 5G applications, such as IoT, demands a massive increase in the number of active connections across the wireless network. The volume of data is then expected to be at least 10 times that of a 4G network. To accommodate this increasing data requirement without significantly system complexity and cost, a combination of macro-cell and small-cell base stations, based on expanding existing network has been considered [4], where the possibility of enabling 5G applications without developing an entirely new backhaul network with associated significant extra cost is explored. Notably, the use of small cells and distributed antenna

systems (DAS) have been reported since 2017 and will be employed widely in 5G system to achieve higher data capacity, as shown in Figure 1.3.

Advanced 5G network will continue “down scaling” towards the user-end from macro-cell to micro-cell or even pico-cell and femto-cell in the future which involves extra network densification [5]. To support massive M2M and D2D applications in 5G, these compact small-cells must be high energy efficient with effective thermal management. Fortunately, gallium nitride (GaN) device technology has been developed by semiconductor industry to the point where commercial devices are readily available. GaN has been recognized as superior candidate in power amplifiers for 5G applications. It has superior features including high breakdown voltage, high power density, and as a result, ease in impedance matching [6]. GaN devices are also able to achieve high efficiency compared to previous technologies, such as laterally diffused metal oxide semiconductor (LDMOS) devices [6], [7]. However, there is still a need to make intellectual choices in terms of system architecture, circuit design and possible optimization. For future 5G system, the ability to support more users with more data on a quick response network depends on an advanced, high spectral-efficiency digital modulation scheme. Normally, this approach results in denser constellations and communication systems using modulation envelopes with higher peak average power ratio (PAPR). This feature brings new challenges in realizing 5G applications, RF designers still need to focus on optimizing existing design approaches as well as proposing new and novel solutions and architectures for wireless transceivers.

1.2 RF Power Amplifiers in Wireless Communication

The radio frequency power amplifier (RFPA) is a key component in the transceiver chain in modern communication system. Actually, when a signal propagates through any RF system, it suffers losses due to material dielectric loss, circuit reflection loss as well as long-distance transmission loss. Furthermore, noise levels become problematic when the signal needs to be precise, of high fidelity, and detectable. Hence, signal amplification is extensively used to overcome the losses and maintain the signal above noise levels. This makes RFPA a significant element in the transceiver chain, necessary in achieving the desired signal propagation.

In general, RFPA design involves selection of amplifier topology, class or mode, bias and impedance optimization and signal linearization [8]. Power amplifier topologies include traditional single-end, Doherty, LINC (linear amplification using nonlinear components), Outphasing, Envelope Elimination and Restoration (EER) and Envelope Tracking (ET). Classically, RFPAs can be categorized into Class A, Class AB, Class B and Class C, distinguished through different conduction angles [8]. By *over-driving* the PA into saturation region, different classes are possible, such as saturated Class-A and -C, Class-D, Class-E and Class-F which depend on both conduction angle, drive level and the output harmonic terminations [8],[9], as shown in Figure 1.4.

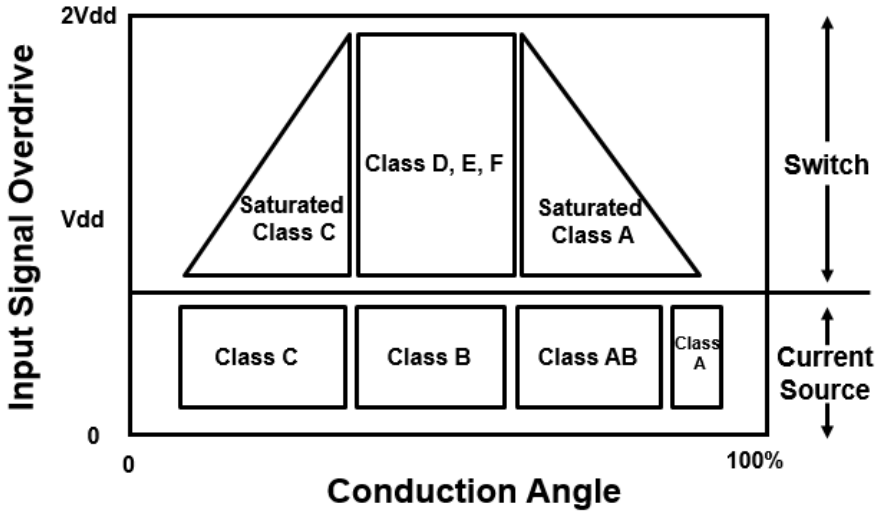


Figure 1.4 Classical definition of RFPA classes

In RFPA design and optimization, DCIV measurements, load-line theory and load-pull measurements are very commonly used tools. These measurements and methodologies guide designers to identify optimized bias and load for the target applications. Signal linearization is normally used to improve the fidelity of the system, and involves different techniques such as feedforward and feedback schemes, as well as analogue and digital predistortions, etc.

1.3 Efficiency and Linearity Challenges in RFPA Design

In mobile communication networks, the base transceiver station (BTS) is the most expensive and power consuming unit. Typically, the power consumption of these base stations account for around 57% of the total energy consumption in the celluar networks [10], [11]. In a typical BTS, PA is the most power consuming component, more than 80% of the input energy is dissipated as heat [10]. Thus, the energy efficiency of the PA becomes a critical factor when considering operational expenditure (OPEX) and cost to the carrier.

In classical RFPA design space, high efficiency is achieved by driving the active device close or beyond its saturation point. Unfortunately, from a linearity perspective, this leads to power compression and signal distortion in the output waveforms. In other words, to be highly power-efficient under a limited supply voltage, the RFPA must operate in or close to its non-linear region. Distortion is always considered as the unwanted components in a communication system. The *in-band* distortion will degrade the signal quality usually characterized by the metric error vector magnitude (EVM), while the out-of-band spectral regrowth may interfere the adjacent frequency band, characterized by adjacent channel power ratio (ACPR). In traditional design approaches, in order to achieve a certain linearity level, the system is forced to operate in a “back-off” region to avoid strong nonlinearities due to signal compression and distortion. This is however at cost of reduced power efficiency. This is knowns as the classical “trade-off” between efficiency and linearity in RFPA design.

To overcome these limitations, many different solutions have been proposed focusing on simultaneously achieving efficiency enhancement and signal linearization. On the one hand, to make the RFPA more efficient, designers have introduced different amplifier topologies to boost power efficiency in the “back-off” region, among which, Doherty (constant-supply, active load modulation) and ET (supply-modulation) are the two most popular schemes. On the other hand, in order to improve the signal linearity, many linearization techniques are exploited in different applications. Among these linearization technique, digital predistortion (DPD) is widely used in modern communication schemes which provides flexibility in circuit design and excellent performance. Although these solutions inevitably increase the system complexity, they are capable of achieving higher specifications compared to the conventional amplifiers. Many researches have focused on developing these approaches over the last few decades, to push the performance into excellence. However, the challenge remains, especially transmitting a modulated signal with increasing PAPR, while limiting physical complexity and cost in 5G or small-cell scenario. An energy/cost efficient solution is in demand [12].

This thesis primarily focuses on ways in which to improve the linearity of RFPA. It will provide analysis for both internal optimization of circuit level PA designs as well as external modification at the system level. The first part of the thesis is focusing on the circuit level optimization, where the impact from circuit environment, such as biasing, drive level and impedance terminations are discussed in detail. In the second part of the thesis, a generalized linearization technique is presented in the envelope domain. Importantly, this technique and

methodology theoretically supports the linearization applied at both input predistortion and output supply modulation.

1.4 Thesis Structure

This thesis is structured as follows:

In Chapter 2, a general literature review has been provided, focusing on linearization of the RFPA. For the motivation background, the requirements and challenges from emerging 5G communication systems are firstly discussed. This is followed by a review of existing and emerging RFPA linearization techniques, such as feedforward, feedback, predistortion and envelope linearization. The operational concept of the well-recognised DPD technique is explained and the performance of different DPD models is summarized, followed by a discussion on emerging challenges. GaN device technology is introduced with comparison to other existing technologies.

Chapter 3 presents the relevant RFPA design methodology, from using IV curves and load-line theory to the introduction of different classes of PA with discussions about their performance in terms of efficiency, power and linearity. Advanced high efficiency topologies are reviewed at the end of the chapter with specific comments on linearity.

In Chapter 4, the intrinsic nonlinear sources / elements are identified within the device. A tunable analytical nonlinear model is then established to allow analysis of the individual distortion contributions of those nonlinear sources. The impacts on RFPA linearity from both trans-conductive and reactive elements are discussed in detail. A number of key experiments / measurements have been

identified to improve the circuit linearity and linearizability. The interaction of all these nonlinear components in the device has been revealed and discussed. At the end of this chapter, an example of circuit level optimization of a commercially available device model is provided and discussed.

In Chapter 5, a pair of envelope domain linearization techniques are presented and discussed. They are then verified with simulation results with analytical device and commercially available device models. The application significance in terms of tackling the challenges in 5G systems are discussed. The performance and advantages of the envelope technique are then summarized.

Chapter 6 summarizes this thesis with notable observations and conclusions. It also discusses a number of future works from further verification and extensive application.

Chapter 2 Literature Review

2.1 Introduction

In this chapter, a review of relevant background knowledge and most recent progress in RFPA linearization is presented. Firstly, an overview of the current requirements in terms of RFPA linearity in wireless communication system is summarized, then the evolution and recent progress in existing linearization techniques is discussed. Towards the end of the chapter, a review on GaN device technology and its associated nonlinear characteristics is presented.

2.2 Signal Linearity Requirements for 5G Communication

In the next few years, the sub-6 GHz “pre-5G” LTE-Advance system will first be employed to pave the way for above-6 GHz “real-5G” [13]. Despite of the difference in spectrum allocation compared to current communication systems, 5G waveforms will tend to be characterized with high PAPR. This feature of 5G waveforms will naturally make the efficiency enhancing RFPA topologies such as Doherty and Envelope Tracking very attractive, as they provide high efficiency in the “back-off” region [14]. However, both these techniques inherently introduce

distortion into the amplified output through mixing effects, from either load or supply modulation, respectively. The trade-offs between efficiency, linearity and bandwidth remain. In general, the limits on allowable levels of distortion in communication systems are normally specified by established standards bodies (for example 3GPP) and enforced by the governing communication agencies, such as OFCOM in the UK. It also varies with different applications that involve RF transmitters. To author's knowledge, there is currently no fixed standard on linearity for 5G, so this thesis makes reference to the linearity requirements in typical 3G/4G base stations, with a general review of current research works and commercial prototypes to define the standard requirement.

As shown in Table 2-1, which lists information mainly from communication society standards documents, the linearity requirements in 3G/4G with respect to different modulation signals are summarized. In the table, EVM refers to the error vector magnitude based on the 16-QAM modulation scheme. ACLR1 refers to the first adjacent channel leakage power ratio (ACLR) while ACLR2 refers to the second.

Table 2-1 Linearity requirements in typical 3G/4G signals, with 5G NR candidates

Specifications	**UMTS**[15]	**WiMAX**[16]	**LTE**[17]	**LTE-A**[18]	**5G NR**
Single-channel bandwidth (MHz)	5	1.25, 5, 10, 20	1.4, 3, 5, 10, 15, 20	20	100
EVM (%)	< 12.5	< 6	< 12.5	< 12.5	
ACLR1 (dBc)	< -45	< -45	< -45	< -45	< -30
ACLR2 (dBc)	< -50	< -50	< -45	< -45	< -30

To keep up to date, as a brief review on current industry standard, Table 2-1 summarizes a number of RFPA examples where DPD is used, ranging from Class AB, Doherty to ET. The performance is evaluated on the basis of ACLR and EVM. To be fair, the chosen RFPA examples are basically either widely used commercial solutions in industry or successful prototypes reported from well-known research groups [19].

Table 2-2 Typical PA performance under DPD

PA Type	Power (dBm)	Bandwidth (MHz)	ACLR (dBc)	EVM (%)	Ref.
Class AB	40	15	-57	0.95	[20]
Class AB	40	5	-58	1.98	[21]
Doherty(dual)	42	5	-55.8	-	[22]
Doherty	47	10	-55	-	[23]
Doherty	50	20	-59	-	[24]
Doherty	38	60	-51	-	[25]
Doherty	40	40	-57	0.48	[26]
Doherty	43	15	-54	-	[27]
Doherty(dual)	39	20	-59	-	[28]
Doherty	54.7	40	-53	-	[29]
ET	40	1.4	-50	-	[30]
ET	45.5	5	-57.3	-	[31]

The PAs in Table 2-2 all comply with the specifications set by the standard in Table 2-1, when linearized using DPD. In fact, among current signal linearization techniques, DPD emerges as a superior choice due to its high flexibility / adaptability, excellent performance and compatibility with digital modulation schemes. It tends to be assumed as a fundamental function block in current and emerging communication systems. However, DPD performance is highly dependent on the accuracy of the RFPA model and requires consideration of implementation which includes computation complexity, such as the number of coefficients and memory tabs and practical hardware limitations. Depending on the requirement, the hardware involved in implementing full DPD systems can be complex and expensive, both in terms of cost and the high-power consumption with increasing signal bandwidth. This may be especially true in 5G scenarios where, in comparison to large marco-cell base stations, the small-cell base stations will have lower transmission power at a single node but more numerous for the entire networks. Thus, the overall power consumption in DPD system is no longer negligible on this scale, and new, compact and low power solutions can be expected to be in demand. In the next section, the author will take a step back to review the evolution of linearization techniques ranging from feedforward linearization to digital / analogue predistortion.

2.3 Existing PA Linearization Techniques

Managing distortion in amplifiers has been a problem for engineers since the advent of telecommunications. To improve the signal quality, signal linearization has been used since the 1920s during the expansion of the telecommunications and radio broadcasting industry. Although feedforward and feedback linearization techniques were first applied nearly 90 years ago, they are, remarkably, still used in applications today and remain a development area in both industry and academic research communities [32].

2.3.1 Feedforward Linearization

Feedforward linearization is recognized as the first form of RF linearization technique. It was first proposed by Howard Black in 1923 at Bell Telephone Laboratories [33]. Illustrated in Figure 2.1, the distortion correction is applied at the output and the opearation is completely RF domain [32],[34]. Feedforward system

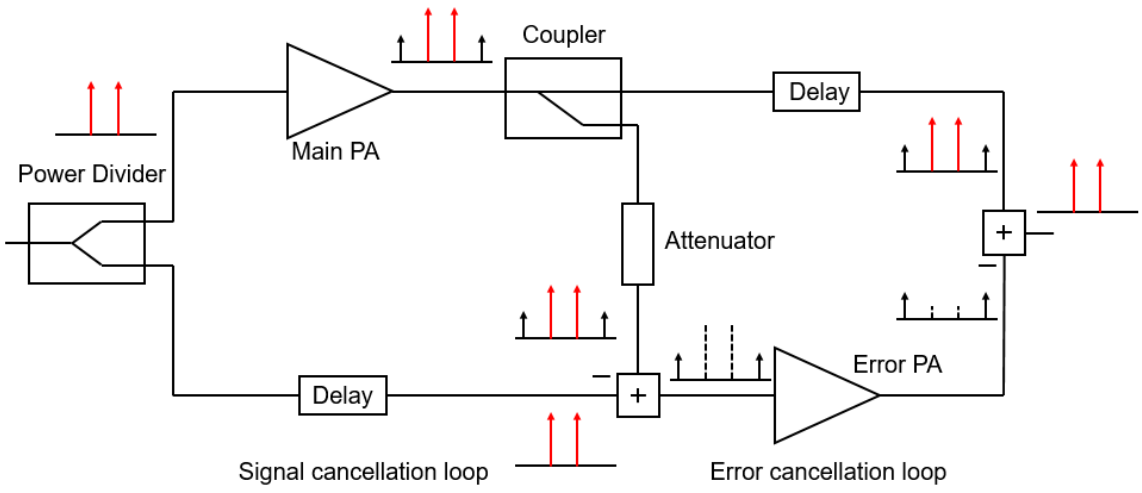


Figure 2.1 Block diagram of feedforward linearization

consists of two loops to achieve linearization: a signal cancellation loop and an error cancellation loop. To provide an illustration, the error from the distortion in main PA is first extracted from the signal cancellation loop. This is done by subtracting an input sample from the output of the main PA. Following this, the error signal is amplified through an error PA and combined with the PA output where the cancellation is achieved. In a practical implementation, to get perfect error extraction and distortion cancellation, this system requires both precise amplitude and phase / delay alignment, in all element loops, which is quite challenging. Manually tuned amplitude and phase matching are usually not sufficient to provide desired distortion suppression, so adaptive methods were introduced and this technique found extensive use from 1980s, when solid-state amplifiers became a commercial reality. A well-designed feedforward PA can be expected to provide up to 30dB of correction [35]. A critical element in this technique is the “error PA”. In the original theory, the error PA needs to remain in its linear regime, generally operating low power, without introducing additional distortion into the error signal. However, the “error PA” needs to deliver relatively high power to overcome the small coupling factor used in the directional coupler. As a consequence, the “error PA” normally ends up being of comparable size as the main PA. Therefore, the power consumption from this linear “error PA“ is not negligible, which degrades the overall efficiency of the system and cost.

Even though this technique was introduced nearly 100 years ago, recent research continues to report successful improvements, meaning feedforward linearization remains as an interesting architecture choice to PA designers in certain

applications [36]–[40]. Notably, instead of a linear error PA in the traditional feedforward technique, in recent work [36], R. N. Braithwaite successfully exploited a nonlinear error amplifier for higher efficiency (system efficiency 25%). This was then extended to the Doherty PA topology and measured to compare with the performance under DPD linearization [40]. With this adaption for feedforward scheme, improved linearity was achieved in comparison with DPD linearization, without significantly sacrificing efficiency. This superior linearity performance re-established feedforward as an attractive choice of linear PA once again.

2.3.2 Feedback Linearization

Feedback linearization was first documented by the same engineer, Howard Black, after his experiments on feedforward [33]. In theory, the operation of a feedback circuit generally involves taking a small sample from output signal and subtracting an attenuated version of the signal sample from the input. This technique becomes very popular and widely used in low-frequency analogue circuits where the time

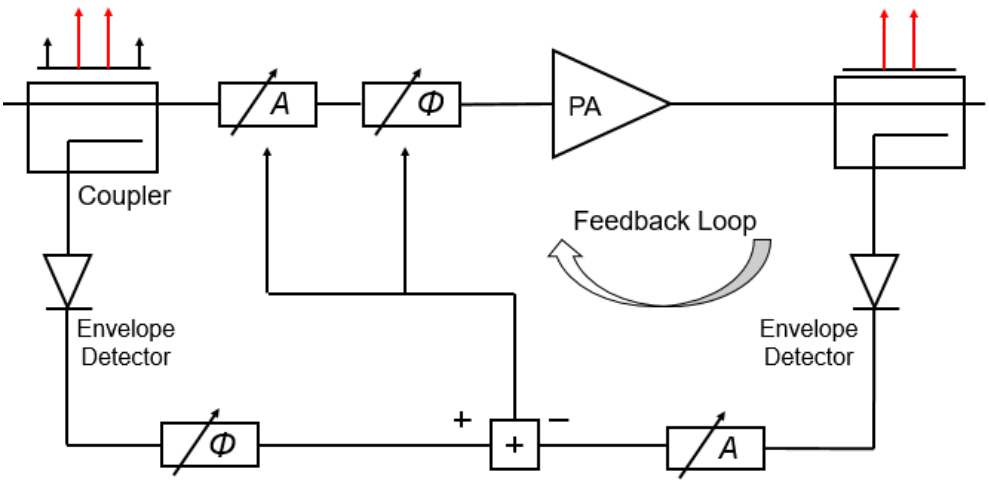


Figure 2.2 Indirect feedback linearization (IFB)

delay around the loop is negligible. However, it raises concern about PA stability and difficulty in making wideband nonlinear feedback loop architectures [41]. Consequently, the feedback in RFPA linearization is usually provided at baseband frequency, with down-converted input and output samples instead of RF. This technique is referred to as indirect feedback (IFB), as shown in Figure 2.2. It is worth mentioning that, this feedback scheme is the precursor of digital pre-distortion. By comparing the sampled signal in different modulation schemes, engineer proposed two different architectures: polar feedback and cartesian feedback.

In polar feedback linearization, Figure 2.3, the phase and amplitude of the baseband envelope are compared individually for distortion correction [35]. In short, the modulation signals are compared in their polar form. A sample of the amplifier output is coupled and down-converted to baseband. The input and output baseband signals are split in two routes, for both phase and amplitude comparison. Illustrated in Figure 2.3, the phase is compared through the phase-lock loop (PLL) and then used to drive a voltage-controlled oscillator (VCO). The envelope

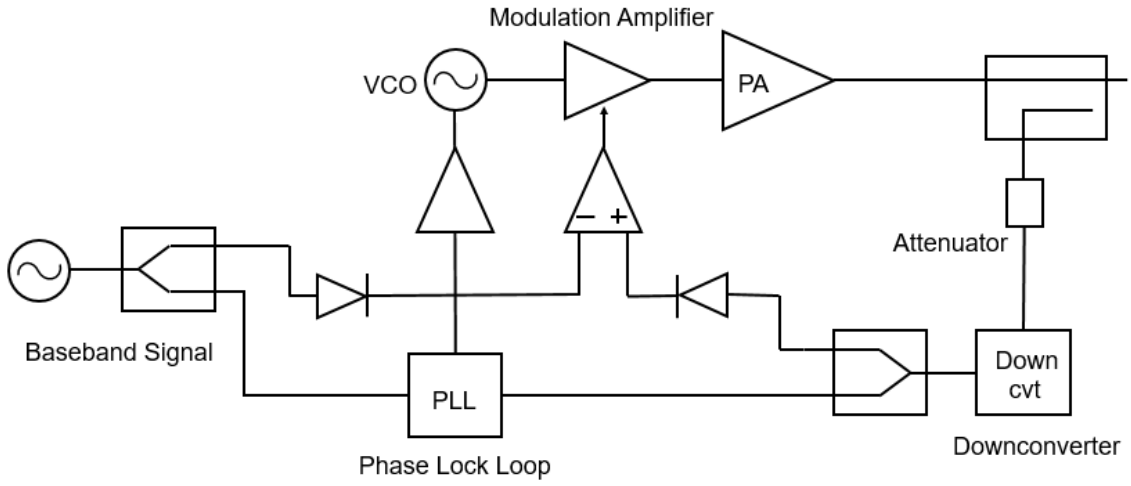


Figure 2.3 Polar feedback linearization

amplitudes are detected and then compared by a differential amplifier. The output of the differential amplifier is then used to drive a modulation amplifier which amplitude modulates the output of VCO. The correction is done by exploiting the mixing effect within the PA.

In Cartesian feedback linearization, the correction is applied in the *I/Q* domain. Illustrated in Figure 2.4, the system is constructed in a quadrature architecture. A sample of the PA output is coupled and down-converted by the *I/Q* demodulator, where the output is compared directly with the input baseband *I* and *Q* components. In theory, this standard quadrature architecture can be applied to any modulated signals. The components in this system, like differential IF amplifiers and *I/Q* modulators and demodulators that are generally available in quadrature RF transmitters, which makes the implementation straightforward and cost effective. In fact, this cartesian feedback scheme is commonly used as the theoretical basis of DPD correction platforms.

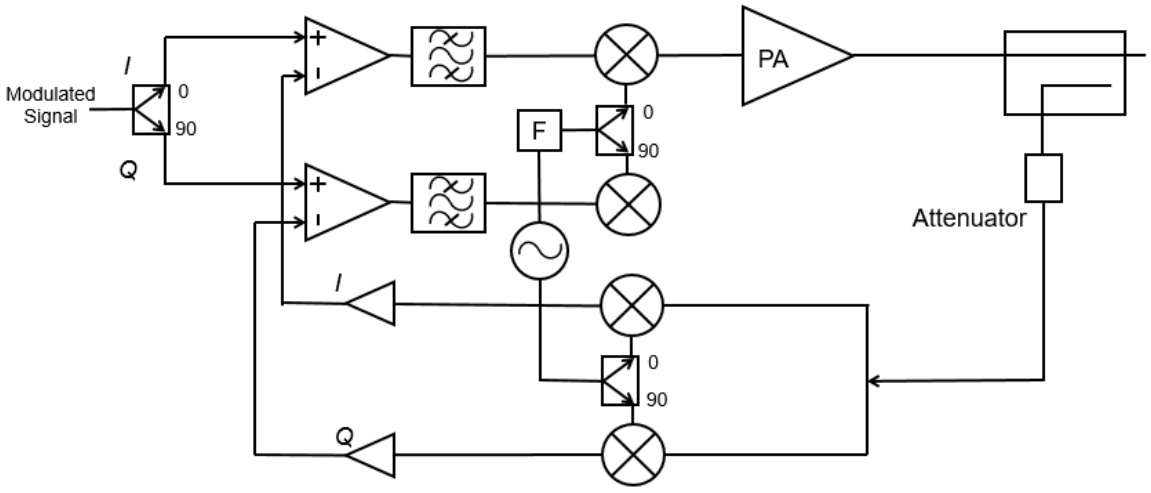


Figure 2.4 Cartesian feedback linearization

2.3.3 Predistortion Linearization

Prior to going into details concerning DPD, this section provides a review of fundamental predistortion theory as well as the general challenges faced when applying predistortion. Different to feedforward and feedback techniques, predistortion linearization focus on generating a nonlinear transfer characteristic to create an inverse nonlinear response of the PA. In terms of implementation, predistortion can be categorized into two groups: analogue predistortion and digital predistortion. Most predistortion linearization is now implemented in digital domain, and as such, can be relatively easily integrated in current base-station hardware and software platforms. However, there are still applications involving where analogue predistortion is more appropriate, as this type of linearizer is capable of operating over wide bandwidth and in a more cost-effective way to their digital counterparts.

Basically, despite of the different implementations, analogue and digtal predistotion share the same fundamental system structure as shown in Figure 2.5. The distortion correction mechanism can be illustrated in both frequency domain

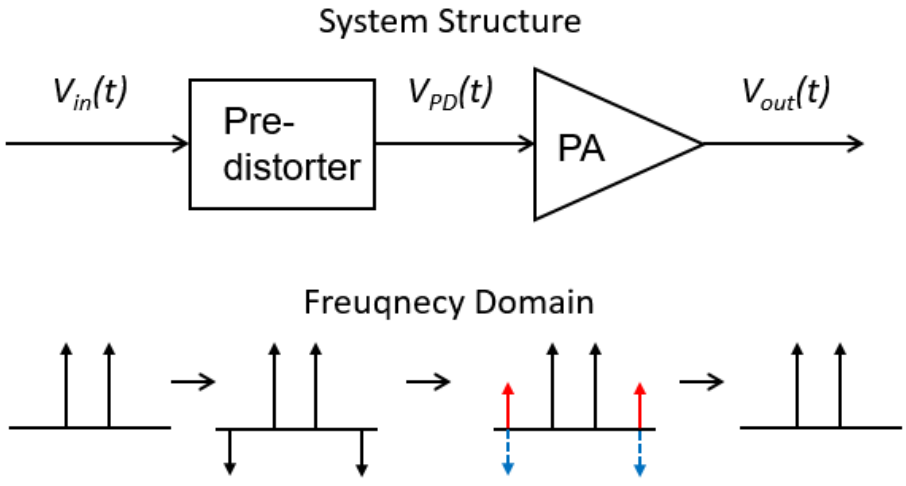


Figure 2.5 Predistortion system structure and frequency domain operation

as shown in Figure 2.5 or gain and AM/AM reponses as shown in Figure 2.6. The PA begins to distort when driven towards compression for high power and high efficiency. A user defined artifitial nonlinear block is inserted as pre-distorter, just before the input of the PA. This block generates the predistoted signal required to compensate the intrinsic distortion, which results in a linear transfer response for overall system behavior. The challenge for achieving high quality correction lies in implementing a very precise pre-distorter, which involves accurate modelling of the PA in algorithm and the output capability such as power and bandwidth considerations in hardware implementation.

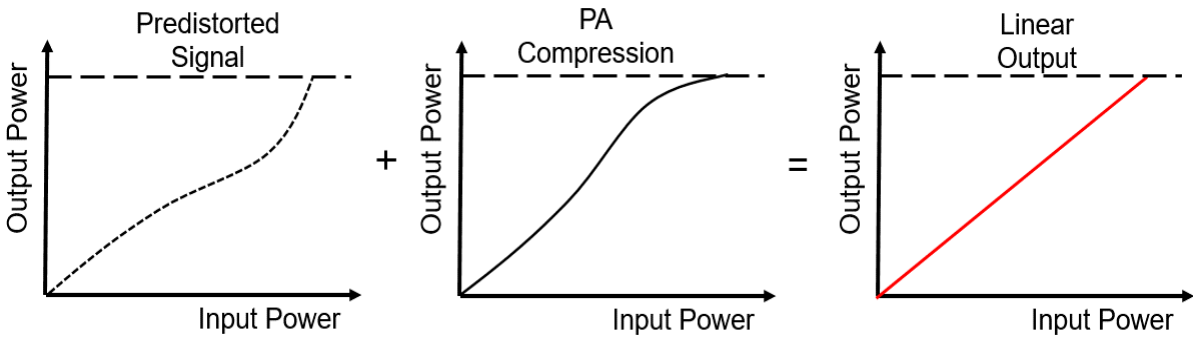


Figure 2.6 Gain or AM/AM responses of predistortion

Notably, predistortion is only capable to compensate for the power levels close to saturation. For example, to compensate a higher power level such as strong 3dB compression, the DPD algorithm will theoritically work out a compensation signal with similarly strong 3dB expansion. This scenario would place unrealistic demands on the ouput capability of the pre-distorter in gerating such an unrealistic and “peaky” signal which eventually heat up the power transistor. This results in a marked deterioration of the linearization performance [42], [43].

Referring to the frequency spectrum shown in Figure 2.5, the function of the pre-distorter is to add anti-phase intermodulation (IM) products to cancel the intrinsic IM distortion in the PA. Therefore, this nonlinear pre-distorter must be able to generate frequency components at those IM frequency. For acceptable levels of linearization, this requires a bandwidth of typically 5 times than the signal modulation bandwidth, or IF (intermediate frequency) bandwidth. Consequently, the predistorter must be designed to accommodate the signal over this frequency bandwidth to achieve the best predistortion results as well as the PA. This bandwidth requirement can place a significant design burden on DPD systems.

Analogue predistortion has been reported to be mostly used in spaceborne communications and early cell-phone handsets due to its advantage in wideband operation and small physical size. Digital predistortion is used widely, later, in cellular communications with the advantage of flexibility, high performances and compatibility with digital modulations, which will be discussed in more details in next section. It's noteworthy that, recently, there are a few reports to exploit analogue predistortion or hybrid solutions for cellular application which are able to reduce the complexity and power consumption of linearization for 5G applications in small-cells [44]–[48].

2.3.4 ET-like Envelope Linearization

While ET is normally focused on efficiency enhancement, there are a few reports exploiting ET modulator to linearize the RFPA [49]–[56]. This technique expands the PA from a two-port model into a three-port model, which utilises the presence

of the ET modulator (supply modulator) to inject baseband distortion correction signals into the output, to cancel the intrinsic distortion in the RFPA. By engineering the mixing effect at the drain of the PA, the supply modulation induced distortion can be chosen to be the same magnitude and anti-phase compare to the intrinsic distortion. In this way, an envelope signal injection through ET modulator is capable of linearizing some of the distortions in RFPA. This technique is able to be integrated into ET system using the correct shaping function [55].

However, different from the envelope elimination and restoration (EER) technique or the polar transmitters, there is no AM and PM separation at the input of the ET architecture. As a result, the envelope injection has limitation to linearize the PM behaviour of the RFPA [53].

2.4 Digital Predistortion

Due to its flexibility, excellent performance and compatibility with digital modulation scheme, DPD has been recognised as the preferred linearization technique for

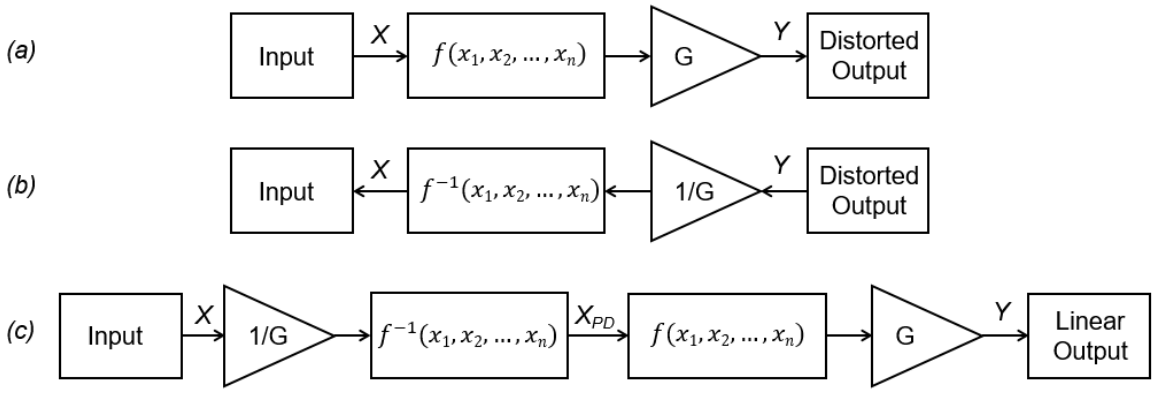


Figure 2.7 Block diagram of DPD concept

modern communication systems. As for the approach used in predistortion, DPD follows the basic operation concept in Figure 2.6. Taking advantage of digital signal processing (DSP), DPD algorithms mainly focus on creating an inverse model of the nonlinear PA. For instance, as shown in Figure 2.7, the nonlinear behaviour of the PA is first modelled using a nonlinear function . Acknowledging the memory effects present in all RFPAs, in practice, the response of the PA is not only dependent on the input signal at that instant, but also on the local history of the input signal, . Thus, the nonlinear PA model becomes . This model is then numerically inverted to get the inverse model . Lastly, an inverse function block is inserted before the nonlinear PA which generates a pre-distorted signal . The output of the PA then equals . In fact, the model inversion is operated in a normalized space to maintain the linear gain of the PA. Ideally, following this approach, the output of the PA should be as linear as the input signal.

As shown in Figure 2.8, the implementation of DPD in a PA architecture involves the conversion between the digital data domain to the RF signal, e.g.

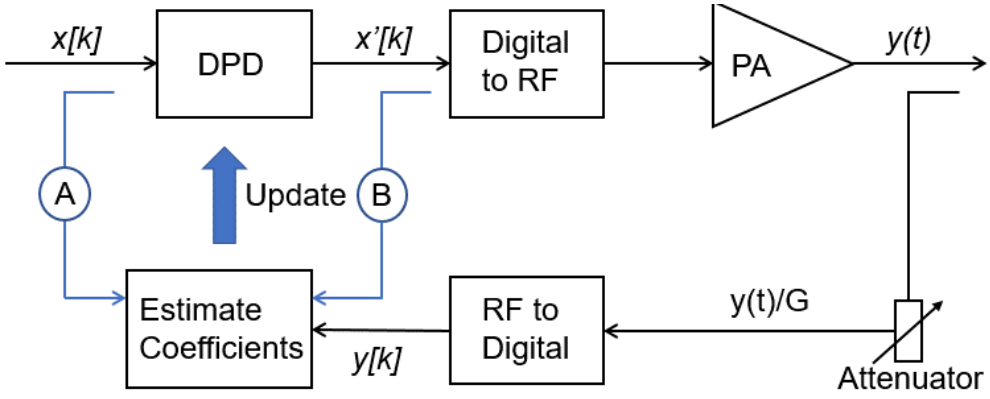


Figure 2.8 Schematic of DPD implementation in the PA transmitter architecture

"Digital to RF" and "RF to Digital". In the schematic, these two subsystems comprise not only digital to analogue converters (DAC) and analogue to digital converters (ADC), but also the modulators for up-conversion and down-conversion. The quality of these subsystem limits the performance of a DPD system in terms of bandwidth, resolution, quantization error and power consumption. In addition to the hardware limitations, it's worth mentioning, the accuracy and stability of the numerical DPD model determines the performance of DPD linearization. Many researches have and are focusing on building accurate models for simple hardware implementation [21], [57], [66]–[70], [58]–[65].

2.4.1 DPD Behaviour Models

From the 1990s, DPD system started to use a lookup table (LUT) model which is simply used to characterize the PA in terms of AM-to-AM and AM-to-PM behaviours [58], an implementation that is normally static or memoryless. Although it has been experimentally verified that, in most cases, the intrinsic nonlinear behaviours contribute the most of distortion instead of memory effects [21], engineers found that the memory effects, which include simple electrical memory and the contribution from nonlinear parasitics, start to play an increased role in wideband systems, and so memory-aware models are desired in wideband DPD.

The Volterra series (VS) model [62] is a very good candidate for accurately modelling nonlinear systems with memory effects. The model can be mathematically represented in Equation (2.1)

$$y_{VS}(n) = \sum_{k=1}^{K} \sum_{m_1=0}^{M} \dots \sum_{m_k=0}^{M} h_k(m_1, \dots, m_k) \prod_{j=1}^{k} x(n-m) \tag{2.1}$$

As the model is inherently in polynomial and linear in parameters (LIP), linear techniques such as least squares estimation (LSE) can be used to determine the coefficients values. This makes the model extraction and inversion straightforward in numerical domain. Unfortunately, in the original VS model, the complexity and the numbers of coefficients significantly increases when higher modelling accuracy is needed, where the nonlinear order and memory length increases. This makes the VS model less favoured in practical hardware implementation. Simplification and pruning are used to reduce the complexity of full VS model.

The memory polynomial (MP) model is a special case of VS model [63],[69], where the complexity is dramatically reduced. The generalized MP (GMP) model expands the modelling capability of the MP model and is currently a widely used DPD model [70]. A practical and efficient pruning technique, called dynamic deviation reduction (DDR) simplifies the VS model considering up-to 2nd order dynamics and maintains good performance [64], [21], [68]. There are also models involves two cascade stages and finite impulse response (FIR) filters, such as Wiener model [66], [67] and Hammerstein model [57], [61], [65]. Artificial neural network (ANN) models are also used due to their universal modelling capability for nonlinear systems [59], [60]. As ANN is not a linear-in-parameter (LIP) model, it requires an iterative training process which increases the computational complexity. The performance of these DPD models is summarized in terms of accuracy, complexity and capability in Table 2-3 [19].

Table 2-3 Performance summary of DPD models

DPD Model	Accuracy	Complexity	Capability	LIP
VS	Excellent	High	Strong	Y
MP	Satisfactory	Low	Medium	Y
GMP	Excellent	Medium	Medium	Y
DDR	Satisfactory	Low	Medium	Y
Wiener	Reasonable	Low	Weak	Y
Augmented Wie-	Satisfactory	Medium	Medium	N
Hammerstein	Reasonable	Low	Weak	Y
Augmented	Satisfactory	Medium	Medium	N
ANN	Excellent	High	Strong	N

2.4.2 DPD Coefficients Update Schemes

Most DPD systems used in base stations now are adaptive. The inverse model coefficients are continually updated to achieve dynamic cancellation for linear operation. Referring back to Figure 2.8, there are typically two paths used in coefficients updating, annotated as Ⓐ for “direct learning” and Ⓑ for “indirect learning”.

In direct learning, the pre-distorter DPD is inside the feedback loop while in indirect learning it is outside. The indirect learning scheme has proven to be an extremely popular adaptive estimator in DPD because of ease of implementation and fast convergence. Recently, in [40], indirect learning is compared with the close loop direct learning approaches, where it is revealed that the simplicity from indirect learning relies on an assumption that the pre-inverse of the nonlinear system is exactly the same as post-inverse. This is only valid when the input and output signals are very close with only minor deviations. It’s not applicable for the PAs in a wideband system or with strong saturation behaviour. Experiments on dual-loop model extraction were introduced [71] to improve model estimation accuracy. However, concluded in [40], indirect learning has inherent limitations such as high ADC sampling rate, coefficient offsets caused by EVM and PA saturation.

2.4.3 5G Challenges for DPD

As discussed earlier in this chapter in Section 2.2, DPD has been recognized as a fundamental linearization technique in current communication systems. However, there remains significant challenges for their use in the emerging 5G networks. Long-term evolution (LTE) carrier aggregation (CA) will push signal bandwidth to 100 MHz and beyond. The higher bandwidth and higher PAPR generally means a more complex DPD algorithm with many more coefficients. Moreover, as the DPD must accommodate bandwidths at least 3 to 5 times to the signal bandwidth, high speed RF ADCs and DACs are required in hardware implementation. Accurate and efficient models are required for further complicated applications such as multiband concurrent PAs and massive MIMO, and a low power and compact adaption for small-cells is in high demand. DPD is a good solution, for now, but it is unlikely to be a complete solution for the linearization of future communication systems and may well need some help from other techniques.

2.5 GaN Device Technology for RFPA

Gallium nitride (GaN) is a binary III/V direct bandgap semiconductor compound which is most commonly used in illumination (LED lights) and Blu-ray players. The atoms in GaN are bonded by a very ionic gallium-nitrogen chemical bond which produces a bandgap of 3.4eV. In comparison with the traditional materials used in RF device, gallium arsenide (GaAs) has a bandgap of 1.4eV and silicon has a bandgap of only 1.1eV. This large bandgap allows GaN device to tolerate a higher breakdown fields which allows the device operating at high supply voltages. Electrons in GaN have a high saturation velocity, which, in turn, delivers high current density. With high voltage and high current density, GaN devices are ideal for high power operation. Quantitatively, GaN devices can operate at typically 5 times higher voltage and can deliver twice as much current than GaAs devices. In other words, GaN devices can deliver 10 times more power density than GaAs devices [72]. High power density results in a high heat flux which requires careful thermal management. Fortunately, when integrated on high thermal conductive materials like silicon carbide (SiC), GaN-on-SiC devices exhibit outstanding thermal properties with thermal conductivities 6 times higher than that of GaAs and 3 times higher than that of silicon (Si). All these features make GaN devices very capable candidates to work under high power and high temperature with high reliability.

Moreover, high power density also means smaller devices in certain applications. In RF circuit design space, this means the transistor has lower intrinsic capacitance and lower combining losses, which suits applications for wider

bandwidth and higher efficiency. Therefore, GaN devices are considered as ideal candidates for future RF applications [6], [72], [73].

Before GaN, LDMOS technology dominated in the base station RFPA applications [74]. Table 2-4, summaries the difference between GaN and LDMOS [75]. Although LDMOS still dominates lower-frequency applications below 4GHz, currently due to the lower cost/Watt, in high frequency applications, where the packaging and assembly become more costly than sole die, the advantage of GaN is expected to be significant which overtakes the markets. As is shown in Table 2-4, parasitic capacitance has big impact on the bandwidth of the matching and then the efficiency. In addition of these, the linearity performance is also affected by the parasitics.

Table 2-4 GaN vs. LDMOS

Parameter	**GaN**	**LDMOS**
Full form	Gallium Nitride	Laterally Diffused MOSFET
Applications	➢ GaN provides high efficiency, power density and higher gain within a smaller package ➢ Suitable for broadband applications due to higher output impedance and lower Cds capacitance ➢ Advantages: GaN transistors have small parasitic capacitance, which results in easy wideband matching comparing to LDMOS transistors of identical power level	➢ LDMOS is used for cellular and broadcast narrowband applications due to high power and efficiency ➢ LDMOS(50V) is used for <1.5 GHz applications while LDMOS (28V) is used for frequencies upto 4GHz ➢ Disadvantages: LDMOS transistor has large Cgs/Cds capacitance due to large peripheral inherently which limits application bandwidth.
F_{max} (GHz)	30 GHz (50V)	22 GHz (28V) 15 GHz (50V)
Power Density (W/mm)	5-10 (50V)	0.8 (28V) 2 (50V)
Efficiency at P_{1dB} (%)	70 (50V)	60 (28V) 55 (50V)
Bandwidth (MHz)	500-2500 (50V)	100-400 (28V) 100-500 (50V)
Cds (pF/W)	¼ smaller for GaN	0.23 for LDMOS (28V) ½ smaller for LDMOS (50V)
Cgs (pF/W)	½ smaller for GaN	0.94 for LDMOS (28V) ½ smaller for LDMOS (50V)

Investigations into the linearity performance of GaN devices due to the non-linear capacitances are summarized in [76]. Semi-analytical expressions are derived for AM/AM and AM/PM distortions, considering all the known nonlinear elements in the equivalent circuits. Notably, with low dielectric constant, GaN devices has smaller capacitances with lower dependence on voltage. This nonlinearity from parasitic capacitance determines AM/PM performance of the RFPA.

The device level modelling and analysis involves understanding the fundamental nonlinear sources, including drain-source current generator I_{ds}, the gate-source capacitance C_{gs},the drain-source capacitance C_{ds} and the gate-drain capacitance C_{gd}. Together, these intrinsic nonlinear sources in the device are the origin of the distortion within RFPA and will ultimately limit the *linearizability* of the circuit.

2.6 Linearizability

Although the term, *linearizability* has been generally used to describe the linearization performance of the RFPA [77]–[80], to the author's knowledge, there is no universally recognised definition within the industrial and research communities. There is therefore a need to clarify, that in this thesis, *linearizability* has a definition that depends on the objective being discussed.

From an RFPA circuit design perspective, *linearizability* describes how well the RFPA responds to a linearization technique. Through certain design

methodologies, such as bias optimization and impedance termination optimization, it is possible to improve the *linearizability* of the RFPA. This emphasises the importance of circuit design and optimization when using a fixed linearization technique.

From a linearization technique perspective, *linearizability* describes the capability of this linearization technique itself. For example, Table 2-3 describes the relevant *linearizability* of different DPD benches. In this case, it emphasises the capability of a certain linearization techniques applied on an established RFPA circuit.

2.7 Conclusion

In order to achieve a linear amplification for a RFPA in a communication system, the designers are required to have a good knowledge of the i) device and circuit level distortion generation and ii) the system level linearization used. This thesis consists of an analysis of both these parts. In the next few chapters, both circuit level analysis and envelope linearization techniques will be presented and an attempt to address *linearizability* on these two levels along with linearity analysis.

Chapter 3 Power Amplifier Design

3.1 Introduction

This chapter will focus on the topics around PA design methodology. It will include conventional PA design strategies, which start with fundamental transistor characteristics and then the use of simple load-line methodology. Finally, different classes of PA are analysed together with their performance in terms of both efficiency and linearity.

3.2 Power Amplification

As shown in Figure 3.1, the block diagram of a PA consists of three key design elements: the input matching network, the transistor or active device (PA core) and

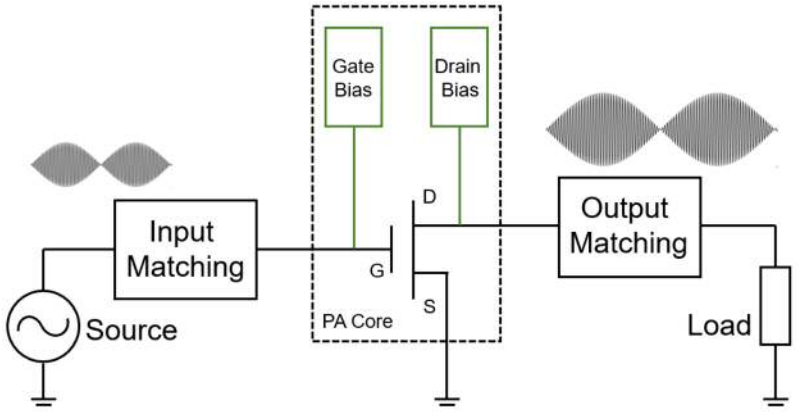


Figure 3.1 Block diagram of a FET power amplifier

the output matching network. The signal from input source is amplified by this circuit and then delivered to the load. The transistor must be biased appropriately to establish the correct quiescent state and to allow the signal to be properly amplified, and provide sufficient output power, gain and efficiency.

From the perspective of energy conversion, power amplification can be viewed as the result of DC to RF power conversion in this circuit responding to an input stimulus. The drain efficiency of the PA can be calculated from Equation (3.1).

$$\eta_{drain} = \frac{P_{RFout}}{P_{DC}} = \frac{P_{RFout}}{V_{DC} \times I_{DC}} \tag{3.1}$$

η_{drain} is effectively the ratio between overall RF output power delivered and the DC power consumed, which describes the efficiency of the power conversion process. Another common figure-of-merit is power added efficiency (PAE), which can be calculated using Equation (3.2).

$$\eta_{PAE} = \frac{P_{RFout} - P_{RFin}}{P_{DC}} = \frac{P_{RFout} - P_{RFin}}{V_{DC} \times I_{DC}} \tag{3.2}$$

The PAE takes the power added by the amplifier in the numerator instead of overall RF output power. In theory, an amplifier with infinite gain will have PAE equal to the drain efficiency. In practical, for a real amplifier which has a gain of 20dB (100 in linear scale), the input power will be 1% of output power. In this case, PAE will be very close to, but a little lower than the drain efficiency. Generally, the

conversion between drain efficiency and PAE is expressed in Equation (3.3), where G is the power gain of the amplifier.

$$\eta_{PAE} = \eta_{drain} \frac{G-1}{G} \tag{3.3}$$

3.3 IV Curve and Load-line in PA Design

In Figure 3.2, the load presented to the transistor Z_{Load} comprises the output matching network and the load. In most communication systems, the actual load is normally diplexer or circulator in 50Ohm. The output matching network is designed to transform the 50Ohm load to the optimum load impedance Z_{Load} presented to the transistor at reference plane-A. As the biasing networks are designed not to impact the RF band of interest, Z_{Load} at reference plane-A should

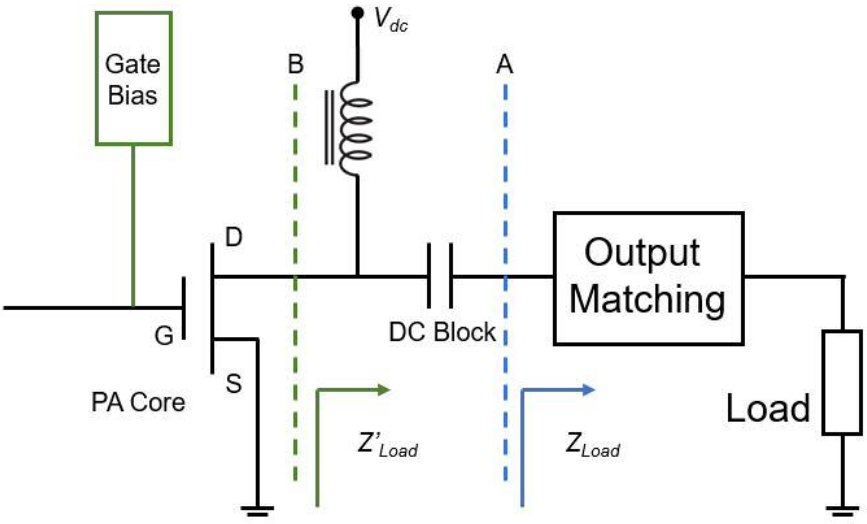


Figure 3.2 Power amplifier output load

completely determine the RF impedance presented to the device. In fact, Z'_{Load} will be used for precise RFPA design, which will include any impact from baseband impedance introduced by the output biasing network, and is of special interest when considering biasing induced electrical memory effects for example. In order to, analyse and optimize circuit performance, there is a need to introduce two important design tools: I-V curves and Load-line.

The DCIV characteristic has become a standard device characterisation measurement in amplifier design. By sweeping the DC voltage at Gate (V_{GS}) and Drain (V_{DS}) and measuring the static drain source current I_{DS}, a DCIV mesh can be plotted, as shown in Figure 3.3. This mesh defines the transistor's operational "space" with three important characteristics: I_{Max}, which defines the highest possible drain current; V_{Max}, which defines the highest achieved drain voltage and "knee" region, which describes a V_{DS} depended transistor turn-on effect and then defines the lowest value of output voltages V_{Min}. When an RF excitation signal is applied to a properly biased transistor, plotting the output current I_{ds} vs. the output voltage V_{ds} forms the so called "load-line" which maps the transistor operation onto the DCIV mesh.

As shown in Figure 3.3, a load-line of a class-B biased transistor is plotted (Red). Notably, referring to the load-line, the turn-on point at V_{bias} is determined by the quiescent drain biasing voltage in the circuit. The slope of the load-line corresponds to the choice of the load. This load-line, along with the DCIV mesh

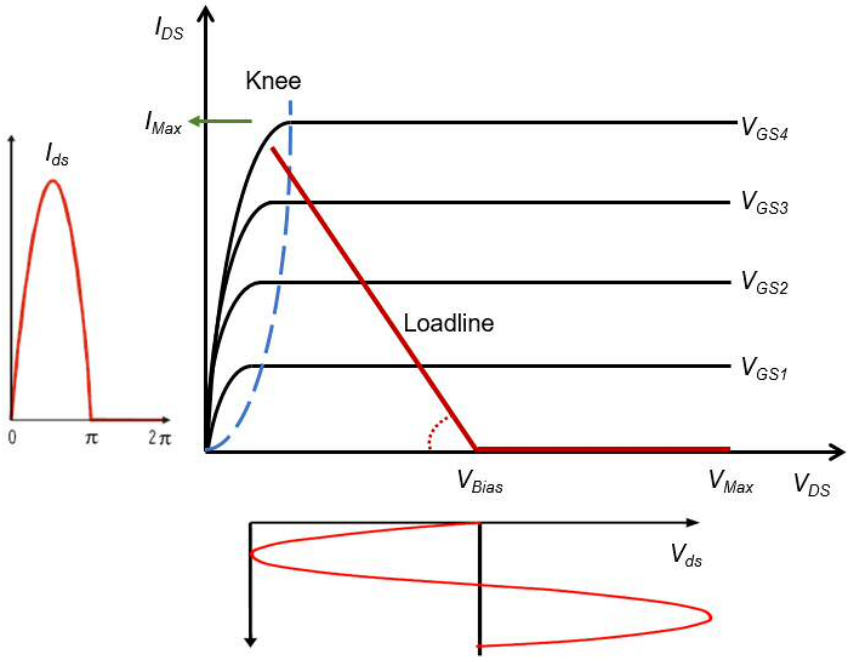


Figure 3.3 DCIV and Load line

describe the operation of this PA within a device's determined design space, or in other words, boundary conditions.

Optimizing the PA for maximum output power requires the load-line to fully occupy the transistor's operational space, which is defined by I_{Max}, V_{Max} and V_{Min} (knee). In theoretical PA design, the DCIV and load-line are also used to predict the efficiency from a waveform perspective. The design strategy exploiting the load-line theory will be discussed in detail in later sections.

Besides power and efficiency, DCIV and load-line analysis can also indicate performance in terms of linearity. As shown in Figure 3.4, when the PA is driven into compression, the load-line extends into knee region and diverts downward. Although this possibly provides higher efficiency, it results in unfavoured clipping and distortion in current and voltage waveforms, generating strong nonlinear

behaviours. In a more practical case, the load impedance is not necessarily pure resistive but contains reactive parts, which results in load-line looping effects, as shown in Figure 3.5. Although reactive loading is often employed intentionally when RFPAs are required to deliver wideband operation while still maintaining high power and efficiency, this involves RF memory (reactive impedance in RF band) and electrical memory (reactive impedance in baseband), which deteriorate signal linearity.

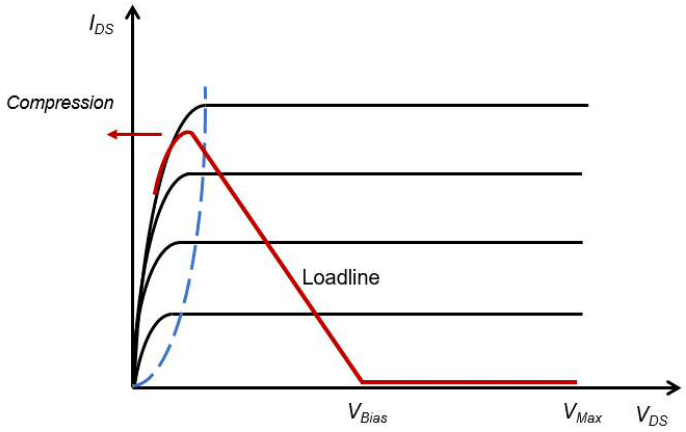


Figure 3.4 Hard drive Load-line

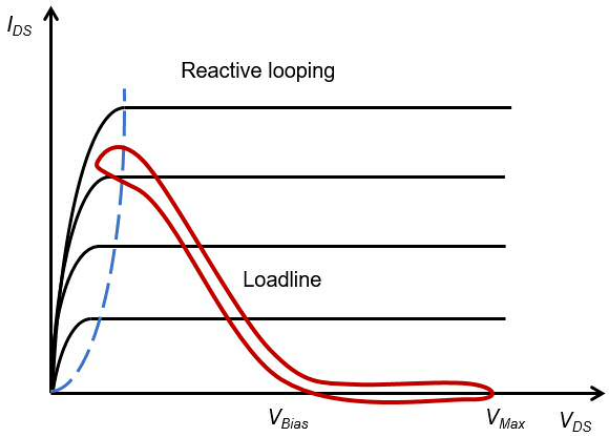


Figure 3.5 Reactive looping Load-line

3.4 Transfer Characteristic and Two-dimensional Function

In a theoretical sense and for the work in this thesis, the core of the power amplifier is a FET transistor, and can be considered as a voltage control current source, the output current is determined by input voltage which can be described by a transfer characteristic function $f(V_{gs})$.

$$I_{ds} = f(V_{gs}) \tag{3.4}$$

As shown in Figure 3.6, the plot of a transfer characteristic can be basically divided into two strongly nonlinear regions: the turn-on and saturation regions, and one weakly nonlinear region between these two. In weakly nonlinear region, the transfer function can be modelled as a power series with nonlinear transconductance as coefficients.

$$I_{ds} = g_{m0} + g_{m1} \cdot V_{gs} + g_{m2} \cdot V_{gs}{}^2 + g_{m3} \cdot V_{gs}{}^3 + ... \tag{3.5}$$

In practical, for FET devices, the drain current does not only depend on the gate voltage V_{gs}, but also the drain voltage V_{ds}. In other words, the FET-based PA needs to be considered as a two-port network which features a two-dimensional (2D) transfer function,

$$I_{ds} = f(V_{gs}, V_{ds}) \tag{3.6}$$

Equation (3.6) shows how the drain current can be better described by a two-dimensional power series, where $K_{2_{gm\&g1}}$, $K_{3_{2gm\&g1}}$, $K_{3_{gm\&2g1}}$ are the cross coefficients,

$$\begin{aligned} I_{ds} &= g_{m0} + g_{m1} \cdot V_{gs} + g_{m2} \cdot V_{gs}^{\ 2} + g_{m3} \cdot V_{gs}^{\ 3} + \ldots \\ &+ g_0 + g_1 \cdot V_{ds} + g_2 \cdot V_{ds}^{\ 2} + g_3 \cdot V_{ds}^{\ 3} + \ldots \\ &+ K_{2_{gm\&g1}} \cdot V_{gs} \cdot V_{ds} + K_{3_{2gm\&g1}} \cdot V_{gs}^{\ 2} \cdot V_{ds} + K_{3_{gm\&2g1}} \cdot V_{gs} \cdot V_{ds}^{\ 2} + \ldots \end{aligned} \tag{3.7}$$

With some simplification, Equation (3.6) can be modelled as the products of two separate functions, which depending on V_{gs} and V_{ds} separately,

$$I_{ds} = f(V_{gs}) \cdot g(V_{ds}) \tag{3.8}$$

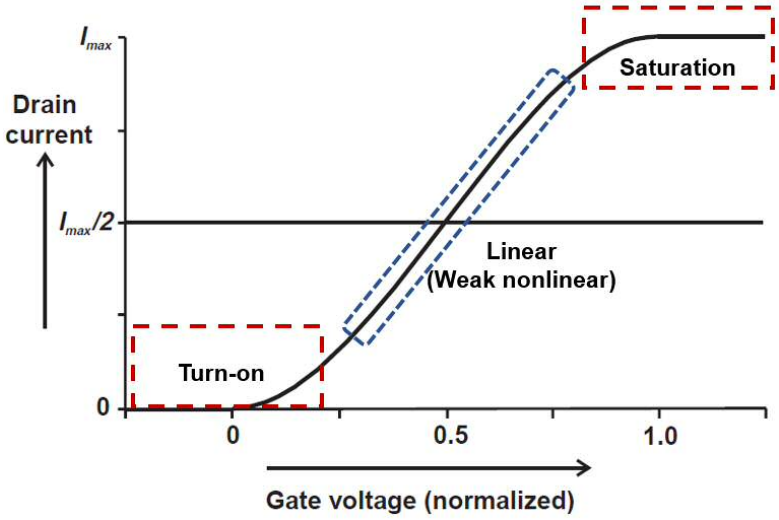


Figure 3.6 PA transfer function

Equation (3.8) is a widely used model for a two dimensional transfer function, where the descriptions of $f(V_{gs})$ and $g(V_{ds})$ are not necessarily limited to polynomial power series, for instance, in the Angelov Model [81], [82].

In the transfer function, for most classes of operation, an abrupt turn-on effect normally results in better efficiency because of less residue overlap between voltage and current waveforms, due to a more ideal rectified current waveform.

However, this idea current behaviour leads to worse linearity as significant higher order terms appear in the power series. In practice, non-switch mode PAs need to operate slightly in the two nonlinear regions to achieve higher efficiency, while needing to remain in the weakly nonlinear region for linearity consideration.

3.5 Classes of Power Amplifier

Generally, RFPA can be categorised into four fundamental classes of operation; class-A, class-AB, class-B and class-C. These classes are established by selection of specific input bias voltages, as shown in Figure 3.7. The definition of bias point can also be summarized using the concept of conduct angle [8]. Figure 3.8

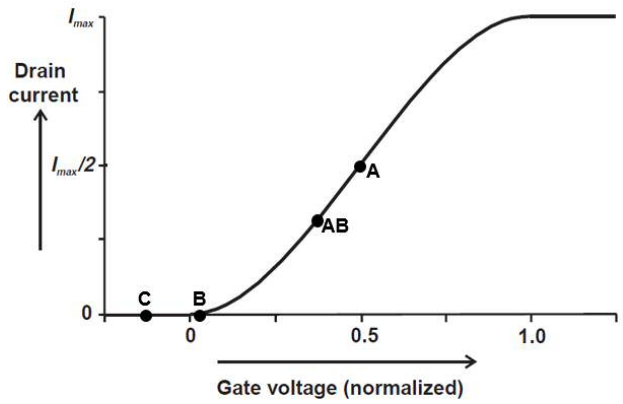


Figure 3.7 Power amplifier classes with different bias point, copy from reference [8]

illustrates the concept of conduction angle with the plot of a sinusoid input gate voltage waveform V_g and in response, an output drain current waveform I_d. The transistor is biased at $V_g = V_q$ and its intrinsic threshold voltage is V_t. In this example, the transistor generates a drain current only when the V_g is above V_t.

The conduction angle α describes the portion of RF cycle where the transistor output current is above zero. The drain current waveform can be analytically written as:

$$\begin{aligned} i_d(\theta) &= I_q + I_{pk} \cdot \cos\theta, && -\alpha/2 < \theta < \alpha/2; \\ &= 0, && -\pi < \theta < -\alpha/2; \alpha/2 < \theta < \pi \end{aligned} \tag{3.9}$$

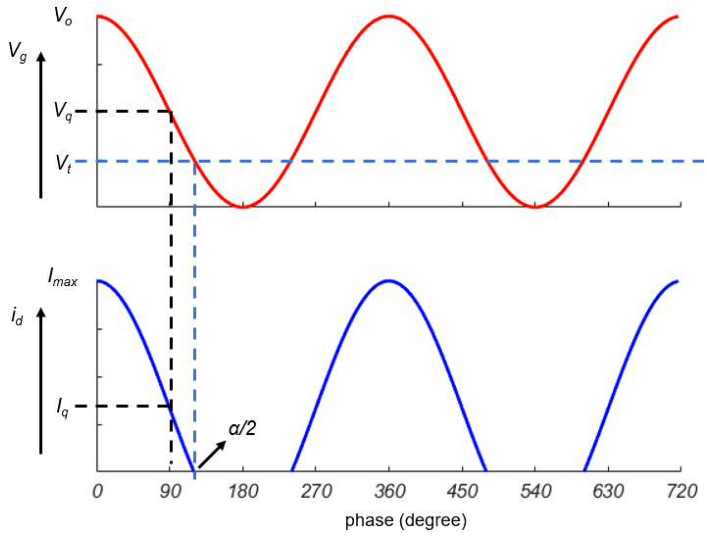


Figure 3.8 Concept of conduction angle

where $\cos(\alpha/2) = -\dfrac{I_q}{I_{pk}}$, and $I_{pk} = I_{max} - I_q$, so

$$id(\theta) = \frac{I_{max}}{1-\cos(\alpha/2)} \cdot (\cos\theta - \cos(\alpha/2)) \tag{3.10}$$

The DC component can be calculated as,

$$I_{dc} = \frac{1}{2\pi} \cdot \int_{-\alpha/2}^{\alpha/2} \frac{I_{max}}{1-\cos(\alpha/2)} \cdot (\cos\theta - \cos(\alpha/2)) \cdot d\theta \tag{3.11}$$

and the magnitude of *n*th harmonic is

$$I_n = \frac{1}{\pi} \cdot \int_{-\alpha/2}^{\alpha/2} \frac{I_{max}}{1-\cos(\alpha/2)} \cdot (\cos\theta - \cos(\alpha/2)) \cdot \cos n\theta \cdot d\theta \tag{3.12}$$

Figure 3.9 shows the spectral behaviour of output drain current as a function of changing conduction angle, up to the 5th harmonic, for a normalized current waveform amplitude. Notably, the DC component decreases as the conduction angle is reduced. As a result, the efficiency increases. The theoretical drain efficiencies from different conduction angles are summarized in Table 3-1.

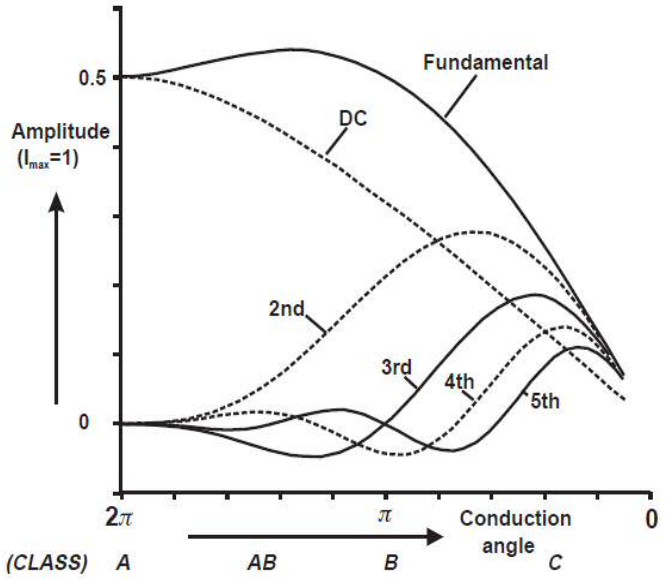


Figure 3.9 Fourier analysis of reduced conduction angle current components, copy from reference [8]

Table 3-1 PA classes with different conduction angle and efficiency

PA Class	Conduction angle	Theoretical Drain Efficiency
Class A	2π	50%
Class AB	$\pi < \alpha < 2\pi$	50-78.5%
Class B	π	78.5%
Class C	$0 < \alpha < \pi$	78.5-100%

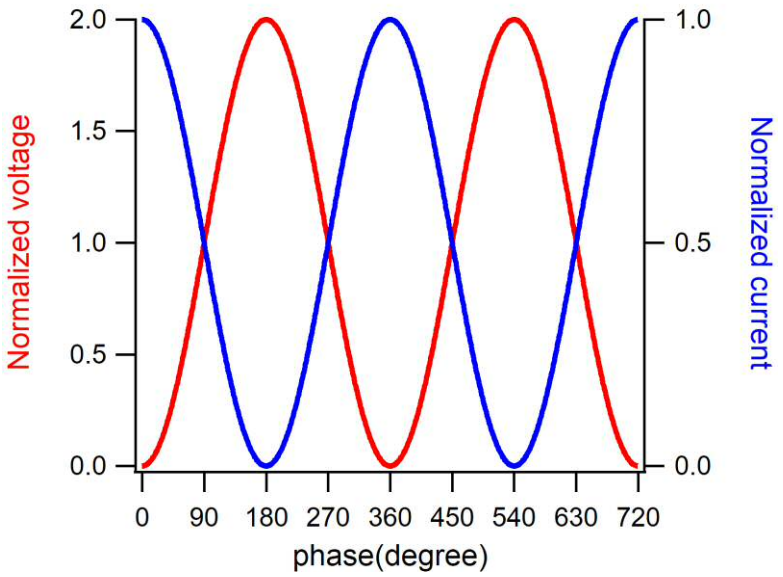


Figure 3.10 Simulated ideal Class-A drain waveforms ($\alpha = 2\pi$)

It is possible to take a closer look at the drain voltage and current waveforms of the RFPA of different classes of operation. The following waveforms for Class A, B, AB and C are obtained from an ideal formulation-based simulation, representing a perfectly trans-conductive device with no knee region. The theoretical

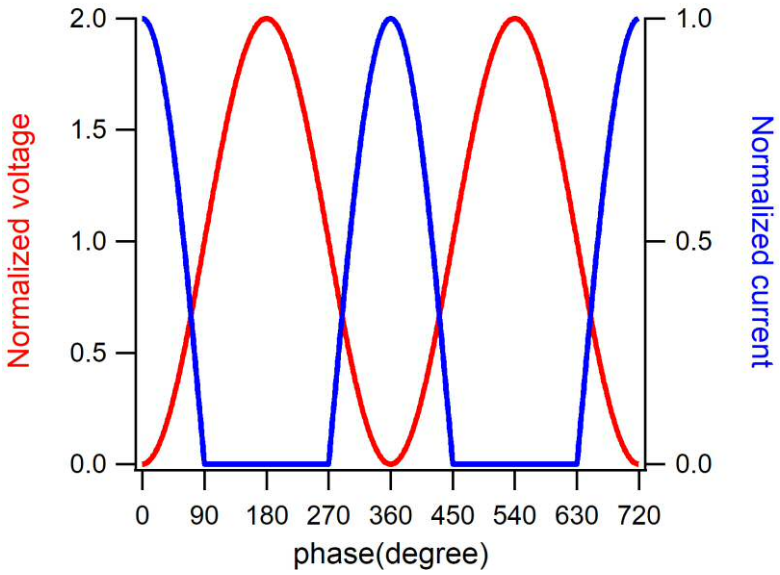


Figure 3.11 Simulated ideal Class-B drain waveforms ($\alpha = \pi$)

waveforms are obtained with an ideal fundamental load and with all other harmonics terminated in a short-circuit, resulting in a sinusoidal voltage waveform.

Class-A output waveforms are shown in Figure 3.10. The device is active all of the time in Class A with a conduction angle of $360°$. Theoretically, as there is no distortion involved in Class A waveform, it has been recognized as the linear PA. This waveform results in an ideal 50% drain efficiency.

In **Class-B**, as the input bias is set at the threshold voltage V_t, the device is active for only half cycle with a conduction angle $180°$. As shown in Figure 3.11, the drain current waveform is half rectified. The efficiency is then increased because of less overlap between voltage and current compared to Class A, which is in theory 78.5%. The increasing efficiency in Class-B is at a cost of gain and output power compared to Class-A. It should be noted that the current waveform is distorted due to the presence of harmonic current components.

Class-AB is a intermediate class in which the bias point is set between Class A and Class B. With reference to Figure 3.9, class-AB offers the highest fundamental current of all modes, and as a result is popular, as it offeres a good trade-off between power, linearity and efficiency. Typical waveforms for Class AB is shown in Figure 3.12, here with an efficiency of 67.6%.

In **Class-C**, V_g is biased below the threshold voltage V_t. Due to its low conduction angle, the drain efficiency of class-C is higher than class-B, but the

fundamental output power and gain is lower, impacting PAE. The waveforms are shown in Figure 3.13, in this case with an efficiency of 86.2%.

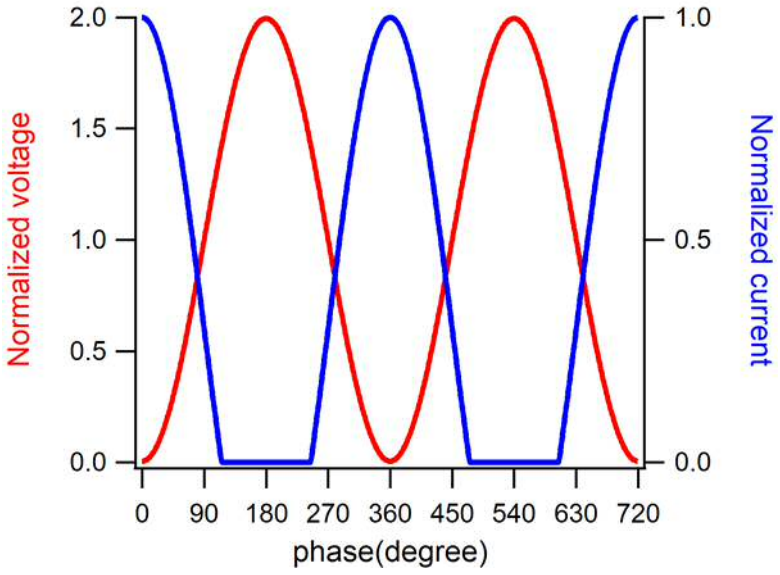


Figure 3.12 Simulated ideal Class-AB drain waveforms ($\alpha = \frac{6}{5}\pi$)

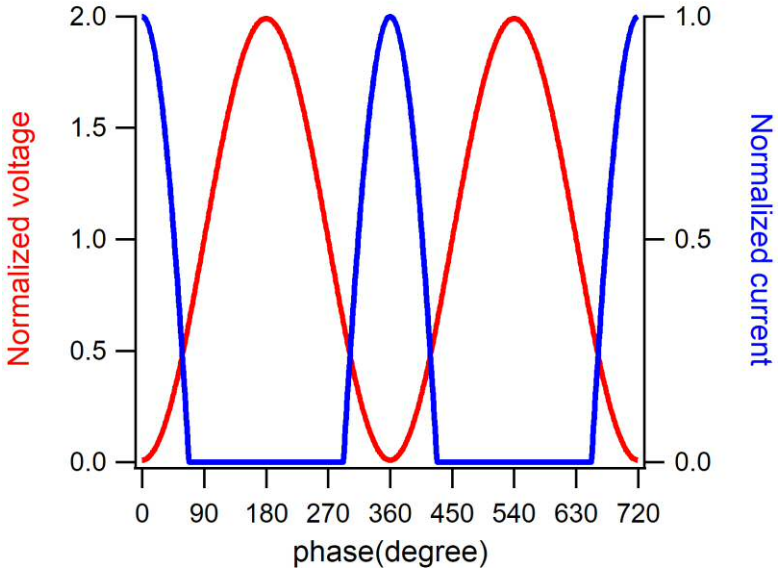


Figure 3.13 Simulated ideal Class-C waveforms ($\alpha = \frac{4}{5}\pi$)

Class-D and **Class-E** power amplifiers work in switching mode. In other words, the voltage and current waveforms are shaped through abrupt switching of the drain current, where ideally no overlap exists. In theory, these modes can achieve 100% efficiency. In practice, however, semiconductor switches have limitations in terms of transition time, although this doesn't impact significantly their benefits in low frequency applications [8]. Some high frequency applications are reported for GaN devices [83].

It is worth mentioning that, when comparing class-A, B, AB and C, the improvement in efficiency is achieved by reducing the conduction angle and then reducing the overlap between current and voltage waveforms. Instead of shaping the current waveform through selection of bias point, these classes use a similar approach where the voltage waveform conduction angle is controlled by changing the load condition. Early work in this area led to the invention of another high efficiency mode - harmonic-tuned Class F.

Class-F is achieved by using bias, drive level and harmonic termination to controlling the harmonic components in the output waveforms of an RFPA [84]–[87]. The current waveform is normally half rectified due to the threshold voltage biasing. The load is designed to be a low impedance (short-circuit) for even-order harmonics (current peaking) and a high impedance (open-circuit) for odd-order harmonics (voltage peaking). In this way, and given sufficient harmonics components in the current waveform, the drain voltage waveform can theoretically be shaped into a square wave, as shown in Figure 3.14. This waveform results in

100% efficiency. In practice, class-F PAs can usually be implemented with up to three harmonics controlled, as shown in Figure 3.15. This "realistic" set of waveforms yield an efficiency of 90.7%.

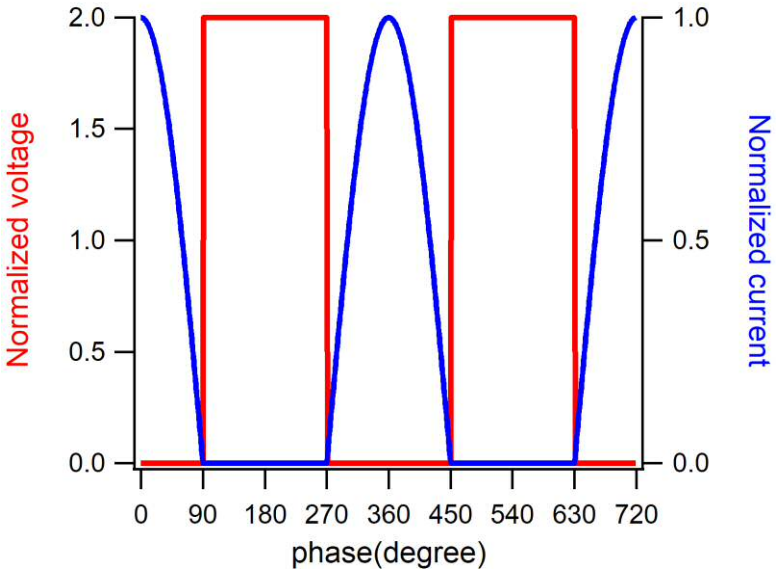


Figure 3.14 Class-F waveforms with infinite harmonics controlled

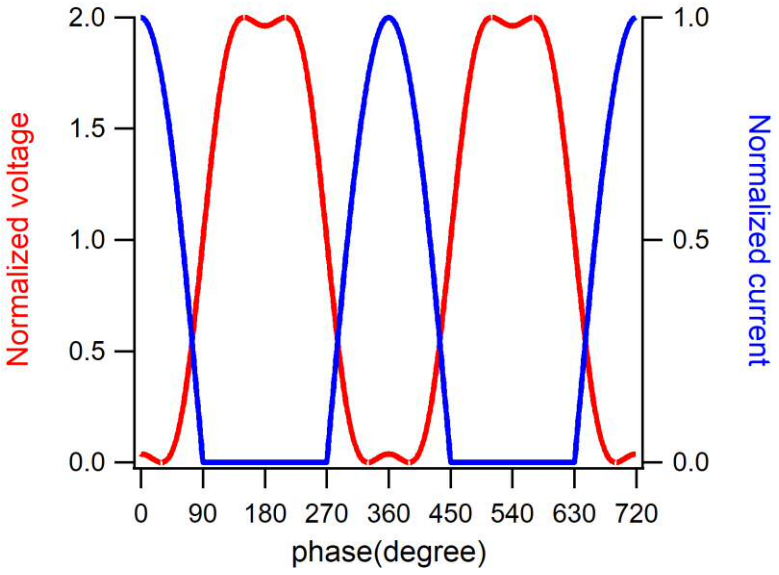


Figure 3.15 Class-F waveforms with up to 3rd harmonic controlled

Class-F^{-1} is similar in terms of its operation to class-F, where however the shape of the current and voltage waveform are reversed, with the current waveform shaped into square wave and voltage waveform shaped into a half rectified sinusoid waveform [85]–[87].

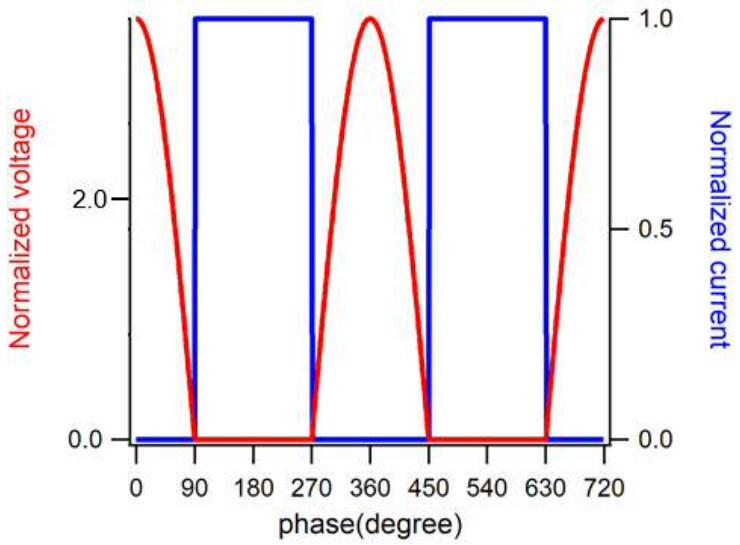


Figure 3.16 Inverse Class-F waveforms with infinite harmonics controlled

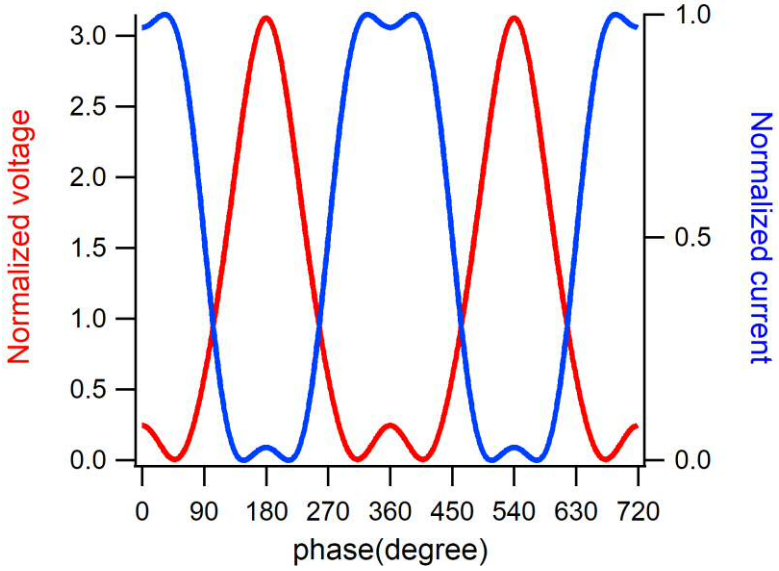


Figure 3.17 Inverse Class-F waveforms with up to 3rd order harmonics controlled

The waveforms with infinite harmonics controlled are shown in Figure 3.16. The high drain voltage is achieved because of the second harmonic peaking effect. The ideal waveform set results in 100% efficiency. Figure 3.17 shows a practical set of waveforms with up to 3rd order harmonic controlled, where the corresponding efficiency is 80.9%.

Class-J is recognised as the precursor of the **Continuous Modes** [88]. This is a specific mode that recognises that class-B efficiency and power can be maintained for a specific combination of reactive fundamental and second harmonic impedances. As an extension to class-B, class-J exploits this property, and with the correct matching network, has been demonstrated to achieve the same efficiency as class-B, but over a much wider bandwidth [88]–[91]. This concept is then developed into continuous class-BJ and continuous class-F for high efficiency broadband operation [92]–[97]. Explaining this in a little more detail, although traditional class-F and class-F^{-1} can achieve very high efficiency, they are bandwidth limited due to their need for singular, harmonic-tuned impedance terminations. Fortunately, exploiting continuous mode concepts, this limitation can be overcome by allowing the load impedance to vary reactively, with frequency, rather than adhering to strict open and short conditions. In theory, continuous operation is best defined by a sine function addition to form the output voltage.

$$V_{ContBJ} = V_{\max}(1-\cos\theta)\cdot(1-\alpha\sin\theta) \tag{3.13}$$

$$V_{ContF} = V_{\max}(1-\frac{2}{\sqrt{3}}\cos\theta+\frac{1}{3\sqrt{3}}\cos 3\theta)\cdot(1-\alpha\sin\theta) \tag{3.14}$$

$$-1 \le \alpha \le 1 \tag{3.15}$$

Considering only up to the 3rd order harmonics as a simplification, the drain voltage can be written as Equation (3.13) for continuous class-J and Equation (3.14) for continuous class-F. α is defined between [-1,1] here, in order to keep the voltage always above zero. Key to this concept, all these waveforms deliver the same efficiency, 78.5% for class-BJ and 90.7% for continuous class-F (3rd order), but become asymmetrical and "peaky", as shown in Figure 3.18. In this analysis, the traditional class-B and class-F modes are special cases of continuous modes, $\alpha = 0$.

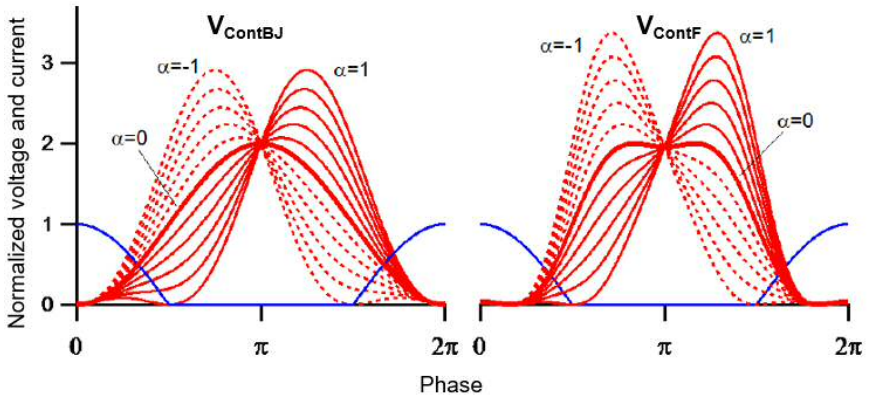


Figure 3.18 Class-BJ and continuous Class-F waveforms

Although Class-J is very attractive in wideband high efficiency application, the linearization towards this class needs more work. There is a need to mention, the waveform in Class-J can be asymmetry ($\alpha \neq 0$) and spend more time close to the knee (peaks). In addition, the reactive load condition may be problematic to simple DPD algorithms. The challenge in linearization of Class-J is beyond the scope of this chapter.

The discussion of the different classes mentioned above focuses primarily on efficiency and bandwidth. In addition to these key parameters, PA linearity also varies for the different classes. Briefly, class-A, -AB, -B and -C have different linearity performance due to the device nonlinear transconductance which is determined by bias point and drive level. Class-J and continuous class-F modes, for example, present IM products and *in-band* distortions due to the secondary mixing from harmonics components. Detailed analysis of this is included in Chapter 4, in discussions concerning circuit-level distortion.

3.6 Advanced Topologies for Power Amplifier

Generally, in conventional RFPA designs, the peak efficiency is achieved at or near the maximum output power level, when the device is in its compression region. This efficiency comes at the cost however, of unwanted distortion in the signal when achieving a desired output power. Achieving a good trade-off between efficiency and linearity usually requires the RFPA to operate at power back-off. This is more common in modern communication systems which have signals of high PAPR, where the RFPA is operating for relatively long periods at reduced power rather than the peak power. As briefly discussed in section 1.3, a number of high efficiency topologies have been introduced to enhance the efficiency in back-off power region, such as the Doherty and ET. This section will explain these two advanced topologies in more detail with a discussion on their relative linearity performances.

The topology of the Doherty power amplifier (DPA) was first introduced by William H. Doherty in the 1930's [98] . It was first used in valve-based AM transmitters where efficiencies of over 60% were achieved [99]. Starting from late 1990's, the DPA has been applied in base transceiver station (BTS) application for high efficiency and reduced operational running costs (OPEX). In theory, the DPA exploits active load modulation to achieve high efficiency over a broad range of power levels [8]. In other words, when the load-line is dynamically changed corresponding to the input drive level, high efficiency can be achieved over a broad range of power levels. The concept of active load modulation in the DPA is

illustrated in Figure 3.19 where two current sources are connected to a common load $R_L/2$. The equivalent impedances seen from each side of the common load are written in (3.16) and (3.17)

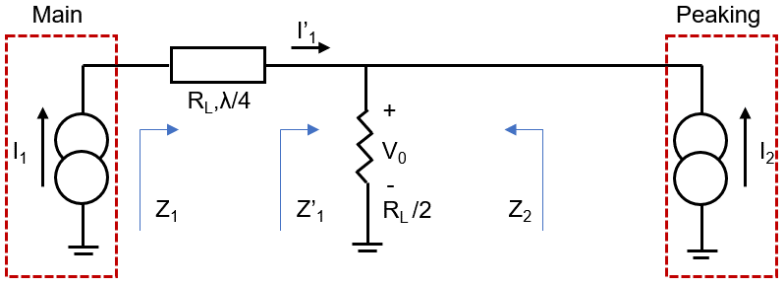


Figure 3.19 Active load modulation

$$Z_2 = \frac{R_L}{2}\left(1+\frac{I_1'}{I_2}\right) \tag{3.16}$$

$$Z_1' = \frac{R_L}{2}\left(1+\frac{I_2}{I_1'}\right) \tag{3.17}$$

For the Doherty to function, an impedance inverter, represented by the quarter wave-length transmission line ($R_L, \lambda/4$), is inserted in order to ensure that the load modulation trajectory of Z_1 is driven to lower values with an increasing I_2, which is not the case of Z_1' presented in (3.17). Note that if $I_2 = 0$, then $Z_1 = 2R_L$; if $I_2 = I_1$, then $Z_1 = Z_2 = R_L$. Figure 3.20 shows the typical topology of a 2-way DPA. This topology consists of two RF amplifiers, labelled as main and peaking devices, with separate input bias. For example, typically a class-AB bias is used for the main device and a class-C bias for the peaking device. The input signal can be obtained by using a signal splitter, as shown or by independently controlled using

dual inputs. As there is a 90-degree impedance inverter on the output of the main device, phase shifting is required at the input of the peaking device, although not shown in Figure 3.20, to properly combine in-phase the output signals from main and peaking devices. The ideal voltage and current behaviours of 2-way symmetry DPA is shown in Figure 3.21, in which V_o represents the magnitude of the RF output voltage as is the case in Figure 3.19. V_{om} represents the maximum output voltage. The transition point, annotated as α, determines the turn-on point of the peaking device. For signal levels above α, both main and peaking device are conducting, thus load modulation occurs. At levels below α, no load modulation occurs and the efficiency of the DPA relies on the main device only.

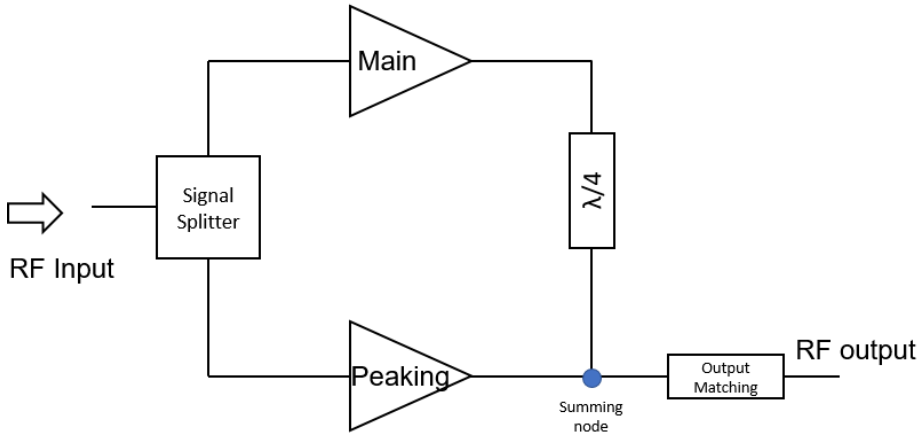


Figure 3.20 Typical topology of DPA

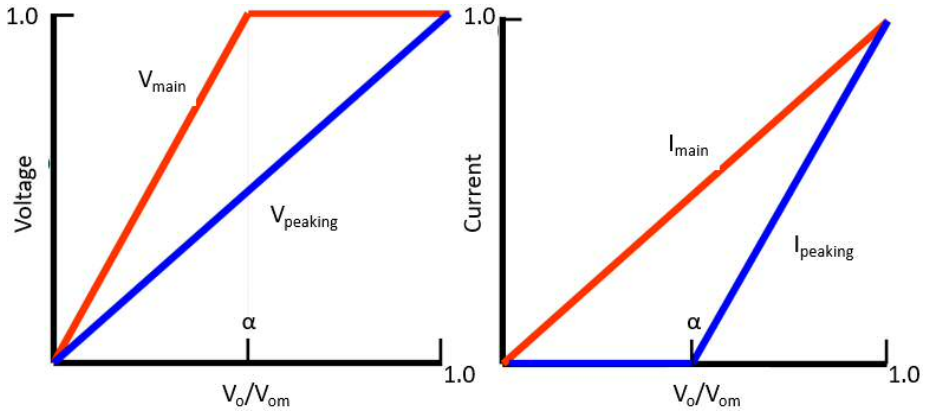


Figure 3.21 Voltage and current behaviours for ideal DPA operation

The efficiency of an ideal 2-way symmetrical DPA is shown in Figure 3.22. It can be seen that the ideal DPA exhibits significantly higher efficiency at all power levels but especially in the back-off region, which is determined by the transition point α and maximum power level. This topology can be then extended into a 3-way Doherty with the similar principle, as well as implemented with asymmetry devices.

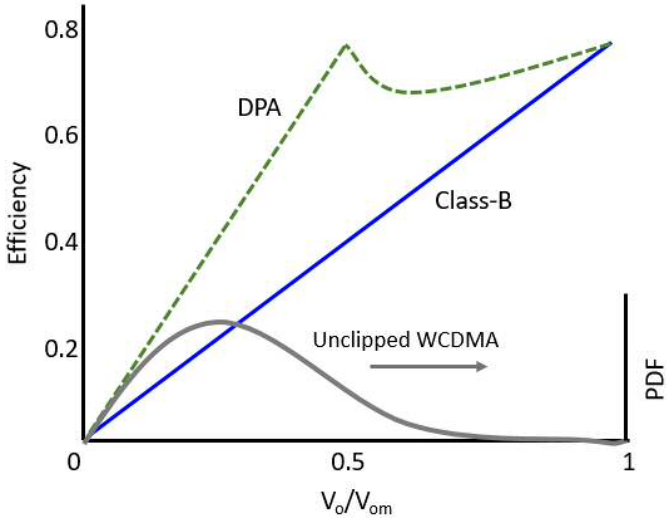


Figure 3.22 Efficiency for an ideal DPA compared with corresponding class-B operation

Although the DPA is very attractive in terms of efficiency enhancement, the actual stand-alone linearity performance cannot meet the requirements by cellular standards and linearization is necessary when applying the DPA in communication systems. In fact, the use of linearization schemes, such as DPD, has been a key enabler for the wide application of the DPA in modern communication systems. Notably, as the DPA are established using two or more devices biased at different operation points, the overall transfer characteristics will be a combination of all active devices. This can result in a much more complicated nonlinear behaviour compared to a stand-alone PA. For example, a transfer characteristic of DPA may involve more than one gain expansions where the main device, then the peaking device is active, as well as gain compression where the main and peaking devices compress at different points whilst maintains high efficiency operation. From a load-line perspective, as shown in Figure 3.23, active load modulation in the DPA causes the dynamic load-line to move relative to the static DC-IV characteristic, and in a way that corresponds with the time-varying amplitude of the modulated input signal. As a result, the “usual” interactions between dynamic load-line with the knee region and other IV boundaries also change. If care is not taken regarding this in the circuit design, unpredictable nonlinear behaviour can result. This can make the linearization, and specifically the linearizability of the DPA a very challenging design goal, especially when tackling higher efficiency and wide bandwidth requirements for the emerging communication systems.

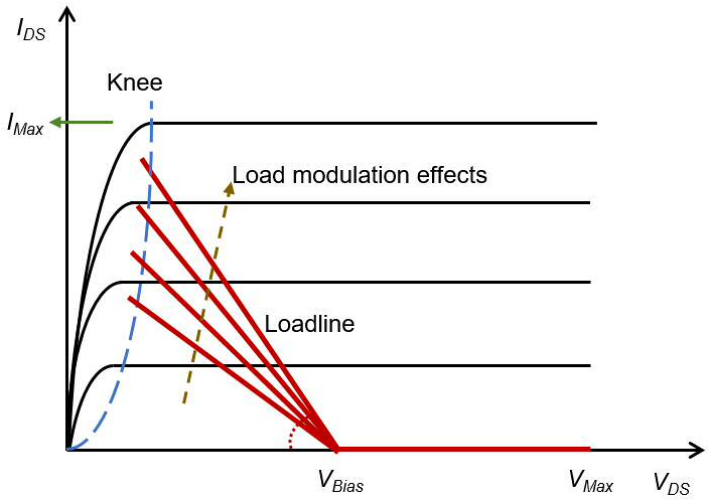


Figure 3.23 Load-lines for active load modulation

Much like the Doherty, the envelope tracking (ET) technique is another effective way of increasing the average efficiency of an RFPA when amplifying modulated signals with high PAPR. ET is closely related to the envelope elimination and restoration (EER) technique, which was introduced in 1952 by Leonard Kahn [100]. A typical topology for an ETPA is shown in Figure 3.24. It consists of a RFPA as well as a supply modulation scheme, which enables the power supply voltage of the device to be actively varied in accordance with the time-varying envelope of the modulated signal being amplified. In this way, the RFPA remains in or near compression for much of the signal amplitude range, and it is thus able to provide high-efficiency amplification. Similar to Figure 3.22, the efficiency performance of an ideal ETPA is illustrated in Figure 3.25. It can be seen that if the power supply voltage is reduced from the point of maximum power operation $V_{dd2} < V_{dd1}$, the efficiency at this reduced power will remain high. The ETPA exploits this principle by providing an optimum, dynamic output bias supply which closely tracks the

magnitude features of the input power envelope, resulting in an efficiency enhancement for a large range of output power levels.

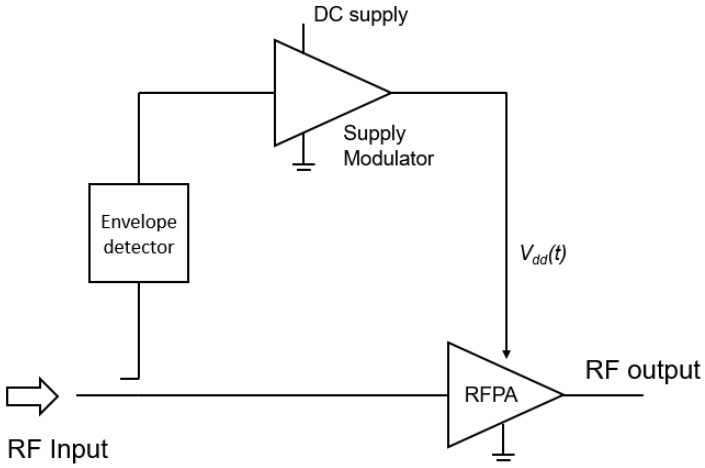


Figure 3.24 Simplified topology of an ETPA

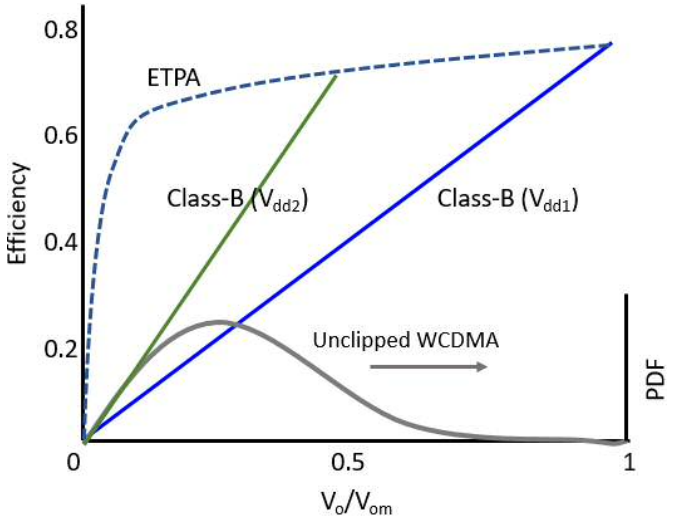


Figure 3.25 Efficiency for an ideal ETPA compared with class-B operation

From a load-line perspective, the ETPA has a load-line that moves according to the variations of the output bias voltage, as shown in Figure 3.26. This approach, much like the Doherty, also involves a change in the interaction with the knee

region and IV boundaries compared to a conventional PA, which can result in un-predictable nonlinear behaviour. However, different from the hardware-defined load modulation in DPA, the supply modulation in ETPA is programmable through a user-defined shaping function that relates the dynamic ET tracking signal to the input modulation envelope. This presents an opportunity, and the possibility of defining an optimized shaping function profile for good linearity, high efficiency as well as a good compromise of the two [50], [55], [101].

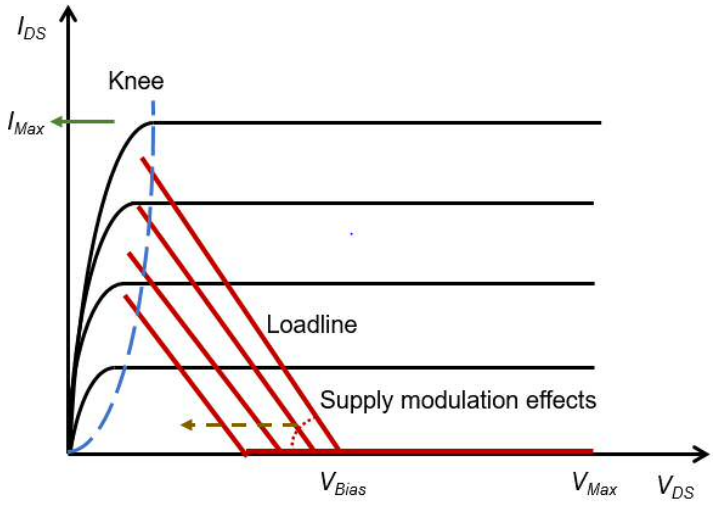


Figure 3.26 Load-lines for active supply modulation

Chapter 4 Nonlinear Sources and Circuit Level Optimization

4.1 Introduction

In order to achieve the best linearity for an RFPA, it is necessary to understand both circuit level and system level approaches to linearization, and the links between the two. In this chapter, a simplified analytical model is built to analysis distortion from intrinsic nonlinear sources within the transistor and to better understand their contributions. Moreover, this chapter includes discussion of circuit level optimization for best linearity and presents circuit level linearization strategies that can be applied before any system linearization is used to further linearize the RFPA. In this sense, circuit level linearization is shown to have the potential to improve the linearity as well as the *linearizability* of the RFPA.

4.2 Nonlinear Sources within a Transistor

The RFPA circuit schematic is simplified and shown in Figure 4.1. The PA core (transistor) sees both Z_{Source} and Z_{Load} as external impedance terminations. To

keep analysis simple, Z_{Source} includes the package parasitic components (not shown), effects of the gate biasing network, the input matching network and the source impedance. Z_{Load} includes the package parasitic components (not shown), the effects of the drain biasing network, the output matching network and the load impedance. In standalone RFPA designs, Z_{Source} and Z_{Load} are all passive impedances, which do not inherently generate distortion components, but can potentially delay and attenuate signals. The nonlinear sources are buried within the active PA core (transistor).

Although the terminations Z_{Source} and Z_{Load} don't generate distortions directly, It should be mentioned that they present terminating impedance to the spectral currents and so generate voltage waveforms V_{gs} and V_{ds}. These waveforms then feed into the transfer function in Equation (3.8), where they are able to affect the signal linearity. A detailed analysis on this impact is discussed in detail in section 4.4.3.

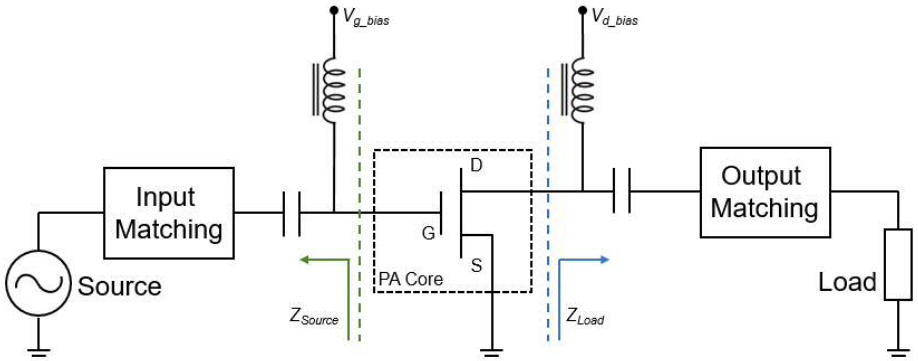


Figure 4.1 Simplified RFPA schematic

The transistor core can be modelled with basic sub-circuit components, as shown in Figure 4.2. The intrinsic behaviour of amplification is achieved by the nonlinear voltage dependent current generator I_{ds}, which is also referred to as the intrinsic nonlinear transconductance. There are then three intrinsic voltage dependent nonlinear capacitance; the gate-source capacitance $C_{gs}(V_{gs})$, the drain-source capacitance $C_{ds}(V_{ds})$ and the feedback capacitance $C_{gd}(V_{gd})$. Theoretically, all these three intrinsic capacitances have nonlinear characteristics that will vary as a function of the voltage across them. However, a further simplification allows

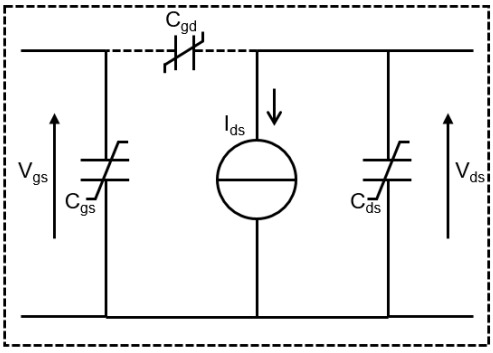


Figure 4.2 FET transistor equivalent circuit model

the nonlinearity in $C_{gd}(V_{gd})$ to be ignored [74], [76]. This is because the gate-drain junction related capacitance C_{gd} is normally reverse biased with a large voltage from drain to gate, and its own nonlinearity only has an impact when the PA device is strongly compressed with a large input signal. For the purposes of this thesis, this is beyond the operation range of the RFPA, so can be ignored. The following sections will discuss how these nonlinear sources are modelled and identify their contributions to total intrinsic distortion.

4.3 Analytical Nonlinear Models

To explore the device non-linear behaviour, a number of analytical models have been developed to represent sub-circuit equivalents [76], [81], [82], [102], [103]. Rather than attempting to model device behaviour for the accurate prediction of device operation, this chapter focuses more on revealing the key distortion generating mechanisms in the transistor. Based on this aim, the simplified device level analytical model will be divided in separate modelling on I_{gen}, $C_{gs}(V_{gs})$ and $C_{ds}(V_{ds})$. These separate models are then organized as shown in Figure 4.2 to be able to present a full model of the transistor. Benefiting from separate modelling on these nonlinear elements, this analytical model is tunable on parameters and used to analyse the individual distortion contribution from different nonlinear sources later in this chapter.

4.3.1 SDD Tunable Device Model

In order to develop a tunable analytical model, a symbolically-defined device (SDD) component is used within Keysight's Advanced Design System (ADS). The SDD is an equation-based component, which allows user to define the input and

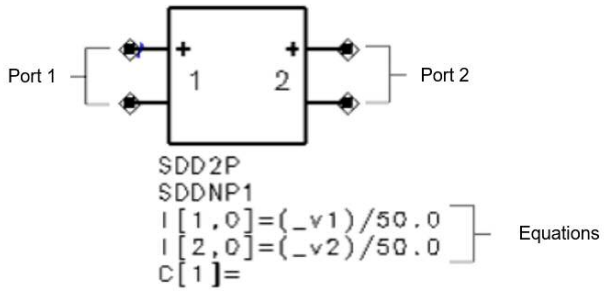


Figure 4.3 Two-port SDD in ADS

output behaviour of a component using a mathematical description. An example is shown in Figure 4.3, where a two-port SDD is used, along with two equations to define the relationship between voltage and current for each port. In theory, an SDD can be defined by specifying equations relating port currents, port voltages, and their derivatives. For example, in Figure 4.3, current $I[n,m]$ present the current at port-n with a weight function m. The weighting function is by definition a frequency-dependent expression used to scale the spectrum of a port current. There are two predefined weighting functions. Weighting function 0 is defined as the identical one when no weighting is required. Weighting function 1 is defined as $j\omega$ and it is used when a time derivative is desired, for example, in reactive components models. For the equations used as definition, they can be either one-dimensional functions, as it's shown in Figure 4.3, or multi-dimensional functions for a more complicate case.

4.3.2.1 Nonlinear I_{gen} Model

To model the transistor's intrinsic transconductance I_{gen}, a simplified Angelov model is used;

$$\begin{aligned} I_{ds} &= I_{pk0} \cdot f_G(V_{gs}) \cdot g_D(V_{ds}) \\ &= I_{pk0} \cdot \tanh(A \cdot V_{ds}) \cdot (1 + \tanh(P_1 \cdot (V_{gs} - V_{pk}))) \end{aligned} \tag{4.1}$$

In this model, hyperbolic tangent functions are used for both $f_G(V_{gs})$ and $g_D(V_{ds})$, which are implemented with using a two-port SDD component, as shown in Figure 4.4. To further clarify the model parameters, it is worth considering the modelled

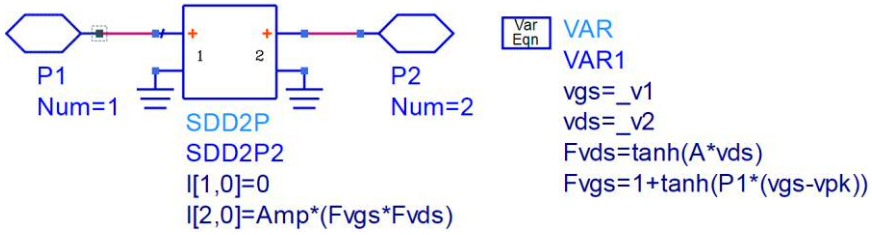


Figure 4.4 I_{gen} model with SDD component

transfer function and DC-IV mesh. Figure 4.5 compares the static transfer characteristic of a 10W GaN (GaN-on-SiC) Cree device (CGH60015D) and the developed analytical model, whereas Figure 4.6 shows the DC-IV mesh for this simplified model alone. In the transfer characteristic, as shown in Figure 4.5, P_1 determines the transfer function slope, while V_{pk} sets the horizon voltage offset and I_{pk0} is used to scale the amplitude of the current. In the DC-IV plot in Figure 4.6,

A is used to define the knee profile (slope). V_{pk} and P_1 will affect the gradient of the horizon mesh and the actual current value.

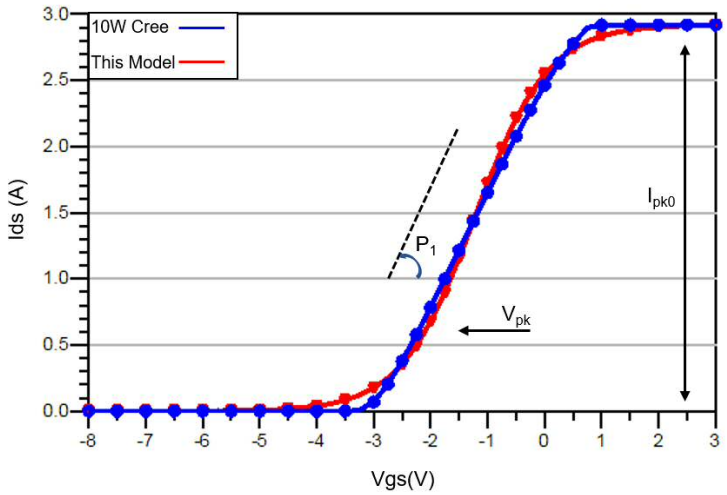


Figure 4.5 Transfer characteristic of a 10W Cree device and the simplified model

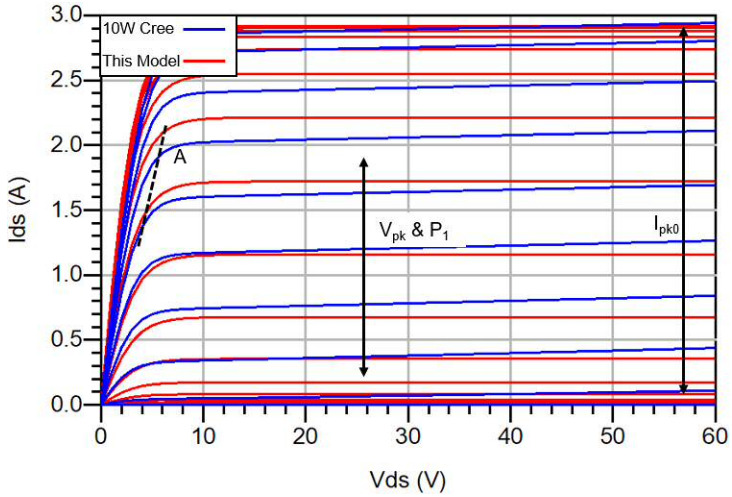


Figure 4.6 DCIV mesh of simplified model

To verify the simplified analytical model, a two-tone simulation is employed to observe the output power, gain, drain efficiency and 3rd order intermodulation distortion (IMD3). It is worth to mention that the two-tone stimulus signal used in this chapter has a tone spacing of 100MHz centred at 1GHz. The schematic is shown in Figure 4.7. The device (current generator) model, here labeled as I_{gen} core is biased for class-B operation. As the SDD has an infinite input impedance, a shunt 50Ohm resistor is added at the input to maintain the correct power match between the power source and device model. The device is then loaded with an ideal termination environment using a one-port impedance component (Z1P_Eqn). This one-port impedance component is defined to present constant impedance over each interested band, such as the frequency spectrals around fundamental, baseband and harmonic impedance separately. In this case, an ideal passive impedance is used as fundamental termination while the baseband and harmonic impedances are set to a short circuit. Gain, efficiency and IMD3 are ploted

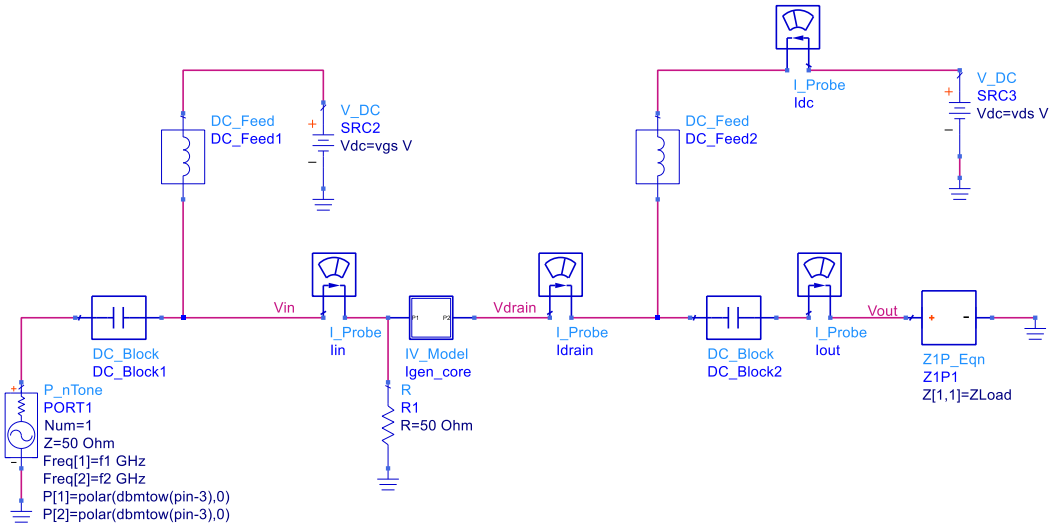


Figure 4.7 Two-tone verification for Igen model

respectively in Figure 4.8, where it is clear that this I_{gen} model is able to produce both signal amplification and distortion at the expected levels. Furthermore, Figure 4.9 presents the load-pull contours for a simulation using this model in a deep class-AB bias. Notably, as there are no reactive components in the device model so far, the power contours follow in a well defined "text-book" shape [8]. The contours are located symmetrically with their centres on the real axis of the Smith Chart, and the shape of power contours align with the constant resistance circles, to the left side, and admittance circles (not shown in Figure 4.9), to the right side.

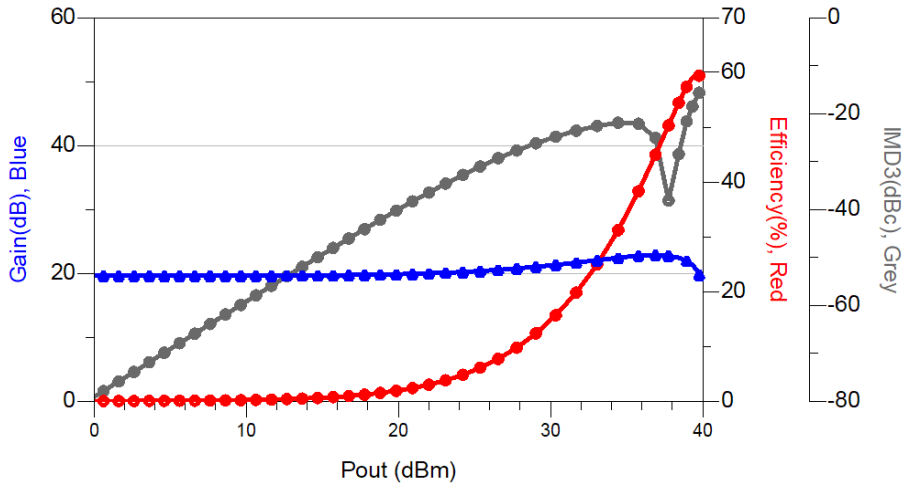


Figure 4.8 Gain, efficiency and IMD3 plots for I_{gen} model with two-tone excitation: 20MHz bandwidth @ 1GHz center frequency, Vg=-3V, Vd=28V

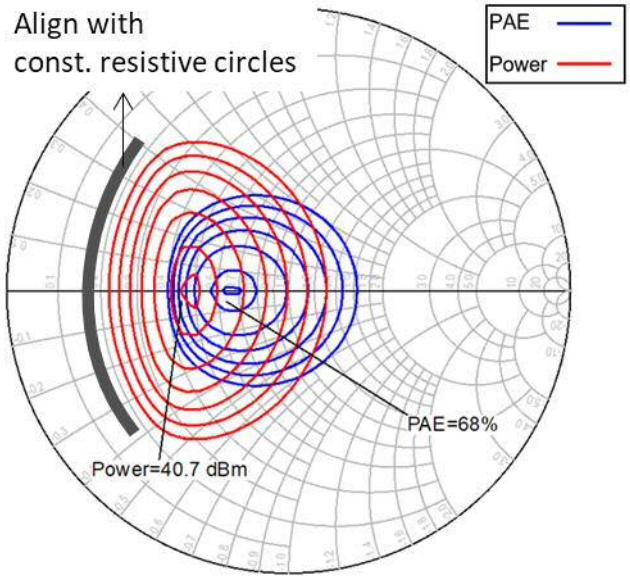


Figure 4.9 Load-pull contours for I_{gen} model: 20MHz bandwidth @ 1GHz center frequency, Vg=-3V, Vd=28V

4.3.2.2 Nonlinear Capacitance Model

As shown in Figure 4.10, a two-port SDD can also be used to model the nonlinear gate capacitance C_{gs} with the time derivative features of the predefined weighting function. In this SDD model, port-1 is used to define the voltage and current relationship of this capacitor while port-2 is used to load a derivative equation into the SDD. To explain the operation of this element in more detail, as port-2 is set as an open port and the current at this port should be zero, which means

$$I[2,0]+I[2,1]=0 \tag{4.2}$$

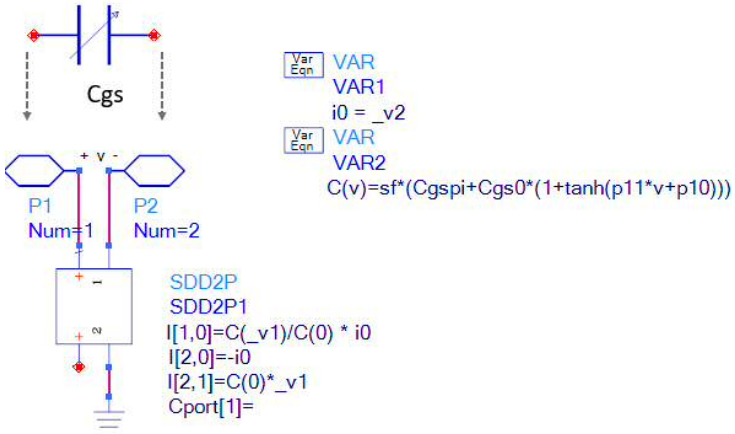


Figure 4.10 Gate capacitance model with SDD

In (4.2), $I[2,0]$ equals the voltage at port-2 through an user defined variable $i0$, $I[2,1]$ is the time derivative components $I[2,1]=C(0)\cdot d(v1)/dt$. In this case, (4.2) can then be written as,

$$i0=C(0)\cdot\frac{d(v1)}{dt} \tag{4.3}$$

So, the current of port-1 is,

$$I[1,0]=\frac{C(v1)}{C(0)}\cdot i0=C(v1)\cdot\frac{d(v1)}{dt} \tag{4.4}$$

where $C(v1)$ is a voltage dependent function,

$$C(v)=C_{gspi}+C_{gs0}\cdot(1+\tanh(P11\bullet v+P10)) \tag{4.5}$$

Actually, (4.4) is a basic voltage-based (instead of charge) model for a capacitor and (4.5) defines its voltage dependent feature. In (4.5), the capacitance-voltage relationship is described with a hyperbolic tangent (tanh) function similar to Angelov model. C_{gspi} and C_{gs0} contribute the linear part in C_{gs}, while $P10$ and $P11$ are the nonlinear coefficients, which determines the zero order and first order nonlinear dependency on V_{gs}, respectively. This model is then compared with a 10W Cree die model, and the results shown in Figure 4.11.

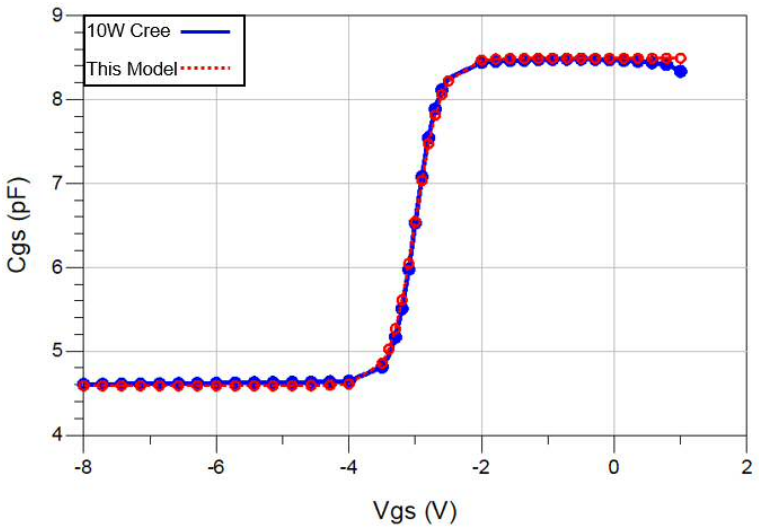


Figure 4.11 Cgs model compared with 10W Cree die model

The nonlinear drain capacitance C_{ds} is modelled using the same SDD capacitor based on (4.4). Instead of the tanh function in (4.5), the voltage dependent feature is easily modelled with a 7th order polynomial series of drain-source voltage V_{ds},

$$C_{ds} = \sum_{n=0}^{n=7} C_n \cdot V_{ds}^{\ n} \tag{4.6}$$

7th order is chosen as it gives good curve-fitting and provides good tracking on the odd order mixing effects up to 7th order. Figure 4.12 compares the model with the 10W Cree die model discussed earlier.

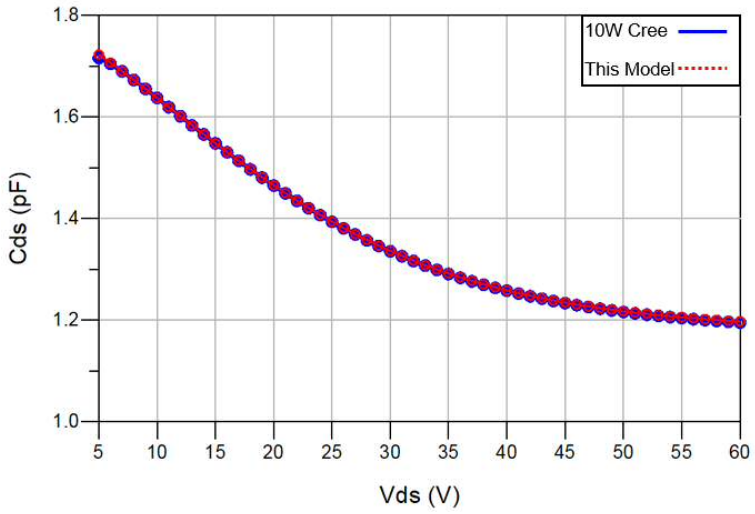


Figure 4.12 Cds model compared with 10W Cree die model

The gate-drain capacitance is modelled as a linear capacitor C_{gd}, 0.2pF in this work, following the discussion in Section 4.2. So now, all the key nonlinear elements in the transistor model have been constructed.

4.3.2 Model Verification

The analytical model elements discussed in the last section are assembled into a single model ,as shown in Figure 4.13. Then this model is verified with a load-pull simulation at the same deep class-AB bias used previously with just the current generator. For the model to be representative of a practical device, gate resistance R_g, gate inductance L_g, drain resistance R_d, and drain inductance L_d were added. The results are compared to those obtained from a Cree GaN 10W bare die model as shown in Figure 4.14 and Figure 4.15. In Figure 4.14, the power contours obtained from the new model are in good alignment with the 10W Cree die model. In Figure 4.15, the PAE contours are offset, but both achieve similar PAE peaks and levels. Notably, the contours are rotated compared to Figure 4.9 due to the presence of the reactive components.

So far, a simple analytical device model has been developed, which is effective in modeling the key features in a Cree device model for the PA application. This model can now be used to investigate the respective impact of the key nonlinear elements in the circuit.

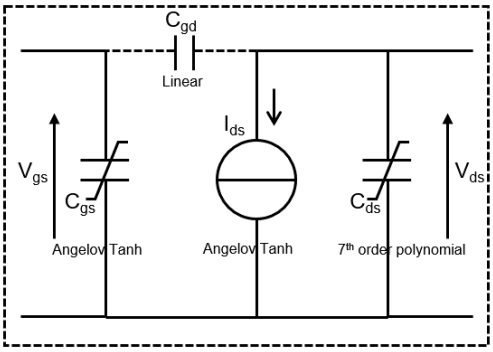


Figure 4.13 Simplified fully analytical model

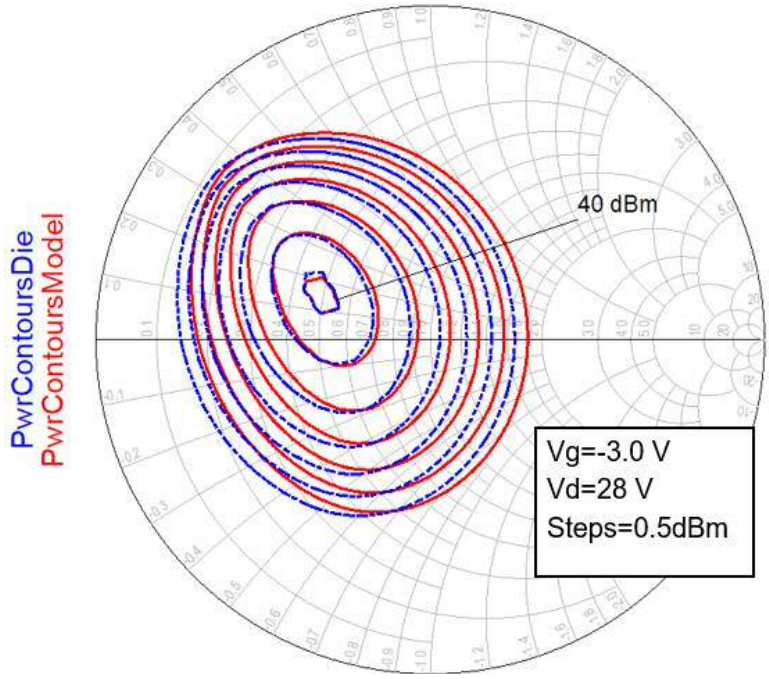


Figure 4.14 Power contours in Load-pull simulation
(Cree 10 W die model: Blue; Analytical model: Red)
20MHz bandwidth @ 1GHz center frequency, Vg=-3V, Vd=28V

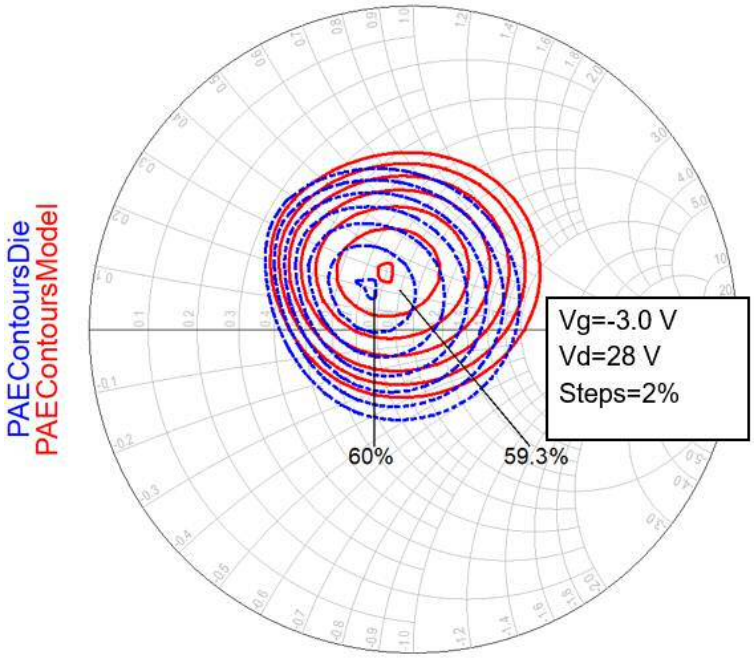


Figure 4.15 PAE contours in Load-pull simulation
(Cree 10 W die model: Blue; Analytical model: Red)
20MHz bandwidth @ 1GHz center frequency, Vg=-3V, Vd=28V

4.4 Distortion Analysis at Device Level

The analytical device model built using SDD components in ADS will now be used to investigate the non-linear behavior of relevant transistors. With this tunable model, it is possible to analysis the individual distortion contributions from different nonlinear sources. The analysis is explained in detail in this section. Within the transistor, I_{gen} is the only active nonlinear components which generates intrinsic distortion. I_{gen} is a two-dimensional function which depends on both V_{gs} and V_{ds}, so any additional nonlinear systems active at the transistor's gate and / or drain will contribute to the overall distortion through the transfer function $I_{gen} = I_{pk0} \cdot f_G(V_{gs}) \cdot g_D(V_{ds})$. In this way, nonlinear capacitance C_{gs} and C_{ds} can contribute to the overall distortion level by generating nonlinear voltages. For simplicity, this section begins by analysing the distortion from the I_{gen} core, and then considers the impact of the intrinsic nonlinear input and output capacitances C_{gs} and C_{ds}. In this section, a two-tone excitation with a tone-spacing of 100MHz centred at 1GHz is used in order to analyse the IM distortion products generation mechanism. The simulations are conducted using a harmonic balance (HB) simulation in the frequency domain.

4.4.1 Distortion from the current generator

As discussed in the previous sections, the current generator I_{gen} core generates distortion current components due to the nonlinear transfer characteristic. It is important to mention that the transfer characteristic itself depends on the circuit environment, which includes bias, drive level and termination impedance. The out-of-band terminations (baseband and harmonics) are considered firstly when being presented with short circuits, preventing nonlinear voltage components being generated at these frequencies and the possible secondary mixing from the device itself. To investigate the impact of bias on the circuit nonlinear behavior, Figure 4.16 shows the transfer characteristics when different gate bias V_g applied to the I_{gen} model, when excited with the same amplitude alternating input voltage drive (1V peak). This results in operational windows located around the different biasing points on the full transfer characteristic. This explains why class A is generally considered as the most linear class as the linearity of the exposed transfer function decreases from A→AB→B→C. This figure also clearly shows why the gain varies for different classes, for example why the gain for class-A is higher than the other classes. Figure 4.17 shows the corresponding dynamic load-line for the same set of input bias points, again when using the same drive level.

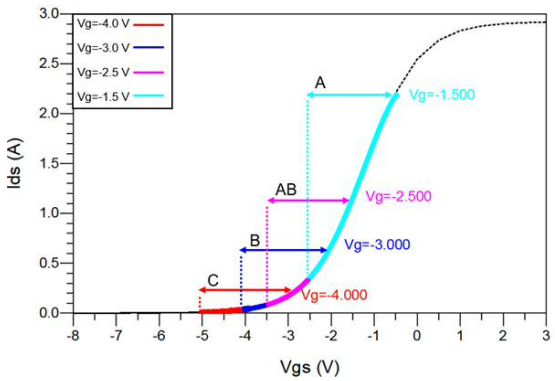


Figure 4.16 Transfer characteristics due to different gate bias

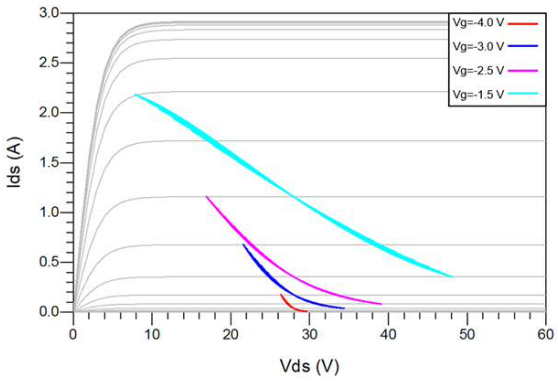


Figure 4.17 Dynamic load-line with different gate bias

In addition to bias, Figure 4.18 shows how the drive level affects the transfer characteristic. Comparing Figure 4.18 and Figure 4.19, it's clear that with a higher drive level, the dynamic IV curve covers a wider range on the full transfer characteristic. Notably, the compressed peak corresponding to class-A bias is due to the knee interaction between the knee region and the load-line which is shown in Figure 4.18. The dispersion in the load-line in Figure 4.19 is due to the peaks of the two-tone AM modulated output signal interacting with the knee region and pinch-off IV-plane boundaries, which causes misshaping of the current waveform, and in turn the observable self-biasing effect.

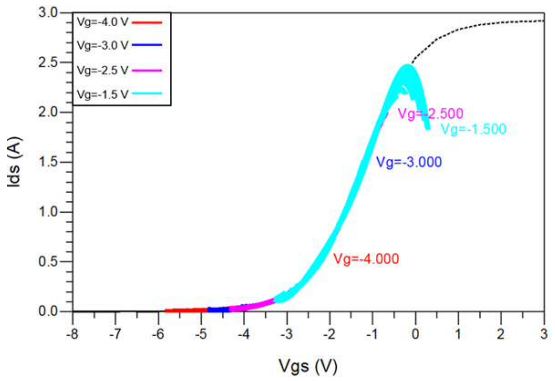


Figure 4.18 Transfer characteristics when the device is driven into compression

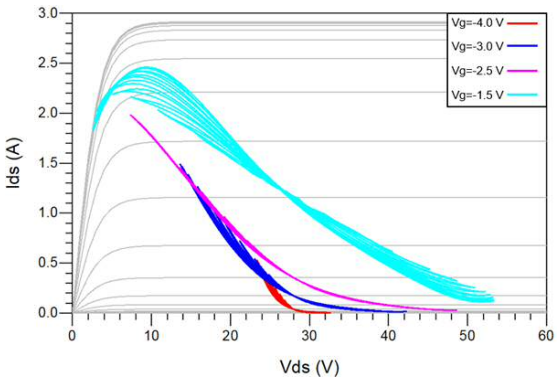
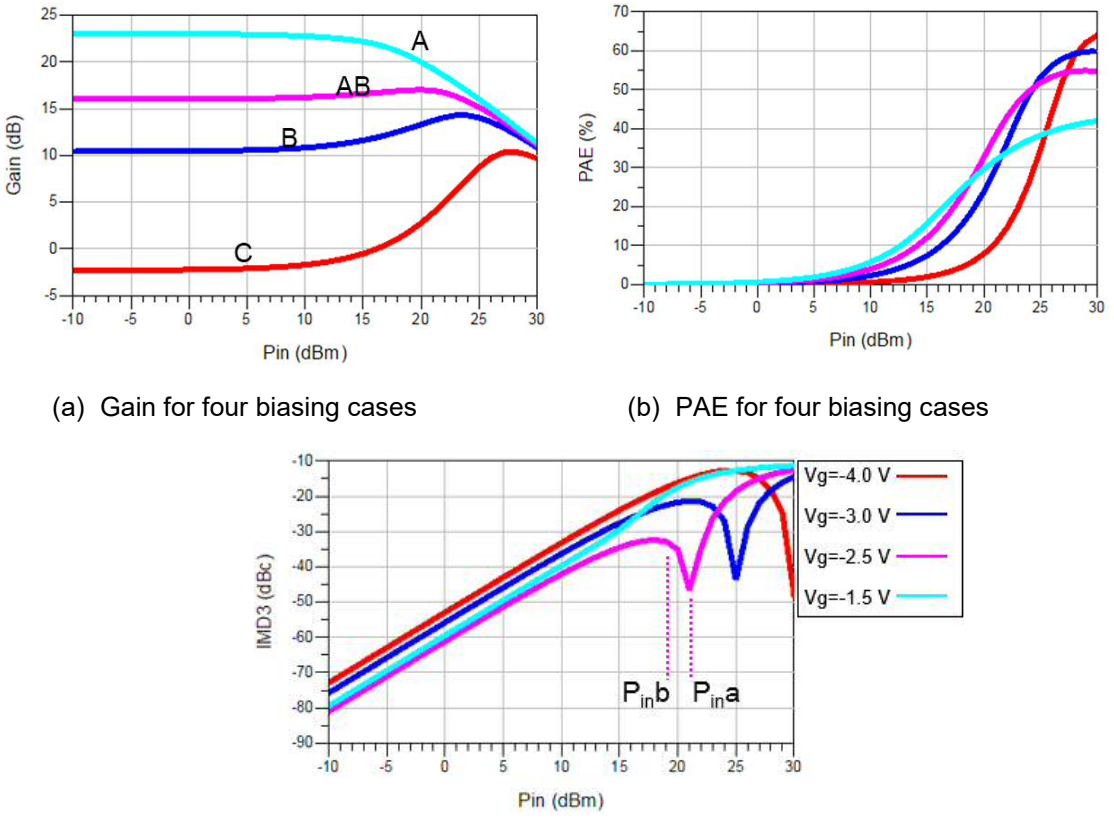


Figure 4.19 Dynamic load-line with different gate bias when the device is driven into compression

While distortion can be observed in Figure 4.17 and Figure 4.19 in a qualitative way, IMD3 is used as a quantitative description of the distortion around carrier (IM products). Figure 4.20 summarizes the performance of the I_{gen} model due to different bias in a power sweep simulation, with relevant Gain, PAE and IMD3 plotted.



(a) Gain for four biasing cases

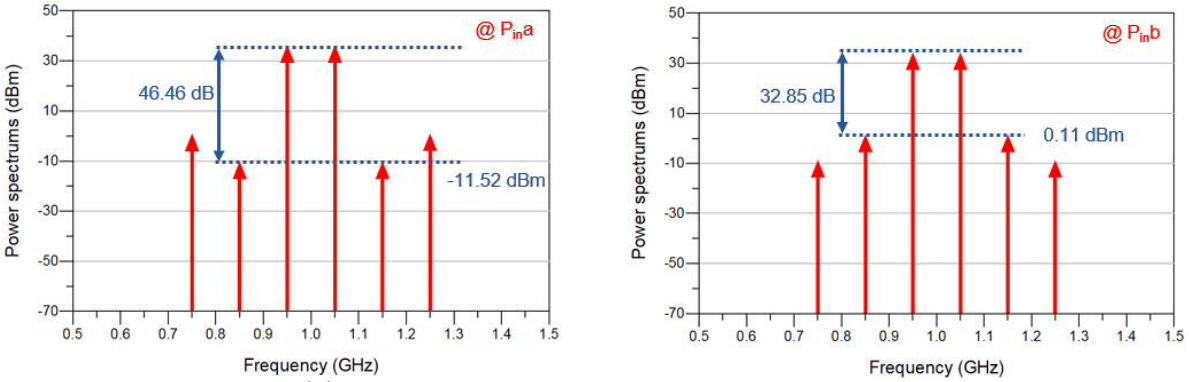
(b) PAE for four biasing cases

(c) IMD3 for four biasing cases

Figure 4.20 Performance of the I_{gen} model with different bias in a power sweep (Gain, PAE and IMD3)

The gain and efficiency shown in Figure 4.20 agree with the previous analysis for different classes of amplification. In the IMD3 plots, there are typical "sweet-spots" observed in the class-AB and class-B curves. In fact, this "sweet-spot" behavior, which is caused by the cancellation of distortion terms from secondary mixing is a familiar feature and utilised in linear amplifier designs in [103]–[110]. Let's now

consider the IM products at this "sweet-spot" in more detail. Figure 4.20 presents a closer look at the spectrums around the carrier at input power level of $P_{in}a$=21dBm and $P_{in}b$=19dBm for Class AB bias.



(a) (b)

Figure 4.21 In-band power spectrums at different input powers: (a) Pina, (b) Pinb

To understand the "sweet-spot" observed at $P_{in}a$, there is a need to discuss the mixing mechanism in nonlinear systems. Firstly, for simplicity, the I_{gen} model is assumed to be a 5th order nonlinear system as described in (4.7),

$$I_{gen}(v_{gs}) = a0 + a1 \cdot v_{gs} + a2 \cdot v_{gs}^{\ 2} + a3 \cdot v_{gs}^{\ 3} + a4 \cdot v_{gs}^{\ 4} + a5 \cdot v_{gs}^{\ 5} \tag{4.7}$$

in which the input signal v_{gs} here is a two-tone signal,

$$v_{gs} = A \cdot \cos(\omega_m t) \cdot \cos(\omega_c t) \tag{4.8}$$

where A represents the magnitude of the signal (drive level). ω_c means the carrier frequency while ω_m is the message frequency. Substituting (4.8) into (4.7), the input signal is mixed by the nonlinear system, and as a result generates frequency

components $\cos((\omega_c \pm 3\omega_m)t)$ (third order, IM3 components) and $\cos((\omega_c \pm 5\omega_m)t)$ (fifth order, IM5 components), which are shown in (4.9), (4.10) respectively.

$$IM3 = (\frac{25}{256} a5 \cdot A^5 + \frac{3}{32} a3 \cdot A^3) \cdot \cos((\omega_c \pm 3\omega_m)t) \tag{4.9}$$

$$IM5 = \frac{5}{256} a5 \cdot A^5 \cdot \cos((\omega_c \pm 5\omega_m)t) \tag{4.10}$$

Note that the IM3 distortion (4.9) involves the contributions from 3rd order mixing ($a3$) and 5th order mixing ($a5$). In reality, both (4.9) and (4.10) should involve higher order mixing from $a7$ and above, which are neglected here. In (4.9), if $a3$ and $a5$ have the same sign, the 5th order mixing generates an additive distortion on top of the intrinsic 3rd order product which increases the distortion level. However, if $a3$ and $a5$ have opposite signs, the 5th order mixing can produce a subtractive effect which suppresses the intrinsic 3rd order product. Hence, theoretically, with proper magnitudes, the resulting IM3 can be forced to zero, at a given drive level, which means no 3rd order distortion is generated. From the mathematic perspective, by solving equation (4.9), the condition for an IM3 nulling can be obtained,

$$\frac{a3}{a5} = -\frac{25}{24} A^2 \tag{4.11}$$

Equation (4.11) shows that, in addition to the opposite signs of $a3$ and $a5$, this IM3 nulling effect depends on the ratio between $a3$ and $a5$, and the magnitude A. While A depends on the signal drive level, $a3$ and $a5$ describing the transfer characteristic which depend on bias, impedance termination and drive level. Note that this cancellation effect is normally observed close to the transistor's compression

as the interaction with the knee region and other boundaries results in a strong nonlinear characteristic with coefficients of opposite sign.

Benefiting from the analytical model established in this work, the "sweet spot" theory is quantitatively verified. As shown in Table 4-1, $a3$ and $a5$ are extracted at $P_{in}a$ and $P_{in}b$ respectively for both class-AB ($V_g = -2.5V$) and class-A ($V_g = -1.5V$). Notably, $a3$ and $a5$ are of opposite sign for class-AB bias while are the same sign for class-A bias. This explains the reason why a "sweet-spot" is visible in class-AB but not in class-A in Figure 4.20 (c). For class-AB bias, $P_{in}a$ yields a lower IM3 compared to that of $P_{in}b$, which explains the lower distortion level "sweet-spot" at $P_{in}a$. In addition to this, the ratio between IM3 components for class-AB recorded in Table 4-1 (9.88e-3/2.56e-3 =11.72dB), which agrees with the 11.63dB difference plotted in Figure 4.21. This supports the fact that IM3 products depends on the bias / operation point.

Table 4-1 Nonlinear coefficients and IM3 behaviour for different bias and drive level

Experiments		**a3**	**a5**	**A**	**IM3**
Class AB	$P_{in}a$	6.6018e-2	-1.4860e-2	2.012	2.5642e-3
(Vg=-2.5 V)	$P_{in}b$	6.5232e-2	-1.4815e-2	1.591	9.8804e-3
Class A	$P_{in}a$	-1.3374e-1	-8.6370e-3	2.012	-1.2993e-1
(Vg=-1.5 V)	$P_{in}b$	-1.0915e-1	-3.3725e-2	1.591	-7.4785e-2
Experiments		**a3/a5**	**A**	**-25/24*A^2**	**IM3**
Class AB	$P_{in}a$	-4.443	2.012	-4.217	2.5642e-3
(Vg=-2.5 V)	$P_{in}b$	-4.403	1.591	-2.637	9.8804e-3

More importantly, taking a closer look at the coefficients $a3$ and $a5$ in the Class AB case, the ratio of $a3/a5$ for the $P_{in}a$ case aligns more closely to the theoretical optimum calculated using equation (4.11) than the $P_{in}b$ case. This explains why the "sweet-spot" is located at $P_{in}a$ with this bias voltage. The analysis above has proven the linearity performance, specifically, the "sweet-spot", depends on both the circuit bias and drive level.

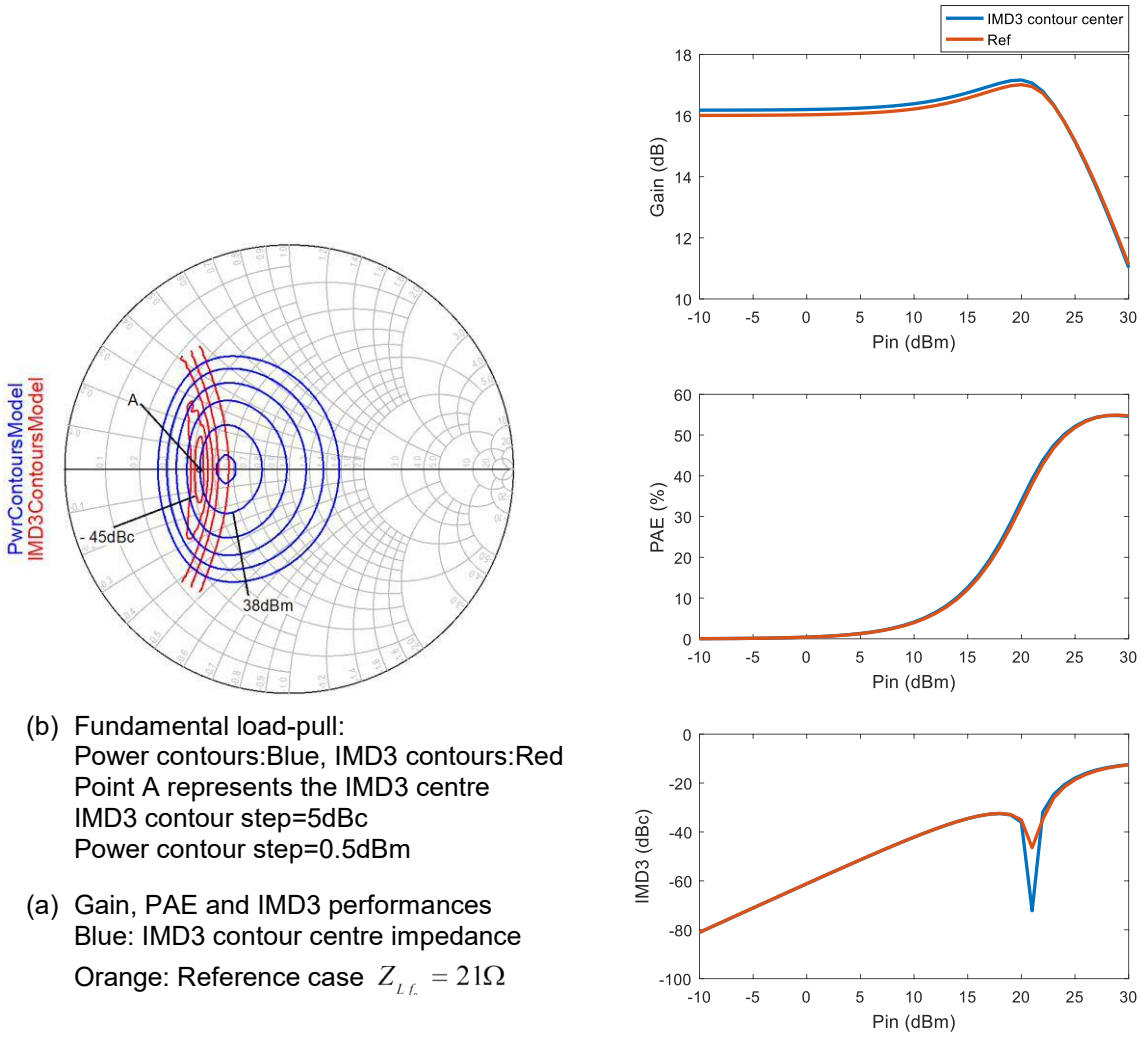
The next step was, to investigate the impact of terminating impedance on the generation of IM3 distortion. Two-tone load-pull simulations have been undertaken for both *in-band* termination (fundamental), and out-of-band terminations (baseband and 2nd harmonic) respectively. The two-tone stimulus signal used has a tone spacing of 100MHz centred at 1GHz.

Firstly, fundamental output load-pull at class-AB bias ($V_g = -2.5V$) is performed at a drive level of $P_{in} = P_{in}a = 21dBm$, while the out-of-band impedances (baseband, 2nd and higher harmonics) are terminated into short circuits. For a reference case, a fundamental load impedance $Z_{Lf_0} = 21\Omega$ is used yielding 10W output power, which is the same case shown in Figure 4.20 with a magenta coloured trace. Then, as shown in Figure 4.22, an optimum impedance case of $Z_{Lf_0} = 21.85\Omega$ for IMD3 optimization is considered, with resulting nonlinear coefficients (a_3, a_5) and IM3 summarized in Table 4-2. This behavior can be explained qualitatively using the load-line theory; when the fundamental load impedance is varied in this experiment, the load-line interaction with the knee region and / or

pinch-off changes, and as a result, the values of a_3 and a_5 also changes. This can now be explained quantitatively as the ratio of $a3/a5$ in the optimum case matches the theoretical optimum, which is calculated by equation (4.11), better than the reference case. This is the reason why the depth of the "sweet-spot" increases. This experiment demonstrates that through optimizing the fundamental load impedance, IM3 level can be minimized.

Table 4-2 Nonlinear coefficients and IM3 for different load fundamental impedance

Experiments		**a3**	**a5**	**A**	**IM3**
Class AB	Z_{Lf0}= 21 Ohm	6.6018e-2	-1.4860e-2	2.012	2.5642e-3
(P_{in}=21 dBm)	Z_{Lf0}= 21.85 Ohm	7.1738e-2	-1.7005e-2	2.012	2.4822e-5
Experiments		**a3/a5**	**A**	**-25/24*A^2**	**IM3**
Class AB	Z_{Lf0}= 21 Ohm	-4.443	2.012	-4.217	2.5642e-3
(P_{in}=21 dBm)	Z_{Lf0}= 21.85 Ohm	-4.219	2.012	-4.217	2.4822e-5



(b) Fundamental load-pull:
Power contours:Blue, IMD3 contours:Red
Point A represents the IMD3 centre
IMD3 contour step=5dBc
Power contour step=0.5dBm

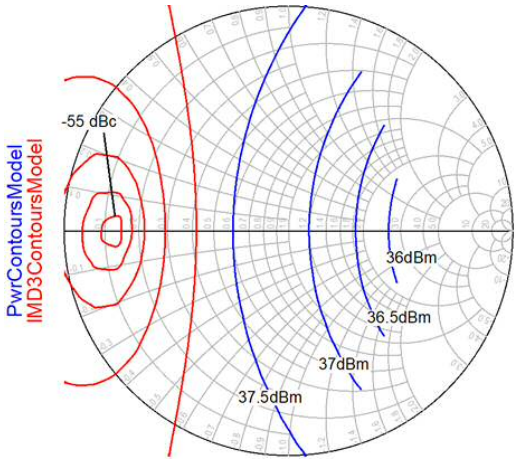
(a) Gain, PAE and IMD3 performances
Blue: IMD3 contour centre impedance
Orange: Reference case $Z_{LF} = 21\Omega$

Figure 4.22 Two-tone fundamental load-pull

A similar investigation has been conducted by controlling the output baseband impedance while fundamental $Z_{Lf_0} = 21\Omega$ and harmonics $Z_{L(2f_0 \& higher)} = 0\Omega$ are kept constant. The corresponding results are shown in Figure 4.23 and Table 4-3. These results can be explained in the same way comparing the resulted $a3/a5$ ratio with the theoretical optimum. This experiment demonstrates an optimization of output baseband impedance.

Table 4-3 Nonlinear coefficients and IM3 for different load baseband impedance

Experiments		**a3**	**a5**	**A**	**IM3**
Class AB	Z_{LBB}= 0 Ohm	6.6018e-2	-1.4860e-2	2.012	2.5642e-3
(P_{in}=21 dBm)	Z_{LBB}= 5Ohm	7.1960e-2	-1.7023e-2	2.012	1.3475e-4
Experiments		**a3/a5**	**A**	**-25/24*A^2**	**IM3**
Class AB	Z_{LBB}= 0 Ohm	-4.443	2.012	-4.217	2.5642e-3
(P_{in}=21 dBm)	Z_{LBB}= 5Ohm	-4.227	2.012	-4.217	1.3475e-4



(a) Baseband load-pull:
Power contours:Blue, IMD3 contours:Red
IMD3 contour step=5dBc
Power contour step=0.5dBm

(b) Gain, PAE and IMD3 performances
Blue: IMD3 contour centre impedance
Orange: Reference case $Z_{L(2f_0 \& higher)} = 0\Omega$

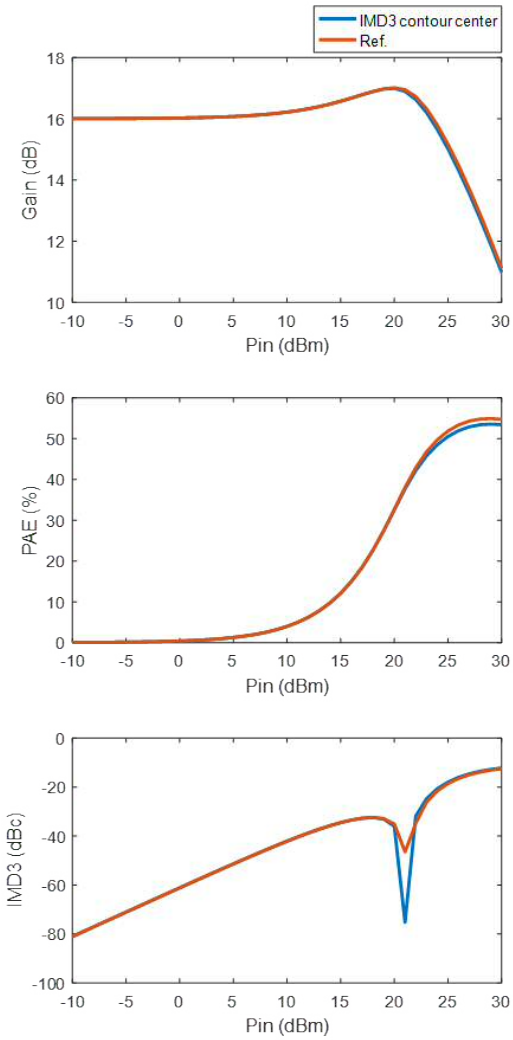
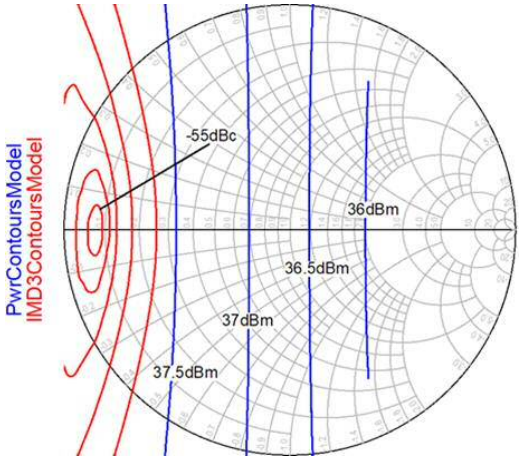


Figure 4.23 Two-tone baseband load-pull

This is followed by a 2nd harmonic load-pull at the output with constant fundamental $Z_{Lf_0} = 21\Omega$, baseband and higher harmonics $Z_{L(BB\&higher)} = 0\Omega$ impedances. The corresponding results are shown in Figure 4.24 and Table 4-4. These results can be explained in the same way comparing the resulted $a3/a5$ ratio with the theoretical optimum. This experiment demonstrates an optimization of output 2nd harmonic impedance.

Table 4-4 Nonlinear coefficients and IM3 for different load 2nd harmonic impedance

Experiments		**a3**	**a5**	**A**	**IM3**
Class AB	Z_{L2nd}= 0 Ohm	6.6018e-2	-1.4860e-2	2.012	2.5642e-3
(P_{in}=21 dBm)	Z_{L2nd}= 3 Ohm	7.1960e-2	-1.7023e-2	2.012	1.3475e-4
Experiments		**a3/a5**	**A**	**-25/24*A^2**	**IM3**
Class AB	Z_{L2nd}= 0 Ohm	-4.443	2.012	-4.217	2.5642e-3
(P_{in}=21 dBm)	Z_{L2nd}= 3 Ohm	-4.203	2.012	-4.217	-1.8105e-4



(b) 2nd harmonic load-pull:
Power contours:Blue, IMD3 contours:Red
IMD3 contour step=5dBc
Power contour step=0.5dBm

(c) Gain, PAE and IMD3 performances
Blue: IMD3 contour centre impedance
Orange: Reference case $Z_{L(BB\&higher)} = 0\Omega$

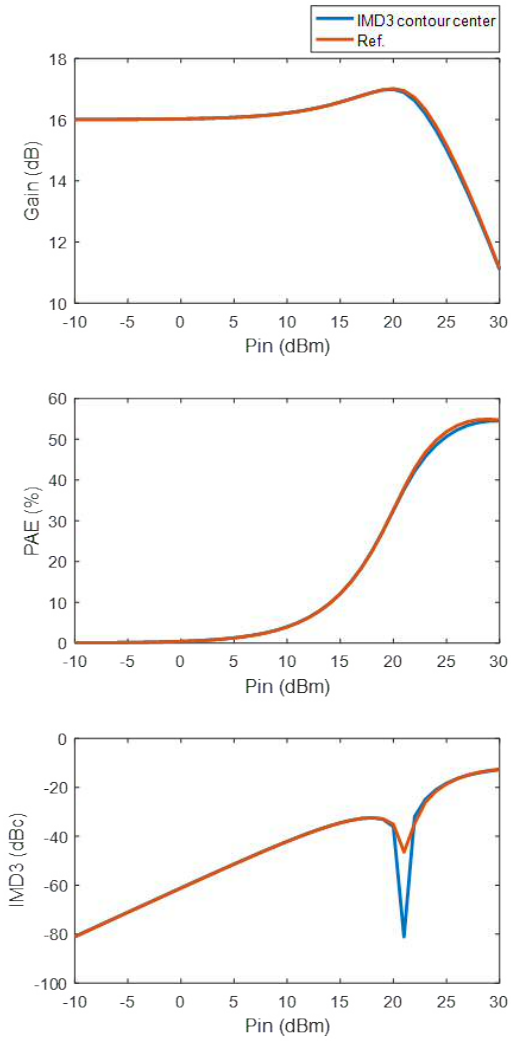


Figure 4.24 Two-tone 2nd harmonic load-pull

To conclude the above analysis, the contour behaviour presented in Figure 4.22, Figure 4.23 and Figure 4.24 indicate that IMD3 is a function of output termination impedance (load) at fundamental, baseband and 2nd harmonic frequencies. It is clear that by presenting the optimum impedance, in all cases, the depth of the IMD3 null "sweet spot" can be increased, compared to the reference case. It should be mentioned that, although the IM3 product depends on the fundamental impedance, in practice, fundamental matching is designed primarily for a good compromise of efficiency and power. It makes sense therefore to focus on optimizing the linearity through the proper choice of out-of-band impedance termination, as indicated in Figure 4.23 for baseband and Figure 4.24 for 2nd harmonic respectively.

The impact from out-of-band impedance is again related in the fact that the a_3 and a_5 coefficients, in the analytical model, are modified because of the modified interaction with the knee and / or pinch-off region. This is better explained by considering the secondary mixing of out-of-band components in the RFPA. Considering the full output current $I_{out} = I_0 \cdot f_G(V_{gs}) \cdot g_D(V_{ds})$, if an out-of-band impedance is presented other than a short circuit, it is possible that both V_{gs} and V_{ds} will have out-of-band components which are dependent on these impedance values. As a result of the mixing processes in the nonlinear transfer function, these out-of-band components can be up/down converted to IM products, and so contribute to the IMD sidebands distortion. Ideally, secondary mixing can be eliminated by controlling the impedance environment and presenting low impedance or short circuits

to out-of-band current components. However, this is not necessarily the best solution as secondary mixing products can be either cumulative (+) or compensating (-) in the way they modify the values of the resulting IMD products. In other words, secondary mixing effects can be beneficial as well as problematic. For instance, the compensating effect can be used to optimize the linearity of the RFPA without changing fundamental matching [111], [112].

The next step was to investigate the impact of drive level on IM3 products. To do this, the load-pull measurements that resulted in the contour plots shown in Figure 4.22(a) for baseband and 2nd harmonic are repeated for three different drive levels $P_{in} = (19dBm, 21dBm, 22dBm)$, as shown in Figure 4.25 and Figure 4.26, respectively, while keeping the fundamental load constant $Z_{Lf_0} = 21\Omega$.

Note that for increasing drive levels at constant bias condition, the IMD3 contours for out-of-band impedance load-pull tend to move outside the unity ($|\Gamma_L| = 1$) Smith Chart region, suggesting the need for active as opposed to passive impedance termination. Although this is observed and reported in [113], the dependence on drive level has not been clearly revealed or discussed before. To analyse this observation in more detail, Table 4-5 summarizes the nonlinear coefficients and IM3 levels with optimum baseband terminations in comparison to those in the reference cases, where the baseband and 2nd harmonic are both shorted with constant fundamental load $Z_{Lf_0} = 21\Omega$. It shows that with proper baseband termination, the effective IM3 levels can be suppressed. Notably, the polarity of

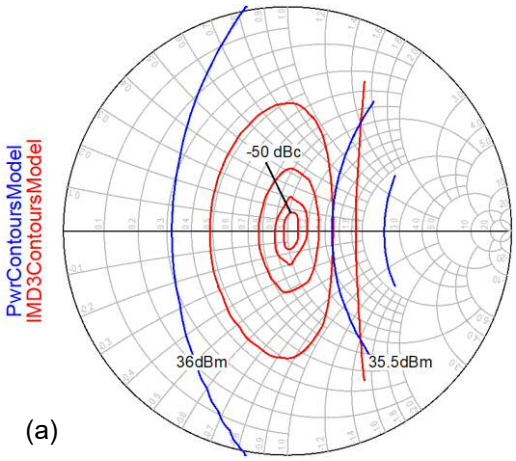
effective IM3 level for $P_{in} = 22dBm$ is opposite sign to the other two cases with lower drive levels. As in the cases for low drive levels (19dBm,21dBm), the effective IM3 components for the reference bias and termination are both positive (+), then the compensating IM3s as the result of optimum termination should be both negative sign (-). This explains why the optimum contour centres are in the same region of the Smith Chart ($|\Gamma_L| < 1$). However, for the high drive levels (21dBm), the effective IM3 is of opposite sign which corresponds to the compensating components with different polarity, thus the optimum goes outside of the Smith Chart ($|\Gamma_L| > 1$). A similar effect can be observed in the 2nd harmonic.

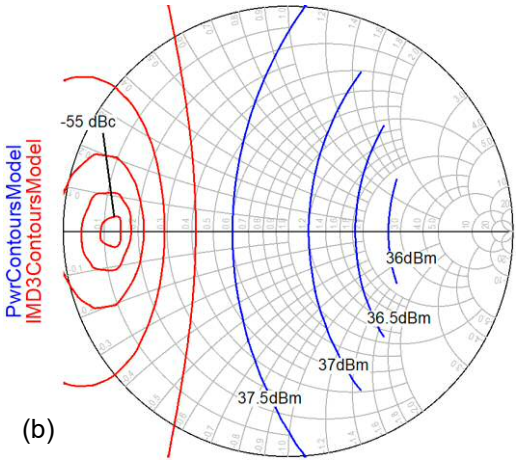
Although the impact of drive level has been discussed before together with the impact of bias, this experiment demonstrates the impact from a different perspective. In practice, it has been found to be very difficult to improve the linearity for a PA design at high compression level, due to the strong nonlinear effects from the interactions with knee and other IV boundaries. This is due to the fact that through these interactions, the nonlinear coefficients of the transfer characteristic will not result in IM3 distortion products that can be corrected with passive terminations, especially in the case that results in the generation of distortion products with opposite sign. However, this is the limitation when operating with passive techniques where one is confined to impedances within the unity Smith Chart ($|\Gamma_L| < 1$). With active injection techniques, such as DPD and baseband injection, it is possible to find baseband impedances that provide optimization, outside the Smith Chart unity circle ($|\Gamma_L| > 1$). This experiment has demonstrated how it is

possible to identify the boundaries where passive optimization is limited to impedances within the Smith Chart. As shown in Table 4-5, the highlighted IM3 data (Pin=22dBm) without impedance optimization has a negative sign and the ratio (a3/a5) > (-25/24*A^2). These are both opposite to the cases with lower drive levels (Pin=19dBm,21dBm) and thus the optimum impedance tends to be out side of the Smith Chart. This is an important note when considering improving linearity for those high power high saturation PAs, where the level of knee boundary interaction has increasing contribution to the distortion.

Table 4-5 Nonlinear coefficients and IM3 for different drive level investigation

Experiments		**Z_{LBB}**	**a3**	**a5**	**A**	**IM3**
Class AB	P_{in}=19 dBm	0	6.5232e-2	-1.4815e-2	1.591	9.8804e-3
(Ref. Case)	P_{in}=21 dBm	0	6.6018e-2	-1.4860e-2	2.012	2.5642e-3
	P_{in}=22 dBm	0	6.5048e-2	-1.4288e-2	2.257	**-1.1606e-2**
Class AB	P_{in}=19 dBm	55Ohm	1.1630e-1	-4.3560e-2	1.591	5.4495e-4
(Opt. BB)	P_{in}=21 dBm	5Ohm	7.1960e-2	-1.7023e-2	2.012	1.3475e-4
	P_{in}=22 dBm	-15Ohm	4.9864e-2	-9.6277e-3	2.257	**-1.3184e-3**
Experiments		**Z_{LBB}**	**a3/a5**		**A**	**-25/24*A^2**
Class AB	P_{in}=19 dBm	0	-4.403		1.591	-2.637
(Ref. Case)	P_{in}=21 dBm	0	-4.443		2.012	-4.217
	P_{in}=22 dBm	0	-4.553		2.257	**-5.306**
Class AB	P_{in}=19 dBm	55Ohm	-2.670		1.591	-2.637
(Opt. BB)	P_{in}=21 dBm	5Ohm	-4.203		2.012	-4.217
	P_{in}=22 dBm	-15Ohm	-5.179		2.257	**-5.306**





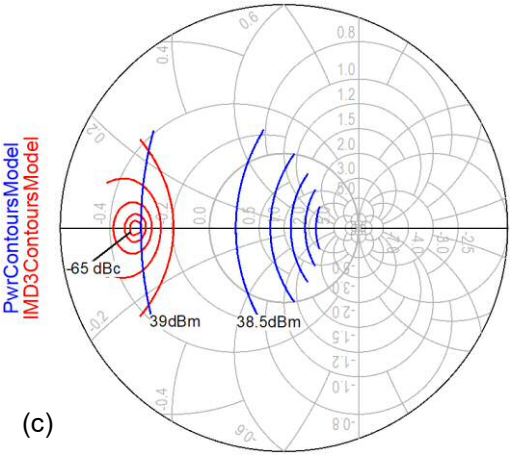
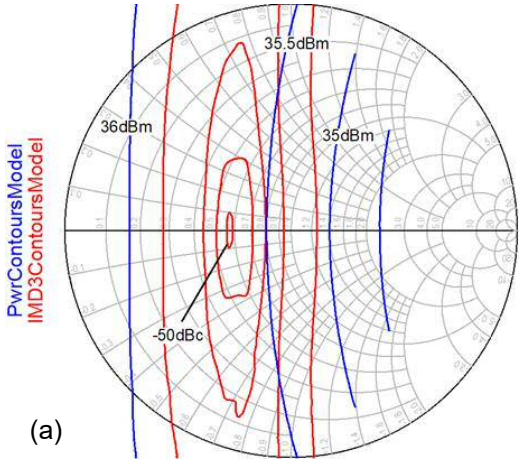
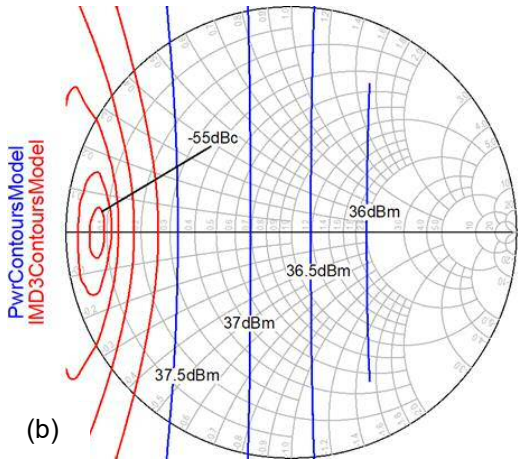


Figure 4.25 Two-tone baseband load-pull with different drive level:
IMD3 contours: Red, step=5dBc
Power contours: Blue, step=0.5dBm
(a)Pin=19dBm,(b)Pin=21dBm,(c)Pin=22dBm





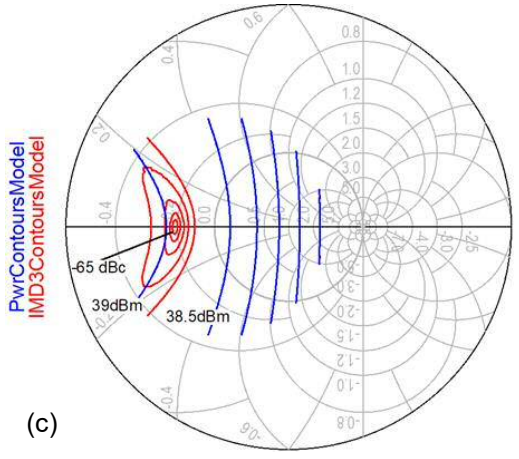


Figure 4.26 Two-tone 2nd harmonic load-pull with different drive level:
IMD3 contours: Red, step=5dBc
Power contours: Blue, step=0.5dBm
(a)Pin=19dBm,(b)Pin=21dBm,(c)Pin=22dBm

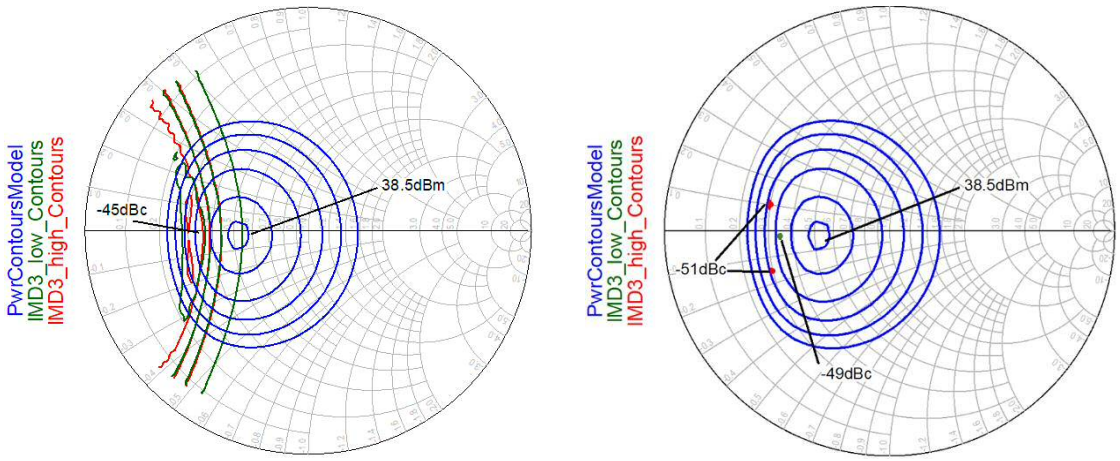
4.4.2 Distortion from parasitic capacitance

In addition to the current generator, the impact from nonlinear parasitic capacitance is analysed in this section. Although the nonlinear capacitance can be simply considered as nonlinear impedance presented at the current generator, which links back to the analysis in the previous section, a real capacitance behaves differently to a single impedance / admittance spot on Smith Chart. It presents, in nature, a frequency dependent impedance. In other words, even a linear reactive parasitic in the device will present different reactance responding across a certain frequency band. So, the reactive components in the circuits normally require careful consideration on wideband designs. Although it is often referred to in terms of efficiency and power, bandwidth also presents a limitation for linearity. In the perspective of this work, reactive parasitic components can lead to problematic IMD asymmetries [114]. A number of two-tone experiments have been conducted to reveal the impact of the parasitic capacitance. Same as in previous experiments, a two-tone excitation with a tone spacing of 100MHz centred at 1GHz is used. The simulations are conducted using HB simulation in the frequency domain.

First a limited number of constant capacitance cases are investigated. The values of the constant capacitance used here are selected to be representative of the linear averages of the nonlinear capacitance corresponding to the 10W Cree bare die device discussed before.

1. Constant gate capacitance only

Figure 4.27 shows the fundamental load-pull (output) result with constant gate-source capacitance $C_{gs(const.)}$ only, where the same impedance is applied to all spectral components around the fundamental tones. There is a need to mention, the contour plot is very useful to present the optimum case, especially in a trade-off design, however, the accuracy of this plot depends on the data interpolation and is limited by the resolution (contour steps). In a theoretical sense, a more precise data interpreting tool is needed. The raw data from the load-pull simulation is re-plotted in Figure 4.28. To explain the results, it can be observed that in Figure 4.27, the power contours are located symmetrically centred on real axis in the Smith Chart, similar to that shown in Figure 4.22. This is because a gate-source capacitance alone, without the feedback mechanism from C_{gd}, can be considered as isolated from the output node. However, the input gate voltage still changes as the capacitance presents a frequency dependent impedance load to the signal source. Thus, there is a difference between IMD3 High and IMD3 Low observed in both Figure 4.27 and Figure 4.28. This behaviour translates into IMD3 asymmetries in practical circuit design, where a certain impedance, normally optimized for efficiency and power, is presented as the output load.

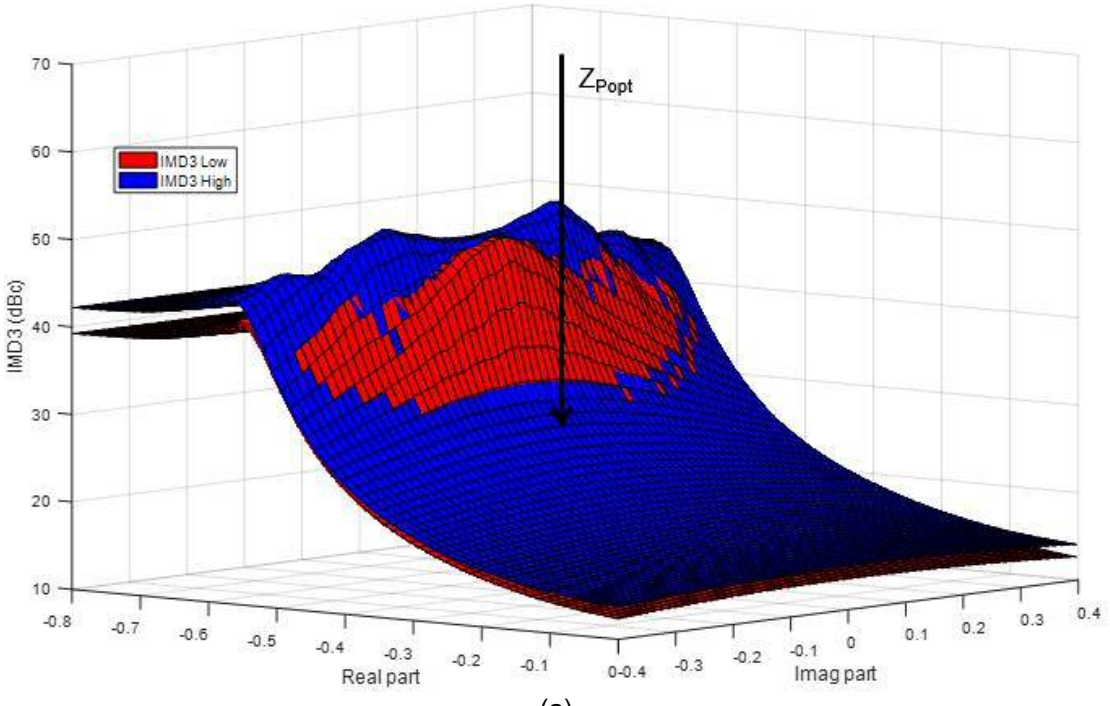


(a) (b)

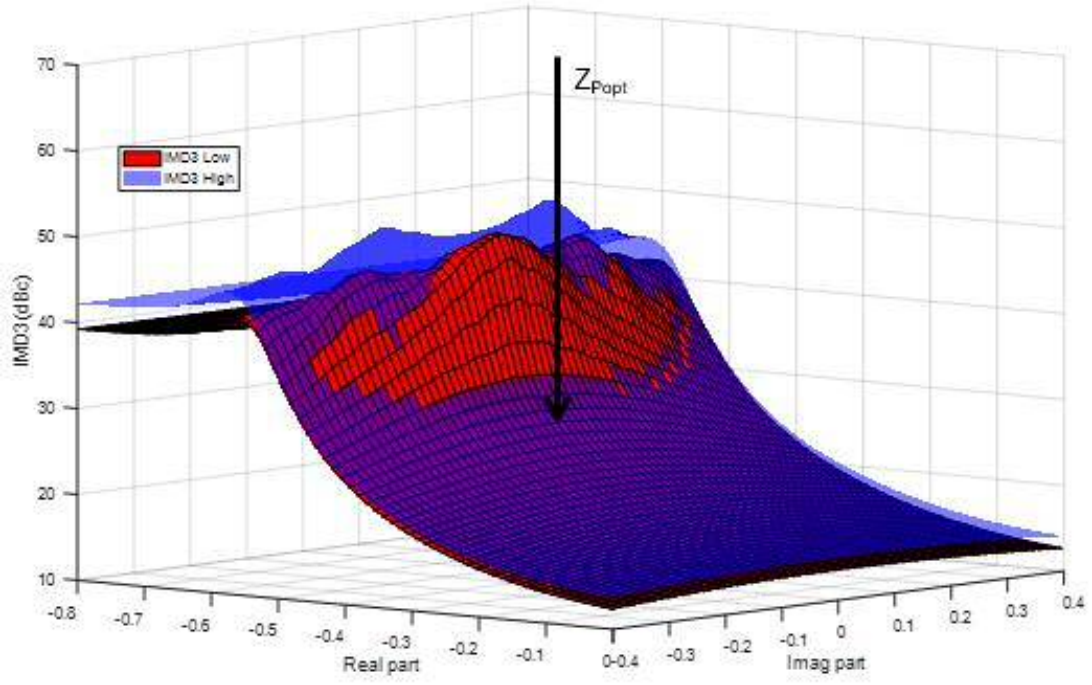
Figure 4.27 Fundamental load-pull with constant $C_{gs(const.)}$:

Two-tone signal: 100MHz @1GHz, Out-of-band impedance $Z_{outband} = 0$

(a) Power contours: Blue, IMD3 contours: Red(high), Green(low)
(b) Power contours: Blue, IMD3 peaks: Red(high), Green(low)



(a)



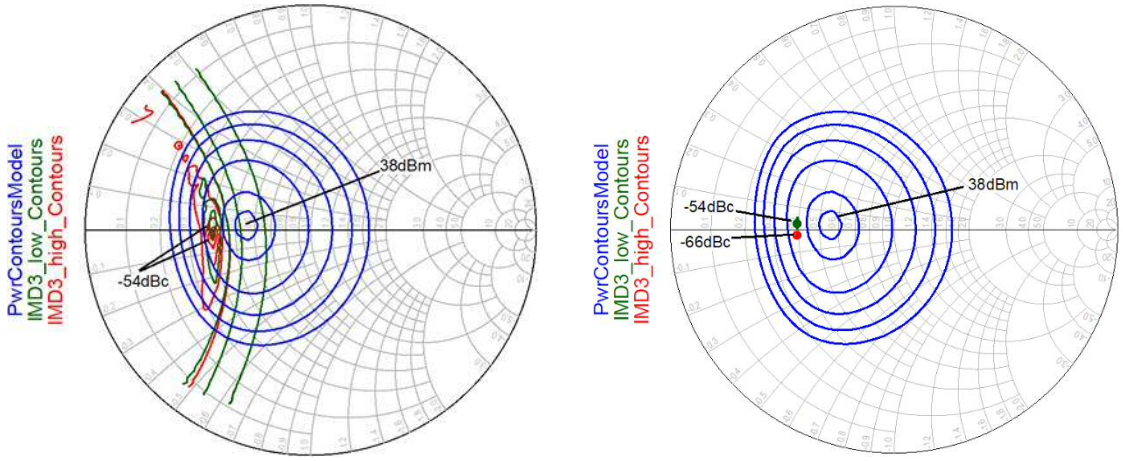
(b)

Figure 4.28 3D re-plots of the fundamental load-pull with constant $C_{gs(const.)}$

(a) Opaque surface
(b) IMD3 High semi-transparent

2. Constant gate-source capacitance + feedback capacitance

Figure 4.29 presents the fundamental load-pull results with constant gate-source capacitance $C_{gs(const.)}$ and feedback capacitance C_{gd}, which is then re-plotted in 3D in Figure 4.30. It can be observed in Figure 4.29 that the power contours and IMD3 contours are now shifted, as the gate capacitance has an impact on the output via the feedback mechanism through C_{gd}. The IMD3 asymmetries can also be observed in this case as the presence of the capacitive load to both input and output.

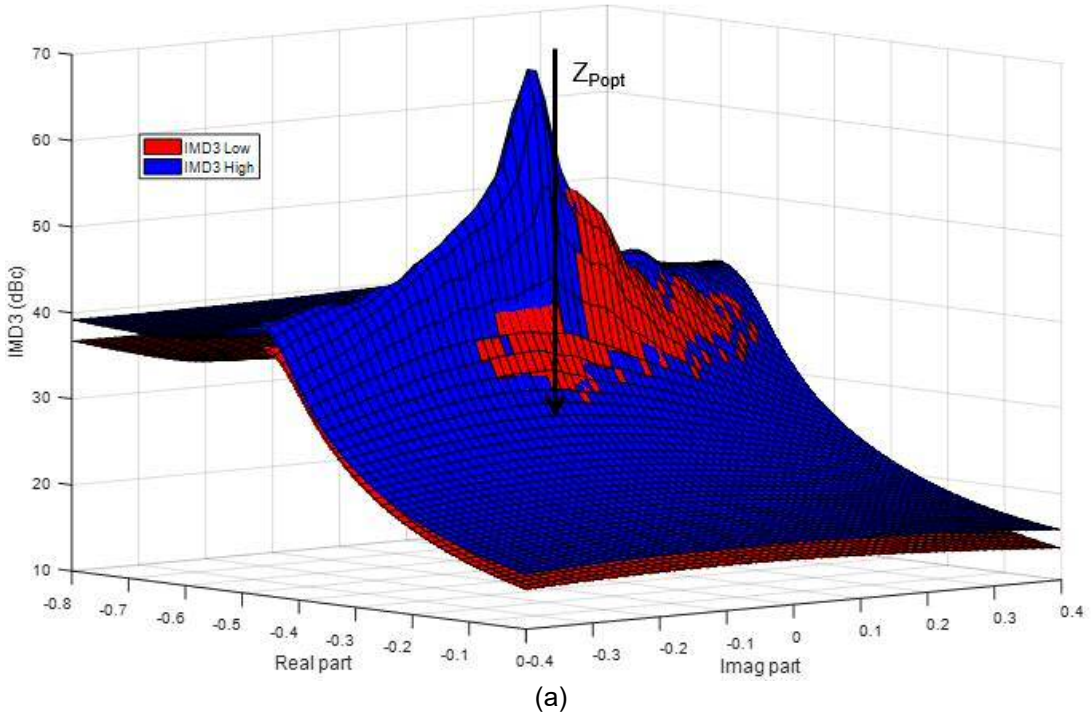


(a) (b)

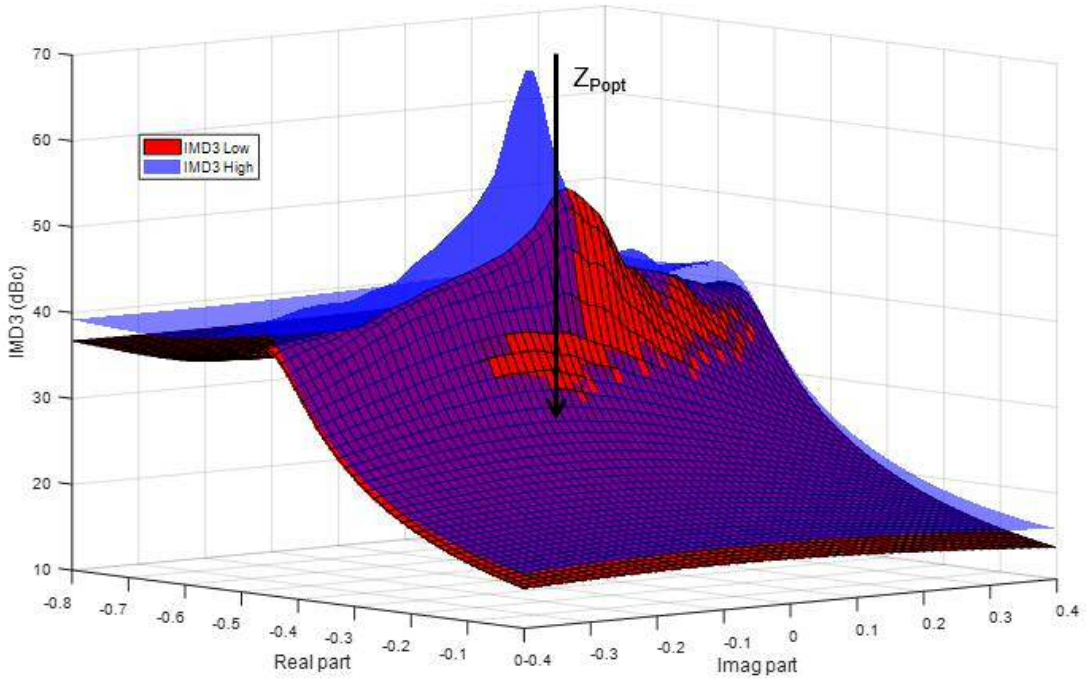
Figure 4.29 Fundamental load-pull with constant $C_{gs(const.)} + C_{gd}$:

Two-tone signal: 100MHz @1GHz, Out-of-band impedance $Z_{outband} = 0$

(a) Power contours: Blue, IMD3 contours: Red(high), Green(low)
(b) Power contours: Blue, IMD3 peaks: Red(high), Green(low)



(a)



(b)

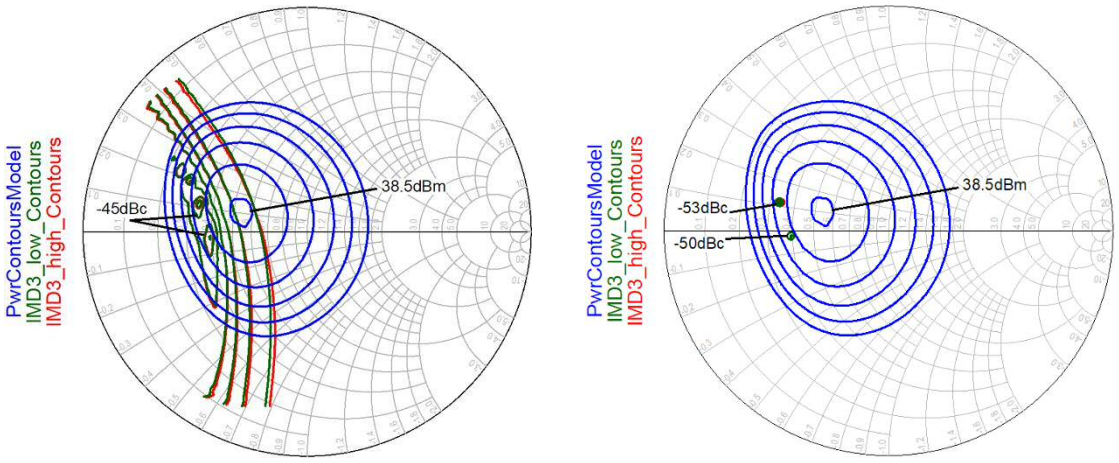
Figure 4.30 3D re-plots of the fundamental load-pull with constant $C_{gs(const.)}+C_{gd}$

(c) Opaque surface
(d) IMD3 High semi-transparent

3. Constant drain-source capacitance only

Figure 4.31 presents the fundamental load-pull results with constant drain-source capacitance $C_{ds(const.)}$ only, which is then re-plotted in 3D in Figure 4.32. It can be observed in Figure 4.31 that the power contours and IMD3 contours are shifted, as the drain-source capacitance has a direct impact on the output terminations presented to the current generator. The difference between IMD3 Low and IMD3 High is not large compared to the cases with input capacitance. This is probably due to the fact that the transfer characteristc's dependency on V_{ds} is not comparable to the dependency on V_{gs} , at this drive level. However, with an incresed drive level, it should be only visible if the nonlinearities from the trans-

conductive current generator is not overwhelmingly dominating the IMDs. This is also linked to a conlusion in [76] that C_{ds} has negligible impact on AM/PM in the GaN HEMT based PAs, however, it is one of the dominant contributors in LDMOS. So there is a need to mention that, the observation in this experiment might be closely linked to the specific device model used in this thesis which itself is dependent on the device technology.

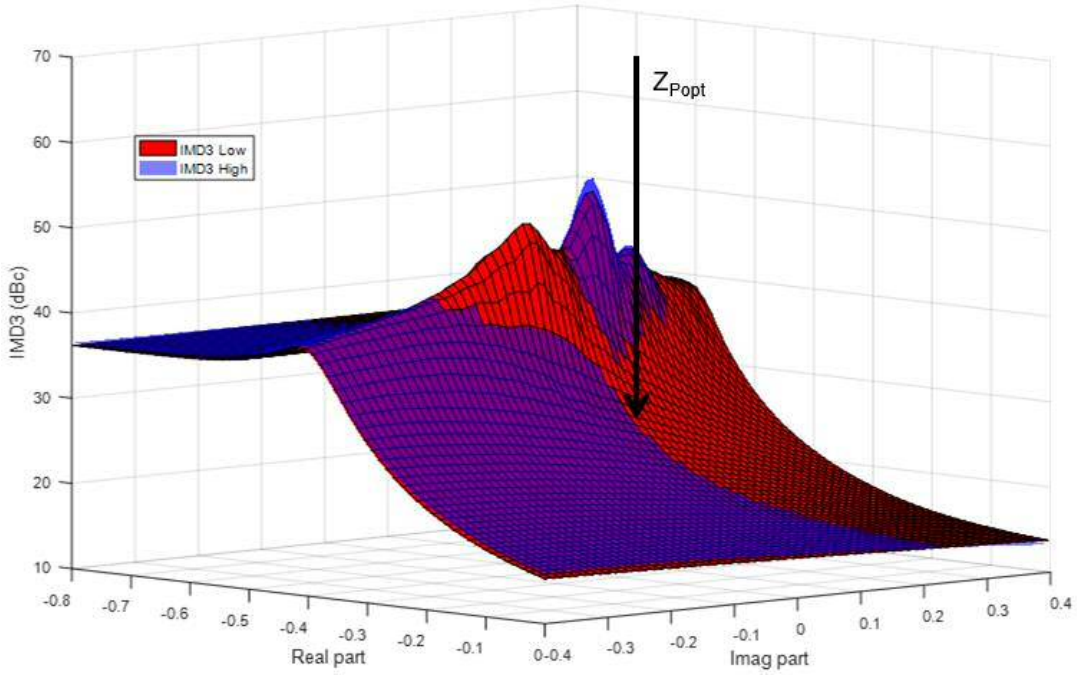


(a) (b)

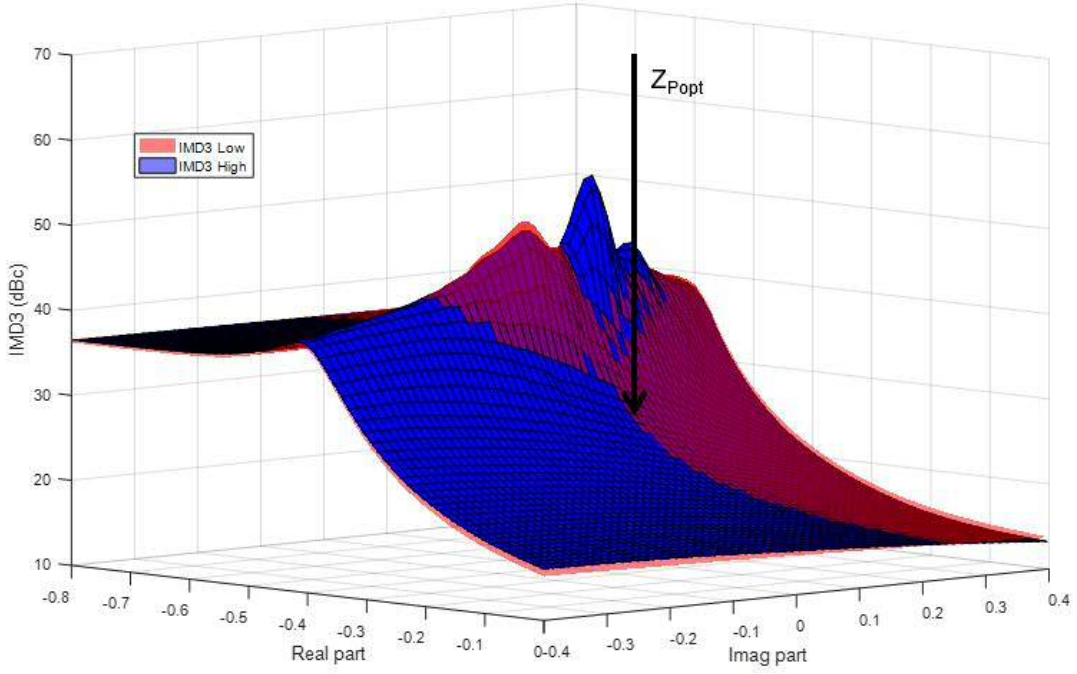
Figure 4.31 Fundamental load-pull with constant $C_{ds(const.)}$:

Two-tone signal: 100MHz @1GHz, Out-of-band impedance $Z_{outband} = 0$

(a) Power contours: Blue, IMD3 contours: Red(high), Green(low)
(b) Power contours: Blue, IMD3 peaks: Red(high), Green(low)



(a)



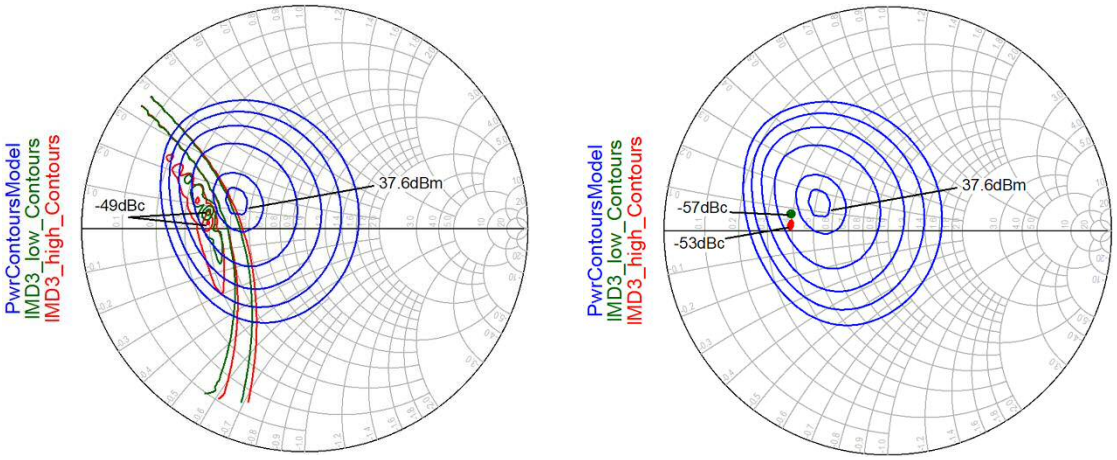
(b)

Figure 4.32 3D re-plots of the fundamental load-pull with constant $C_{ds(const.)}$

(a) IMD3 High semi-transparent
(b) IMD3 Low semi-transparent

4. Full constant capacitance model

Figure 4.33 presents the fundamental load-pull results with a full constant capacitance model with $C_{gs(const.)} + C_{gd} + C_{ds(const.)}$, which again is then 3D re-plotted in Figure 4.34. The results can be concluded as when the current generator is exposed to a capacitive loaded environment. The presence of a frequency dependent impedance load (capacitance), in this case, has different impact on IMD3 low and high sidebands, which notably translates to sideband asymmetry. The optimum load-pull centres (output) for the best IMD3 performance of each sideband can split due to this effect.

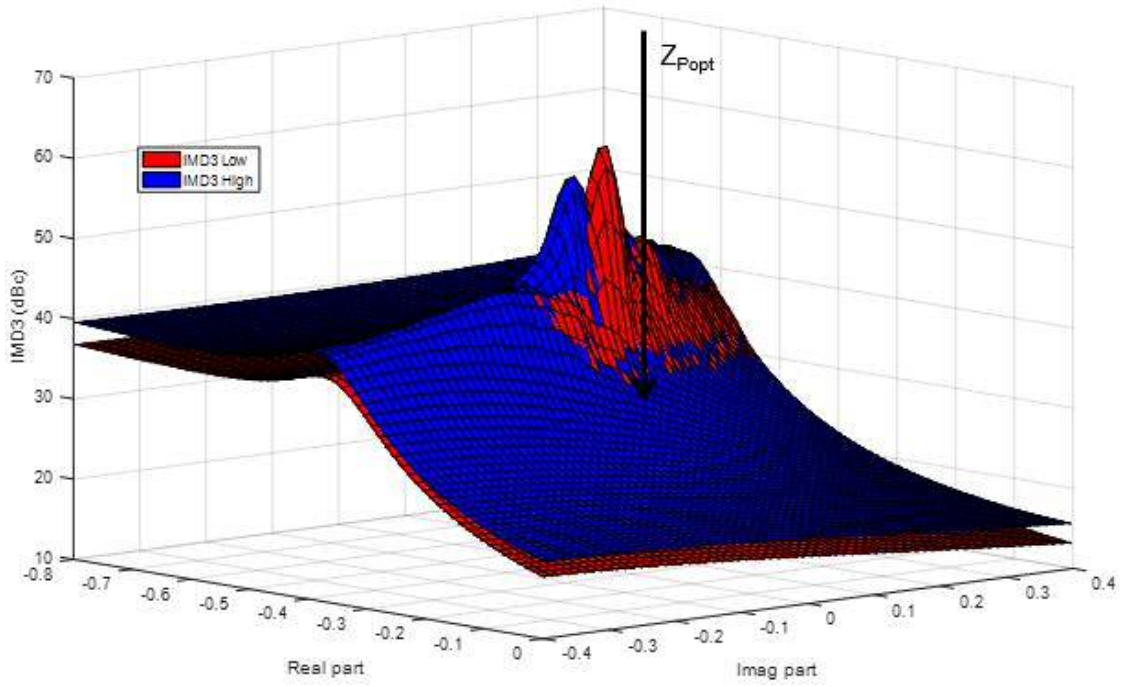


(a) (b)

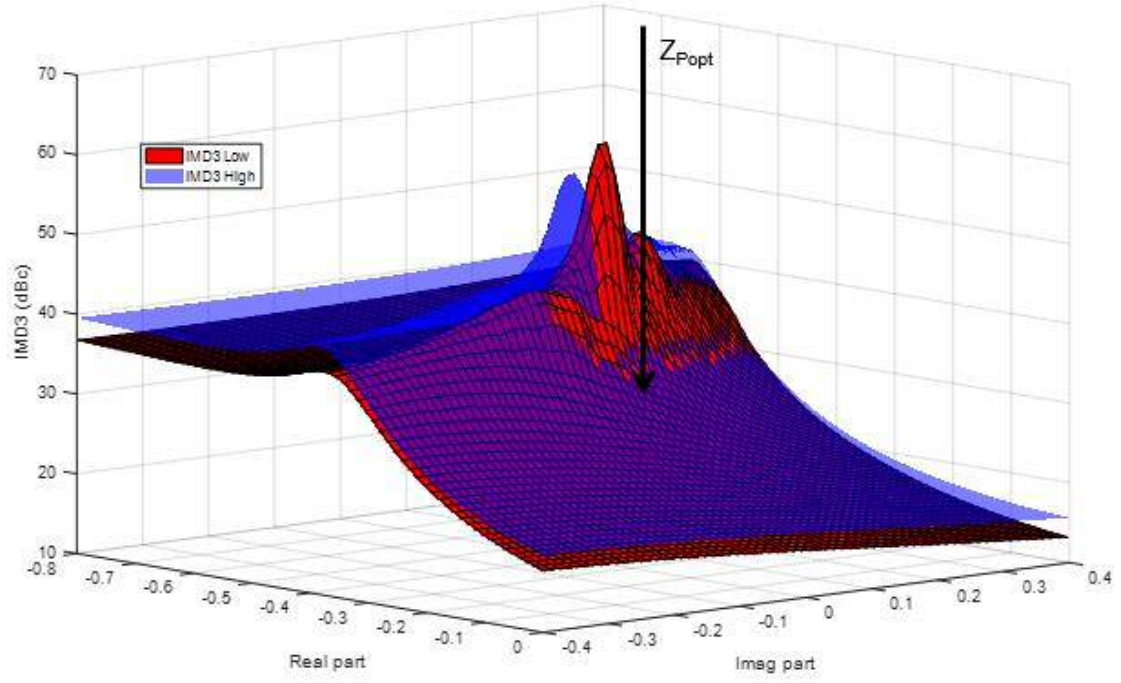
Figure 4.33 Fundamental load-pull with full constant capacitance model $C_{gs(const.)} + C_{gd} + C_{ds(const.)}$:

Two-tone signal: 100MHz @1GHz, Out-of-band impedance $Z_{outband} = 0$

(a) Power contours: Blue, IMD3 contours: Red(high), Green(low)
(b) Power contours: Blue, IMD3 peaks: Red(high), Green(low)



(a)



(b)

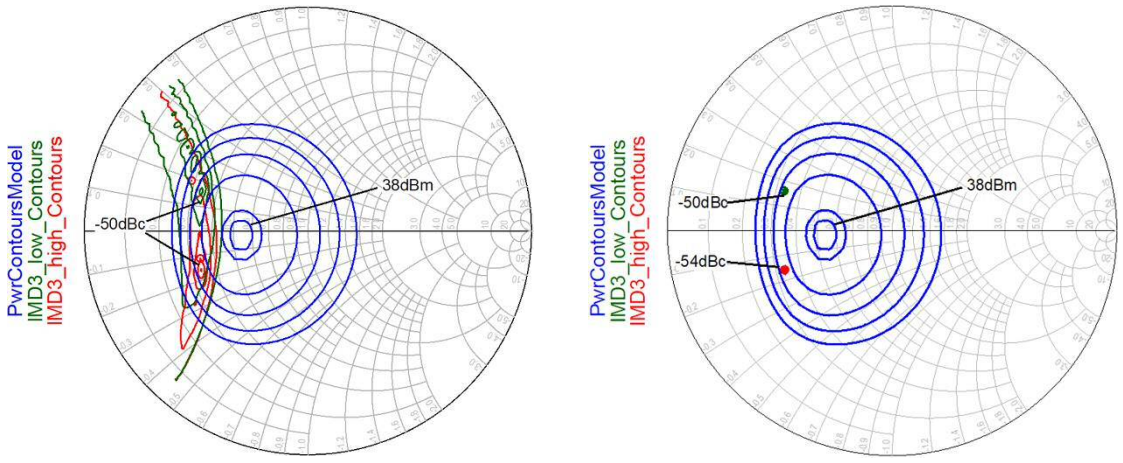
Figure 4.34 3D re-plots of the fundamental load-pull with full constant capacitance model $C_{gs(const.)} + C_{gd} + C_{ds(const.)}$

(a) Opaque surface

(b) IMD3 High semi-transparent

5. Nonlinear gate-source capacitance only

Figure 4.35 shows the results with nonlinear gate-source capacitance $C_{gs(nlinear)}$ only. It can be observed that in addition to the impact from a linear constant gate-source capacitance shown in Figure 4.27, the nonlinear capacitance has a more obvious impact on the IMD3 contours which splits the IMD3 low and high. This is due to the nature that the nonlinear capacitance generates additional distortion at the input node while the constant capacitance simply delays the signal without necessarily distorting it. Although the power contours still symmetrically align on real axis, as there is no feedback mechanism between input and output, the linearity behavior is significantly changed with the presence of a nonlinear input capacitance.



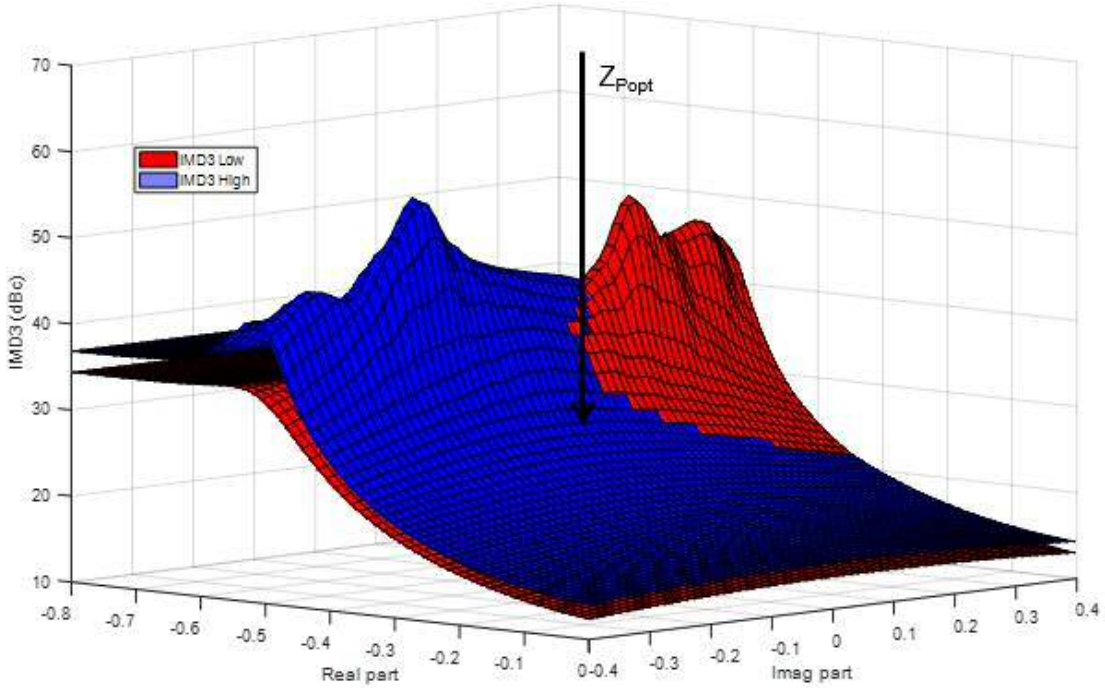
(a) (b)

Figure 4.35 Fundamental load-pull with nonlinear $C_{gs(nlinear.)}$:

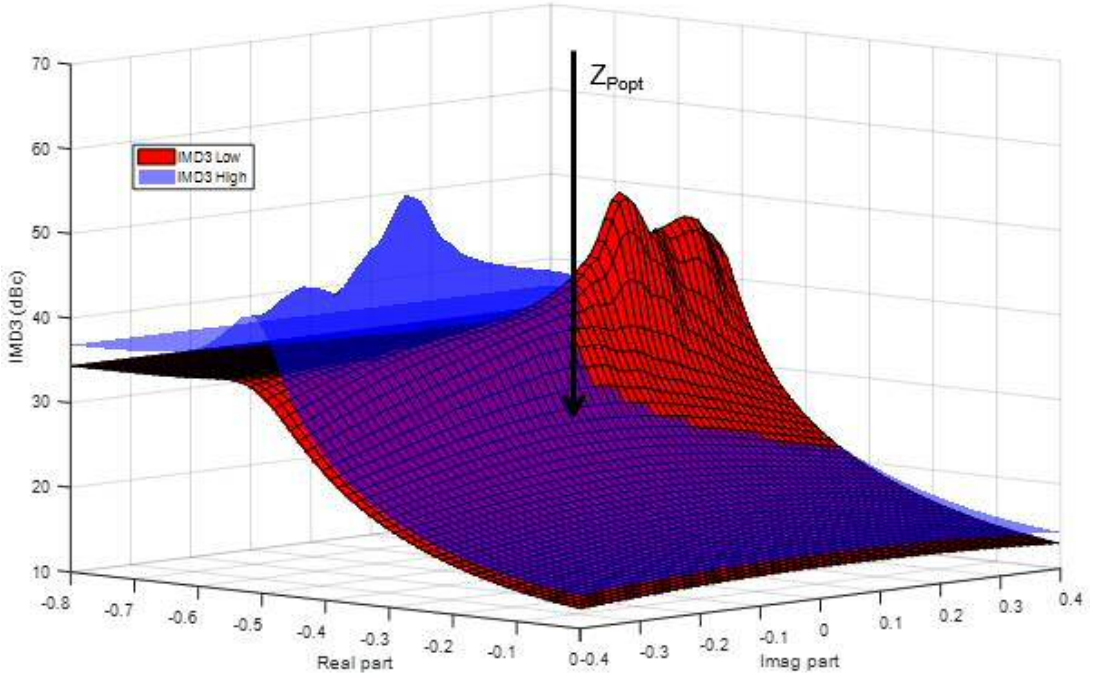
Two-tone signal: 100MHz @1GHz, Out-of-band impedance $Z_{outband} = 0$

(a) Power contours: Blue, IMD3 contours: Red(high), Green(low)

(b) Power contours: Blue, IMD3 peaks: Red(high), Green(low)



(a)



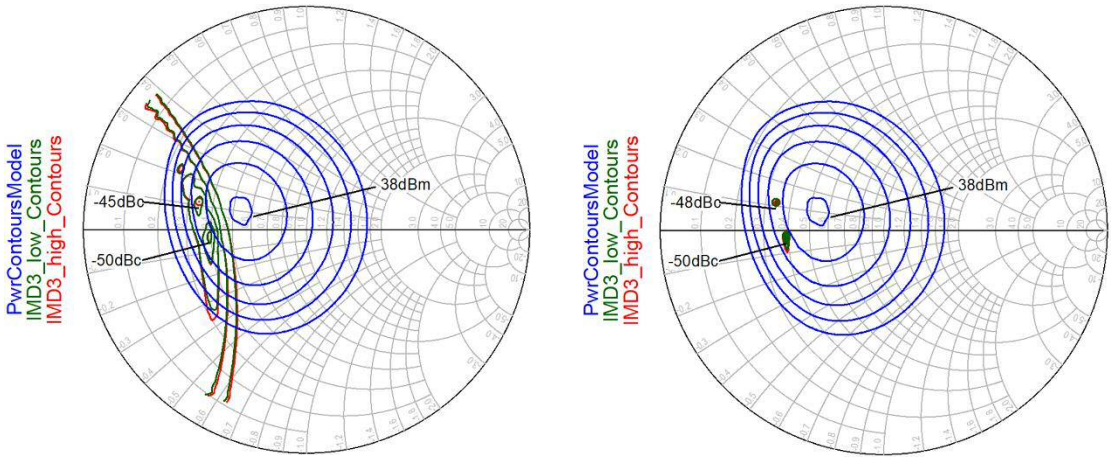
(b)

Figure 4.36 3D re-plots of the fundamental load-pull with nonlinear $C_{gs(nlinear.)}$:

(a) Opaque surface
(b) IMD3 High semi-transparent

6. Nonlinear drain-source capacitance only

Although it has been discussed before, C_{ds} has limited impact on the linearity of GaN based devices, it is worth to look at the impact from the nonlinear part in C_{ds}. The results are shown in Figure 4.37 and Figure 4.38. It can be observed that IMD3 low and IMD3 high are still very close with a twin-peaks feature, which is similar to the one observed in constant drain-source capacitance case, as shown in Figure 4.32.

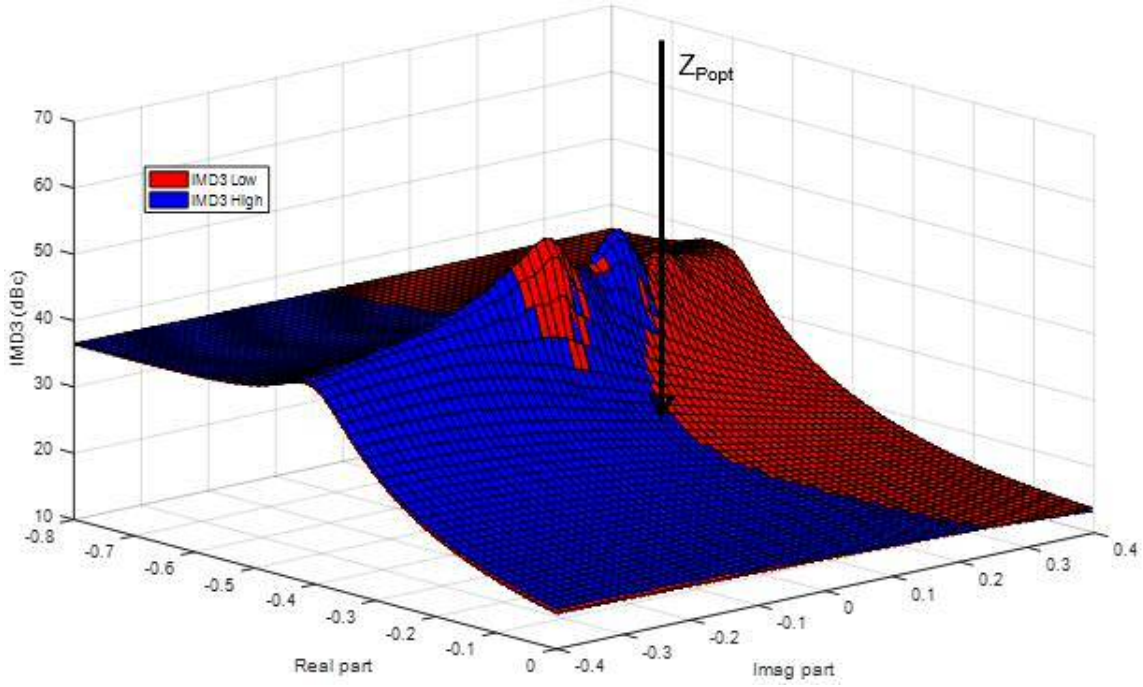


(a) (b)

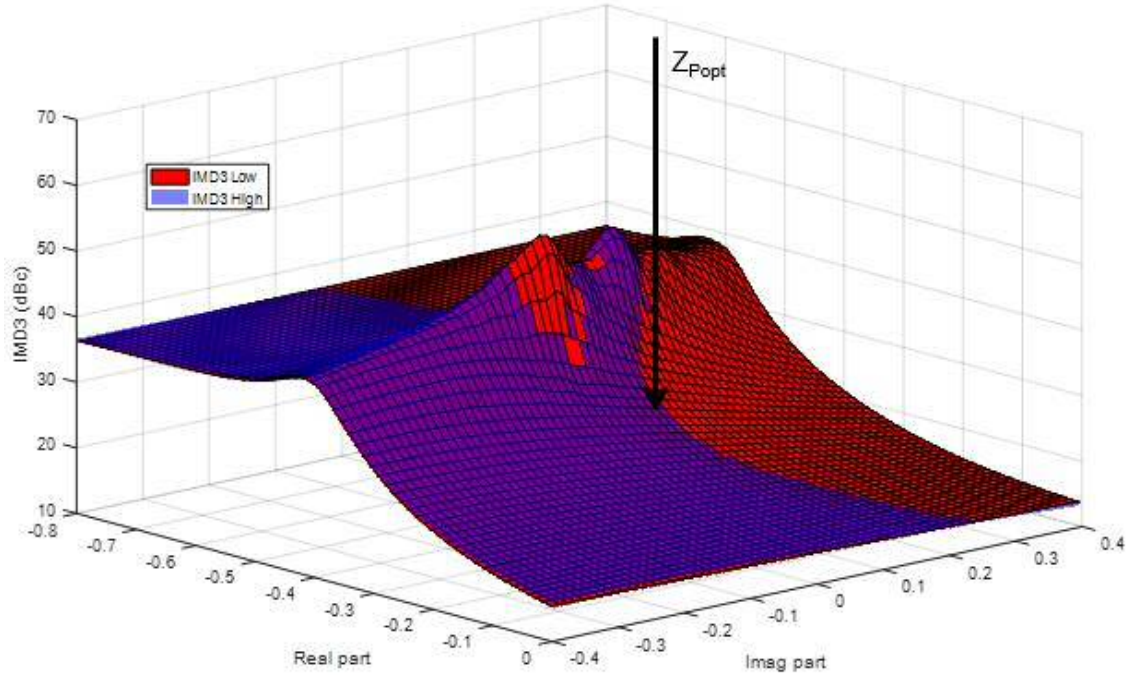
Figure 4.37 Fundamental load-pull with nonlinear $C_{ds(nlinear.)}$:

Two-tone signal: 100MHz @1GHz, Out-of-band impedance $Z_{outband} = 0$

(a) Power contours: Blue, IMD3 contours: Red(high), Green(low)
(b) Power contours: Blue, IMD3 peaks: Red(high), Green(low)



(a)



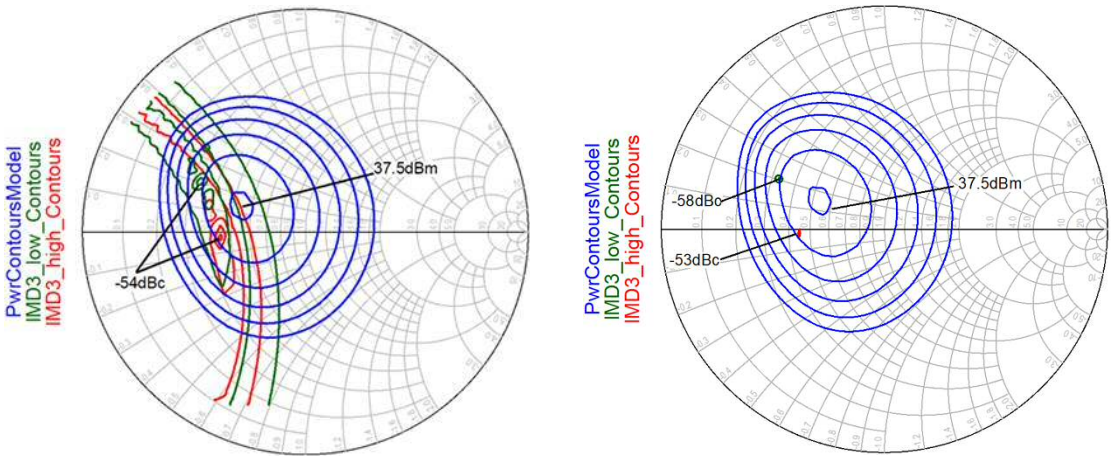
(b)

Figure 4.38 3D re-plots of the fundamental load-pull with nonlinear $C_{ds(nlinear.)}$

(a) Opaque surface
(b) IMD3 High semi-transparent

7. Full nonlinear capacitance model

The results for a full nonlinear model, which is the analytical model established earlier in this chapter are shown in Figure 4.39 and Figure 4.40. As the overall linearity behavior is a combination of all the nonlinear contributors, the splitting of IMD3 peaks, the difference between IMD3 low and high and the rotation of the contours can all be observed in this case.

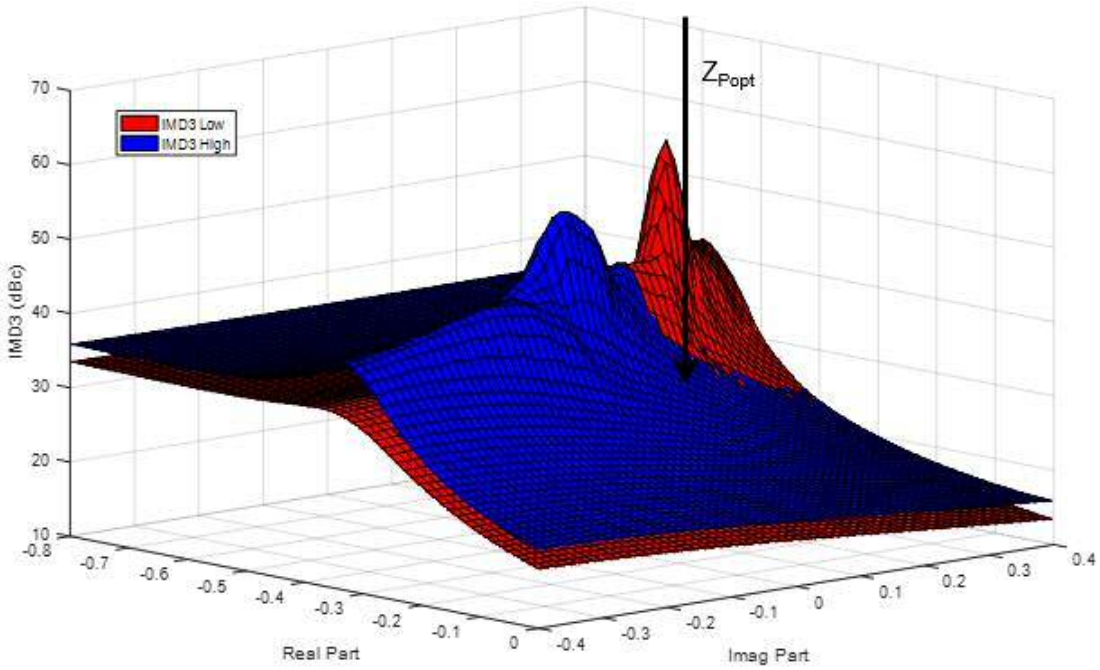


(a) (b)

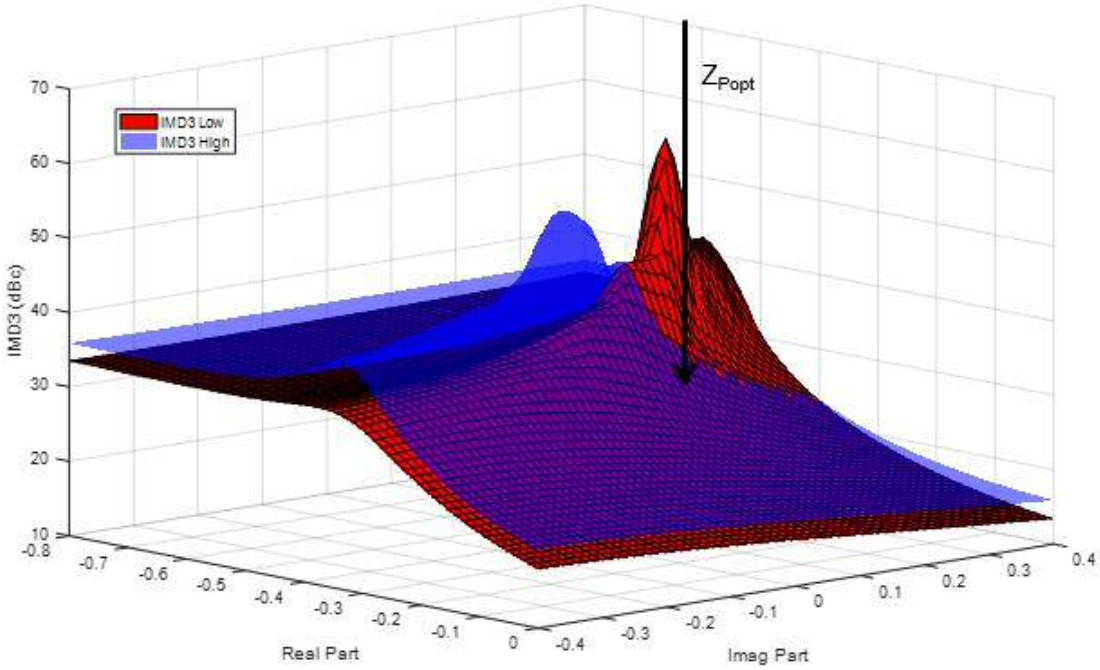
Figure 4.39 Fundamental load-pull with full nonlinear model $C_{gs(nlinear.)} + C_{gd} + C_{ds(nlinear.)}$:

Two-tone signal: 100MHz @1GHz, Out-of-band impedance $Z_{outband} = 0$

(a) Power contours: Blue, IMD3 contours: Red(high), Green(low)
(b) Power contours: Blue, IMD3 peaks: Red(high), Green(low)



(a)



(b)

Figure 4.40 3D re-plots of the fundamental load-pull with full nonlinear model $C_{gs(nlinear.)} + C_{gd} + C_{ds(nlinear.)}$:

(a) Opaque surface

(b) IMD3 High semi-transparent

All these experiments together demonstrate the nonlinear contributions from the parasitic capacitance in a device, specifically a GaN based device. To summarize:

1. C_{gs} alone, without a feedback path, has no direct impact on the output side of the device. However, as the capacitor is a frequency dependent impedance that loads the signal source, it still generates distortion in the form of frequency dependent voltages at the gate. Thus, the difference between IMD3 low and high can be observed even with a single, constant C_{gs} present. When this is compared to the behaviour of a nonlinear C_{gs}, the impact from gate-source capacitance becomes more visible.

2. C_{gd} is the feedback capacitance connecting input and output nodes. As discussed earlier, it is assumed to be a linear capacitance in this analytical model. The contribution of C_{gd} is not analysed in detail in this work. However, it has been shown that when it connects the C_{gs} to the drain side, together, they rotate the contours and affect the overall linearity of the circuit.

3. C_{ds} is found to be less relevant to the linearity behaviour of this model, especially for GaN technology [76]. This should be checked with various technology in the future work.

The contribution from all these components together with the current generator, IMD3 asymmetry, the splitting of IMD3 peaks and rotation of the contours have been observed and discussed. It is worth to mention that, although the IMD3 low and high "peaks" split or separate in the presence of these capacitances, which identify different ideal optimum terminating impedances, there are still *crossing-over* regions that are able to provide sideband symmetry $IMD3low=IMD3high$. However, these symmetry cases have to be achieved at compromised raw IMD3 levels and involves a discussion on the raw circuit level linearity and linearizability. In conclusion, these experiments have proven to be very useful in providing design optimization strategies at both circuit level and device technology level.

Although it is not discussed in detail in this thesis, in addition to the factors of IMD3 separation observed in this section, the bandwidth of the signal also has an impact. The separation is more obvious with a wideband excitation, while the separation distance depends on the signal bandwidth.

In a similar way, further load-pull simulations can be carried out to explore the effect of out-of-band impedances in baseband and 2nd harmonic frequencies. The results are shown in Figure 4.41 (a) and (b) respectively when using the full nonlinear analytical model. These experiments show that the out-of-band impedance can be used to optimize the linearity without affecting the optimum fundamental matching [111], [112]. Furthermore, and similar to the observation earlier, it is possible to achieve symmetrical sidebands with optimum out-of-band impedance at a compromised distortion level.

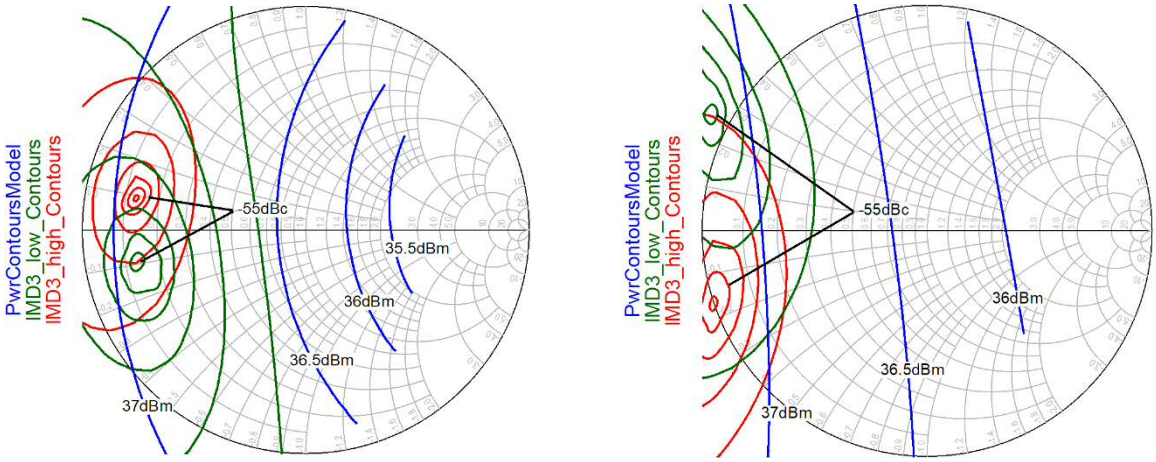


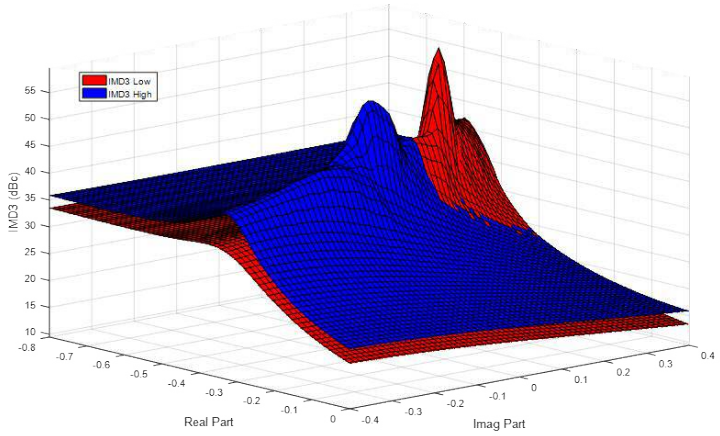
Figure 4.41 Baseband and 2nd harmonic load-pull with nonlinear $C_{gs(nlinear)}+C_{gd}+C_{ds(nlinear)}$:
Power contours: Blue, IMD3 peaks: Red(high), Green(low)
Two-tone signal: 100MHz @1GHz

(a) Baseband load-pull: $Z_{Lf_0}=21\Omega$, $Z_{L(2f_0\&higher)}=0\Omega$

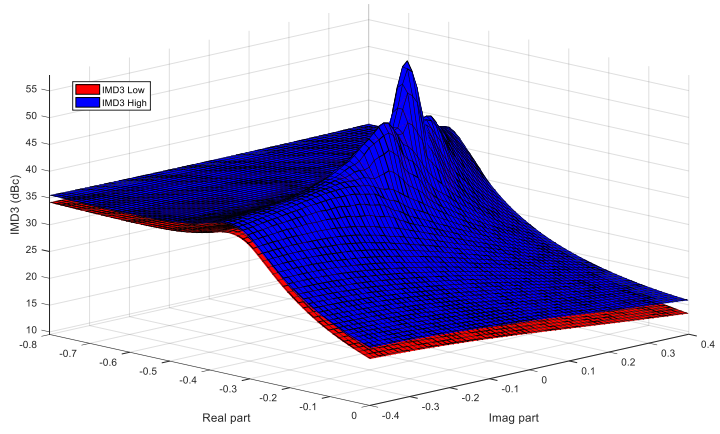
(b) 2nd harmonic load-pull: $Z_{Lf_0}=21\Omega$, $Z_{L(BB\&higher)}=0\Omega$

4.4.3 Impact of source baseband impedance

There is a need to mention that the results shown above are obtained from load-pull simulations where the load is swept while source impedance is fixed. It is an important observation to highlight, where, when the source baseband impedance is reduced, the splitting of the two optimum impedance points becomes less obvious.



(a)



(b)

Figure 4.42 3D plots comparison of fundamental load-pull with full nonlinear model:
IMD3 Low: Red, IMD3 High: Blue
(a) Z_{SBB} =50Ohm, (b)Z_{SBB}=0.01Ohm

As shown in Figure 4.42, when baseband source impedance is reduced in this case, the splitting of the IMD3 peaks becomes invisible whereas, the IMD3 low and high remains relatively close in the same shape and overlapping each other.

This can be explained through closer examination of the circuit at the input node. Ideally, the input signal should be purely linear at the gate before it is amplified by the device. However, when the input voltage is 'loaded' with a nonlinear input capacitance, C_{gs} in this case, the nonlinear behaviour of the capacitor will up / down convert the fundamental signal to the baseband or harmonic frequencies. This down converted nonlinear current is only able to generate a voltage if it is loaded with a relevant-band non-zero impedance, so when the source baseband impedance Z_{SBB} is reduced close to a short circuit, no distortion voltage is developed. As a result, and referring to the equivalent circuit shown in Figure 4.43, the input node is not able to support any baseband voltage components in the gate voltage V_{gs}.

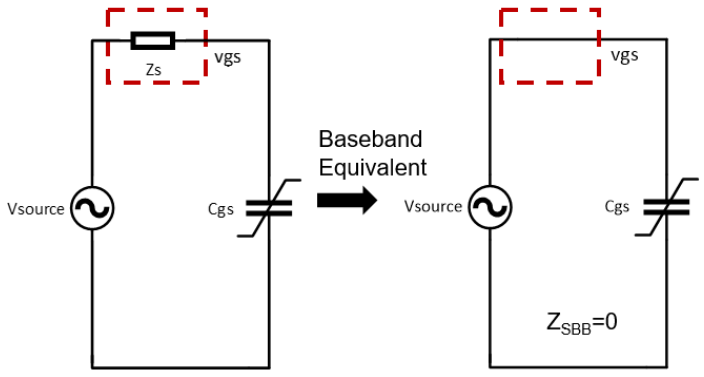


Figure 4.43 Baseband equivalent circuit for input node

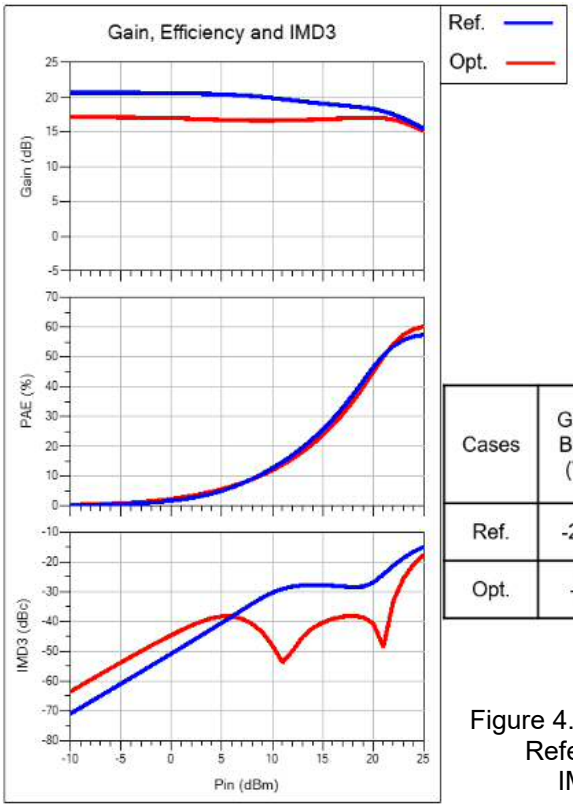
Generally, this baseband voltage should then, through the up-conversion / mixing process in the active nonlinear device, contribute to the IM products at the output. However, in this case, the contribution is greatly suppressed. Thus, the nonlinear contribution from C_{gs} is limited and the splitting effect of the IMD3 peaks becomes less visible. As linearization techniques, such as DPD, respond more effectively to symmetrical sidebands, 'shorting' the baseband as a technique can be extensively used to improve the linearizability of a RFPA, although it might be easily ignored when searching the local optimum of the raw linearity of RFPA. Initially, aiming to increase the video bandwidth of the RFPA, the benefits to linearity from a baseband short circuit at the input has been experimentally proven to be optimal in circuit design [115]. It has been observed that, with an input baseband short circuit, the designed RFPA performs with better linearity when corrected by a DPD system. To develop a sophisticated way to explore the actual optimum, especially for a non-short condition, requires an extensive work to establish a measurement system that supports source-pull at the input node, similar to the one developed in this thesis at the output side. This is beyond the scope of this work, however, potentially becomes an important measurement to fully address the linearizability of an RFPA circuit.

4.5 Circuit Level Optimization

To summarize the previous analysis based on the analytical model, the nonlinear distortion in an RFPA circuit is affected by:

1. Bias- which determines the nonlinear transconductance of the device.
2. Drive Level- as it affects the compression level and interaction with the knee and other boundaries.
3. Impedance Termination – as it affects the dynamic load-line, its interaction with knee, secondary mixing from out-of-band components and sideband asymmetry.

To verify the impact of these observations, the factors discussed above are applied in the optimization to improve linearity for an RFPA design using a commercial



Cases	Gate Bias (V)	Fundamental Termination (Ω)	Baseband Termination (Ω)	2nd harmonic Termination (Ω)
Ref.	-2.7	21	0	0
Opt.	-3	21	-j*0.5	0

Figure 4.44 Optimization for 10W Cree die device.
Reference case: Blue; Optimization: Red
IMD3 here presents only IMD3_low

10W GaN device model (Cree – CGH60015D) with results presented in Figure 4.44. Firstly, the optimum bias is selected to exploit the "sweet spot", close to compression at a small sacrifice of gain. The optimum fundamental termination is then determined using load-pull identifying the efficiency / power trade-off. The optimum baseband termination is then selected for minimizing the IMD3 sideband magnitude as well as asymmetry, which is shown in Figure 4.45. In this example, 2nd harmonic and higher harmonics were kept short circuit. However, they all have impact on IM products through the secondary mixing effect.

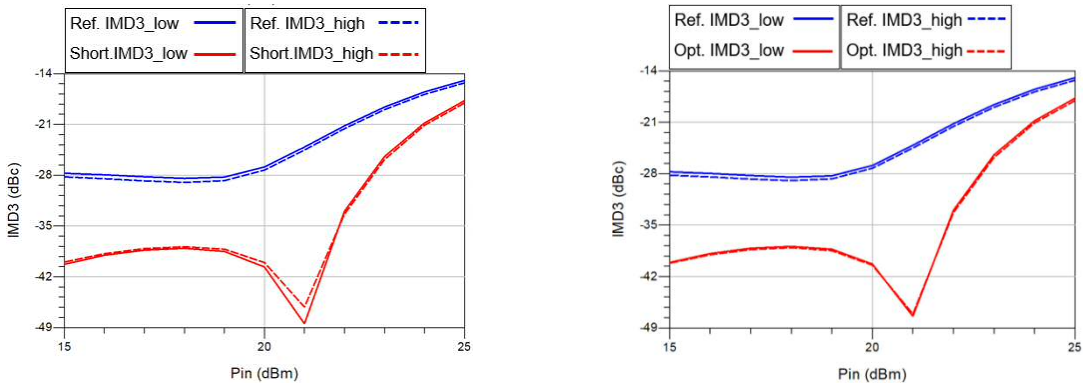


Figure 4.45 Optimization for IMD3 sideband asymmetry
Two-tone signal: 100MHz @1GHz
(a) Baseband short circuit, (b) Baseband optimum termination

This application example provides a simple application for some of the previous analysis. A more sophisticated design approach in terms of both circuit level optimization and its linearizability responding to the DPD system need to be verified in the future.

4.6 Conclusion

To conclude this chapter, an analytical device model has been established to investigate the distortion contributions from different nonlinear elements in a GaN based active device, including the trans-conductive current generator and the parasitic nonlinear capacitances. Then, a number of simulations are conducted to explore the impact on linearity from device bias, drive level and impedance terminations. In a nonlinear system with only a trans-conductive current generator, IM distortion is symmetrical and can be significantly reduced with the appropriate bias and out-of-band terminations. At higher drive levels, these terminations can necessitate active injection of signals. IM asymmetry is observed when reactive elements are exposed to the current generator. This effect is more significant in the case when nonlinear capacitors are introduced. While the symmetrical IM levels can represent an achievable compromise in favour of DPD systems to improve the circuit linearizability, the split IMD3 peaks represent local minima for raw linearity.

It is worth mentioning that the load-pull simulation bench used in this chapter is capable of control the baseband, fundamental and 2nd harmonic impedance environment simultaneously. To the author's knowledge, although a measurement system providing this level of capability is not yet available, the novel work presented in this chapter demonstrates through robust simulation, its significance in optimizing an RFPA. A number of key measurements have been identified for future linearity consideration and will be included in more detail in the future work section of this thesis. For the first time, a systematic study of all main nonlinear

components has been done in this chapter. The distortion contributions from those key nonlinear circuit components including the intrinsic current generator I_{gen}, input capacitance C_{gs}, output capacitance C_{ds} and feedback capacitance C_{gd} have been analysed holistically and verified with the simulation bench developed in this thesis. For the first time, the interaction of all these nonlinear components has been revealed and discussed. The observation in this chapter can be used not only to guide the circuit designers to achieve optimum circuit environment for high linearity and linearizability, but also to guide the device technology development to optimize the critical parasitics.

Chapter 5 Envelope Domain Analysis for Distortion in PA Circuits

5.1 Introduction

In the last chapter, an analysis on linearity at circuit level in the frequency domain is provided. Although it is possible to improve the linearity of RFPA by optimizing the circuit environment, circuit level linearization is very sensitive to the device characteristics and less flexible when compared to system level approaches. A pair of system level linearization approaches are presented in this chapter which are based on an envelope domain analysis of distortion generation and cancellation. The application of envelope linearization at both input / gate and output / drain are presented and discussed.

5.2 Complex Envelope Signals in Amplifier Circuits

Firstly, in this section, a review of the fundamental knowledge of envelope domain analysis is provided. Modern communication signals can be represented by a sum

of RF frequency carriers with complex amplitudes, which vary in time. In mathematical form, this can be described as:

$$S(t) = \mathrm{Re}\left\{ \sum_n A_n(t) e^{j\omega_n(t)} \right\} \tag{5.1}$$

in which, $A_n(t)$ is the time-varying complex amplitude for *n*th carrier with frequency ω_n. For instance, a two-tone sinusoidal signal, assuming an equal magnitude A_1 for both tones, can be represented in the envelope domain as,

$$\begin{aligned} S(t) &= A_1\cos(\omega_1 t) + A_1\cos(\omega_2 t) = 2A_1\cos(\frac{\omega_1-\omega_2}{2}t)\cos(\omega_c t) \\ &= \mathrm{Re}\left\{ 2A_1\cos(\frac{\omega_1-\omega_2}{2}t) e^{j\omega_c t} \right\} \\ &= \mathrm{Re}\left\{ A(t) e^{j\omega_c t} \right\} \end{aligned} \tag{5.2}$$

in which, the time-varying amplitude is in sinusoidal from,

$$A(t) = 2A_1\cos(\frac{\Delta\omega}{2}t) \tag{5.3}$$

An example two-tone waveform is shown in Figure 5.1, where the centre frequency is 1GHz and the bandwidth (tone-spacing) is 20MHz (f_1=990MHz, f_2=1010MHz).

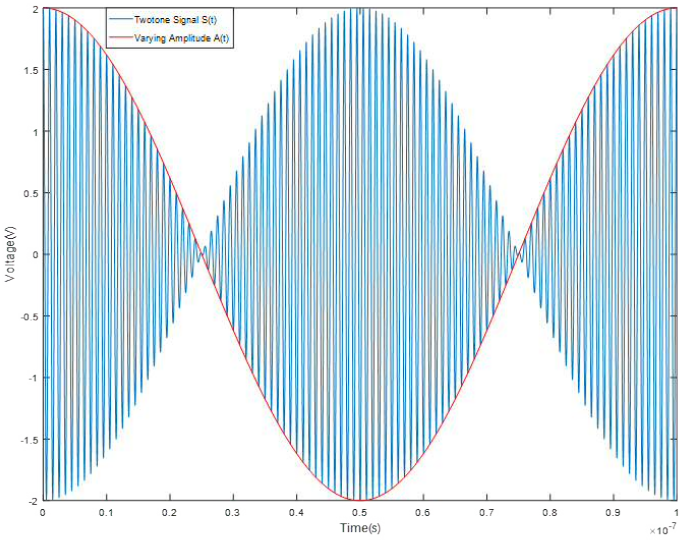
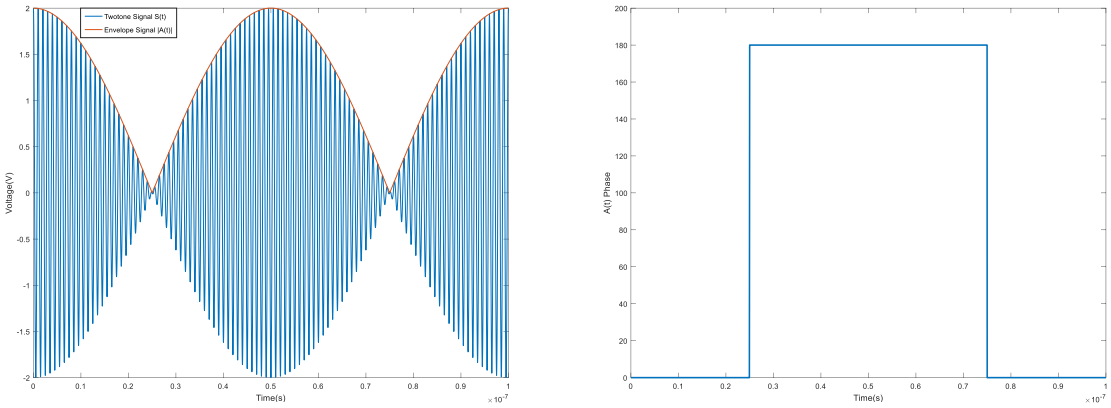


Figure 5.1 Two-tone waveform
Two-tone signal: 10MHz @1GHz
Two-tone signal S(t): Blue, Time-varying amplitude A(t): Red

The time-varying amplitude can be represented in terms of its envelope $|A(t)|$ and phase $\phi(A(t))$, as shown in Figure 5.2.



(a) (b)

Figure 5.2 Two-tone envelope signal:
(a) Envelope $|A(t)|$, (b) Phase $\phi(A(t))$

Furthermore, complex signals can be expressed in either rectangular or polar form [116]. The rectangular form is given below where x_I and x_Q are the amplitudes of two orthogonal sinusoidal signal. This form is also referred to as orthogonal *IQ* modulation.

$$S(t) = \mathrm{Re}\left\{A(t)e^{j\omega_c t}\right\} = x_I(t)\cos(\omega_c t) - x_Q(t)\sin(\omega_c t) \tag{5.4}$$

The spectral representation in terms of positive and negative frequency space can be written as;

$$S(t) = \mathrm{Re}\left\{A(t)e^{j\omega_c t}\right\} = \frac{1}{2}A(t)e^{j\omega_c t} + \frac{1}{2}A^*(t)e^{-j\omega_c t} \tag{5.5}$$

which is also referred to as complex analytic signal. There is a need to mention that complex analytic signal form is used later in this thesis, as it is more convenient for sorting and rearranging algebraic terms after massive expansion due to the signal mixing process.

5.3 Distortion Generation and Elimination in the Envelope Domain

To keep the mixing terms analytically manageable, for this work, a one-dimension (1-D) 5th order polynomial is used to describe the nonlinear RFPA,

$$y(t) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 \tag{5.6}$$

This 1-D function describes the input-output transfer characteristic. It is shown in last chapter that the values of these coefficients $(a_0, a_1, a_2, a_3, a_4, a_5, ...)$ are a function of how the transistor is biased, driven and the respective impedance terminations. This will be extended into a 2-D function when considering the drain node as the 2nd input of the RFPA which is the case later in Section 5.5.

When this nonlinear system is excited with a modulated signal $S(t)$,

$$S(t) = \frac{1}{2} A(t) e^{j\omega_c t} + \frac{1}{2} A^*(t) e^{-j\omega_c t} \tag{5.7}$$

the output signal will be

$$y(t) = a_0 + a_1\left(\frac{A(t)e^{j\omega_c t} + A^*(t)e^{-j\omega_c t}}{2}\right) + a_2\left(\frac{A(t)e^{j\omega_c t} + A^*(t)e^{-j\omega_c t}}{2}\right)^2 + a_3\left(\frac{A(t)e^{j\omega_c t} + A^*(t)e^{-j\omega_c t}}{2}\right)^3$$
$$+a_4\left(\frac{A(t)e^{j\omega_c t} + A^*(t)e^{-j\omega_c t}}{2}\right)^4 + a_5\left(\frac{A(t)e^{j\omega_c t} + A^*(t)e^{-j\omega_c t}}{2}\right)^5$$

(5.8)

Based on binomial theorem, this produces 6+5+4+3+2+1=21 terms. However, the most interesting terms for IM products will be those at $e^{j\omega_c t}$ or $e^{-j\omega_c t}$. After some rearrangement, the coefficient of the *in-band* term $e^{j\omega_c t}$ is

$$Y_{inband\_p} = \frac{(8a_1 + 6a_3 \mid A(t)\mid^2 + 5a_5 \mid A(t)\mid^4)A(t)}{16} \tag{5.9}$$

Symmetrically, the coefficient of the $e^{-j\omega_c t}$ term is

$$Y_{inband\_n} = \frac{(8a_1 + 6a_3 \mid A(t)\mid^2 + 5a_5 \mid A(t)\mid^4)A^*(t)}{16} \tag{5.10}$$

It is worth noting that the a_1 term of the *in-band* products presents the linear gain of the RFPA, while a_3 and a_5 introduce 3rd and 5th order nonlinear mixing products. It should also be mentioned that even order terms such as a_2 and a_4 will not fall into the IM products via this simple nonlinear process. Similar to the analysis presented in chapter 4, based on equation (5.9) it is possible to achieve an IMD null, just like the "sweet spot" if the device is biased with a_3 and a_5 having opposite signs. However, in the case of complex communication signals, the spectrum of

the time-varying envelope $|A(t)|^2$ is not the same as the spectrum of $|A(t)|^4$, so nulling of all the IM3 spectral lines as in the two-tone stimulus is not possible. This provides a motivation to introduce compensating mixing mechanisms by injecting an appropriate auxiliary signal as a function of same time-varying envelope. This can be done from both the input side (gate) and the output side (drain) exploiting different mixing mechanism from V_{gs} and V_{ds} respectively.

5.4 Input Envelope Linearization

In this section, the application of envelope linearization at input node is analysed and verified in the simulation environment. It contains the initial concept of input envelope linearization (IEL), the design challenges and then the simulation performance.

5.4.1 Input Envelope Linearization

Basically, the concept of input envelope linearization (IEL) is to introduce the compensating mixing terms into the transfer characteristic by manipulating input voltage V_{gs}. This can be done by either changing the input signal itself or injecting a small signal via the input bias network (adaptive gate biasing). As the whole nonlinear mathematical description will be dynamically modified by a varying bias, adaptive gate biasing is very difficult to include in a simple theoretical derivation. For this reason, this work focuses only on the theoretical analysis involved with manipulating the input signal. The aim is to achieve a simple mathematic solution for nonlinear correction which is predictable, rather than applying a *brute-force* solution involving exhaustive parameter sweeping.

To derive the desired input correction signal, firstly, we need to take a closer look at the *in-band* product of the nonlinear system. Assuming the RFPA is modelled with a polynomial series as in (5.6), then it follows that the *in-band* output components are as follows;

$$\begin{aligned} Y_{inband} &= Y_{inband\_p} + Y_{inband\_n} \\ &= \frac{8a_1 + 6a_3 \mid A(t)\mid^2 + 5a_5 \mid A(t)\mid^4}{8} \cdot \frac{1}{2}(A(t)e^{j\omega_c t} + A^*(t)e^{-j\omega_c t}) \\ &= (a_1 + \frac{6a_3 \mid A(t)\mid^2 + 5a_5 \mid A(t)\mid^4}{8}) \cdot S(t) \end{aligned} \tag{5.11}$$

The undesired mixing components are in the form of products from even order envelope terms $6a_3 \mid A(t)\mid^2$, $5a_5 \mid A(t)\mid^4$ and the original input signal $S(t)$, while the linear signal amplification is described by a_1. (5.11) forms an *in-band* mathematic description of this nonlinear system. Now, we introduce a new input signal, quantified as a product of even order envelope terms and the original signal,

$$\begin{aligned} S'(t) &= (b_0 + b_2 \cdot B + b_4 \cdot B^2) \cdot S(t) \\ &= (b_0 + b_2 \mid A(t)\mid^2 + b_4 \mid A(t)\mid^4) \cdot S(t) \end{aligned} \tag{5.12}$$

where $B = \mid A(t)\mid^2$. (b_0, b_2, b_4) are correction coefficients which are optimized for nonlinear cancellation. Substituting (5.12) into (5.6), the output mixing products will now be

$$y(t) = (\frac{a_1 b_0}{2} + \frac{3a_3 b_0^{\,3} + 4a_1 b_2}{8} \mid A(t)\mid^2 + \frac{5a_5 b_0^{\,3} + 18a_3 b_2 b_0^{\,2} + 8a_1 b_4}{8} \mid A(t)\mid^4 + \ldots) \cdot S(t) \tag{5.13}$$

In (5.13) the mixing products higher than 5th order are omitted. Thus, we successfully introduce a set of external parameters (b_0, b_2, b_4) into the nonlinear mixing products. The solution of equation (5.14) and (5.15) will give the proper (b_0, b_2, b_4) to eliminate the 3rd order and 5th order mixing products. Note in this case all the spectral distortion terms associated with spectrum of the time-varying envelope

$|A(t)|^2$ and $|A(t)|^4$ are nulled in this case. This includes the *in-band* terms which generating the EVM error. Basically, by solving equation (5.14), the 3rd order envelop is nulled instead of only IM3 spectral in frequency domain analysis.

$$3a_3b_0^3 + 4a_1b_2 = 0 \tag{5.14}$$

$$5a_5b_0^5 + 18a_3b_2b_0^2 + 8a_1b_4 = 0 \tag{5.15}$$

IEL technique will introduce the correction coefficients through the correction route.

Solution of equations (5.14) and (5.15), if possible, will give the values of the coefficients (b_0, b_2, b_4) to provide for perfect IEL in terms of the device nonlinearity. However, in practice, higher order terms and the impact of self-biasing will modify this solution resulting in the need for an iterative, optimization approach.

5.4.2 Complex Coefficients and Normalization

Note that the above analysis attempts to maintain generality and simplicity. There are two facts which have been simplified and omitted in the derivation of the expression.

The first is that the coefficients (b_0, b_2, b_4) should be complex for modulated signals including both magnitude and phase corrections. As IEL targets manipulation of the input signal, it has to be structured in a generic *I/Q* domain to be compatible with the modulated signal. The proposed application schematic of this IEL technique is shown in Figure 5.3. Following the correction route, the input signal $S(t)$

is first *I/Q* demodulated and then translated into envelope signal $|A(t)|^2$ by the ET squarer. This signal is then modulated by two polynomial functions in both in-

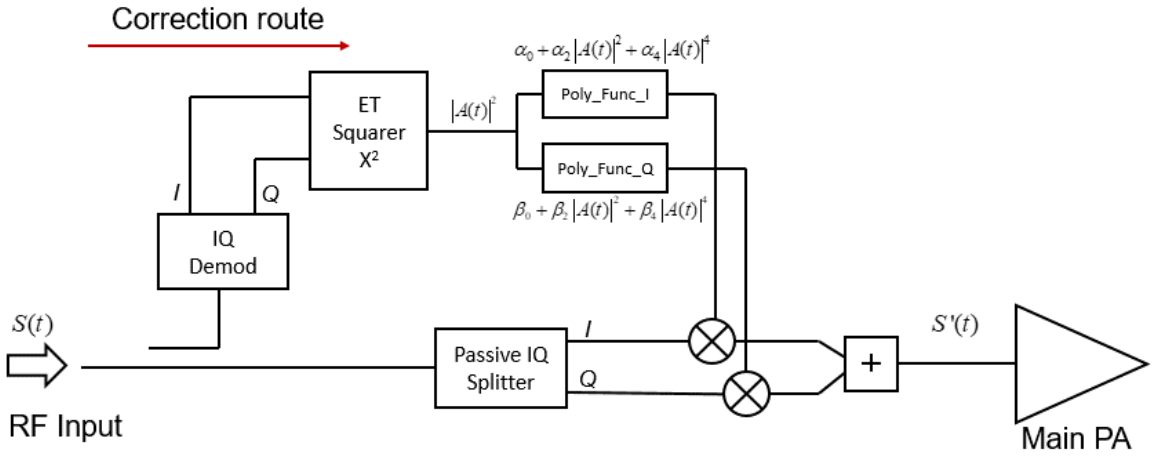


Figure 5.3 Application Schematic of Input Envelope Linearization (IEL)

phase (I) and quadrature (Q), separately. Later, these correction terms are mixed into the original signal which generates a new corrected input signal $S'(t)$. The full transformation from $S(t)$ to the new input signal $S'(t)$ in Figure 5.3 is shown in (5.16).

$$S'(t)=\left[\alpha_0-j\beta_0+(\alpha_2-j\beta_2)|A(t)|^2+(\alpha_4-j\beta_4)|A(t)|^4\right]\cdot S(t) \tag{5.16}$$

Comparing (5.16) and (5.12),

$$\begin{aligned} b_0&=\alpha_0-j\beta_0\\ b_2&=\alpha_2-j\beta_2\\ b_4&=\alpha_4-j\beta_4 \end{aligned} \tag{5.17}$$

The correction coefficients (b_0, b_2, b_4) are complex values, while $(\alpha_0, \alpha_2, \alpha_4)$, $(\beta_0, \beta_2, \beta_4)$ are real numbers.

Note the DC offset coming from the squarer and even order operation. The transfer characteristic is captured and modelled into a polynomial expression. The ET squarer and even order terms will generate a signal-dependent offset. This is highlighted in (5.18) for the example case of a single tone;

$$\begin{aligned}(A(t)\cos(\omega_0 t))^2 &= \frac{A^2(t)}{2}(\cos(2\omega_0 t)+1) \\ &= \frac{A^2(t)}{2}\cos(2\omega_0 t) + \boxed{\frac{A^2(t)}{2}}\end{aligned} \tag{5.18}$$

The DC offset components $A^2(t)/2$ is a time varying DC offset which depends on the signal envelope. This is an unwanted component which expands if higher order polynomial operations are included. For example, (5.19) shows a 4th order operation,

$$\begin{aligned}(A(t)\cos(\omega_0 t))^4 &= (A(t)\cos(\omega_0 t))^2 \cdot (A(t)\cos(\omega_0 t))^2 \\ &= A^4(t)\cdot\left(\frac{\cos(4\omega_0 t)}{8} + \boxed{\frac{\cos(2\omega_0 t)}{2}} + \boxed{\frac{3}{8}}\right)\end{aligned} \tag{5.19}$$

where the $3/8$ is the sum DC offset of two successive squaring functions and $\cos(2\omega_0 t)/2$ is the expanded cross-term from the squaring offset. The higher the order introduced, the larger the DC offset. This DC offset and the expanding effect with higher order operations will accumulate and change the power reference level of the polynomial model. The nonlinear behavior responding to this new signal

can deviate drastically from the model extraction reference, so requires a further normalization for the correction coefficients (b_0, b_2, b_4).

5.4.3 Input Envelope Linearization Simulation

To verify this IEL concept, a circuit level simulation bench is implemented in ADS as shown in Figure 5.4. Connection are made using node names so that wire connections can be omitted thus keeping the schematic simple. The simulation in this section uses a circuit envelope simulator.

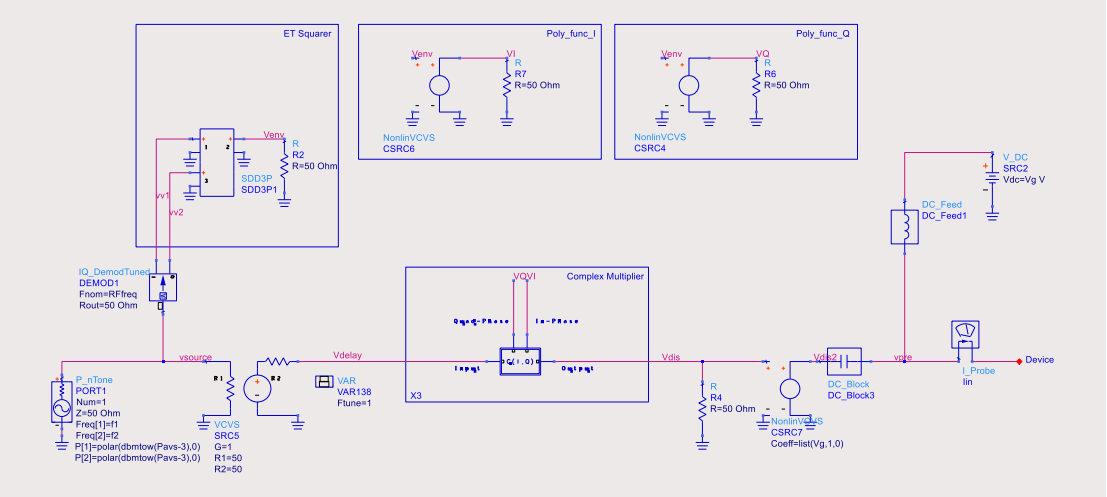


Figure 5.4 IEL Simulation Schematic

The polynomial function formulation is implemented with nonlinear voltage-controlled voltage source (VCVS) which are able to generate an output voltage as a polynomial function of the input voltage. The central complex multiplier block is implemented as shown in Figure 5.5. The In-Phase and Quad-Phase nodes are then loaded with polynomial functions as correction terms.

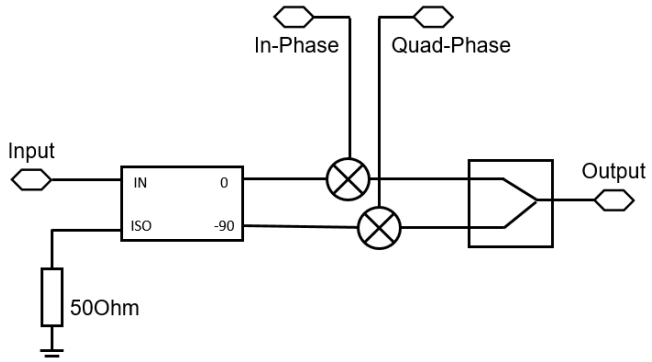


Figure 5.5 Schematic of complex multiplier

The simulation bench is capable of importing IEL coefficients here up to 7th order (b_0, b_2, b_4, b_6), however, theoretically, it can be configured to include higher orders, as required, at a cost of implementation complexity.

The IEL coefficients are calculated and normalized in MATLAB. The algorithm in the MATLAB scripts follows the analysis presented in the last section 5.4.1. To analyse this IEL technique experimentally, a number of simple modulated signal (two-tone) simulations are conducted firstly with the analytical model established in chapter 4 and then with a Cree 10W GaN device model. The first set of simulations are performed with current generator model only. The input signal and circuit biasing data are summarized in Table 5-1.

Table 5-1 Simulation setup for IEL verification (I_{gen} *only*)

Device Models	**Input Signal**	**Carrier Frequency**	**Bandwidth (tong spacing)**	**Power level**	**Gate Bias**	**Drain Bias**
I_{gen} only	Two-tone	2 GHz	20 MHz	15 dBm (6dB OBO)	- 2.9 V	28 V

There are then a set of simulations to check the IEL performance including:

Sim-1. Non-IEL reference case, poly-model extraction;

Sim-2. Upto 3rd order nonlinear IEL correction;

Sim-3. Upto 5th order nonlinear IEL correction;

Sim-4. Upto 7th order nonlinear IEL correction.

The Non-IEL reference case is a simulation for the original two-tone response of the system. This case is used to present a reference nonlinear behaviour without IEL, and extract the original nonlinear coefficients $(a_0, a_1, a_2, a_3, a_4, a_5, ...)$ as shown in equation (5.6). In this case, only zero order coefficients are enabled;

$$\begin{cases} \alpha_0 = 1 \\ \alpha_2 = 0 \\ \alpha_4 = 0 \\ \alpha_6 = 0 \end{cases} \cdots \begin{cases} \beta_0 = 1 \\ \beta_2 = 0 \\ \beta_4 = 0 \\ \beta_6 = 0 \end{cases} \tag{5.20}$$

Substituting (5.20) into (5.16), it can be found that the input signal is rotated due to the implementation of the symmetrical quadrature based complex multiplier for signal modification in *IQ* domain. This case is taken as reference in this section, as the observed symmetric rotation is happening before the signal passed through the device and can be considered as linear for *in-band* investigations. The original transfer characteristics is extracted from this reference case. It is fitted into a polynomial model and then used to calculate the required correction coefficients. The input and output signal spectrals for this reference case are shown in Figure 5.6.

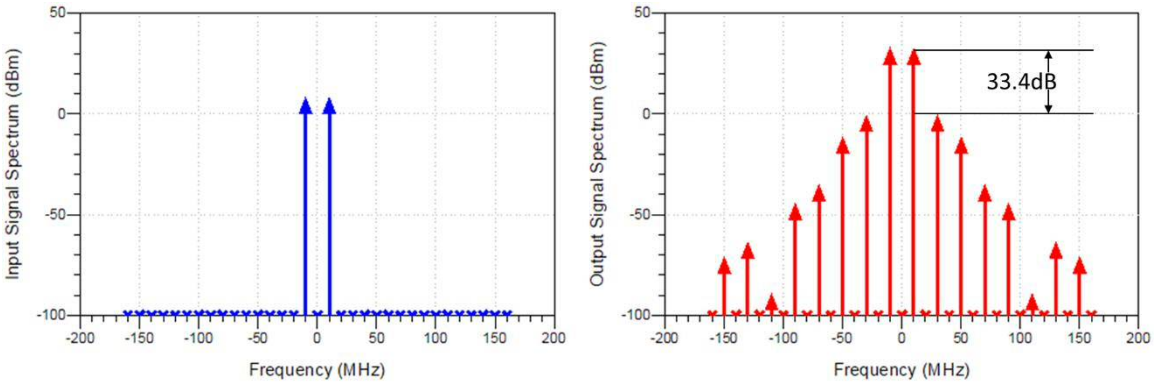


Figure 5.6 Input (gate) and output signal spectrals (I_{gen} only, Sim-1)

Then the 3rd order IEL correction is simulated with the results shown in Figure 5.7,

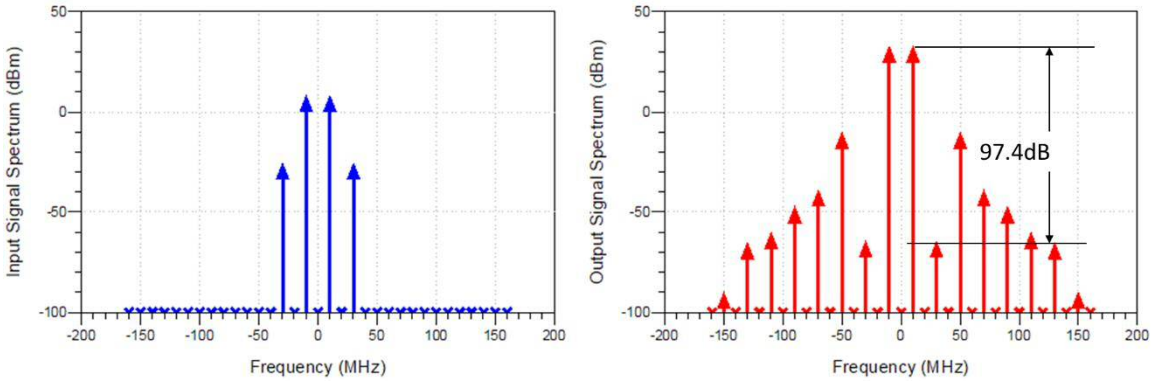


Figure 5.7 Input (gate) and output signal spectrals (I_{gen} only, Sim-2)

It can be seen in Figure 5.7 that the IM3 spectrals has been suppressed by 64dB, however, the spectrals for IM5 on both sides haven't been suppressed which is expected as only 3rd order correction is applied.

The results in Figure 5.8 and Figure 5.9 are obtained from 5th order and 7th order IEL corrections respectively. It can be observed in all these three simulations with different order of IEL corrections, the IM levels haven been suppressed. Especially, with a higher order, the sideband IM levels can be suppressed over a wider band.

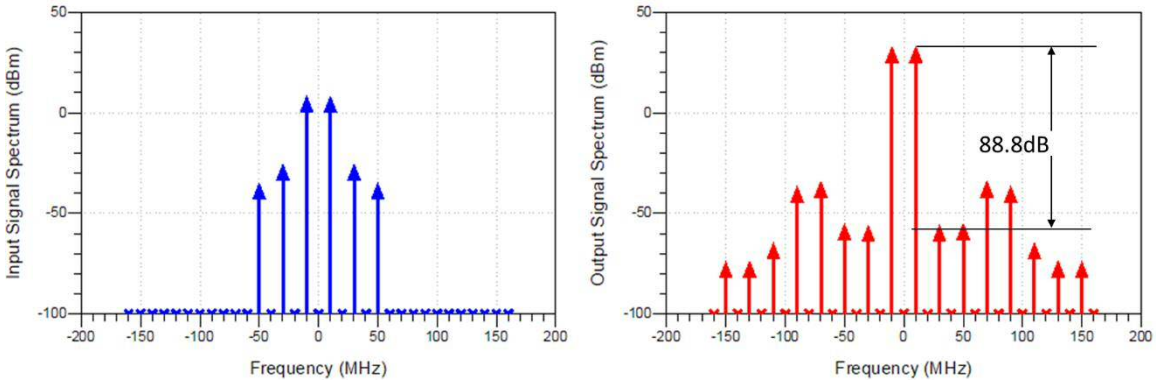


Figure 5.8 Input (gate) and output signal spectrals (I_{gen} only, Sim-3)

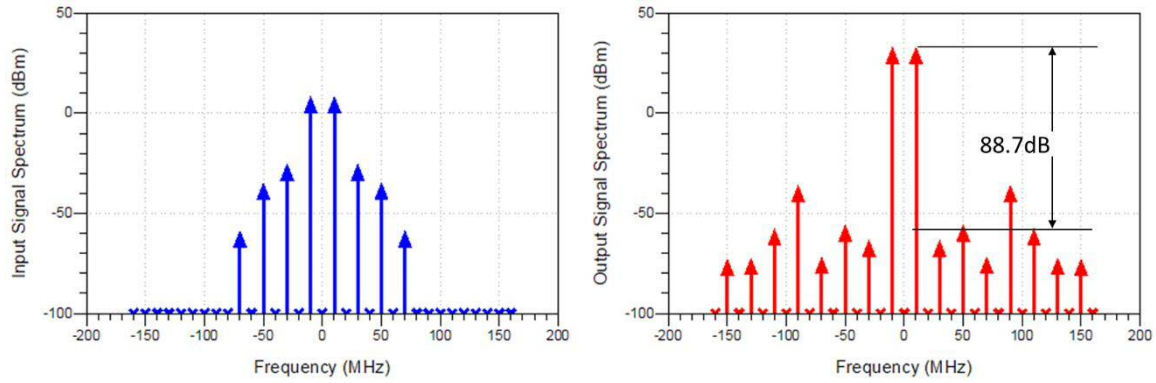


Figure 5.9 Input (gate) and output signal spectrals (I_{gen} only, Sim-4)

It is worth to mention that, in this investigation with I_{gen} model only, the sideband spectrals are symmetrical. This agrees with the observation in chapter 4 where IMD3 low and high contours are overlapping in this pure trans-conductive nonlinear system. Table 5-2 summarizes the original nonlinear coefficients and the correction coefficients in this case. The original nonlinear coefficients (a_1, a_3, a_5, a_7) are real values in this case, and the resulted IEL correction coefficients $(\alpha_{0,2,4,6}, \beta_{0,2,4,6})$ are symmetrical pairs due to the symmetrical complex multiplier.

Table 5-2 Original coefficients and correction coefficients for IEL (I_{gen} *only*)

Device Model		Current generator I_{gen} only			
Simulations		Sim-1	Sim-2	Sim-3	Sim-4
Original a_1		-9.8639e-1	-9.2332e-1	-8.9251e-1	-8.9538e-1
Original a_3		-	-8.2465e-2	-2.0439e-1	-1.8156e-1
Original a_5		-	-	9.6470e-2	5.1137e-2
Original a_7		-	-	-	2.5711e-2
IEL Coeff	**α_0**	1	-1.0764	-1.1114	-1.1146
	β_0	1	-1.0764	-1.1114	-1.1146
IEL Coeff	**α_2**	-	1.0384e-2	2.8108e-2	3.1302e-2
	β_2	-	1.0384e-2	2.8108e-2	3.1302e-2
IEL Coeff	**α_4**	-	-	-1.7745e-3	-2.5790e-3
	β_4	-	-	-1.7745e-3	-2.5790e-3
IEL Coeff	**α_6**	-	-	-	5.7672e-5
	β_6	-	-	-	5.7672e-5

Next, the full analytical model with nonlinear capacitance $C_{gs(nlinear)}$ and $C_{ds(nlinear)}$ is investigated using a similar set of simulations under the same two-tone excitation. The results are shown in the following figures, from Figure 5.10 to Figure 5.13. It can be observed from these results that, although IEL is able to successfully suppress the sideband spectrals, there are residues shown as asymmetry, for example, as shown in Figure 5.11. Table 5-3 sumarises the original nonlinear coefficients and the correction coefficients in this case, where, notably, the original nonlinear coefficients become complex values due to the presence of parasitic capacitance compared to those for I_{gen} model only.

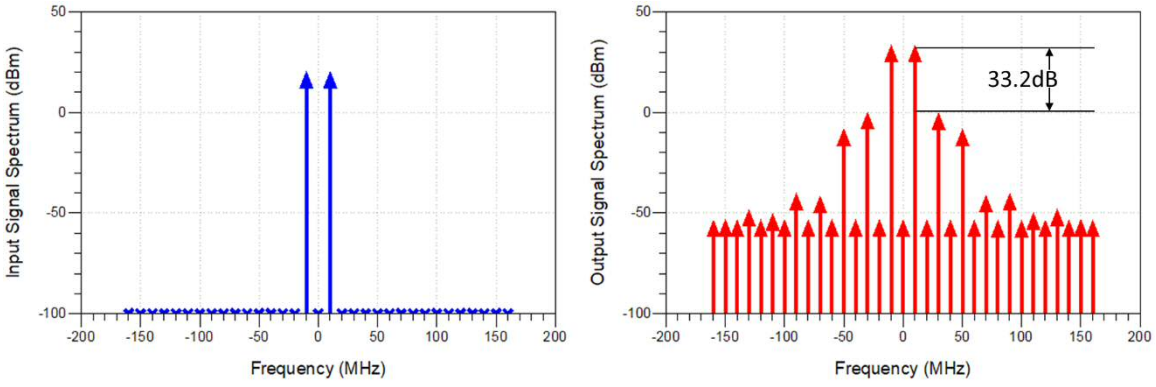


Figure 5.10 Input (gate) and output signal spectrals (full analytical model, Sim-1)

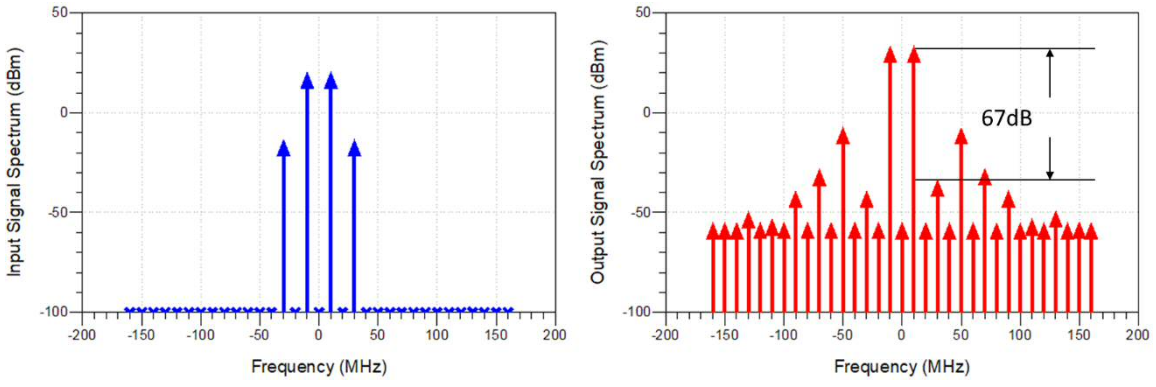


Figure 5.11 Input (gate) and output signal spectrals (full analytical model, Sim-2)

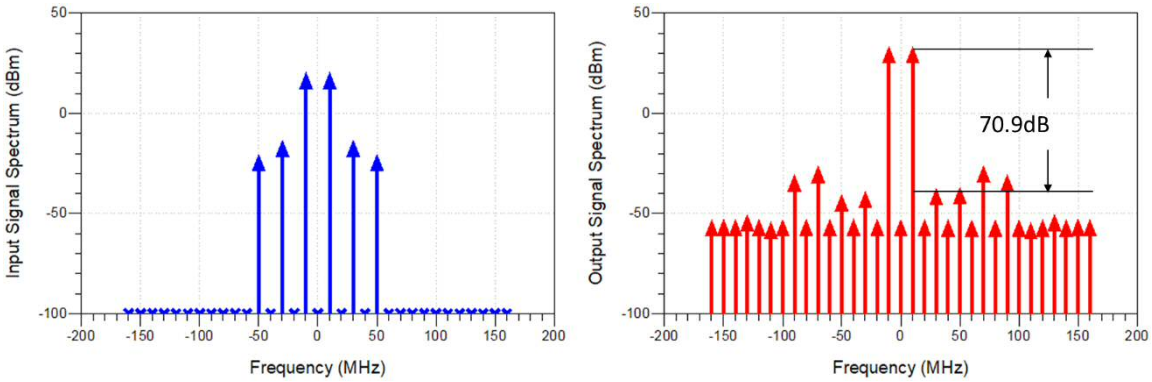


Figure 5.12 Input (gate) and output signal spectrals (full analytical model, Sim-3)

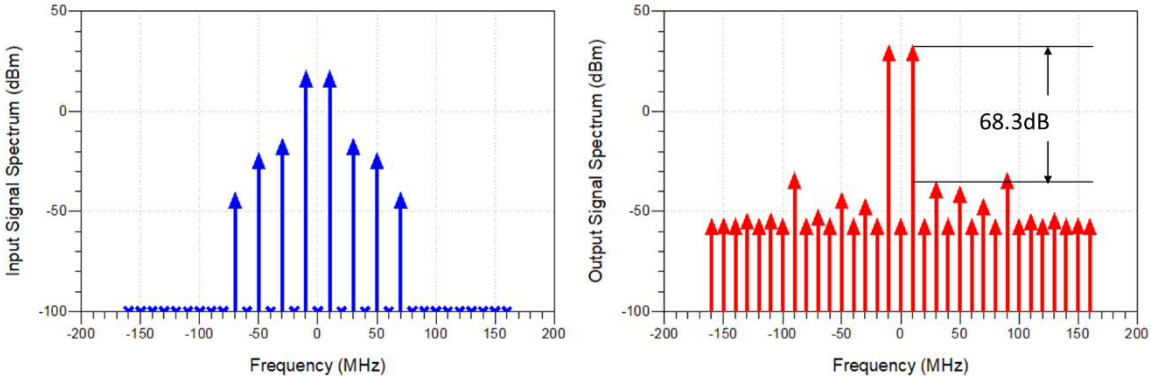


Figure 5.13 Input (gate) and output signal spectrals (full analytical model, Sim-4)

Table 5-3 Original coefficients and correction coefficients for IEL (*full analytical model*)

Device Model		Full analytical model with nonlinear capacitance			
Simulations		Sim-1	Sim-2	Sim-3	Sim-4
Original a_1	**Real**	-9.1480e-1	-8.5317e-1	-8.1342e-1	-8.1317e-1
	Imag	3.7388e-1	3.5942e-1	3.4434e-1	3.4570e-1
Original a_3	**Real**	-	-8.0593e-2	-2.3788e-1	-2.3982e-1
	Imag	-	1.8913e-2	7.8587e-2	6.7796e-2
Original a_5	**Real**	-	-	-1.2445e-1	1.2830e-1
	Imag	-	-	-4.7214e-2	-2.5793e-2
Original a_7	**Real**	-	-	-	-2.1836e-3
	Imag	-	-	-	-1.2149e-2
IEL Coeff	**α_0**	1	-1.4055	-1.4657	-1.4748
	β_0	1	-5.7413e-1	-5.9617e-1	-5.9655e-1
IEL Coeff	**α_2**	-	8.9821e-4	2.8172e-3	3.4060e-3
	β_2	-	2.1307e-4	9.1638e-4	9.4056e-4
IEL Coeff	**α_4**	-	-	-1.2114e-5	-2.1473e-5
	β_4	-	-	-4.4399e-6	-4.8241e-6
IEL Coeff	**α_6**	-	-	-	4.2321e-8
	β_6	-	-	-	1.7376e-9

Then a similar set of two-tone simulations are conducted for a Cree 10W GaN device with different drive levels as shown in Table 5-4.

Table 5-4 Simulation setup for IEL verification (Cree 10W model)

Device Models	Input Signal	Carrier Frequency	Bandwidth (tong spacing)	Power level	Gate Bias	Drain Bias
Cree 10W	Two-tone	2 GHz	20 MHz	15 dBm (6dB OBO)	- 2.9 V	28 V
Cree 10W	Two-tone	2 GHz	20 MHz	21 dBm (1dB Comp)	- 2.9 V	28 V

Firstly, the spectrum results for the 6dB power back-off case are shown from Figure 5.14 to Figure 5.17.

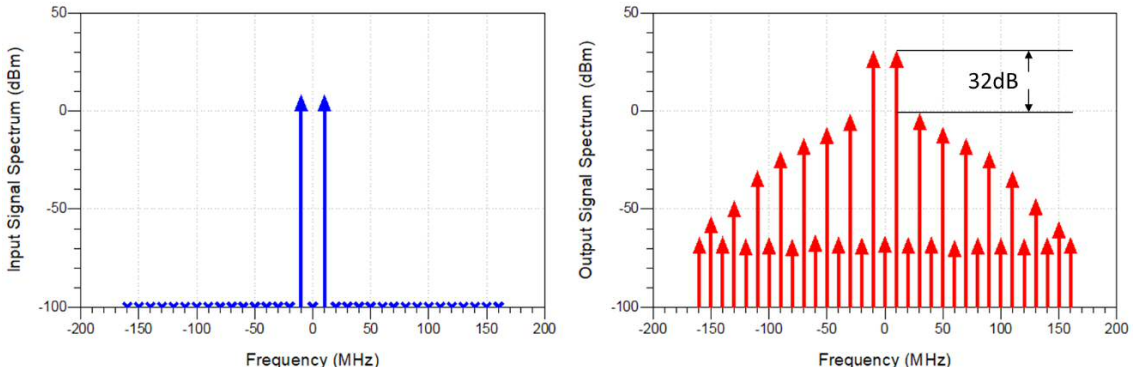


Figure 5.14 Input (gate) and output signal spectrals with 6dB back-off power (Cree 10W, Sim-1).

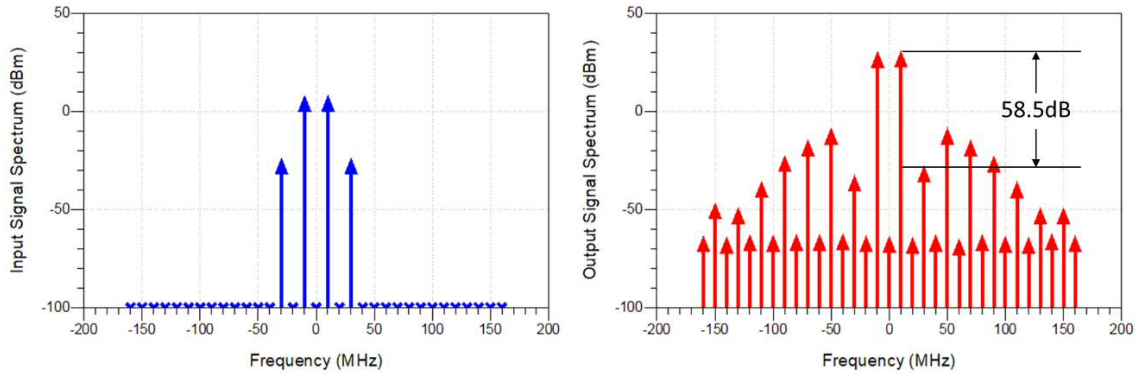


Figure 5.15 Input (gate) and output signal spectrals with 6dB back-off power (Cree 10W, Sim-2).

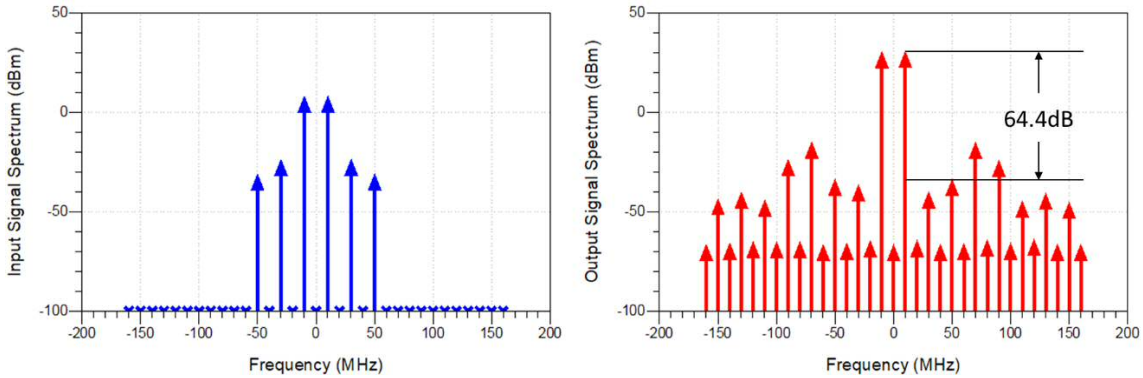


Figure 5.16 Input (gate) and output signal spectrals with 6dB back-off power (Cree 10W, Sim-3).

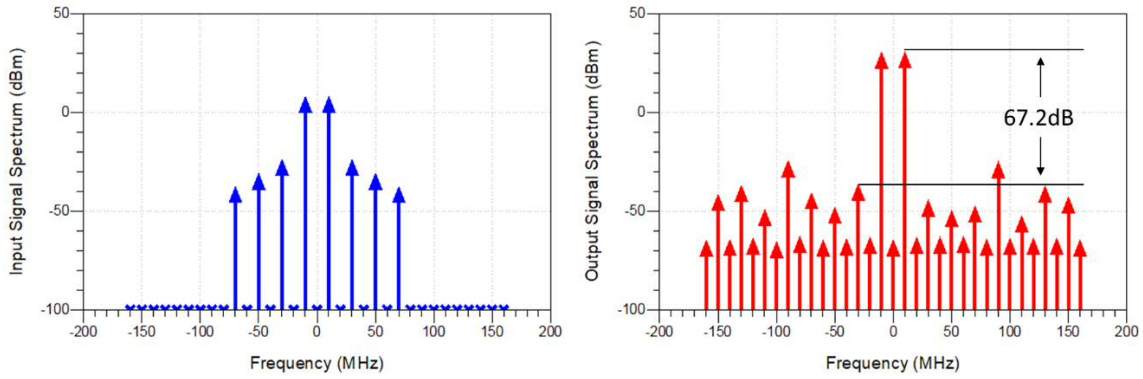


Figure 5.17 Input (gate) and output signal spectrals with 6dB back-off power (Cree 10W, Sim-4).

Then the results for the 1dB compression case are shown in the figures below.

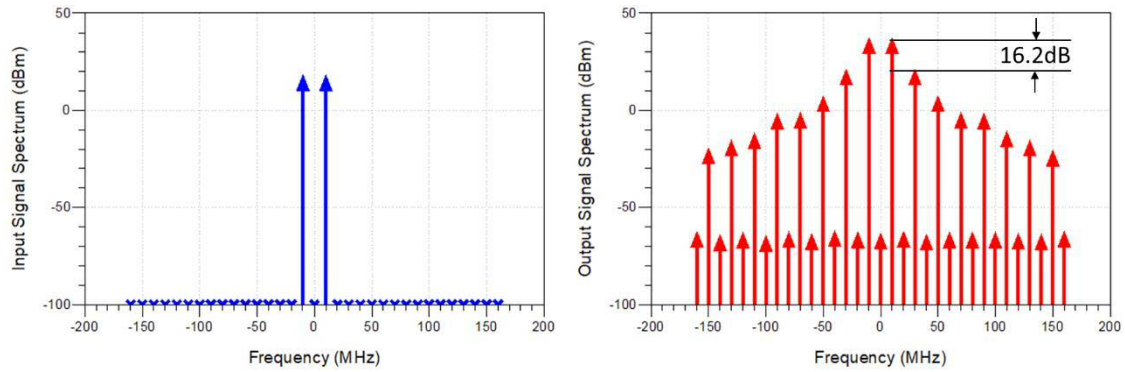


Figure 5.18 Input (gate) and output signal spectrals with 1dB compression power (Cree 10W, Sim-1).

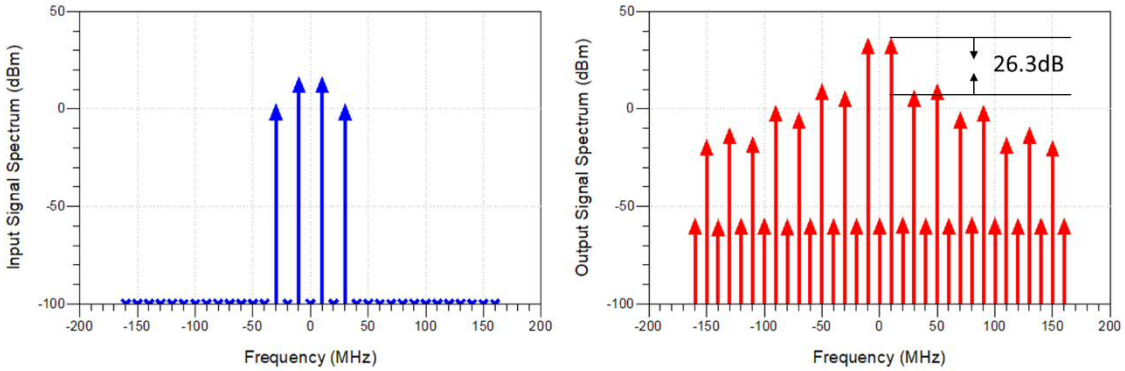


Figure 5.19 Input (gate) and output signal spectrals with 1dB compression power (Cree 10W, Sim-2).

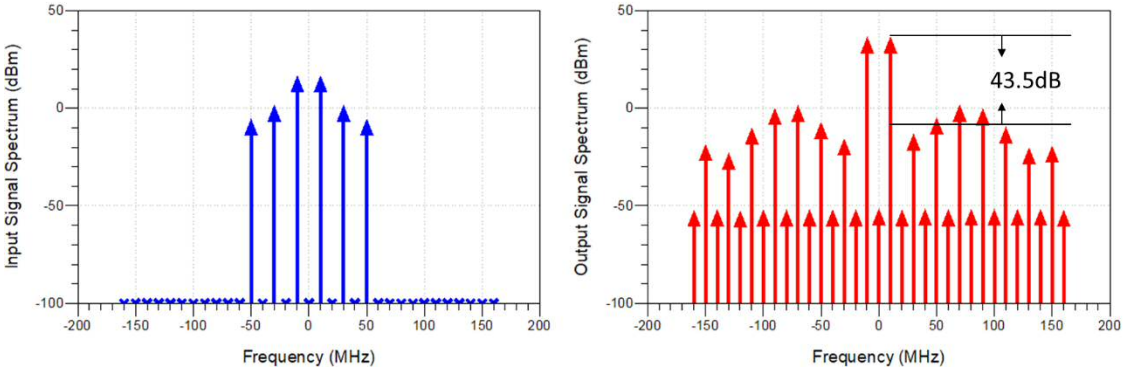


Figure 5.20 Input (gate) and output signal spectrals with 1dB compression power (Cree 10W, Sim-3).

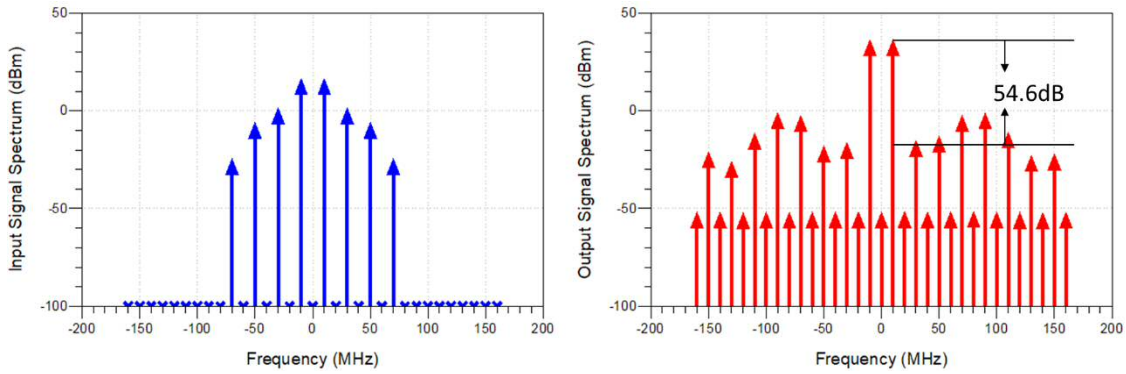


Figure 5.21 Input (gate) and output signal spectrals with 1dB compression power (Cree 10W, Sim-4).

To summarize, simulations on both analytical models and a commercially available model have been performed which are able to show that the required IEL signal has been computed and can be used to linearize the RFPA for both small signal weak nonlinear behavior in the power back-off region as well as in the mild compression region. In theory, this technique depends highly on the transfer characteristics of the RFPA, which determines the correction coefficients. Once the transfer characteristic is successfully extracted and modelled, a set of correction coefficients can be generated and used to predictably linearize the PA for the same quasi-stable operation.

To differentiate the IEL technique from the DPD solutions, here gives a short review to the current DPD system. Typically, in a transceiver with DPD capability, the digital baseband block is followed by a DPD processor. This requires a DAC component to convert the digital signal into analogue domain and then followed by an up-converter converting the signal to RF. As the linearization bandwidth is typically around 5 times larger than the signal bandwidth, which is used to

accommodate at least the IM5 components, this DAC component and up-converter are required to be high speed, large bandwidth resulting in high cost and power consumption. In the IEL technique, the system is featured as “RF-in RF-out”. The only digitized part is the implementation of the algorithm to calculate the coefficients, where may involve ADC and DAC components. However, they are not required to support the full linearization bandwidth, as they are just parameter optimizing units here. Based on this, IEL can potentially save the cost on high speed signal processing units for large bandwidth signal linearization. This makes IEL an attractive solution in tackling the emerging challenges in wideband PA linearization.

5.5 Output Envelope Linearization

Similar to IEL, the concept of output envelope linearization (OEL) is to introduce the compensating mixing signal through manipulation of the output biasing voltage V_{DS}. The cancellation for undesired distortion is achieved by optimizing the transfer function focusing the coefficients of V_{ds} terms instead of V_{gs} in IEL. As discussed in section 5.3, the RFPA is now a 2-D system considering the drain as another input signal. Similar linearization approaches have been reported in [49]–[56], which is categorized as envelope tracking "ET-like" envelope linearization in Chapter 2. Although this has been observed before, a clear mathematic description of this approach has, until now, been absent because of the complicated secondary mixing mechanism in this 2-D system.

A 2-D description is commonly used when it comes to the supply modulation architectures like ET. In this case, the nonlinear system can be described back to origin as

$$y(t) = f(V_{gs}) \cdot g(V_{ds}) \tag{5.21}$$

In IEL analysis, $f(V_{gs})$ is fitted into a polynomial model. If $g(V_{ds})$ can be fitted into a polynomial description, (5.21) can be written as,

$$\begin{aligned} y(t) &= f(V_{gs}) \cdot g(V_{ds}) \\ &= (a_0 + a_1 V_{gs} + a_2 V_{gs}^{\ 2} + a_3 V_{gs}^{\ 3} + a_4 V_{gs}^{\ 4} + a_5 V_{gs}^{\ 5}) \\ &\cdot (k_0 + k_1 V_{ds} + k_2 V_{ds}^{\ 2} + k_3 V_{ds}^{\ 3} + k_4 V_{ds}^{\ 4} + k_5 V_{ds}^{\ 5}) \end{aligned} \tag{5.22}$$

Instead of shaping V_{gs} into an even order polynomial to introduce new mixing terms in IEL, we formulate V_{ds} in OEL. Relatively easy manipulation on the drain voltage V_{ds} is possible when using ET architectures as the required baseband injection route is essentially already there, and the correction function can be achieved by the use of the proper shaping function, relating the ET tracking signal to the magnitude of the input modulated signal. Similar to IEL envelope analysis, the even order polynomial based on $B = |A(t)|^2$ is used to describe the required signal envelope. The modulated supply voltage is then formulated as

$$V_{DS} = V_0 + b_2 \cdot B + b_4 \cdot B^2 \tag{5.23}$$

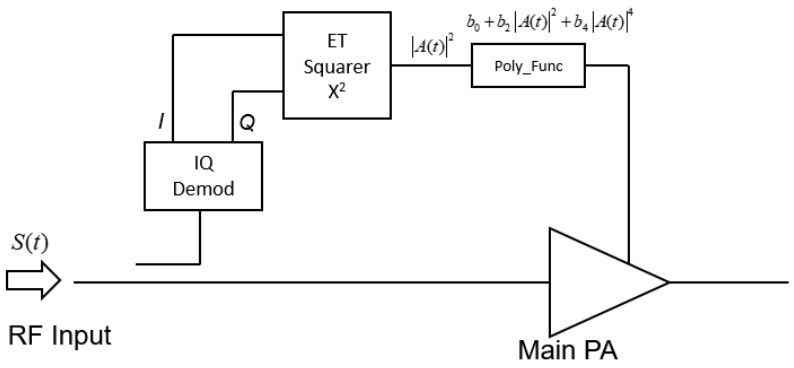


Figure 5.22 Application Schematic of Output Envelope Linearization (OEL)

where V_0 is the DC offset of the bias voltage. (5.23) can be then fed into (5.22) to generate the similar cancellation as in IEL. However, this results in an exploded mathematic expression which is not computable with simple analytical analysis neither convenient to use in practical engineering. For the reasons of simplicity, the desired correction coefficients are found by parameter sweeping and optimization. The OEL application schematic is shown in Figure 5.22. This is then implemented in ADS as shown in Figure 5.23. The formulated output biasing voltage is then fed into the drain of the device.

The correction order of OEL used in this bench is limited up-to 5th order correction. Note that in a practical implementation, a higher order correction translates to a drastically increased bandwidth requirement for the supply modulator, which may well be impractical, so it is reasonable to check the linearization performance with a limited correction order.

A set of simulations at 1dB compression power are investigated for the proposed OEL technique using a Cree 10W GaN device model (CGH60015D). The signal and circuit biasing conditions are shown in Table 5-2. Correction coefficients are

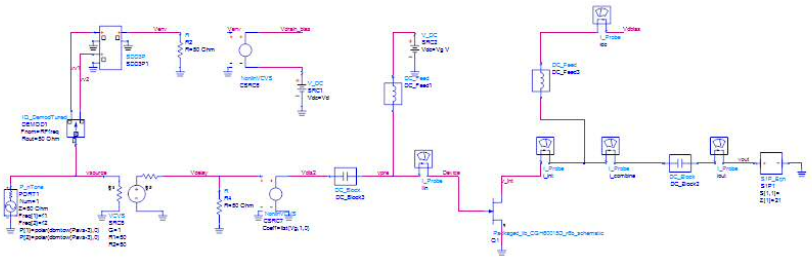


Figure 5.23 OEL Simulation Schematic

recorded in Table 5-3 for up-to 5th order correction. These coefficients are found by an iterative parameter optimization. It is worth to mention that the correction coefficients (b_0, b_2, b_4) are now only real values as the drain bias with time-varying envelope while the original nonlinear coefficients (a_1, a_3, a_5, a_7) are complex. This limits OEL can only deal with AM/AM distortion but no AM/PM correction as no phase manipulation is possible through the ET route.

Table 5-5 Simulation setup for OEL verification

Device Model	Input Signal	Carrier Frequency	Bandwidth	Power level	Gate Bias	Drain Bias
Cree 10W	Two-tone	2 GHz	20 MHz	21 dBm	- 2.9 V	28 V

Table 5-6 Three cases for OEL verification

Simulations	Correction order	b_0	b_2	b_4
Sim-1	Non	0	0	0
Sim-2	3rd	-14	0.7	0
Sim-3	5th	-15.2	0.97	0.0009

The results are shown in Figure 5.24, Figure 5.25 and Figure 5.26, respectively for 3 cases.

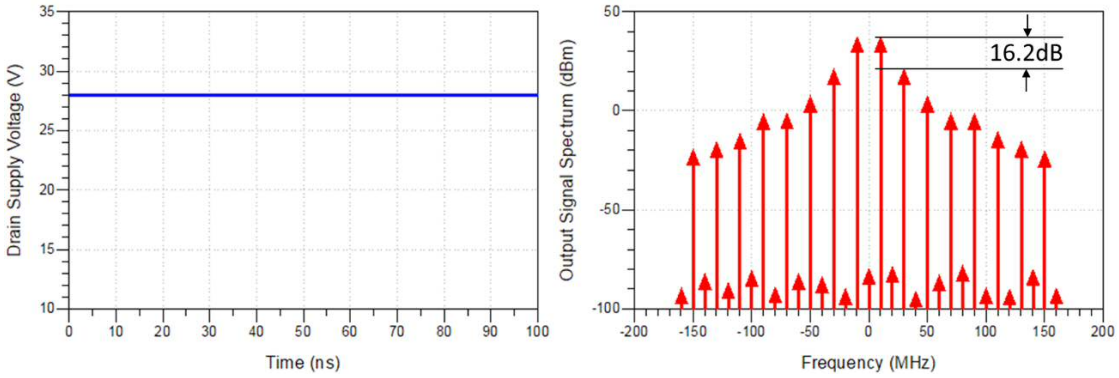


Figure 5.24 Drain supply voltage and output signal spectrals for Sim-1 (reference case).

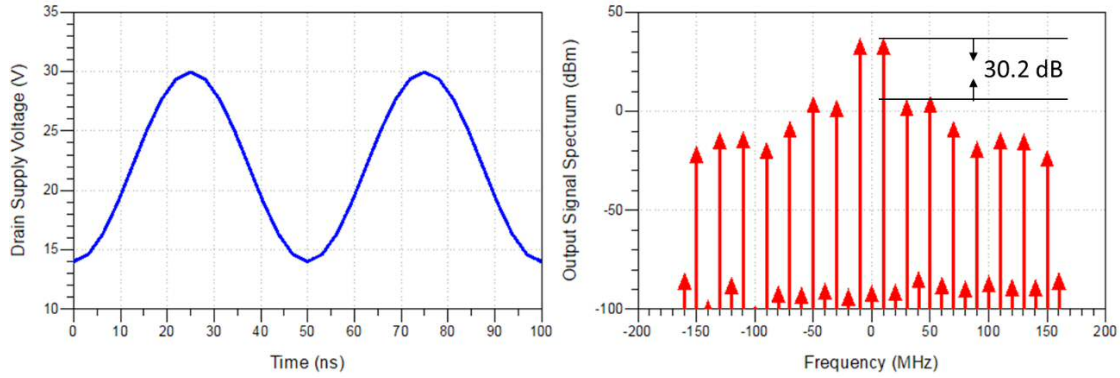


Figure 5.25 Drain supply voltage and output signal spectrals for Sim-2 (3rd order correction).

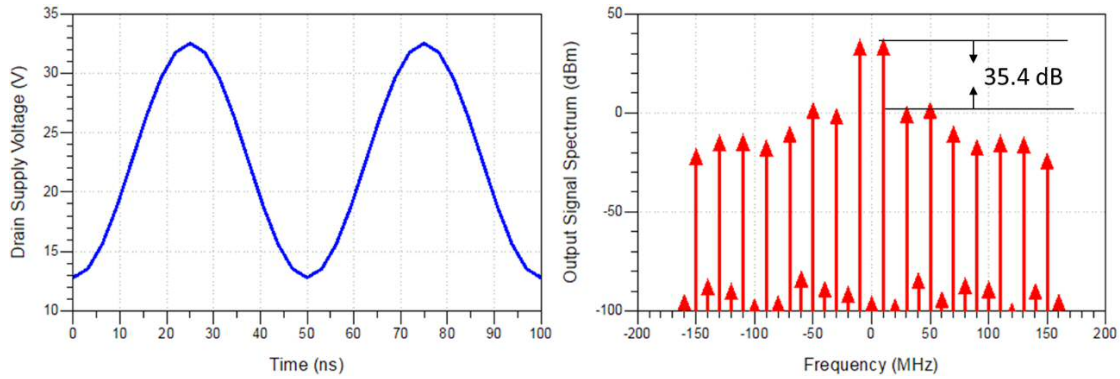


Figure 5.26 Drain supply voltage and output signal spectrals for Sim-3 (5th order correction).

The results have verified that by manipulating the biasing voltage using supply modulation scheme, the linearity of the RFPA can be improved at mild compression level. In this experiment, OEL improves the IMD3 level by 19dB using 5th order correction compared to the reference case. Although higher order correction is promising to get further linearization, it can lead to a higher complexity and cost for the high-speed supply modulator.

OEL technique demonstrated above can potentially simplify the linearization scheme with an ET architecture. Although the performance is limited as no AM/PM corrections, the simple structure can find various applications, such as in the low-power handsets.

5.6 Conclusion

In this Chapter, a pair of system level linearization techniques are introduced. Derived from envelop domain analysis, it has been verified that by manipulating both V_{gs} at gate and V_{ds} at drain, it is possible to introduce artificial mixing terms to cancel the intrinsic nonlinear distortion in the RFPA.

There is a need to mention that, both IEL and OEL techniques are based on an envelope analysis and working completely in envelope domain. It can be observed in equation (5.14) and (5.15) that to achieve an IMD cancellation with an external injection which is formulated with even order envelope terms, the nulling condition has no dependency on the signal envelope. This means this linearization

technique can be widely used to linearize a broad range of signals with varied signal envelope. The nonlinear coefficients only depend on the quasi-stable operation condition defined by the bias, drive level and relevant termination impedance.

Another advantage of the envelope linearization technique is the capability to linearize the *in-band* components. For example, deriving from equation (5.13), the envelope linearization aims to nulling the 3rd order envelope rather than nulling only the spectral at IM3 frequency, as shown in Figure 5.27. In this way, it also suppresses the spectral regrowth within the two-tone bandwidth, which has a desirable feature for modulated systems that improves EVM at the same time.

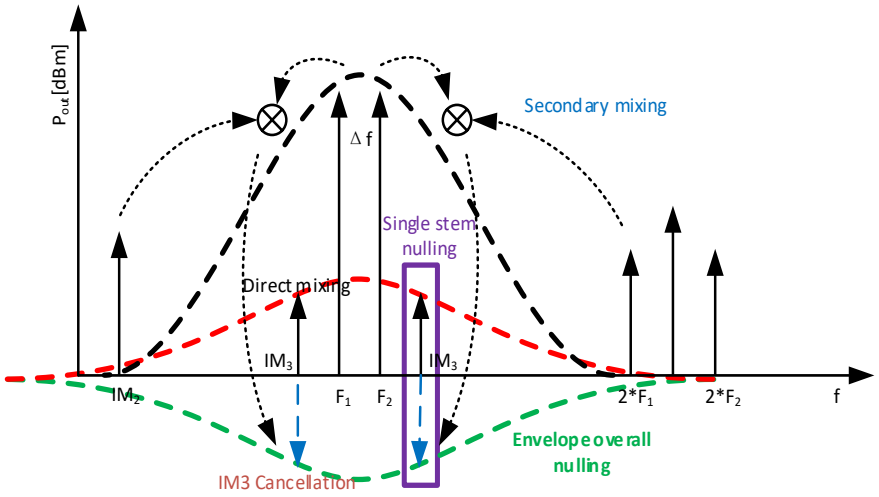


Figure 5.27 Envelope linearization compared to single stem IM3 cancellation

The proposed IEL technique is compared with DPD architectures, where the advantage has been discussed regarding to the challenges in emerging 5G wideband applications. The OEL technique proves to be a linearization with simpler architecture. However, due to the topology limitation, AM/PM correction is not possible with this scheme.

Chapter 6 Conclusion and Future Works

6.1 Conclusion

Improving the linearity of RFPA to meet the emerging communication standards requires both the optimization at circuit design level and the system level linearization. Although DPD is now considered near a universal method of choice for linearization in BTS, it is very challenge to meet the aggressive specifications, such as the fact that the increasing wide bandwidths high speed signal processing units (ADC and DAC). The current DPD system can be too large and power hungry in the various applications of emerging 5G communication systems, such as the linearization of low-power PAs in multi-antenna arrays or picocell applications for instance. So, both the new standards and markets are calling for novel solutions for signal linearization. Thus, this thesis is focusing on exploring the key distortion mechanisms in the RFPA and demonstrating novel optimization methodologies and linearization techniques.

With the analytical model established in this thesis, the individual distortion contributions from the key nonlinear sources / elements have been discovered. The circuit level optimization for linearity, such as bias optimization, impedance termination optimization are discussed using a powerful load-pull bench in

simulation environment, which has the control over baseband, fundamental and harmonics simultaneously. Although, to the author's knowledge, this kind of load-pull system has not been available in hardware, this work has proven the potential use of this system and identified the key measurements to optimize the circuit linearity and *linearizability*. Some of the observations in this thesis have never been revealed and analysed before, however, they prove to be very closely related to the challenges in PA design.

A novel envelope linearization concepts have been presented and discussed in this thesis. This linearization implemented either at the input node (IEL) or output node (OEL) has the potential to linearize the RFPA through a complete envelope operation, which suppresses the distortion shown at IM products as well as *in-band* spectral regrowth. As a result, it improves the ACPR as well as the EVM of the nonlinear RFPA. Both IEL and OEL have their own application significance in terms of tackling the linearization challenges in 5G systems.

With the analysis on both circuit level and system level, this thesis has provided a bottom-to-top level linearization strategy on PA, which involves the consideration on linearity as well as linearizability.

6.2 Future works

This research has been completed and demonstrated mostly in a simulation environment due to the hardware limitation at the moment. Future extensive works should be considered;

1. Extend the circuit level analysis with different device technology.

   As it has been discussed before in chapter 4, some of the observation may highly vary with different device technology which presents different characteristics in terms of the current generator transconductance and the nonlinear capacitance. The analysis methodology presented in this work should still be viable for those technologies. So, it is important to verify those observations using different devices and then address the limitation of them. This can be significant in technology optimizations.

2. Linearity optimized input and output network for class-AB power amplifier.

   The work in this thesis starts from a theoretical and measurement perspective. The demonstrated linearity optimization methodology such as baseband load-pull and 2nd harmonic load-pull simultaneously with a fundamental impedance control is challenge in current measurement system. However, it is worth to continue the verification with a hardware fabrication which is guided directly from the circuit simulation. Further work is required to verify the optimum baseband and even harmonic terminations for both input matching and output matching networks regarding to both linearity and linearizability.

3. Extend the IEL with memory tabs to enhance the performance

IEL has shown its capability to linearize simple modulated signals, such as two-tone, in this work. In theory, it is able to linearize any signal as the nature of envelope domain operation. However, the formulation to derive the correction coefficients requires a coefficient extraction process at a quasi-stable operation. In addition to this, the formulation used to describe the transfer characteristic is based on memoryless polynomial. This is the weakness of the IEL technique presented in this work, in tackling the complex modulated systems which involves various memory effects, such as thermal, RF and electrical memory. It is worth to extend this work with memory polynomials and further investigation with complex modulated signal. A proposed implementation of memory tabs in IEL is shown in Figure 6.1. The polynomial function can be extended into several patches to include the memory tabs.

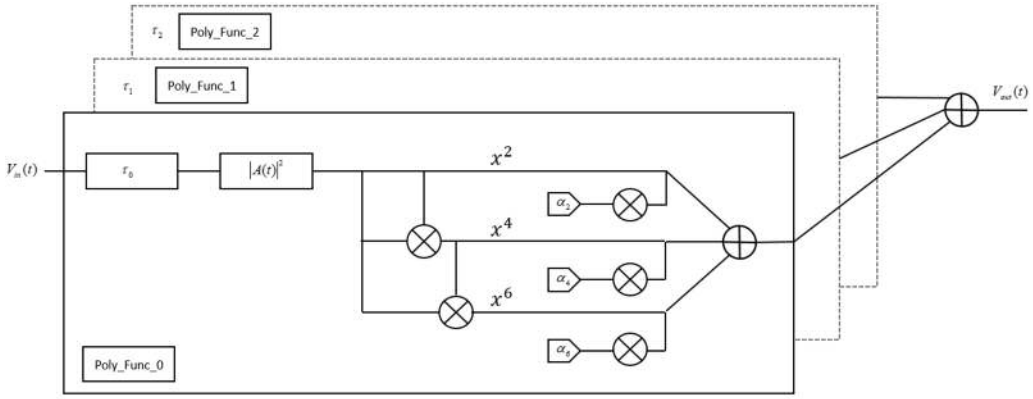


Figure 6.1 IEL with memory tabs

4. Using shaping function in ET bench to demonstrate OEL

As discussed in chapter 5, the correction coefficients and relevant injection process can be implemented with a shaping function in ET system. It is worth to

engineer a shaping function which linearizes the RFPA using OEL principle and maintains relatively high efficiency at the same time. Although a trade-off between efficiency and linearity may still exist in the choice of the shaping function, the potential linearization benefits is able to contribute to system simplification in compact applications, for example, the ETPA in handsets without DPD.

5. Envelope linearization for varied modulated signals

   As mentioned in chapter 5, one advantage of the envelope linearization is the fact that this technique is not limited by the signal type. Further verifications should be done regarding to different signal modulations. This also initiates a discussion on the performance in terms of improving the EVM.

Bibliography

[1] I. M. T. Vision, “Framework and overall objectives of the future development of IMT for 2020 and beyond.”

[2] Ofcom, “Enabling 5G in the UK.”

[3] “5G vs 4G - Speed, Latency and Coverage Compared.” [Online]. Available: https://5g.co.uk/guides/4g-versus-5g-what-will-the-next-generation-bring/. [Accessed: 24-Jun-2019].

[4] Anritsu Ltd., “Ten 5G Challenges for Engineers to Overcome,” 2016.

[5] Qorvo Inc., “Defining What’s Possible - The Path to 5G.”

[6] Qorvo Inc., “GaN – A Critical Technology for 5G.”

[7] R. Ma, K. H. Teo, S. Shinjo, K. Yamanaka, and P. M. Asbeck, “A GaN PA for 4G LTE-Advanced and 5G: Meeting the Telecommunication Needs of Various Vertical Sectors Including Automobiles, Robotics, Health Care, Factory Automation, Agriculture, Education, and More,” *IEEE Microw. Mag.*, vol. 18, no. 7, pp. 77–85, 2017.

[8] S. C. Cripps, *RF power amplifiers for wireless communications*, 2nd ed. Boston: Boston : Artech House, 2006.

[9] D. K. Su and W. J. McFarland, “An IC for linearizing RF power amplifiers using envelope elimination and restoration,” *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 2252–2258, 1998.

[10] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, “Energy-Efficient Base-

Stations Sleep-Mode Techniques in Green Cellular Networks: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 17, no. 2, pp. 803–826, 2015.

[11] M. H. Alsharif, J. Kim, J. H. Kim, M. H. Alsharif, J. Kim, and J. H. Kim, "Green and Sustainable Cellular Base Stations: An Overview and Future Research Directions," *Energies*, vol. 10, no. 5, p. 587, Apr. 2017.

[12] C. Han *et al.*, "Green radio: radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, Jun. 2011.

[13] D. Y. C. Lie, J. C. Mayeda, and J. Lopez, "Highly efficient 5G linear power amplifiers (PA) design challenges," in *2017 International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*, 2017, pp. 1–3.

[14] D. Kimball, T. Nakatani, J. Yan, and K. Mukai, "Analog & digital envelope tracking power amplifier reduced bandwidth techniques for 5G Nr," in *Asia-Pacific Microwave Conference Proceedings, APMC*, 2019, vol. 2018–November, pp. 55–57.

[15] "Universal Mobile Telecommunications System Base Stations Radio Transmission and Reception." ETSI Standard TS 125 104 V11.5.0, 2013.

[16] "Air Interface for Broadband Wireless Access Systems." IEEE Standard 802.16, 2012.

[17] "E-UTRA Base Station Radio Transmission and Reception." 3GPP Standard TS 36.104 V11.4.0, 2013.

[18] "E-UTRA Radio Frequency System Scenarios." 3GPP Standard TR 36.942 V11.0.0, 2012.

[19] L. Guan and A. Zhu, “Green Communications: Digital Predistortion for Wideband RF Power Amplifiers,” *IEEE Microw. Mag.*, vol. 15, no. 7, pp. 84–99, 2014.

[20] J. Kim, Y. Y. Woo, J. Moon, and B. Kim, “A New Wideband Adaptive Digital Predistortion Technique Employing Feedback Linearization,” *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 2, pp. 385–392, 2008.

[21] A. Zhu, P. J. Draxler, J. J. Yan, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, “Open-Loop Digital Predistorter for RF Power Amplifiers Using Dynamic Deviation Reduction-Based Volterra Series,” *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 7, pp. 1524–1534, 2008.

[22] J. C. Cahuana, P. Landin, D. Gustafsson, C. Fager, and T. Eriksson, “Linearization of dual-input Doherty power amplifiers,” in *International Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits, INMMiC 2014*, 2014.

[23] R. N. Braithwaite and S. Carichner, “An improved Doherty amplifier using cascaded digital predistortion and digital gate voltage enhancement,” in *2009 IEEE MTT-S International Microwave Symposium Digest*, 2009, pp. 1073–1076.

[24] NXP Semiconductors, “BLF7G20LS-200 Doherty 1.805 to 1.88 GHz RF Power Amplifier.” 2010.

[25] A. Kwan, F. M. Ghannouchi, O. Hammi, M. Helaoui, and M. R. Smith, “Look-up table-based digital predistorter implementation for field programmable gate arrays using long-term evolution signals with 60 mhz

bandwidth," *IET Sci. Meas. Technol.*, vol. 6, no. 3, pp. 181–188, 2012.

[26] L. Guan and A. Zhu, "Optimized Low-Complexity Implementation of Least Squares Based Model Extraction for Digital Predistortion of RF Power Amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 3, pp. 594–603, 2012.

[27] S. Boumaiza and H. Golestaneh, "Joint circuit-level and digital predistortion strategies for enhancing the linearity-efficiency tradeoff of Doherty power amplifiers," in *IEEE MTT-S International Microwave Symposium Digest*, 2016, vol. 2016–August.

[28] P. Singerl, T. Magesacher, and M. Mataln, "Predistortion- and development-platform for multi-input RF power amplifiers," in *EuMIC 2016 - 11th European Microwave Integrated Circuits Conference*, 2016, pp. 29–32.

[29] S. Bensmida, O. Hammi, A. Kwan, M. S. Sharawi, K. A. Morris, and F. M. Ghannouchi, "Extending the Characterization Bandwidth of Dynamic Nonlinear Transmitters With Application to Digital Predistortion," *IEEE Trans. Microw. Theory Tech.*, vol. 64, no. 8, pp. 2640–2651, Aug. 2016.

[30] S. Bensmida *et al.*, "Nonlinear envelope tracking for efficiency optimization of power amplifiers," in *2012 7th European Microwave Integrated Circuit Conference*, 2012, pp. 480–483.

[31] Anding Zhu *et al.*, "Digital Predistortion for Envelope-Tracking Power Amplifiers Using Decomposed Piecewise Volterra Series," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 10, pp. 2237–2247, Oct. 2008.

[32] A. Katz, J. Wood, and D. Chokola, “The Evolution of PA Linearization: From Classic Feedforward and Feedback Through Analog and Digital Predistortion,” *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 32–40, 2016.

[33] H. S. Black, “Inventing the negative feedback amplifier: Six years of persistent search helped the author conceive the idea ‘in a flash’ aboard the old Lackawanna Ferry,” *IEEE Spectr.*, vol. 14, no. 12, pp. 55–60, 1977.

[34] J. Vankka, *Digital Synthesizers and Transmitters for Software Radio*. Springer, 2005.

[35] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers*. Artech House, 2014.

[36] R. N. Braithwaite and A. Khanifar, “High efficiency feedforward power amplifier using a nonlinear error amplifier and offset alignment control,” in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, 2013, pp. 1–4.

[37] R. N. Braithwaite, “Positive Feedback Pilot System for Second Loop Control in a Feedforward Power Amplifier,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 55, no. 10, pp. 3293–3305, 2008.

[38] H. Choi, Y. Jeong, C. D. Kim, and J. S. Kenney, “Efficiency Enhancement of Feedforward Amplifiers by Employing a Negative Group-Delay Circuit,” *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 5, pp. 1116–1125, 2010.

[39] R. N. Braithwaite and M. J. Hunton, “A positive feedback pilot system for a wide bandwidth feedforward amplifier,” in *The 40th European Microwave Conference*, 2010, pp. 1778–1781.

[40] R. N. Braithwaite, “A Comparison for a Doherty power amplifier linearized using digital predistortion and feedforward compensation,” in *2015 IEEE MTT-S International Microwave Symposium*, 2015, pp. 1–4.

[41] H. S. Black, “Stabilized feedback amplifiers,” *Bell Syst. Tech. J.*, vol. 13, no. 1, pp. 1–18, 1934.

[42] C. Nader, P. N. Landin, W. Van Moer, N. Bjorsell, and P. Handel, “Performance Evaluation of Peak-to-Average Power Ratio Reduction and Digital Pre-Distortion for OFDM Based Systems,” *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 12, pp. 3504–3511, 2011.

[43] C. Nader, P. N. Landin, W. Van Moer, N. Bjorsell, P. Handel, and D. Ronnow, “Peak-Power Controlling Technique for Enhancing Digital Pre-Distortion of RF Power Amplifiers,” *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 11, pp. 3571–3581, 2012.

[44] J. Xia, A. Islam, H. Huang, and S. Boumaiza, “Envelope Memory Polynomial Reformulation for Hardware Optimization of Analog-RF Predistortion,” *IEEE Microw. Wirel. Components Lett.*, vol. 25, no. 6, pp. 415–417, 2015.

[45] K. Gumber and M. Rawat, “A Modified Hybrid RF Predistorter Linearizer for Ultra Wideband 5G Systems,” *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 7, no. 4, pp. 547–557, 2017.

[46] R. N. Braithwaite, “Compensation of delay within an analog predistortion module used to linearize a dual-band RF transmitter,” in *2018 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless*

Applications (PAWR), 2018, pp. 51–53.

[47] F. Roger, “A 200mW 100MHz-to-4GHz 11th-order complex analog memory polynomial predistorter for wireless infrastructure RF amplifiers,” in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 94–95.

[48] H. Huang, A. Islam, J. Xia, P. Levine, and S. Boumaiza, “Linear filter assisted envelope memory polynomial for analog/radio frequency predistortion of power amplifiers,” in *2015 IEEE MTT-S International Microwave Symposium*, 2015, pp. 1–3.

[49] Z. Yusoff, J. Lees, J. Benedikt, P. J. Tasker, and S. C. Cripps, “Linearity improvement in RF power amplifier system using integrated Auxiliary Envelope Tracking system,” in *2011 IEEE MTT-S International Microwave Symposium*, 2011, pp. 1–4.

[50] F. L. Ogboi *et al.*, “A LSNA configured to perform baseband engineering for device linearity investigations under modulated excitations,” in *Microwave Conference (EuMC), 2013 European*, 2013, pp. 684–687.

[51] F. L. Ogboi *et al.*, “High bandwidth investigations of a baseband linearization approach formulated in the envelope domain under modulated stimulus,” in *European Microwave Conference (EuMC), 2014 44th*, 2014, pp. 1305–1308.

[52] F. L. Ogboi *et al.*, “Investigation of Various Envelope Complexity Linearity under Modulated Stimulus Using a New Envelope Formulation Approach,” in *Compound Semiconductor Integrated Circuit Symposium (CSICs), 2014*

IEEE, 2014, pp. 1–4.

[53] F. L. Ogboi *et al.*, “Sensitivity of AM/AM linearizer to AM/PM distortion in devices,” in *ARFTG Microwave Measurement Conference (ARFTG), 2014 83rd*, 2014, pp. 1–4.

[54] Yu Zhu *et al.*, “Analysis, simulation, and measurement of envelope tracking linearization,” in *2016 Asia-Pacific Microwave Conference (APMC)*, 2016, pp. 1–4.

[55] Y. Zhu *et al.*, “Novel shaping function for envelope tracking linearization,” in *2017 IEEE Asia Pacific Microwave Conference (APMC)*, 2017, pp. 402–405.

[56] Y. Zhu *et al.*, “Envelope Tracking Linearizability of Power Amplifier,” in *2018 Asia-Pacific Microwave Conference (APMC)*, 2018, pp. 998–1000.

[57] Taijun Liu, S. Boumaiza, and F. M. Ghannouchi, “Augmented hammerstein predistorter for linearization of broad-band wireless transmitters,” *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1340–1349, Jun. 2006.

[58] J. K. Cavers, “Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements,” *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, pp. 374–382, 1990.

[59] M. Rawat, K. Rawat, and F. M. Ghannouchi, “Adaptive Digital Predistortion of Wireless Power Amplifiers/Transmitters Using Dynamic Real-Valued Focused Time-Delay Line Neural Networks,” *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 1, pp. 95–104, Jan. 2010.

[60] T. Liu, S. Boumaiza, and F. M. Ghannouchi, “Dynamic Behavioral Modeling

of 3G Power Amplifiers Using Real-Valued Time-Delay Neural Networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.

[61] F. Mkadem and S. Boumaiza, "Extended Hammerstein Behavioral Model Using Artificial Neural Networks," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 4, pp. 745–751, Apr. 2009.

[62] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Malabar, FL: Reprint Krieger, 2006.

[63] L. Ding *et al.*, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[64] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic Deviation Reduction-Based Volterra Behavioral Modeling of RF Power Amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[65] P. L. Gilabert, G. Montoro, and E. Bertran, "On the Wiener and Hammerstein Models for Power Amplifier Predistortion," in *2005 Asia-Pacific Microwave Conference Proceedings*, vol. 2, pp. 1–4.

[66] Taijun Liu, S. Boumaiza, and F. M. Ghannouchi, "Deembedding static nonlinearities and accurately identifying and modeling memory effects in wide-band RF transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 11, pp. 3578–3587, Nov. 2005.

[67] Hyun Woo Kang, Yong Soo Cho, and Dae Hee Youn, "Adaptive precompensation of Wiener systems," *IEEE Trans. Signal Process.*, vol.

46, no. 10, pp. 2825–2829, 1998.

[68] L. Guan and A. Zhu, “Simplified dynamic deviation reduction-based Volterra model for Doherty power amplifiers,” in *2011 Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits*, 2011, pp. 1–4.

[69] J. Kim and K. Konstantinou, “Digital predistortion of wideband signals based on power amplifier model with memory,” *Electron. Lett.*, vol. 37, no. 23, p. 1417, 2001.

[70] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, “A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers,” *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, 2006.

[71] L. Guan and A. Zhu, “Dual-loop Model Extraction for Digital Predistortion of Wideband RF Power Amplifiers,” *IEEE Microw. Wirel. Components Lett.*, vol. 21, no. 9, pp. 501–503, Sep. 2011.

[72] A. Moore and J. Jimenez, *GaN RF Technology For Dummies*. John Wiley & Sons, Inc, 2015.

[73] K. Yuk, G. R. Branner, and C. Cui, “Future directions for GaN in 5G and satellite communications,” in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017, pp. 803–806.

[74] S. J. C. H. Theeuwen and J. H. Qureshi, “LDMOS Technology for RF Power Amplifiers,” *Microw. Theory Tech. IEEE Trans.*, vol. 60, no. 6, pp. 1755–1763, 2012.

[75] “Difference between GaN and LDMOS.” [Online]. Available:

http://www.rfwireless-world.com/Terminology/Difference-between-GaN-and-LDMOS.html. [Accessed: 04-Mar-2019].

[76] L. Cotimos Nunes, P. M. Cabral, and J. C. Pedro, “AM/AM and AM/PM Distortion Generation Mechanisms in Si LDMOS and GaN HEMT Based RF Power Amplifiers,” *Microw. Theory Tech. IEEE Trans.*, vol. 62, no. 4, pp. 799–809, 2014.

[77] O. Hammi, S.-C. Jung, and F. M. Ghannouchi, “Design for linearizability of GaN based multi-carrier Doherty power amplifier through bias optimization,” in *2012 19th IEEE International Conference on Electronics, Circuits, and Systems (ICECS 2012)*, 2012, pp. 492–495.

[78] M. Naseri Ali Abadi, H. Golestaneh, H. Sarbishaei, and S. Boumaiza, “Doherty Power Amplifier With Extended Bandwidth and Improved Linearizability Under Carrier-Aggregated Signal Stimuli,” *IEEE Microw. Wirel. Components Lett.*, vol. 26, no. 5, pp. 358–360, May 2016.

[79] Y. Zhu *et al.*, “Envelope Tracking Linearizability of Power Amplifier,” in *2018 Asia-Pacific Microwave Conference (APMC)*, 2018, pp. 998–1000.

[80] A. Alt and J. Lees, “Improving efficiency, linearity and linearisability of an asymmetric doherty power amplifier by modulating the peaking Amplifier’s supply voltage,” in *2017 47th European Microwave Conference (EuMC)*, 2017, pp. 464–467.

[81] I. Angelov, H. Zirath, and N. Rosman, “A new empirical nonlinear model for HEMT and MESFET devices,” *Microw. Theory Tech. IEEE Trans.*, vol. 40, no. 12, pp. 2258–2266, 1992.

[82] I. Angelov, L. Bengtsson, and M. Garcia, “Extensions of the Chalmers nonlinear HEMT and MESFET model,” *Microw. Theory Tech. IEEE Trans.*, vol. 44, no. 10, pp. 1664–1674, 1996.

[83] K. Shi, D. A. Calvillo-Cortes, L. C. N. de Vreede, and F. van Rijs, “A compact 65W 1.7–2.3GHz class-E GaN power amplifier for base stations,” in *2011 6th European Microwave Integrated Circuit Conference*, 2011, pp. 542–545.

[84] T. Sharma, R. Darraji, F. Ghannouchi, and N. Dawar, “Generalized Continuous Class-F Harmonic Tuned Power Amplifiers,” *IEEE Microw. Wirel. Components Lett.*, vol. 26, no. 3, pp. 213–215, 2016.

[85] S. Gao *et al.*, “Microwave Class-F and Inverse Class-F Power Amplifiers Designs using GaN Technology and GaAs pHEMT,” in *2006 European Microwave Conference*, 2006, pp. 1719–1722.

[86] Y. Y. Woo, Y. Yang, and B. Kim, “Analysis and experiments for high-efficiency class-F and inverse class-F power amplifiers,” *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 5, pp. 1969–1974, 2006.

[87] A. Grebennikov and F. H. Raab, “History of Class-F and Inverse Class-F Techniques: Developments in High-Efficiency Power Amplification from the 1910s to the 1980s,” *IEEE Microw. Mag.*, vol. 19, no. 7, pp. 99–115, 2018.

[88] S. C. Cripps, P. J. Tasker, A. L. Clarke, J. Lees, and J. Benedikt, “On the Continuity of High Efficiency Modes in Linear RF Power Amplifiers,” *IEEE Microw. Wirel. Components Lett.*, vol. 19, no. 10, pp. 665–667, 2009.

[89] F. Yang, J. Lees, H. Choi, H. Yu, and P. J. Tasker, “Continuous class BJ

power amplifier with DGS output matching," in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, 2014, pp. 1–4.

[90] P. Wright, J. Lees, J. Benedikt, P. J. Tasker, and S. C. Cripps, "A Methodology for Realizing High Efficiency Class-J in a Linear and Broadband PA," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 12, pp. 3196–3204, 2009.

[91] P. Wright, J. Lees, P. J. Tasker, J. Benedikt, and S. C. Cripps, "An efficient, linear, broadband class-J-mode PA realised using RF waveform engineering," in *2009 IEEE MTT-S International Microwave Symposium Digest*, 2009, pp. 653–656.

[92] T. Li and H. Wang, "A Continuous-Mode 23.5-41GHz Hybrid Class-F/F-I Power Amplifier with 46% Peak PAE for 5G Massive MIMO Applications," in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2018, pp. 220–230.

[93] N. Poluri and M. M. De Souza, "High-Efficiency Modes Contiguous With Class B/J and Continuous Class F<inline-formula> <tex-math notation="LaTeX">$^{-1}$ </tex-math></inline-formula> Amplifiers," *IEEE Microw. Wirel. Components Lett.*, vol. 29, no. 2, pp. 137–139, 2019.

[94] E. Aggrawal, K. Rawat, and P. Roblin, "Investigating Continuous Class-F Power Amplifier Using Nonlinear Embedding Model," *IEEE Microw. Wirel. Components Lett.*, vol. 27, no. 6, pp. 593–595, 2017.

[95] P. Chen and T. J. Brazil, "Gaussian processes regression for optimizing harmonic impedance trajectories in continuous class-F power amplifier

design," in *2015 IEEE MTT-S International Microwave Symposium*, 2015, pp. 1–3.

[96] B. Merrick, J. King, and T. Brazil, "A simplified procedure for the design of continuous Class-F power amplifiers," in *2013 European Microwave Integrated Circuit Conference*, 2013, pp. 508–511.

[97] N. Tuffy, L. Guan, A. Zhu, and T. J. Brazil, "A Simplified Broadband Design Methodology for Linearized High-Efficiency Continuous Class-F Power Amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 6, pp. 1952–1963, Jun. 2012.

[98] W. H. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proc. IRE*, vol. 24, no. 9, pp. 1163–1182, Sep. 1936.

[99] W. H. Doherty and O. W. Towner, "A 50-Kilowatt Broadcast Station Utilizing the Doherty Amplifier and Designed for Expansion to 500 Kilowatts," *Proc. IRE*, vol. 27, no. 9, pp. 531–534, Sep. 1939.

[100] L. Kahn, "Single-Sideband Transmission by Envelope Elimination and Restoration," *Proc. IRE*, vol. 40, no. 7, pp. 803–806, Jul. 1952.

[101] P. Asbeck and Z. Popovic, "ET Comes of Age: Envelope Tracking for Higher-Efficiency Power Amplifiers," *IEEE Microw. Mag.*, vol. 17, no. 3, pp. 16–25, Mar. 2016.

[102] J. Haedong, P. Roblin, and X. Zhijian, "Model-Based Nonlinear Embedding for Power-Amplifier Design," *Microw. Theory Tech. IEEE Trans.*, vol. 62, no. 9, pp. 1986–2002, 2014.

[103] C. Fager, J. C. Pedro, N. B. De Carvalho, and H. Zirath, "Prediction of IMD

in LDMOS transistor amplifiers using a new large-signal model," *Microw. Theory Tech. IEEE Trans.*, vol. 50, no. 12, pp. 2834–2842, 2002.

[104] Y.-S. Lee, S.-Y. Lee, and Y.-H. Jeong, "Effects of temperature on the IMD sweet spots in an LDMOSFET RF power amplifier," in *2005 Asia-Pacific Microwave Conference Proceedings*, 2005, vol. 2, p. 4 pp.

[105] S. Dardenne, C. Duvanaud, F. Robin, F. Huin, J.-. Paillot, and R. Allam, "Cancellation of third order intermodulation currents in a two-stages amplifier topology," in *IEEE MTT-S International Microwave Symposium Digest, 2003*, 2003, vol. 1, p. A193–A196 vol.1.

[106] Y. Lee, M. Lee, S. Jung, and Y. Jeong, "Analog Predistortion Power Amplifier Using IMD Sweet Spots for WCDMA Applications," in *2007 Asia-Pacific Microwave Conference*, 2007, pp. 1–4.

[107] R. Quaglia, V. Camarchia, M. Pirola, J. J. M. Rubio, and G. Ghione, "Linear GaN MMIC Combined Power Amplifiers for 7-GHz Microwave Backhaul," *IEEE Trans. Microw. Theory Tech.*, vol. 62, no. 11, pp. 2700–2710, 2014.

[108] P. M. Cabral, J. C. Pedro, and N. B. Carvalho, "Nonlinear device model of microwave power GaN HEMTs for high power-amplifier design," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 11, pp. 2585–2592, 2004.

[109] C. Lee, W. Ma, and N. L. Wang, "Averaging and Cancellation Effect of High-Order Nonlinearity of a Power Amplifier," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 54, no. 12, pp. 2733–2740, 2007.

[110] N. B. De Carvalho and J. C. Pedro, "Large- and small-signal IMD behavior of microwave power amplifiers," *Microw. Theory Tech. IEEE Trans.*, vol.

47, no. 12, pp. 2364–2374, 1999.

[111] M. Spirito, M. J. Pelk, F. van Rijs, S. J. C. H. Theeuwen, D. Hartskeerl, and L. C. N. de Vreede, “Active Harmonic Load–Pull for On-Wafer Out-of-Band Device Linearity Optimization,” *Microw. Theory Tech. IEEE Trans.*, vol. 54, no. 12, pp. 4225–4236, 2006.

[112] L. C. N. de Vreede and M. P. van der Heijden, “‘Linearization Techniques at the Device and Circuit Level’ (Invited),” in *Bipolar/BiCMOS Circuits and Technology Meeting, 2006*, 2006, pp. 1–8.

[113] M. Akmal *et al.*, “Linearity enhancement of GaN HEMTs under complex modulated excitation by optimizing the baseband impedance environment,” in *Microwave Symposium Digest (MTT), 2011 IEEE MTT-S International*, 2011, pp. 1–4.

[114] N. B. de Carvalho and J. C. Pedro, “A comprehensive explanation of distortion sideband asymmetries,” *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 9, pp. 2090–2101, 2002.

[115] R. Wilson, H. Jang, and B. Arigong, “Enhanced instantaneous bandwidth LDMOS RF power transistor using integrated passive devices,” in *2016 IEEE MTT-S International Microwave Symposium (IMS)*, 2016, pp. 1–4.

[116] K. W. Martin, “Complex signal processing is not complex,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 51, no. 9, pp. 1823–1836, 2004.