Enhanced characterization and calibration techniques for millimeter-wave active devices

Jos van ’t Hof

TU Delft

Enhanced characterization and calibration techniques for millimeter-wave active devices

by

Jos van 't Hof

to obtain the degree of Master of Science
with a specialization in RF/Wireless Circuit Design
at the Delft University of Technology,
to be defended publicly on Wednesday September 30, 2020 at 15:30.

Student number:	4624181	
Project duration:	September 1, 2019 – September 30, 2020	
Thesis committee:	Dr. M. Spirito,	TU Delft (ELCA), supervisor
	Dr. M. Babaie,	TU Delft (ELCA)
	Dr. D. Cavallo,	TU Delft (Terahertz Sensing)
	MSc. C. De Martino,	TU Delft (ELCA)

This thesis is confidential and cannot be made public until March 31, 2020.

TU Delft

Preface

Commercial interest for millimeter-wave wireless communication has increased significantly over the past years due to the appearance of large volume applications such as automotive radar and telecom applications, i.e. as 5G systems operating in the FR2 band, ISM applications in the 60 GHz E-band, and beyond 100 GHz in the W-band and D-band. Developments in this frequency range are enabled by the continuous improvements in silicon-based technologies over the past decades and are driven by frequency congestion in low-GHz bands. To support development for these applications there is a demand for new millimeter-wave test equipment, de-embedding/calibration methods and device characterization approaches. Three topics related to challenges in millimeter-wave on-wafer characterization are addressed in this work.

A novel characterization method is developed for in-band linearity estimation that overcomes the limitation of harmonically generated spectral content by the frequency extension modules in the millimeter-wave active load-pull setups. The method is able to estimate trends in the EVM across input drive level, frequency and loading condition from vector gain measurements of the active device. On-wafer measurements were performed to benchmark the method against true EVM measurements in two sub-30 GHz bands. A proof of concept demonstration at 165 GHz showed the ability of the method to estimate trends for in-band linearity across drive level and loading condition at millimeter-wave.

A new parametrized 16-term error model and calibration algorithm was developed to overcome probe-to-probe cross-coupling effects during millimeter-wave on-wafer characterizations. Conventional 16-term calibrations are not practical as these are only effective if the probe-to-probe spacing is not changed after calibration. This new approach allows for interpolation of the error terms for devices under test that have a pad spacing not included in the set of calibration standards. This makes the method powerful and practical as it would only require one set of calibration structures for large-scale on-wafer measurements of devices with a different pad spacing.

Finally, a preliminary investigation was performed into active device behaviour when operating close to the maximum oscillation frequency of the technology. This investigation is conducted using on-wafer millimeter-wave large signal measurements on a 22-nm CMOS FD-SOI technology. A new experimental method is proposed that could show different operating behaviours when moving towards the maximum oscillation frequency. This approach uses large-signal (load-pull) measurements of the device across power and supply voltage. Under the hypothesis that a slew rate like operation is obtained when the maximum oscillation frequency is approached, this characterization may show the transition point in frequency for the onset of such device limiting behaviour.

This thesis was performed in the Electronic Circuits and Architectures (ELCA) group at the Faculty of Electrical Engineering, Mathematics & Computer Science (EEMCS) of the Delft University of Technology (TU Delft) in the Netherlands. The committee for this thesis consists of dr. Marco Spirito (responsible associate professor), dr. Masoud Babaie (assistant professor) and Carmine De Martino (PhD. candidate) from the ELCA group, and dr. Daniele Cavallo (assistant professor) from the Terahertz Sensing group.

I would like to thank my supervisors Marco Spirito and Carmine De Martino for their advice and support in assisting this project. Satoshi Malotaux, Luca Galatro and Rafaelle Romano for their insightful discussions and help in using the millimeter-wave active load-pull setups, including those of Vertigo Technologies. Michele Squillante and Mauro Marchetti for their support for my measurements on the Anteverta MW systems, and Rob Bootsman and Motahhareh Estebsari for their ADS models.

A special thanks and remembrance goes out to prof. Earl McCune for kick-starting this thesis project. His advice during the project was highly appreciated, and his humour, enthusiasm and kindness will be missed.

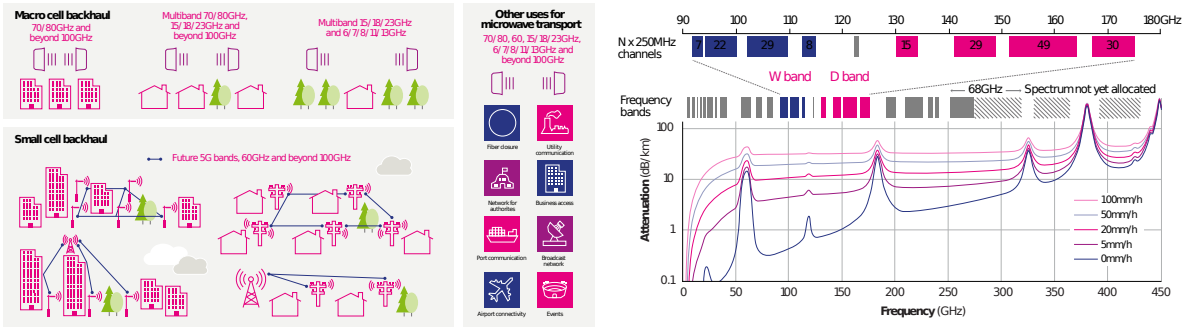
Jos van 't Hof
Seattle, September 2020

Contents

1 Introduction

1.1. Motivation

Commercial interest for millimeter-wave wireless communication has increased significantly over the past years, due to the appearance of large volume applications such as automotive radar and telecom applications (i.e. 5G systems operating in the FR2 band, 60 GHz ISM applications and back-haul in E-band). Millimeter-wave (mmWave) systems operate with Radio Frequency (RF) signals having a wavelength from ten to one millimeter, 30 GHz to 300 GHz. Developments in this frequency range are enabled by the continuous improvements in silicon-based technologies over the past decades and are driven by frequency congestion in low-GHz bands. Two bands currently under consideration for future millimeter-wave communication (i.e. 6G) are W-band and D-band [1] with available frequency slots assigned from 90 GHz to 180 GHz, see fig. 1.1b, where the low atmospheric adsorption windows allow for medium to long distance communication [2]. The increase in available bandwidth, and high frequency reuse capability due to atmospheric attenuation makes millimeter-wave particularly suited for dense urban environments using small-cells fig. 1.1a.



(a) Micro-wave (macro-cell) and millimeter-wave (small-cell) back-haul (b) Atmospheric attenuation in W-band and D-band

Figure 1.1: Figures from the 2017 Ericsson Technology review [1, p. 25-36] showing opportunity for millimeter-wave systems in back-haul networks, and the spectrum allocation and atmospheric attenuation beyond 100 GHz

To exploit the available frequencies and maximize the system throughput, modern communication standards use non-constant envelope modulation schemes and thus require linear operation of the power amplifier (PA) present in the transmitter chain. The trade-off between linearity and power efficiency becomes paramount for millimeter-wave, and proper transistors models, compact or behavioural [3], [4], are required to accurately predict all the non-linear contributions in order to properly tackle this trade-off early in the design cycle. Reducing the number of design iterations of millimeter-wave PAs therefore requires accurate characterization of active devices and validation of these transistor models employed in the circuit design.

To support development for these applications there is a demand for new millimeter-wave test equipment, de-embedding/calibration methods and device characterization approaches. State-of-the-art on-wafer calibration/de-embedding techniques [5] and millimeter-wave active load-pull setups [6] allow for characterization of active devices at the targeted millimeter-wave frequencies. In this thesis new methods are explored for linearity estimation, calibration and behaviour characterization for active devices that are compatible with millimeter-wave active load-pull setups.

1.2. Research themes and objectives

This thesis was performed in the ELCA group at the Faculty of Electrical Engineering, Mathematics & Computer Science (EEMCS) of the Delft University of Technology (TU Delft) in the Netherlands. The work was done in collaboration with the measurement and characterization specialists of Anteverta-MW and Vertigo Technologies, both spin-offs of the ELCA group. The themes addressed in this research include:

- **Millimeter-wave active load-pull characterization** with a focus on using the state-of-the-art systems for linearity characterization of active devices. Active load-pull characterization allows to overcome the limitations in measurement speed and tuning capability of passive load-pull systems. Both mixed-signal active load-pull systems [7] and millimeter-wave active load-pull systems [6] are used in this thesis.
- **Millimeter-wave/sub-Terahertz calibration techniques** including on-wafer calibration and de-embedding kits [5], EM-based calibration [8] and 16-term cross-coupling correcting calibrations [9]–[11].
- **Device limiting effects** when operating active devices at frequencies close to their f_T/f_{max}, focussing on novel measurement approaches utilizing millimeter-wave active load-pull characterization and on-wafer calibration/de-embedding to show physical limitations for high speed devices.

The objectives for this thesis include:

- **Develop a technique for millimeter-wave linearity prediction and evaluation** that utilizes the capabilities of state-of-the-art active load-pull setups.
- **Develop a probe-spacing-parametrized calibration algorithm** that can correct for probe-to-probe cross coupling in millimeter-wave and sub-Terahertz on-wafer measurements.
- **Investigate characterization methods that can show device limiting effects** of active devices that are operated at frequencies close to the $f_T/f_{max}$of the technology.

1.3. Report structure

The remainder of this report is structured into six chapters. Chapter 2 provides a brief overview of the state-of-the-art in millimeter-wave architectures and PAs, it discusses the active load-pull test systems and calibration/de-embedding techniques that were used in this work, and provides an overview of the test characterization dies that have been made prior to this work. Chapter 3 presents a novel characterization method developed for in-band linearity estimation that overcomes the limitation of harmonically generated spectral content by the frequency extension modules in the millimeter-wave active load-pull setups. Chapter 4 presents the development and validation of a novel 16-error term calibration algorithm that may be used in a parametrized fashion to correct for probe spacing dependent cross-coupling in on-wafer measurements. Chapter 5 presents preliminary investigations that have been performed into a characterization method that could show device limiting effects when operating active devices at frequencies close to the $f_T/f_{max}$of the technology. This report concludes with a summary of the conclusions and recommendations in chapter 6.

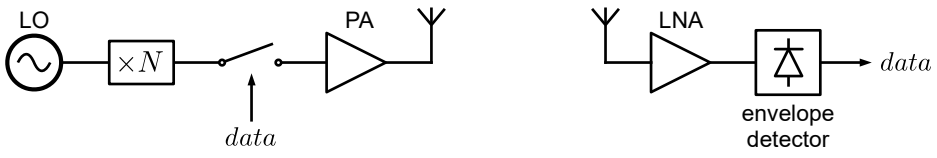
2

State-of-the-art in millimeter-wave characterization

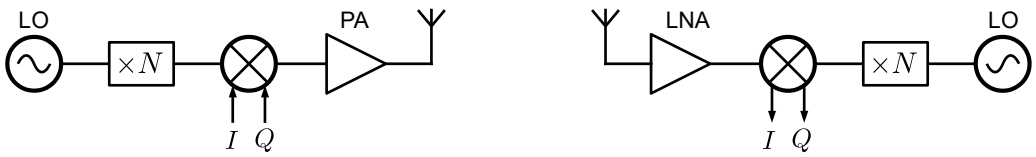
This chapter provides background material on millimeter-wave architectures and PAs. It will highlight what techniques are currently being researched to support development for millimeter-wave applications, with a focus on active load-pull characterization. State-of-the-art calibration/de-embedding techniques are briefly discussed to provide the necessary context for the work presented in chapter 3 and chapter 5. A more extensive review of 16-error term calibration algorithms is presented in chapter 3. This background chapter will conclude with a short summary of the testing and characterization dies that were developed in the ELCA group specifically for technology characterization.

2.1. Architectures and power amplifiers beyond 100 GHz

Millimeter-wave communication beyond 100 GHz is a novel field of research where currently architectures are proposed, developed and tested on small-scale with link distances from centimetres up to several meters. Two distinct directions can be identified in the transceiver architectures in terms of the used technology and modulation approach. In the first category, several publications show the use of low cost CMOS technologies for on-off-keying (OOK) or amplitude-shift-keying (ASK) based communication link [12]–[15]. This results into a simple transceiver architecture as shown in fig. 2.1a consisting on the transmitter side of a frequency multiplied Local Oscillator (LO), a switching circuit for the modulation, a PA and a transmitting antenna, and on the receiving side of an antenna low noise



(a) OOK/ASK based millimeter-wave communication architecture



(b) Direct conversion based millimeter-wave communication architecture

Figure 2.1: Two common millimeter-transceiver architectures in publications: the OOK based architecture a the direct-conversion based architecture

amplifier (LNA) and an envelope/power detector for the demodulation. Using this simple and low cost architecture data rates of 10 Gbps over a distance of 3 meters may be achieved [13][1] which would suit indoor applications. In the second category, several publications use a direct-conversion architecture in higher speed and higher cost Gallium Arsenide (GaAs) and Indium Phosphide (InP) technologies [16]–[18]. Here the same building blocks apply but with the modulators and demodulators replaced by an IQ-mixer allowing for quadrature amplitude modulation (QAM) schemes with rates up to 48 Gbit/s [18][2] over 40 cm distance.

In both cases more (directed) power is required to increase the range of the link. Possible ways of doing so include increasing the transmitter power and/or applying beam forming/lensing techniques using antenna arrays, the latter being well suited for millimeter-wave as the small wavelength results into physically small antennas [19]. In this work a focus is put on the active devices required for the PA in the millimeter-wave system. Because these devices now operate at a fraction of their maximum oscillation frequency f_{max} the achievable gain at these frequencies is low, therefore several stages are required to boost the signal power level, further decreasing the power efficiency of the amplifier.

An extensive PA performance survey has been performed by H. Wang et. al. [20] from peer-reviewed publications such as those from IEEE ISSCC, JSSC and RFIC. Figure 2.2 shows one of the results from this survey, showing the saturated PA output power P_{sat} across frequency for different transistor technologies.

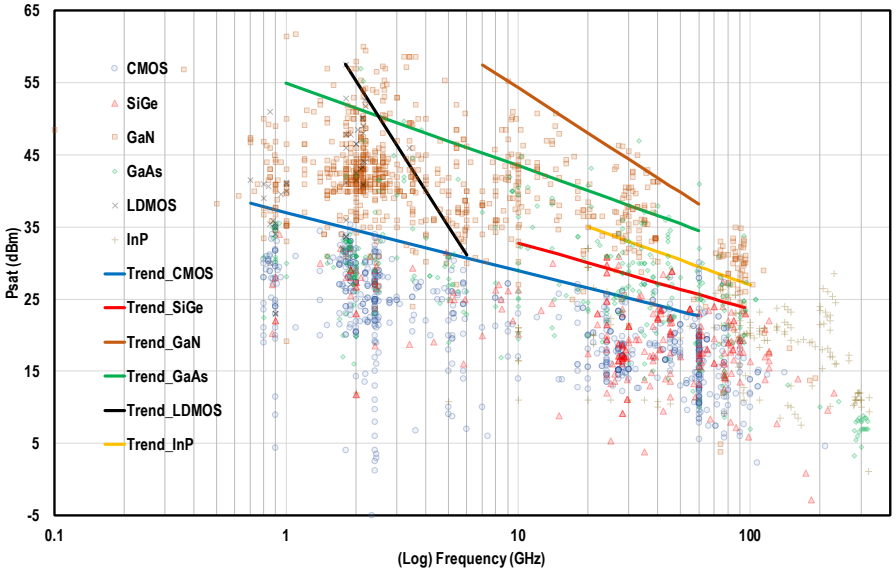


Figure 2.2: PA saturated output for different technologies across frequency. Trends are drawn for publications with $PAE \geq 30\%$ up to the frequency for which this PAE is still obtained, (adapted from [20])

Also in fig. 2.2 trends are drawn for each technologies based on the publications that achieve a 30% PAE or higher. From these trends it can be seen that their exist significant differences in achievable saturated output power among the different technologies. The III-V technologies[3], Gallium Nitride (GaN), GaAs and InP, offer significant advantages in terms of performance. These compound semiconductors are composed of one element from group III with three valence electrons (B, Al, Ga, In or Ti) and one element from group V with five valence electrons (N, P, As, Si, Bi). GaN has the advantage of a higher breakdown voltage than Silicon (Si), allowing it to withstand a larger voltage and therefore may sustain larger voltage swings to achieve higher power levels [21], [22]. GaAs and InP offer higher electron mobilities [23, tab. IV] than Si which offers an advantage for high frequency operation. These differences may be recognized in the trends of fig. 2.2, where GaN offers the highest output powers up to 60 GHz, followed by GaAs and InP.

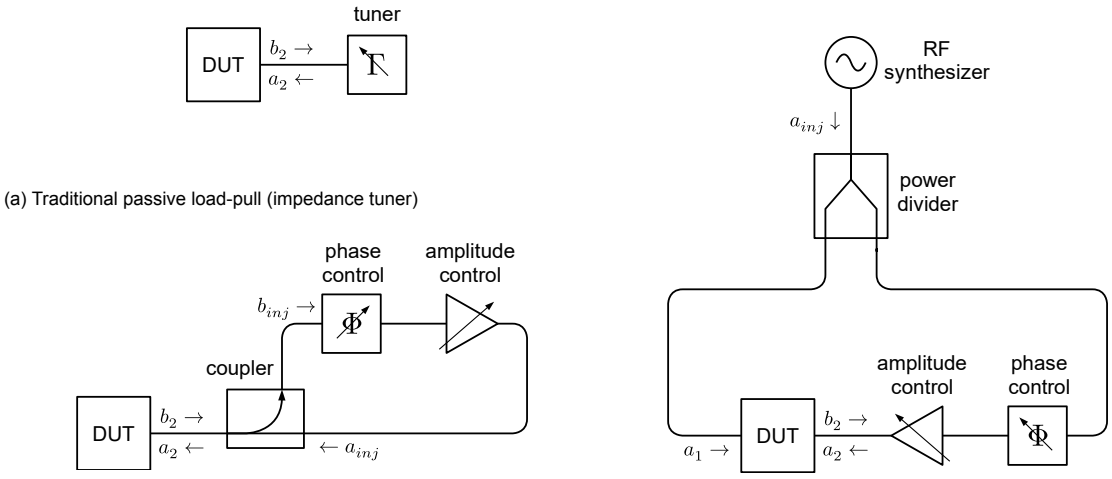
[1] The operating bandwidth in [13] is 18 GHz centred at 130 GHz and uses an oscillator without frequency multiplication
[2] The operating bandwidth in [18] is 10 GHz centred at 120 GHz and uses 64-QAM
[3] Pronounced ”three-five”

The trends seen in the saturated output power in fig. 2.2 follow the Johnson's figure of merit [24] which states that the ultimate performance limits of a transistor are set by the product of the charge carrier saturation velocity[4] and the electric breakdown field strength (see [23, tab. IV] for tabulated values of the technologies). Semiconductor materials with a higher Johnson's figure of merit are able to withstand higher voltages and reach higher speeds which therefore makes them more suitable to create high power at high frequencies. When operating closer to the f_T or f_{max} of the device the achievable power gain G_P of the device decreases. The decreasing output power and power gain result into low efficiency PAs solutions. As shown by the trend lines in fig. 2.2, none of the technologies is able to maintain a PAE higher than 30% beyond 100 GHz. These efficiency numbers reduce even further when these PAs are to be operated in power back-off to achieve the required linearity for a particular modulation scheme. Therefore, it becomes increasingly important to be able to predict linearity performance of active devices in early stages of the design cycle.

2.2. Active load-pull characterization

Load-pull characterization is a measurement technique for large signal device characterization and device model validation [26, chp. 13]. During a load-pull characterization the loading conditions of the Device Under Test (DUT) are varied while a large signal response of the device is measured. In traditional passive load-pull a mechanical impedance tuner is used to set the loading condition on the DUT fig. 2.3a. This can be done at the input (source pull) and/or the output (load-pull) of the device. To measure the loading condition applied to the DUT bi-directional couplers are used to measure the incident and reflected waves to the two-ports of the device. Passive load-pull has two main disadvantages [27, p. 78]: 1) mechanical impedance tuners are slow therefore requiring a lot of measurement time for a large sweep and 2) can only achieve limited range of reflection coefficients because of the losses between the reference plane and the impedance tuner, i.e. the probes, cables and couplers.

The two problems with passive load-pull can be circumvented by an active system that can overcome the losses in the network. By injecting an a_2 wave into the output of the device with controlled a phase and amplitude, set by the ratio $\Gamma_L = a_2/b_2$, any loading condition may be generated. However, to successfully do so a phase coherence should be maintained between the b_2 and a_2 wave [27, p. 78-79]. This phase coherence can be maintained by a closed loop fig. 2.3b or open loop system fig. 2.3c. In a closed loop-system, the b_2 wave is sampled by a coupler and fed into a phase and amplitude controlled feedback loop that injects the a_{inj} wave back into the DUT. The resulting a_2/b_2 ratio



(a) Traditional passive load-pull (impedance tuner)

(b) Closed-loop active load-pull

(c) Open-loop active load-pull

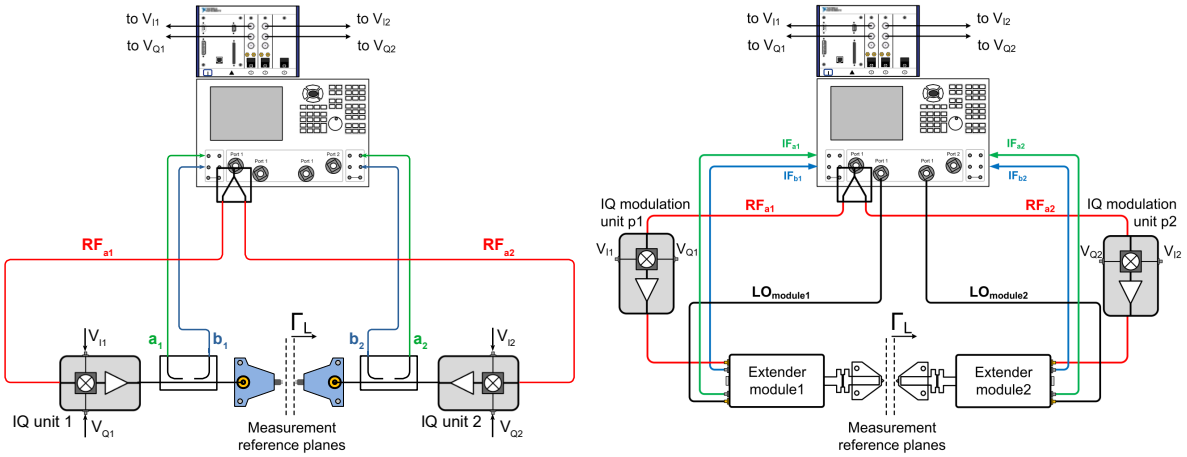
Figure 2.3: Schematic representation of load tuning techniques [27, p. 79]

[4]Note that for high field strengths the saturation velocity is non-linearly dependent on the electric field, i.e. therefore not governed by the electron mobility. Different energy band structures of the materials dictate the actual saturation velocity for different materials [25, p. 171]. This is why GaAs and InP may have a higher (low-field strength) electron mobility but a lower saturation velocity than GaN (see tabulated values in [23, tab. IV]).

can be measured by an bi-directional coupler attached to the receivers of a vector network analyser (VNA). An outer feedback loop can then tune the phase shift and amplitude of the loop to achieve the desired loading condition. Alternatively, the phase coherence may be achieved in an open-loop system by using the same reference signal for the input a_1 wave and the injected b_2 wave.

The state-of-the-art active load-pull setups used in this thesis use open-loop systems. A VNA based open-loop active load-pull system is shown in Figure 2.4a. In this test setup IQ-mixers are used for the amplitude and phase control of the injected waves similar to IQ-modulation. The IQ-mixer at the input is used to set the input drive level of the DUT. The voltages on the I and Q inputs are set by analog waveform generators and the incident and reflected waves to the DUT are measured using external couplers and the receivers of the VNA. Power amplifiers are added after the IQ-mixers to overcome the losses of the cables, probes and couplers. Attenuators may be added in front of the IQ-mixers and amplifiers to ensure correct drive levels and maximizing the dynamic range of the setup. This setup may be regarded as a fundamental frequency VNA based active load-pull setup as it operates in the fundamental frequency range supported by the VNA. Current state-of-the-art VNAs support frequencies up to 67 GHz[5], beyond this range frequency extension modules may be used.

This setup is not necessarily restricted to using a VNA for the RF synthesis and measurement. Inside the VNA, depending on the architecture used, the received a&b-waves may be down-converted before their power is measured. In fact, the architecture inside the VNA, i.e. the mixers for up/down-conversion can be made external, and the signal generation and sampling replaced by Digital-to-Analog converters (DACs) and Analog-to-Digital converters (ADCs). With now full control over the generated and sampled baseband signals, a mixed-signal active load-pull can be performed, first demonstrated by M. Marchetti et. al. [28] and commercialized by Anteverta MW. Mixed-signal active load-pull allows for characterization under realistic modulated signals. When additional loops are added to the system also a harmonic load-pull may be performed where the loading conditions of the harmonics are also controlled. For further information into the operation of mixed-signal active load-pull systems the reader is invited to read [27]–[30].



(a) VNA based active load-pull

(b) VNA frequency extender based millimeter-wave active load-pull (from [6])

Figure 2.4: VNA frequency extender based millimeter-wave active load-pull

The frequency range of the VNA active load-pull setup may be extended by commercially available frequency extension modules. This concept was first demonstrated by C. De Martino et. al. [6] and commercialized by Vertigo technologies. The extension modules contain $\times N$ frequency multipliers in the RF up-conversion paths and a&b-wave down-conversion paths to synthesize the frequencies at the millimeter-wave band of interest. A schematic overview of such a setup is shown in fig. 2.4b. Depending on the multiplication factor of the extension modules the Intermediate Frequency (IF) source signal of the VNA may now lie significantly lower, i.e. between 10 - 20 GHz, thereby also lowering the required frequency range of the IQ-mixer and amplifiers. A second LO signal is generated from the VNA to drive

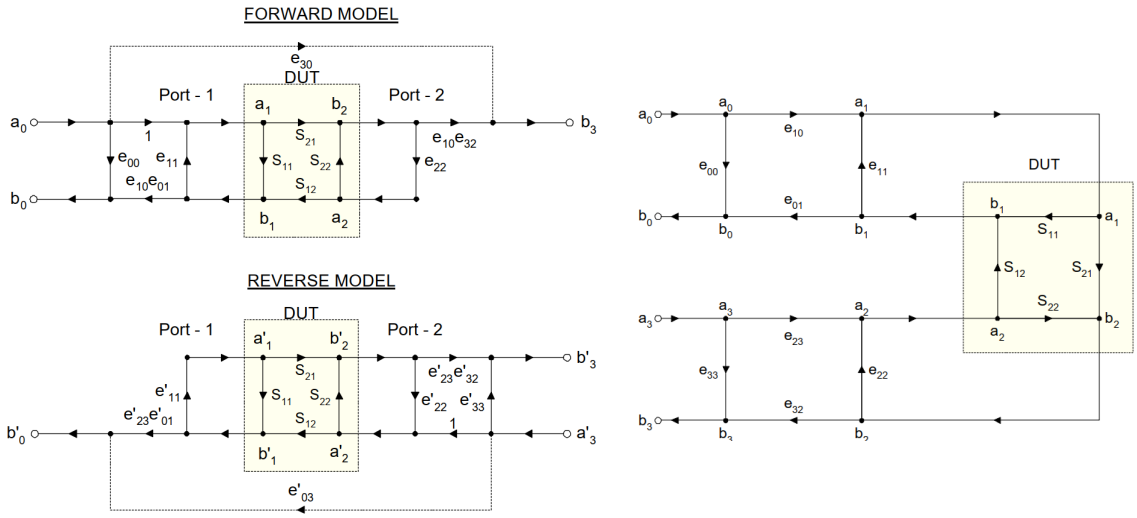
[5] approximately the cut-off frequency of 1.85 mm coaxial cables

the down-converters for the incident and scattered waves. For further information into the operation of millimeter-wave active load-pull systems the reader is invited to read [6], [27].

2.3. Calibration / De-embedding techniques

During characterization of a DUT it is important to quantify and remove the effects of the systematic errors (cable losses, mismatch, directivity, etc.) during the measurement. Doing so requires setting the measurement reference planes to a user defined location at which the behaviour is to be characterized, for example at the connectorized interface of the DUT or at the probe tips for an on-wafer measurement. Using calibration, the systematic errors that are present in the measurement setup are mapped to an error model. Once the parameters of the error model are known, they can be used to correct for the systematic errors in the measurement.

Two error models typically employed for calibration are the 12-terms error model and the 8-terms error model. The models represent the systematic errors as flow graphs that show mismatch, losses, leakage and coupling. Figure 2.5 show these flow graphs of the two models, with the e_{xy} terms being the error terms of the model. Apart from the two cross-terms in the 12-term error model[6], the two models describe exactly the same system and can be interchanged when modern VNAs are used [31][27].



(a) 12-term error model

(b) 8-term error model

Figure 2.5: The two typically used error models for VNA calibration 12-term and 8-term (from [31])

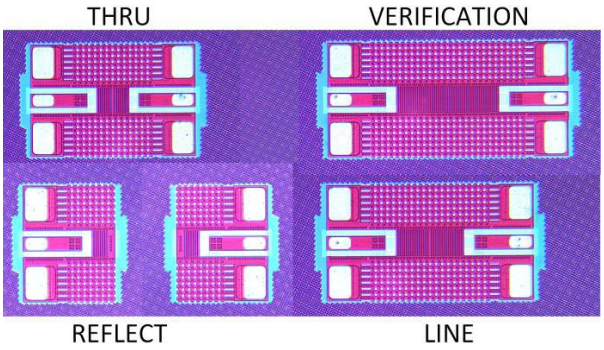
Calibration is the process of measuring known standards in order to determine the error terms in the error model. Different standards may be used to determine the error terms. For two-port measurements the calibration procedure is often named after the standards used [27, p. 18]:

- Short-Open-Load-Thru (SOLT)
- Reciprocal-Short-Open-Load (RSOL) [32]
- Line-Reflect-Match (LRM) [33]
- Thru-Line-Reflect (TRL) [34]

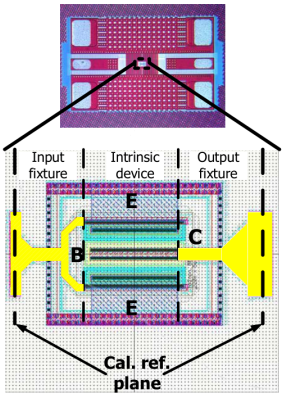
The calibration standards should be defined at the location where the reference planes are to be defined. For example if the reference plane is a the ends of a coaxial cable, coaxial standards should be used that are defined at their coaxial port. Similarly, if the reference plan is at the probe pads of a chip, probe tip calibration standards should be used. However, when it is not possible to physically create the standards at the desired reference planes a de-embedding step should be performed. By de-embedding, additional structures are measured after calibration, i.e. just the on-wafer interconnect structure from the pads to the DUT, which are then later used to remove the effects.

[6] without the two cross terms the model is referred to as the 10-error term model

When it is possible to create the calibration standards at the desired reference plane positions the de-embedding step can be combined with the calibration step. For on-wafer device characterization, the reference planes should be physically as close as possible to the intrinsic device. An calibration / de-embedding technique developed by L. Galatro [5] allows for calibration at the first metallization layer, moving the reference planes as close as possible to the intrinsic device. The capacitively loaded inverted co-planar waveguide (CPW) de-embedding technique has a high cut-off frequency allowing for characterization of the behaviour of the intrinsic device up to 300 GHz. Figure 2.6 shows the on-wafer calibration kit and the location of the reference planes[7]. This technique is used on all the testing and characterization dies used during this thesis (described in section 2.4).



(a) Micrograph of the de-embedding kit on the Infineon B11HFC characterization die



(b) Micrograph (top) and detailed view (bottom) of device layout for the transistor

Figure 2.6: The capacitively loaded inverted CPW de-embedding technique allows for device characterization as close as possible to the intrinsic device (from [5])

2.3.1. Generalized calibration procedure for active load-pull characterization

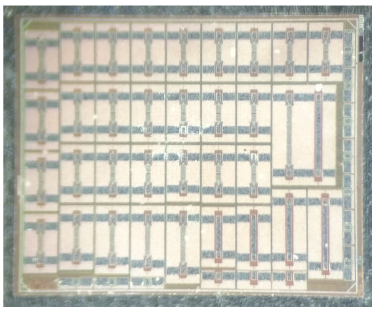
Several calibration steps are required in order to perform an on-wafer active load-pull measurement for active device characterization. The exact steps may differ between the test-setups, but four general steps may be considered [35]–[37]:

- **First tier calibration**
  A first tier two-port calibration, i.e. SOLT, TRL, LRM, is performed at the last connectorized interface, i.e. coaxial or waveguide. This determines the error terms of the error model up until the probes and allows relating the measured a&b power waves an absolute power measured by the power meter in the next step.
- **Power calibration**
  During the power calibration step, the power measured at the coaxial or waveguide reference plane is related to the measured a&b power waves at the receivers/digitizers. For large signal measurements it is important to include this absolute power reference to know the values of quantities such as P_{in} and P_{out}.
- **Power levelling**
  During the power levelling step, the magnitude of the I and Q signals send to the mixers are varied to map their values to the output power level. This allows for accurate power control during the active load-pull measurement, i.e. to set the correct drive level, or the magnitude of the loading condition $|\Gamma_L|$
- **Second tier calibration**
  Finally, a second tier calibration is performed on-wafer, using an on-wafer calibration kit. This calibration corrects for the probes and the on-wafer structures. The power calibration and levelling information may be propagated through the new error model to achieve power measurement and control at the new on-wafer reference planes.

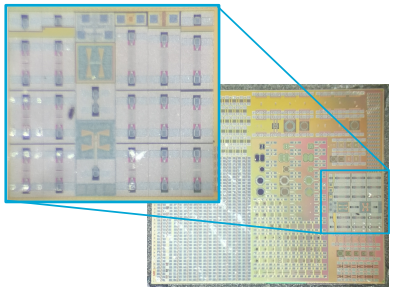
[7] In this work the second de-embedding step that de-embeds the input and output fixtures is not performed.

2.4. Devices under test

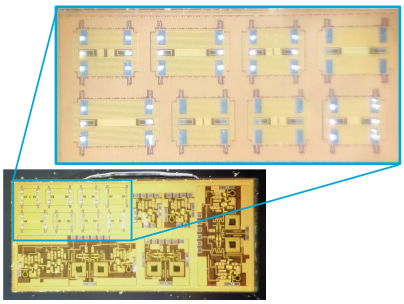
Four technologies were available for on-wafer characterization during this project, two Complementary MOS (CMOS) technologies, a 22nm fully depleted silicon on insulator (FD-SOI) by Global Foundries [38] and a 28nm CMOS-Bulk by TSMC, and two heterojunction bipolar transistor (HBT) technologies, a 130nm Silicon Germanium (SiGe) by IHP [39] and a 130nm SiGe by Infineon [40]. Specialized characterization dies were developed previously by the works of L. Galatro, C. De Martino and S. Malotaux. These dies are shown in fig. 2.7. On each die there are one or multiple DUTs in common-source or common-emitter configuration. The Global Foundries and TSMC dies have on-wafer SOLT and TRL CPW calibration kits. The IHP and Infineon dies only have on-wafer TRL CPW calibration kits available.



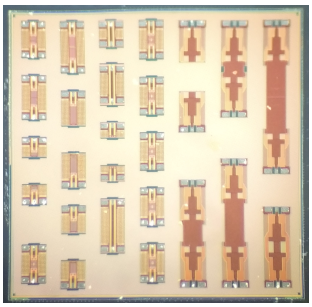
(a) Global Foundries 22nm CMOS FD-SOI



(b) TSMC 28nm CMOS Bulk



(c) IHP 130nm SiGe HBT



(d) Infineon 130nm SiGe HBT

Figure 2.7: Micro-graphs of the characterization dies that were available for this project.

This report shows primarily the results obtained with the 22nm Global Foundries dies, in particular on the super low voltage threshold NMOS2 (slvt) transistors. This transistor was of preference as some characterization on this transistor was performed in previous work and it has the lowest saturated output power of the available DUTs. In addition, the inclusion of the SOLT calibration kit allows for characterization from low frequencies. An overview of the parameters of the four characterization dies is shown in table 2.1.

Table 2.1: Overview of parameters of the four characterization dies

Technology	**22 nm Si CMOS-SOI**	**28 nm Si CMOS-Bulk**	**130 nm SiGe HBT SG13G2**	**130 nm SiGe HBT B11HFC**
Foundry	Global Foundries	TSMC	IHP	Infineon
f_T/f_{max}	347 / 371 GHz		300 / 500 GHz	250 / 370 GHz
V_{be} or V_{gs} max	0.8 V	0.8 V	0.95 V	0.95 V
V_{ce} or V_{ds} max	0.8 V	0.8 V	1.8 V	1.6 V
Probe-pitch	$100\ \mu m$	$100\ \mu m$	$125\ \mu m$	$100\ \mu m$
On-wafer kit	M1-SOLT, M1-TRL, TM-TRL	M1-SOLT, M1-TRL	M1-TRL	M1-TRL
Literature	[38]	[41]	[39]	[5], [40]

3

Vector gain based EVM estimation

In this chapter a novel characterization method is presented for in-band linearity estimation that overcomes the limitation of harmonically generated spectral content by the frequency extension modules used in the millimeter-wave active load-pull setups described in section 2.2. The state-of-the-art millimeter-wave active load-pull setups use commercially available VNA frequency extension modules [6] to generate frequencies above 75 GHz. These extension modules operate using non-linear $\times N$ frequency multipliers in the RF up-conversion path as shown in fig. 3.1. This non-linear element in the path makes conventional non-linear characterization methods such as the two-tone test challenging. This problem is illustrated in fig. 3.2, when two-tones are applied to the input of the module with a $\times 8$ multiplier, the actual excitation of the DUT will contain intermodulation products created by the non-linear multiplier, including problematic in-band intermodulation products.

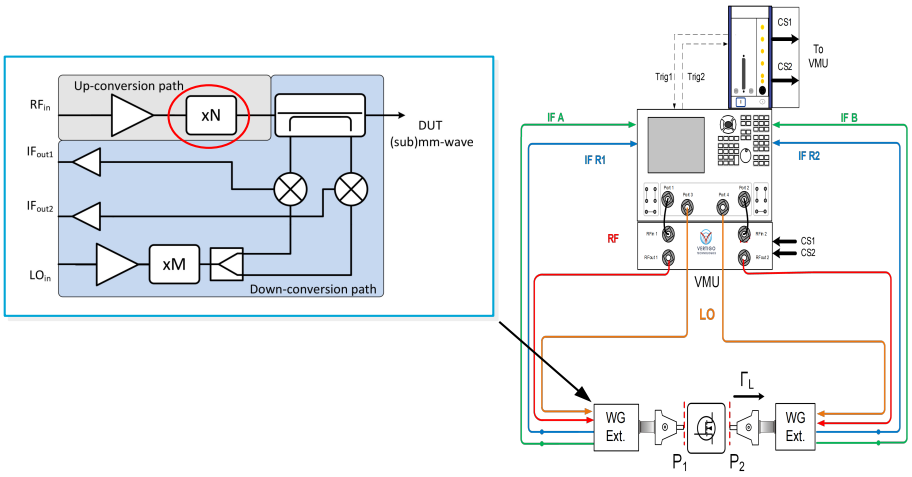


Figure 3.1: Millimeter-wave active load-pull setups use VNA frequency extension modules that have non-linear frequency multipliers in the RF up-conversion path making linearity characterizations such as the two-tone test challenging.

Different frequency extension modules exist such as the VDI dual-source frequency extension module for inter modulation distortion (IMD) measurements [42]. This module combines two RF up-conversion paths with individual frequency multipliers and therefore do not create the problematic IMD products. These dual-source modules however are much more expensive and less readily available. The linearity metrics that can be extracted by current millimeter-wave active load-pull systems are therefore limited to single tone linearity measurements such as amplitude to amplitude (AM-AM) and amplitude to phase (AM-PM) responses. This limits the characterization capabilities for modulated signal linearity in the early stages of the millimeter-wave design process. Since millimeter-wave systems are being intensively investigated for large bandwidth data transmission it is paramount to benchmark

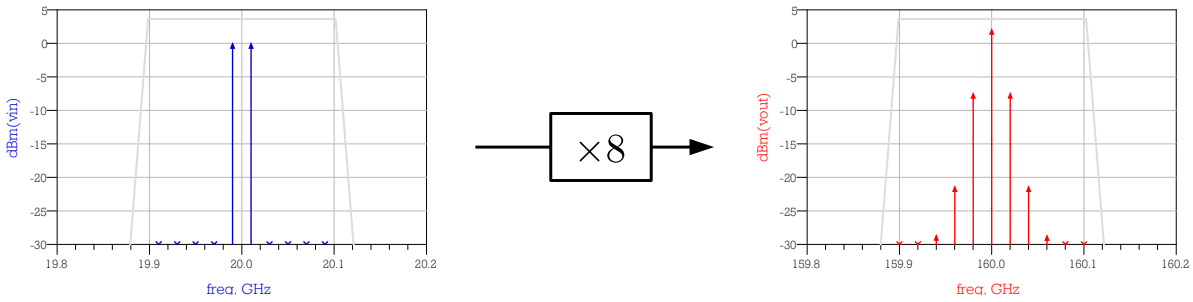


Figure 3.2: Example of in-band intermodulation products created by a 8th order VNA frequency extension module when a two-tone signal is applied.

the transistor model prediction for these metrics. Therefore, improved techniques to allow the extraction of the in-band linearity at device level need to be developed.

The paper by Vanaverbeke et. al. [43] describes how component level AM-AM and AM-PM behaviour can be extracted during a traditional passive or active load-pull characterization from the power waves (or a&b-waves). The AM-AM and AM-PM behaviour is captured in the ratio of the (down-converted) b_2 and a_1 waves also called the vector gain[1] of the DUT [37]. The vector gain can be extracted without requiring precise knowledge of the source-impedance as the AM-AM and AM-PM effects of the source mismatch can later be added in post-processing. Using this approach millimeter-wave active load-pull may become a powerful characterization technique also for linearity evaluation, as first order component level linearity behaviour can be characterized at different bias conditions, drive levels, frequencies and loading conditions.

The obtained AM-AM and AM-PM responses from the active load-pull measurements may be evaluated for modulate signals in the envelope domain as is done during Circuit Envelope Simulations in Keysight's Advanced Design System (ADS) [44] or during X-parameter based characterization and modelling [45]. In envelope domain based simulations, the modulation envelope is represented in the time-domain as a carrier modulated by an envelope: $A(t) \cdot e^{j\phi(t)}$. The envelope of the signal over time can then be used to apply a magnitude and phase transfer according to the AM-AM and AM-PM curves at the carrier frequency. During this process also the contents at the harmonics may be considered as is done in the X-parameter behavioural modelling approach [45]. Envelope domain based analysis may also be performed on the single tone AM-AM and AM-PM responses extracted from millimeter-wave active load-pull characterization systems, however such analysis effectively assumes a narrow-band response of the device and therefore does not fully utilize the ability of the setup to characterize across a large frequency band of operation. Similarly, harmonic characterization is difficult to perform because the millimeter-wave setups use wave-guide transmission lines with one octave wide bandwidths (i.e. WR-5.1 140 - 220 GHz or WR-3.4 220 - 330 GHz) and therefore harmonics cannot be observed during the active load-pull measurement. In this chapter a novel frequency domain based characterization method is described that fully utilizes the capabilities of state-of-the-art millimeter-wave active load-pull systems and can be used to evaluate in-band linearity metrics such as Error Vector Magnitude (EVM).

3.1. Goal & Research question & Structure

The main goal of the work described in this chapter is to develop and validate a method that can extract linearity metrics for modulated signals within the current capabilities of millimeter-wave active load-pull systems. To do so an approach is to be formulated that uses frequency domain behaviour characterized within the desired bandwidth of operation. The method should be validated for on-wafer measurements and should be applicable for use in on-wafer measurements using active load-pull systems operating beyond 75 GHz. The research question associated to this chapter is formulated as followed:

How can linearity metrics for modulated signals be extracted within the current capabilities of millimeter-wave active load-pull systems operating beyond 75 GHz?

[1] In the paper by Vanaverbeke et. al. [43] the vector gain is called the complex-power gain $\hat{G}_p$ and the transducer vector gain is called the complex transducer gain $\hat{G}_T$

In section 3.2 the vector gain based approach to estimate EVM for a modulated signal is described and an overview of the steps that need to be performed in during active load-pull characterization and post-processing of the data is described. In section 3.3 ADS simulations are used test the vector gain based method for EVM estimation using simulation representing the on-wafer characterizations. In section 3.4 the vector gain is measured on-wafer at 5 GHz and 26 GHz for a single stage CMOS amplifier and used to estimate the EVM for a modulated excitation. This result is benchmarked against on-wafer measurements using a active load-pull setup capable of making true modulated wave EVM measurements. In section 3.5 the vector gain based method is demonstrated at D-band where non-linear and modulated-wave characterization methods are challenging. After a discussion on some practical challenges encountered during long duration vector gain measurements in section 3.6 this chapter concludes with a summary, discussion and recommendations in section 3.7.

3.2. Vector gain based approach

As was shown in [43], the AM-AM and AM-PM characteristics of a DUT can be measured from the vector gain during (active) load-pull measurement. The vector gain based approach employed here is an extension to the single tone AM-AM and AM-PM distortion by including multiple measurements across the frequency band of operation. In this approach it is assumed that the measured DUT, i.e. an amplifier at a particular loading condition can be represented by a non-linear complex transfer function that includes a power and frequency dependency, as shown in fig. 3.3. As will become clear, during the approach the power dimension is evaluated individually for each spectral component of the input signal.

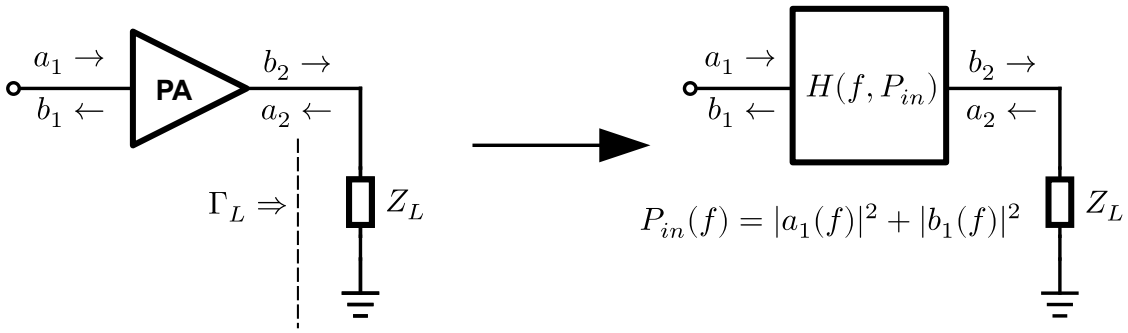


Figure 3.3: The vector gain based approach assumes that the behaviour of the DUT under a specific loading condition can be represented by a complex transfer function that includes a power and frequency dependency.

The vector gain b_2/a_1 of the DUT is measured during an active load-pull characterization for several loading conditions, drive levels and frequencies. It may also be extended to evaluate different bias conditions. For each loading condition that is characterized, the power and frequency dependent complex transfer function is set up. This complex transfer function may then be used in the frequency domain to estimate the (distorted) output spectrum of an modulated signal and could be converted back to time domain to make estimations on metrics such as EVM.

The vector gain based approach differs from being a (behavioural) device model as it only predicts non-linear behaviour at the characterized conditions during the load pull measurements, i.e. the obtained complex transfer functions are only valid under the specific combination of loading condition, frequency and drive level. As the approach is not able to predict behaviour outside what is characterized it cannot be considered a (behavioural) device model.

As explained in the paper by Vanaverbeke et. al. [43] the effects of a source mismatch (when assumed linear) may be added in post processing having the vector gain ratio is defined as:

$$\hat{G}_P = \frac{b_{2,Z_L}}{a_{1,Z_{in}^*}} \tag{3.1}$$

then the approach can be extended to include the influence of the source impedance by including a mismatch factor to obtain the transducer vector gain:

$$\hat{G}_T = \hat{M}_{in} \cdot \hat{G}_P \tag{3.2}$$

with

$$\hat{M}_{in} = \frac{2\sqrt{Z_{S,r}} \cdot \sqrt{Z_{in,r}}}{Z_{in} + Z_S}. \tag{3.3}$$

From eq. (3.2) and eq. (3.3) an optimum may be found that either maximizes the transducer gain G_T or minimizes AM-PM-distortion or a compromise between both [43]. Therefore by taking the component level vector gain metric in the approach the results may later be extended or optimized for a specific application by including the system level source match conditions in post-processing.

3.2.1. Power and frequency dependent complex transfer function

The vector gain transfer function that represents the device behaviour under the characterized operating conditions takes the shape of:

$$H(f, P_{in}) = \frac{b_2}{a_1} = A(f, P_{in}) \cdot e^{j\Phi(f, P_{in})}. \tag{3.4}$$

where $A(f, P_{in})$ is magnitude and $\Phi(f, P_{in})$ the phase of the transfer against input frequency and input power. This transfer function is obtained by multiple single tone Continious Wave (CW) large signal measurements of the DUT, sampling the device response across the frequency band of operation for several power levels. This characterization is illustrated in fig. 3.4.

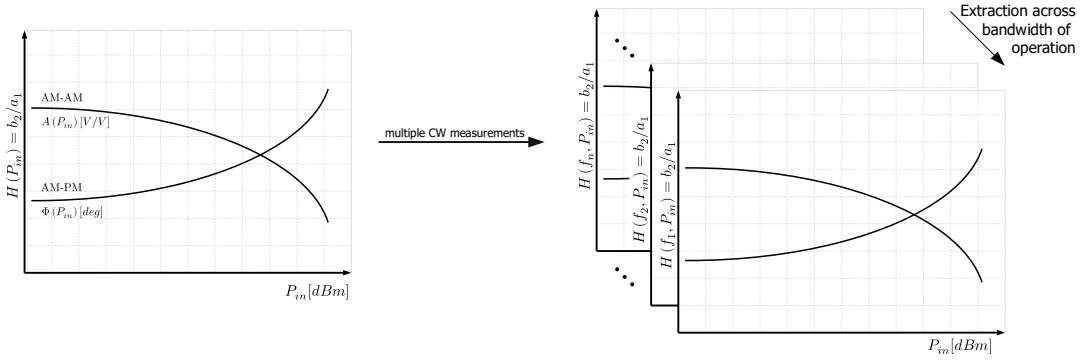


Figure 3.4: To obtain the vector gain transfer function, the For the vector gain based EVM estimation method the vector gain is measured in a large signal (CW or pulsed-RF) measurement at multiple frequencies in the band of interest.

When a fine sampling is used in the frequency and power dimensions a smooth two-dimensional function is obtained describing the device response in the frequency domain at the characterized operating condition (loading & bias). When the obtained two-dimensional function is smooth, interpolation may be used to obtain the transfer of power or frequency points in between what is sampled during the device measurements. Active load-pull setups such as described in section 2.2 automate the large signal measurements, allowing for automated sweeps of drive level, frequency, loading conditions and bias conditions. Preferably a pulsed-RF measurement mode is used to avoid self-heating effects during a measurement if supported by the measurement system[2].

[2]Currently the Vertigo millimeter-wave active load-pull system is not capable to perform pulsed-RF measurements and therefore the vector gain transfers are obtained using a regular CW measurements.

3.2.2. Continuous wave power to spectral power mapping

While the non-linear device response is acquired under CW excitation at different frequencies, the input to output behaviour for modulated signals is computed (numerically) for a multi-tone signal with a frequency grid that accurately captures the in-band spectral content of the modulated signal. For this reason, a relation between the power level in the CW excitation and the power level in each spectral bin of the spectrum in the modulated signal needs to be defined. The need for this mapping becomes clear when considering the power of a bandwidth limited signal:

$$P_{band} = 2\int_{f_1}^{f_2} S_x(f)\, df \tag{3.5}$$

where $f_1 = f_c - BW/2$ and $f_2 = f_c + BW/2$, and f_c the center frequency of the modulated signal, BW the (occupied) bandwidth of the modulated signal and $S_x(f)$ the power spectrum of the modulated signal. Therefore, on average across the band each spectral component has a bin power of:

$$P_{bin} = P_{band}/BW. \tag{3.6}$$

A modulated signal with an band power of for example $P_{band} = -20\ dBm$, in a bandwidth of $BW = 20\ MHz$ has a bin power of less than $-90\ dBm$. This causes two issues, first a (millimeter-wave) active load-pull setup can reach power levels down to around $-40\ dBm$ before becoming noise limited, and second, when characterizing a PA around the -1 dB compression point, a back off of $70\ dB$ will be in the flat region of the gain compression curve. Therefore simply evaluating eq. (3.4) for each spectral component with frequency f and power $P_{in} = S_x(f) \cdot \Delta f$ will show no AM-AM or AM-PM distortion.

The proposed method therefore relates the constant power (CW case) to the average bin power $P_{in} = P_{bin}$ as shown in fig. 3.5. Each spectral bin component is then compared to the bin average power (i.e., P_{avg}/BW) and the relative power offset in the vector gain transfer function is used to apply the effective distortion (AM-AM and AM-PM) to that frequency bin. The motivation behind this particular mapping is that during modulated wave excitation the transistor will experience, on average, a power level equal to P_{avg}, but some spectral components exist with a higher power (i.e. $P_{avg} + \Delta P$) and could therefore be compressed when operating close to the -1 dB compression point.

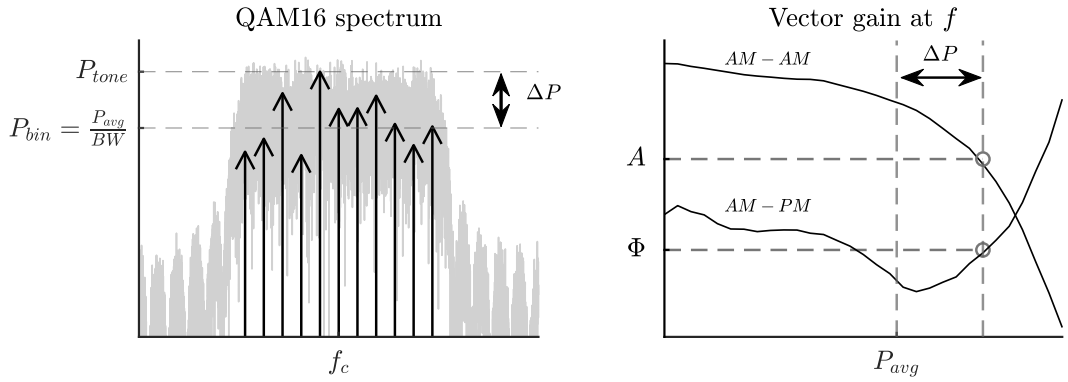
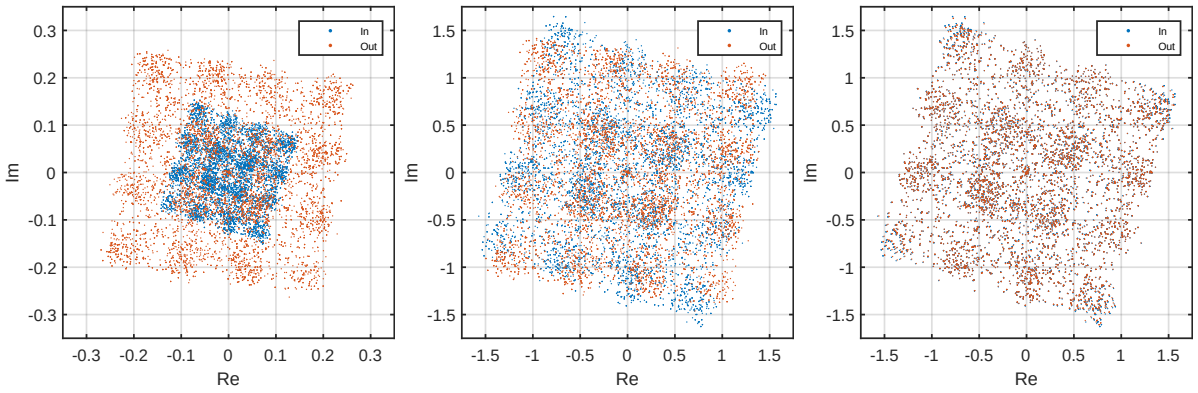


Figure 3.5: Each spectral component of the input signal is scaled in amplitude (A) and phase shifted (Φ) as determined by the measured vector gain at that frequency depending on the offset (ΔP) of the power in the frequency component from the average power in the signal.

3.2.3. Normalization and EVM calculation

After transformation of the modulated signal input spectrum to its output spectrum, the signal is converted back to time domain using the inverse Fourier transform. Before the EVM is calculated the signal is re-normalized to remove the overall gain and phase rotation from the input to the output. Different normalization techniques may result into a different EVM estimations. Here a two step normalization process employed shown in fig. 3.6.



(a) Original before normalization (b) root mean square (RMS) normalization (c) Phase rotation

Figure 3.6: Normalization steps after the estimated output spectrum is obtained. Each dot is a sample of a square root raised cosine (SRRC) QAM-16 filtered signal.

First, the input and output signals are normalized to their RMS value (fig. 3.6b):

$$x_n|_{rms,norm} = \frac{x_n}{\sqrt{\frac{1}{N}\sum_n |x_n|^2}} \tag{3.7}$$

where x_n is the n-th sample, i.e. $x_n = x[n]$. Then, the phase of the output signal is shifted to correspond that of the input signal. To do so the average phase shift among all samples is taken and used to realign the phase of the output signal (fig. 3.6c):

$$\phi_n = arg(x_n) - arg(y_n) \tag{3.8}$$

$$y_n|_{phase,norm} = y_n \cdot exp\left(j\frac{1}{N}\Sigma_n\phi_n\right). \tag{3.9}$$

This phase alignment is only possible when there is no delay from input to output of the system. After normalization the RMS EVM value is calculated as defined in [46] on the a (baseband) sample-by-sample basis before matched filtering:

$$EVM_{rms} = \sqrt{\frac{\frac{1}{N}\sum_n |y_n - x_n|^2}{\frac{1}{N}\sum_n |x_n|^2}} \tag{3.10}$$

where x_n are the RMS normalized samples of the input signal, and y_n the RMS normalized and phase aligned samples of the output signal. The EVM value can be expressed in decibels as:

$$EVM_{rms}|_{dB} = 20log_{10}(EVM_{rms}) \tag{3.11}$$

3.2.4. Overview

Figure 3.7 shows an overview of the steps in the vector gain based approach for EVM estimation. The approach consists of steps performed during on-wafer measurement of the DUT and steps performed during post processing.

The active load-pull measurement is started by setting the first bias, frequency and available input-power. Then the automated active load-pull process converges to the first loading condition and the vector gain is measured. Then it moves on to the next loading condition and repeats. If all loading conditions in the set are measured it moves to the next power level. Similarly after those it moves to the next frequency point. This process may be repeated for different bias levels in an automated fashion.

The results after the measurements are multiple AM-AM and AM-PM curves as was illustrated in fig. 3.4. Next, in the post-processing stage, the spectrum of a modulated signal of interest, i.e. a QAM signal of several symbols, is obtained by applying the Fourier transform on the in-phase (I) and quadrature (Q) symbols. The power spectrum, average power P_{avg} and average bin power P_{bin} are calculated and used to calculate the power offsets ΔP for each frequency component in the input spectrum. Then each frequency component is applied a magnitude amplification and phase shift as according to the mapping in fig. 3.5. The resulting (distorted) output spectrum may be transformed back to the time domain using the inverse Fourier transform. And finally, after normalization to remove the overall gain and phase shift, the EVM may be calculated. These steps, i.e. 6 to 11 in fig. 3.7, may be repeated for modulated signals with different power levels, for each loading condition and bias level that was characterized.

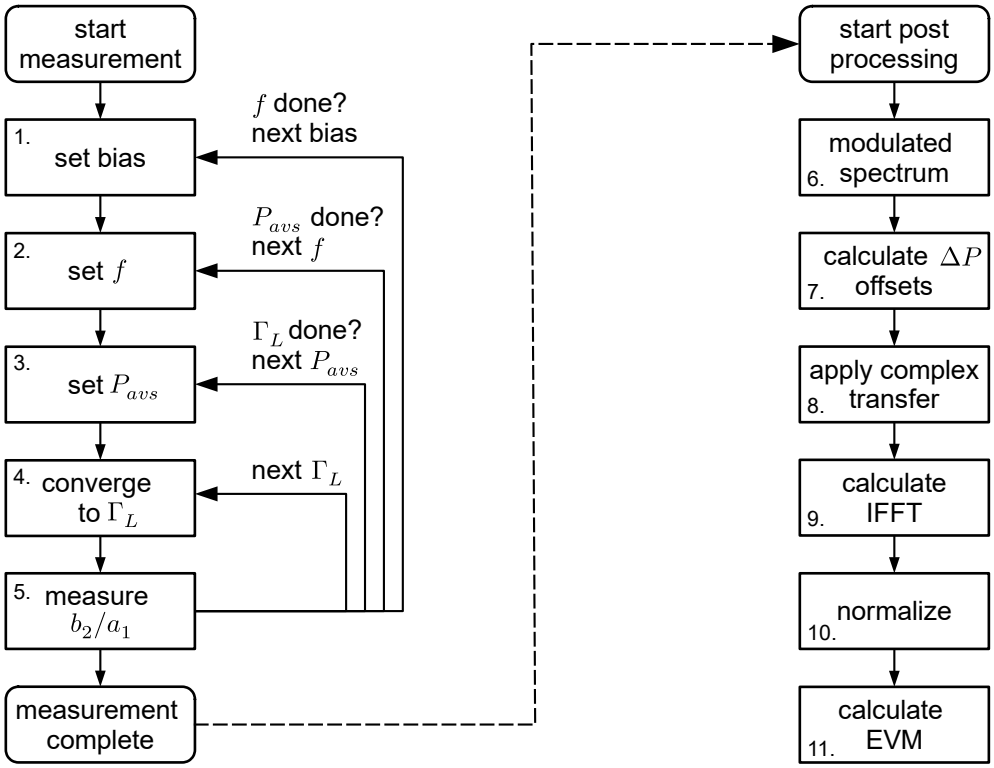


Figure 3.7: Overview of steps in the vector gain based approach for EVM estimation consisting of a measurement part and a post-processing part.

3.2.5. Fundamental limitations

Some fundamental limitations to the vector gain based approach for linearity evaluation exist that should be mentioned. Before all else, the approach does not provide an alternative for compact or behavioural transistor (simulation) models as it only can provide first order predictions of linearity behaviour within the conditions that are characterized during the active load-pull characterization. It should

therefore be considered as a tool to evaluate operating conditions (bias, frequency, drive-level & loading) of a device or circuit during the technology characterization or device/circuit validation phase of the development cycle.

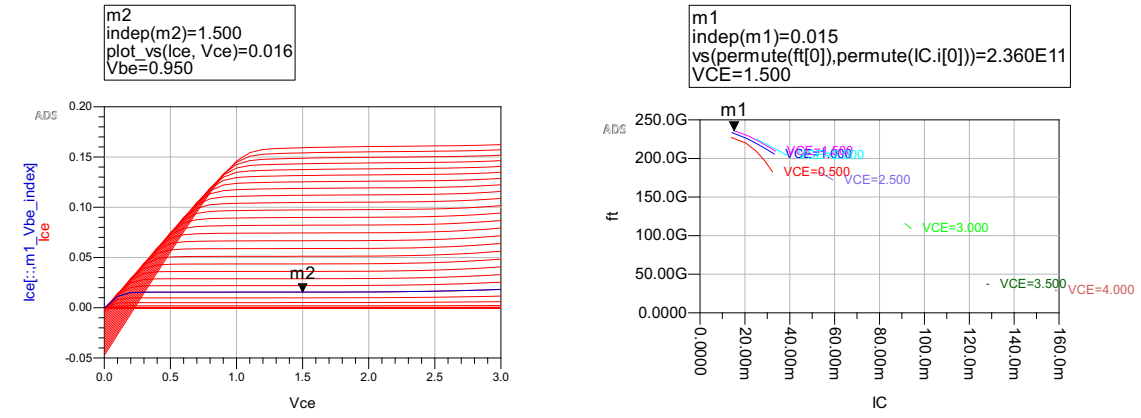
Because of the limitations in millimeter-wave active load-pull systems the vector gain based approach is based on single tone AM-AM and AM-PM responses it therefore does not take intermodulation distortions into account. Even though the input signal is represented as a spectrum of frequencies, only the frequency bin self non-linearities are accounted for, the cross-frequency bin non-linearities, i.e. those that would represent two tone intermodulations, are not accounted for. Therefore, the estimated EVM performance should be considered as a lower bound estimate under the tested operating conditions and modulations. In addition, as the intermodulation products are not considered in the approach, second or third intermodulation (IM2/IM3) estimates or adjacent channel power ratio (ACPR) estimates can not be made using this method.

3.3. Simulation results

The vector gain based approach was tested in Keysight ADS for a single stage amplifier. A SiGe transistor is used in a common emmiter configuration to create the single stage amplifier. A MEXTRAM 504 model of the transistor is used that has an f_T/f_{max} of 236/169 GHz. The EVM estimation capability of the vector gain based approach is benchmarked to the results obtained from an envelope simulation. First, the transistor bias and loading conditions were found for maximum speed and output power. Then a simulation was run representing an (on-wafer) load-pull measurement of the vector gain against frequency drive level and loading condition. Then an envelope simulation is performed to obtain the input and distorted output signals for modulated signals at the same loading conditions and drive levels. Finally, the vector gain from the first simulation is used to make an estimation of the EVM for each of the modulated input signals and loading condition combinations, and the results are compared with the envelope simulation.

3.3.1. Bias and loading conditions

The device was biased for maximum f_T using voltage based biasing. The maximum f_T is found at a collector-emitter voltage $V_{ce} = 1.5V$ and a collector current $I_{ce} = 15mA$ as shown in fig. 3.8b. This collector current is reached when a base-emitter voltage of approximately $V_{be} = 0.95$ is applied as shown in fig. 3.8a.

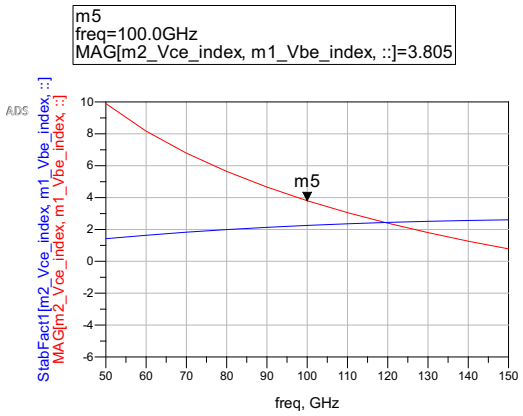


(a) IV curve of the SiGe transistor

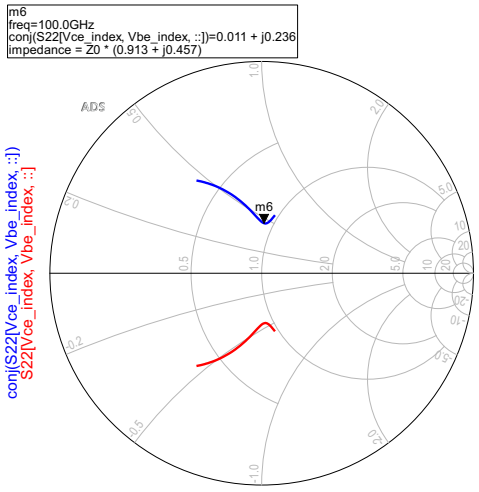
(b) Maximum f_T of the SiGe transistor

Figure 3.8: The SiGe transistor was biased for maximum f_T at $V_{be} = 0.95V$ $V_{ce} = 1.5V$

Figure 3.9a shows the maximum available gain MAG (red) and stability factor (blue) for the device against frequency for these bias conditions. A center frequency of 100 GHz was chosen allowing a maximum achievable gain of more than 3 dB for the single stage amplifier. The conjugate of the output reflection coefficient (S_{22}), shown in blue in fig. 3.9b, was used as a starting point for the load-pull simulation. As the output impedance of the stage is capacitive an inductive load was to be applied to resonate out the reactive component resulting only in the resistive loading of the load.



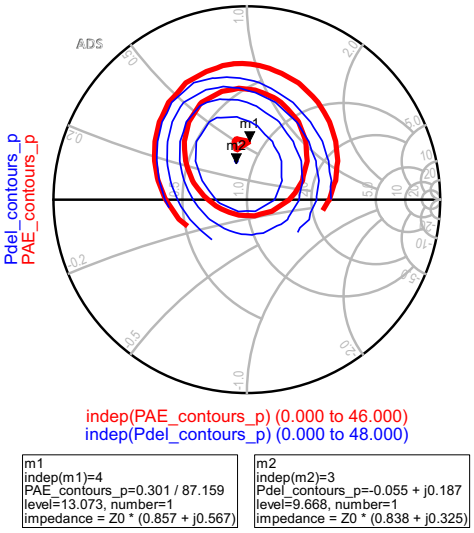
(a) Maximum available gain (red) and stability factor (blue) of the transistor



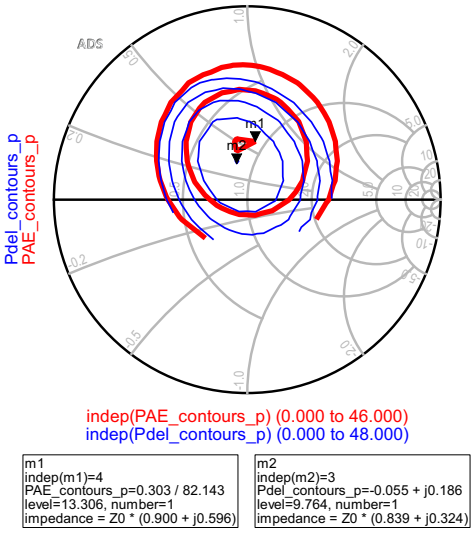
(b) Output impedance (red) and its complex conjugate (blue) of the transistor

Figure 3.9: An operating frequency of 100 GHz was chosen, allowing for a gain of around 3.8 dB. The loading condition is found close to the conjugate of the output impedance.

A load-pull simulation was performed to find the optimal loading condition for maximum output power at the -1 dB compression point. Figure 3.10a shows the contour plots for power added efficiency PAE and output power P_L at -1 dB using 50Ω terminations at the second to fifth harmonic. As the load pull schematic allows for tuning the harmonic loading conditions it was investigated if the second and third harmonic loading provided by wave-guide based probes shows a significant difference in the outcome of the load pull. Wave-guide based probes have an octave wide bandwidth and therefore do not provide a 50Ω loading condition at the harmonics during a large signal or active load-pull measurement. The harmonic input impedance of WR-10 probes were measured previously by C. de Martino and for 200 GHz and 300 GHz are found to be $\Gamma_{L,2f_0} = 0.155 + 0.329j$ and $\Gamma_{L,3f_0} = 0.383 + 0.0806j$. These values



(a) Contours of PAE (thick-red) and P_L at -1 dB compression (blue), 50Ω loading at the second to fifth harmonic



(b) Contours of PAE (thick-red) and P_L at -1 dB compression (blue), with probe-like second and third harmonic loading

Figure 3.10: A load-pull simulation is performed to find the loading condition for maximum output power at the -1 dB compression point. The effects of a probe-like loading on the 2nd and 3rd harmonic do not significantly influence the results

were used in the same load-pull schematic. It was found that the inclusion of these harmonic loading conditions did not significantly influence the output power at the -1 dB compression point, the PAE or the point of optimal loading as is shown in fig. 3.10b.

3.3.2. Vector gain extraction from simulation

The circuit shown in fig. 3.11 shows an ideal representation of a load-pull measurement. The incident and reflected waves are measured by the bi-directional couplers at the input and output port of the device. The vector gain is measured as the ratio between the b_2 over a_1 wave which are measured from the bi-directional couplers. The schematic was simulated using a harmonic balance simulation at the optimal loading condition found in the previous section and five other loading conditions in a radius of r = 0.25 around this optimum.

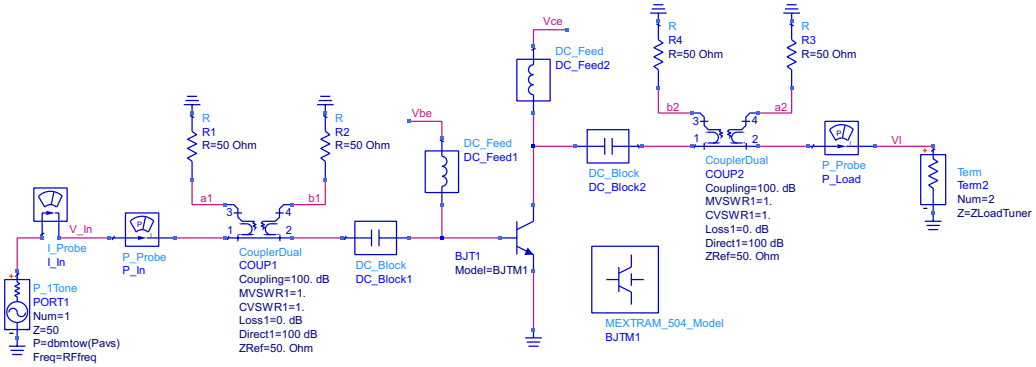
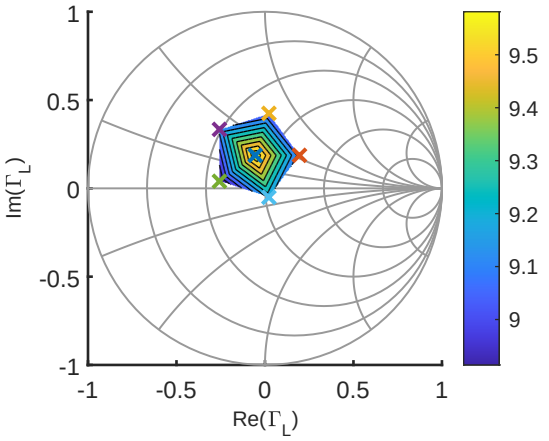


Figure 3.11: Contours of PAE (thick-red) and P_L at -1 dB compression (blue), 50Ω loading at the second to fifth harmonic

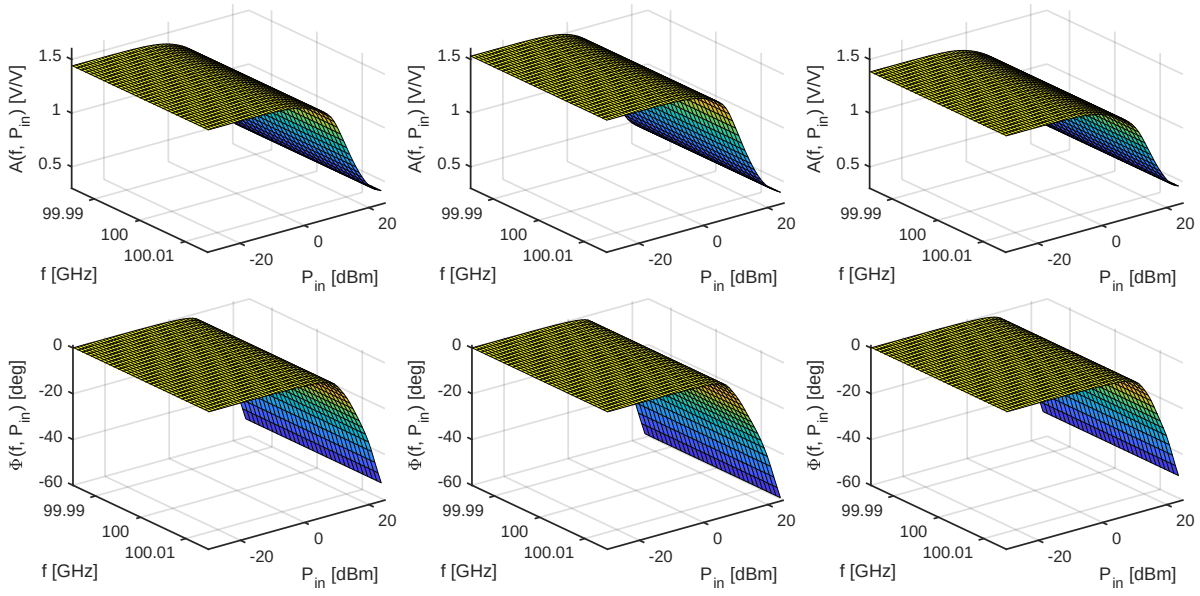
For reference, the gain compression curves and PAE curves for these six loading conditions are shown in fig. 3.12b.



(a) $P_{L,-1dB}$ contours around the optimum loading condition for maximum -1 dB compression point.

(b) Power gain (solid, left axis) and PAE (dot-dash, right axis) for the six loading conditions. Black '+' symbols show the -1 dB compression points.

Figure 3.12: $P_{L,-1dB}$ contours, gain compression and PAE for SiGe common-emitter amplifier. The loading conditions in the Smith chart (left) match by colour with the gain compression & PAE curves (right).



(a) $\Gamma_L = -0.0550 + 0.1870j$ (b) $\Gamma_L = 0.0223 + 0.425j$ (c) $\Gamma_L = -0.257 + 0.040j$

Figure 3.13: Simulated vector gain curves for the SiGe single stage common-emitter amplifier at 100 GHz

For each loading condition the available source power P_{avs} is swept from -30 to 25 dBm (1 dB step size) across a 5 GHz band (100 MHz step size). The obtained two-dimensional vector gain curves from the simulation are shown in fig. 3.13 for three of the six loading conditions as an example of the expected results during an on-wafer measurement. The vector gain curves contain similar information as single tone AM-AM (gain-compression) and AM-PM (phase-variation) plots. They differ only from having a third dimension (frequency) and having the magnitude transfer plotted as an amplitude transfer instead of a power transfer. The AM-AM behaviour of the amplifier can be seen in the magnitude transfer plots (top figures in fig. 3.13) and the AM-PM behaviour can be seen in the phase transfer plots (bottom figures in fig. 3.13). Both magnitude and phase transfers are shown across the characterized band of operation, but in this idealized case they show almost no variation across the band. Only a small decrease in magnitude and a negative phase shift with increase frequency exists (not visible from fig. 3.13). However, as will come clear from the on-wafer measurements in section 3.4 and section 3.5 this is not necessarily the case for a real device.

3.3.3. EVM simulation and comparison

The ADS Verification Test Bench (VTB) [47] was used to generate modulated signals and find the device response using an envelope simulation. Figure 3.14 shows the schematic of the circuit used in this simulation. The complex baseband input and output signals are simulated for the same loading conditions and power levels, and were then exported to MATLAB. The vector gain based approach uses the samples of complex baseband modulated input signal from the test bench and applies the post-processing steps as described in section 3.2 in MATLAB to estimate the distorted output signal. Lastly, the EVM is calculated for both the output signal from the envelope simulation and from the vector gain based approach.

Figure 3.15a shows a comparison between the EVM directly from the envelope simulation and the EVM estimated by the vector gain based approach. The modulated signal that is used is a 16-QAM signal with a symbol rate of $R_s = 1000\ Msyms/s$ and has a total of $N_s = 200\ symbols$. As the EVM among the loading conditions is quite similar, the difference between the simulation and estimation is also shown in fig. 3.15b. The estimation by the vector gain based approach stays quite close to that of the test bench, having an estimated EVM value mostly below that of the test bench up until $P_{in} = 0\ dBm$. As the power in the drive signal is increased, the spectral components having a higher power are given an amplification and phase shift corresponding to a signal far into compression, i.e. the $P_{avg} + \Delta P$ in fig. 3.5 is (much) higher than the -1 dB compression point. These spectral components therefore get amplified less due to the gain compression. This results into a spectral flattening as the

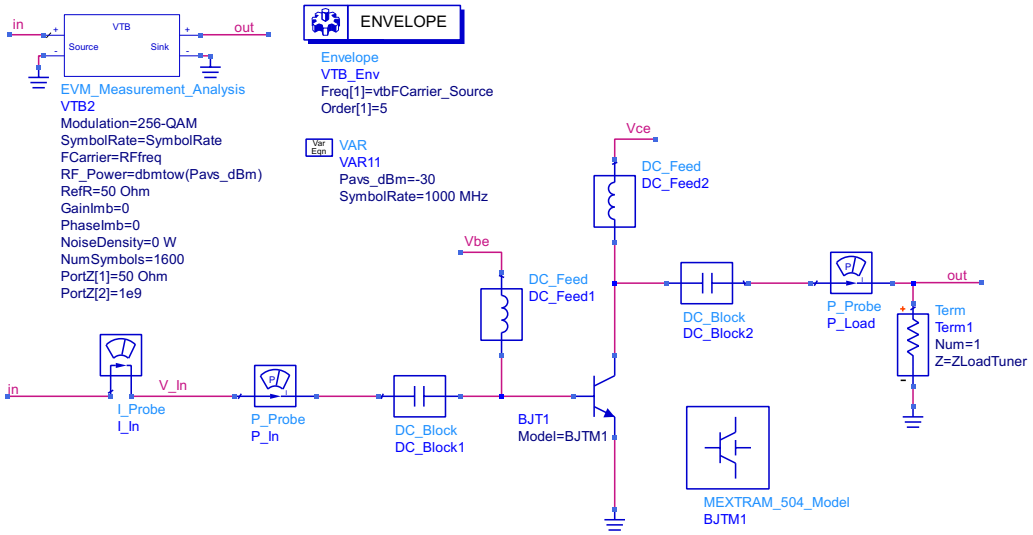
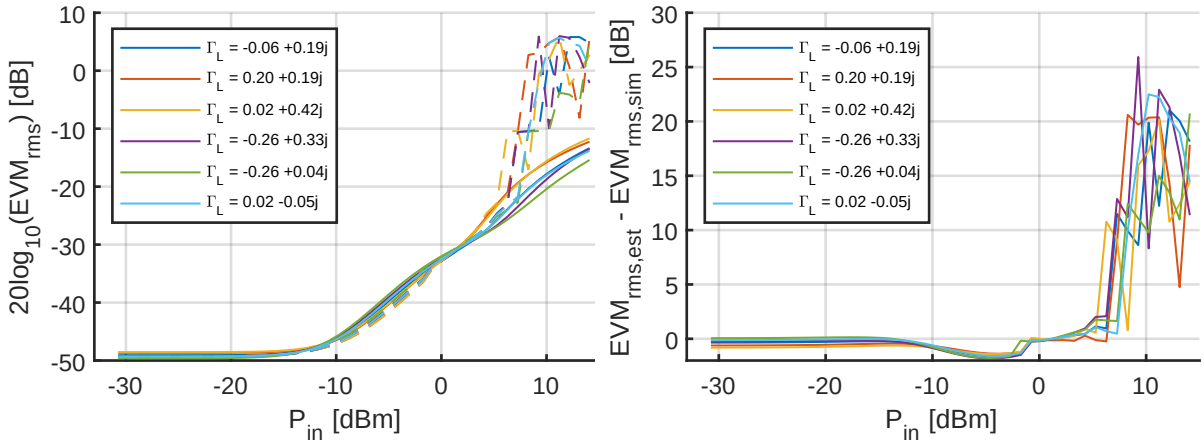


Figure 3.14: Schematic used to simulate modulated signal response using the ADS Verification Test Bench

power increases as is shown in fig. 3.16. This is an intended effect of the approach as each spectral component is treated as its own single tone having an individual AM-AM and AM-PM transfer. However, from the EVM estimates in fig. 3.15 it is clear that this assumption creates a large error close and after the -1 dB compression point in this case.

This error by the approach might limit the usability for modulation schemes that are intended to be able to operate close to the -1 dB compression point, i.e. schemes that do not need a low EVM in order to achieve a low bit-error-rate, mostly low order schemes. However, when a particular EVM target is to be met for a modulation scheme, i.e. and EVM of -30 dB or lower for QAM-16, the vector gain based method could allow for a first order validation to see at which frequencies, bias levels, drive-levels and loading conditions this target can be achieved. In this simulation, this target is possible for all loading conditions up to an input power of $P_{in} = -3\ dBm$.



(a) EVM from the ADS Verification Test Bench (solid lines) and the vector gain based approach (dashed lines)

(b) Difference between EVM from the ADS Verification Test Bench and the vector gain based approach

Figure 3.15: Comparison between EVM from the ADS Verification Test Bench and the vector gain based approach for a 16-QAM signal with a symbol rate of $R_s = 1000\ Msyms$ and a total of $N_s = 200\ symbols$.

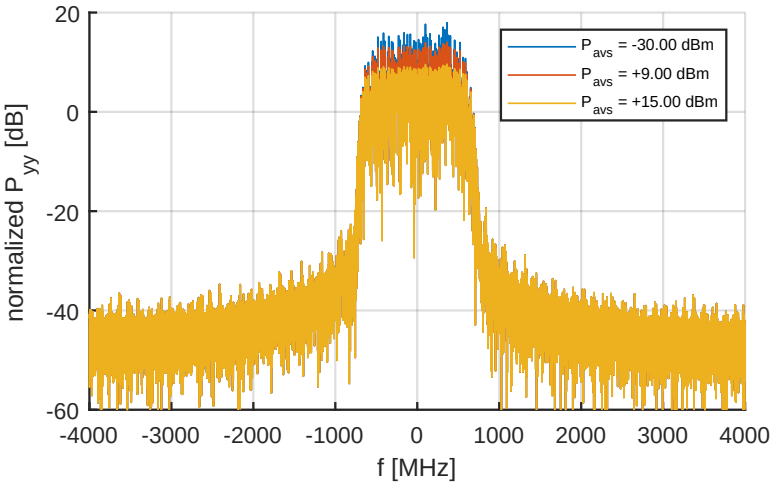


Figure 3.16: Output spectrum flattening estimated by the vector gain based approach when operating towards the -1 dB compression point

3.4. Experimental results of 5 GHz and 26 GHz benchmarking

To benchmark the results of the vector gain based method for EVM estimation experiments were conducted in the 5 GHz and 26 GHz (5G) bands. At these frequencies it is still possible to perform modulated wave active load-pull measurements [29]. The method is benchmarked against the true EVM measurements of the DUT under modulated signal excitation. The Global Foundries 22nm test die is used (see section 2.4) during these experiments. On the die, the metal 1 calibration/de-embedding structures are used with the slvt-nmos2 as the DUT.

3.4.1. Setup and method

To validate the vector gain method, two different test setups were used as shown in fig. 3.17. Using setup 1, the vector gain is measured using CW excitation and post-processed to obtain an EVM estimate using the process as described in section 3.2. The results of the vector gain based EVM estimation method are compared with true EVM measurements from the second setup. Using setup 2, the DUT is excited with a complex modulated signal under the same loading conditions and drive levels as in setup 1. For the 5 GHz and 26 GHz measurements an Anteverta MT-2000 Series Mixed-Signal Active Load Pull system [7] is used with only fundamental loading and excitation to conduct both the CW and modulated wave measurements.

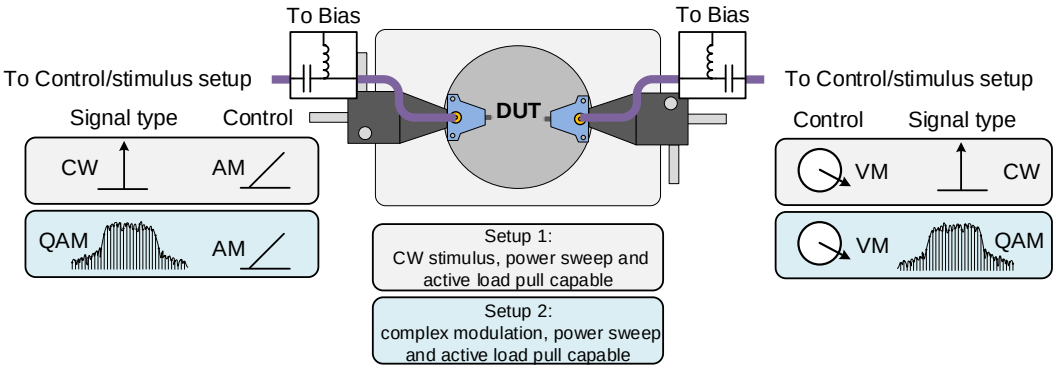


Figure 3.17: Two measurement setups are used for benchmarking of the vector gain based method for EVM estimation at 5 GHz and 26 GHz. A single tone load-pull measurement is used in setup 1 to measure the vector gain of the DUT and estimate EVM performance in post-processing. This result is benchmarked against a measurement using setup 2 that uses a complex modulated wave excitation.

Table 3.1: Overview of measurement parameters during 5 GHz and 26 GHz vector gain benchmarking tests

Parameter			Setup 1		Setup 2		Unit
			5 GHz	26 GHz	5 GHz	26 GHz	
Device Under Test							
Foundry			Global Foundries		Global Foundries		
Technology			CMOS FD-SOI		CMOS FD-SOI		
Device			slvt-nmos2		slvt-nmos2		
Wafer #			a44de872 [1)]		a44de872 [1)]		
Die #			4-3 [2)]		4-3 [2)]		
Bias / Supply							
Gate bias		V_{gs}	0.65		0.65		V
Drain bias		V_{ds}	0.8		0.8		V
Frequency							
Points		N_{freq}	21	25	n.a.	n.a.	
Sweep	start	f_1	4.99	25.985	5	26	GHz
	Δ	Δf	1	1	n.a.	n.a.	MHz
	stop	f_2	5.01	26.015	n.a.	n.a.	GHz
Bandwidth		BW	n.a.	n.a.	20	20	MHz
Drive level							
Points		$N_{P_{avs}}$	31	26	36	31	
Sweep	start	$P_{avs,1}$	-30	-25	-30	-20	dBm
	Δ	ΔP_{avs}	1	1	1	1	dB
	stop	$P_{avs,2}$	0	0	5	10	dBm
Max compr.			4	2	4	1	dB
Loading condition							
Strategy			load-pull for max P_{L-1dB}				
Method			RT	CW	MW	MW	
Points		N_{Γ_L}	6	16	6	16	
Sweep	type		circular	radial	circular	radial	
	center	Γ_L	0.5 + 0.0j	0.4 - 0.1j	0.5 + 0.0j	0.4 - 0.1j	
Modulated signal							
Modulation			n.a.	n.a.	QAM16	QAM16	
Roll-off		α	n.a.	n.a.	0.25	0.25	
Filtering			n.a.	n.a.	SRRC	SRRC	
Calibration							
1st-tier			LRM-ideal	LRM-ideal	LRM-ideal	LRM-ideal	
2nd-tier			M1 LRM-ideal	M1 LRM-ideal	M1 LRM-ideal	M1 LRM-ideal	

n.a. - Not applicable

[1)] - Wafer send to Anteverta was not given an ID

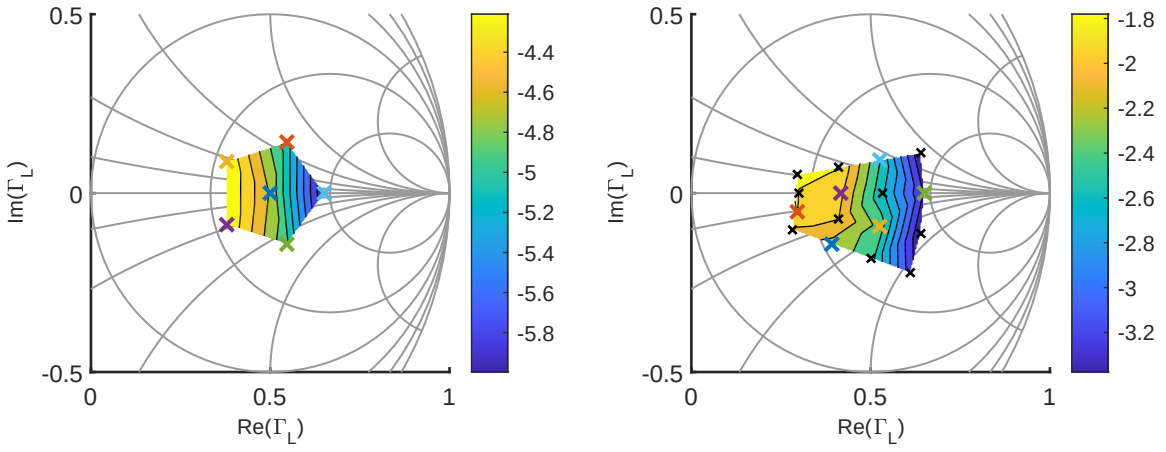
[2)] - Not tracked during characterization

Table 3.1 shows the parameters that were used during the vector gain benchmarking measurements. The transistor was biased for its highest f_T at $V_{gs} = 0.65V$ and $V_{ds} = 0.8V$, previously found during a small signal characterization. Using setup 1, a CW load-pull measurement was performed to find the maximum output power P_L at the -1 dB compression point at the center frequency of the band. During this load-pull an initial constellation at the conjugate of the S_{22} of the device was moved around manually until the optimum loading condition was captured within its center. Then a constellation was chosen around this point to capture a diverse range of gain compression shapes of the device to test the vector gain approach.

For both the 5 GHz and 26 GHz measurements the same calibration approach was used. A first tier LRM calibration was performed using the 3.5 mm Rosenberger calibration kit. A power calibration was performed on the source port (port-1) using the Keysight U8488A power meter. Power levelling is performed using the loads of the Rosenberger calibration kit. The second tier calibration was performed on-wafer on the same die as the DUT using the metal-1 LRM calibration kit[3]. The standards on this second tier calibration kit were assumed to be ideal. The thru may be considered to have zero-length as the distance from the probe pads to the device intrinsic reference planes is equal to the that from the probe pads to the center of the thru.

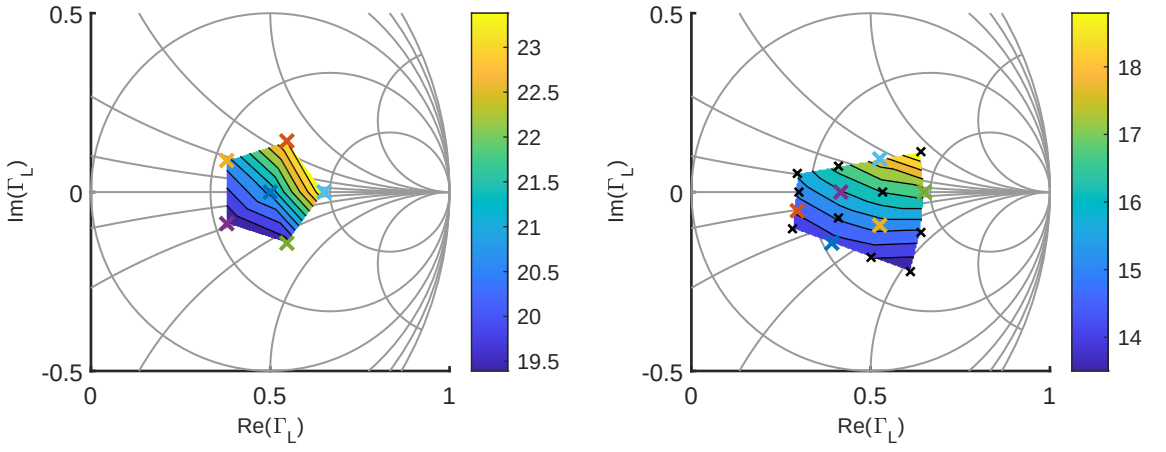
3.4.2. Loading conditions and characterization range

Figure 3.18 shows the contour plots for power gain G_P and output power P_L at the -1 dB compression point at 5 GHz and 26 GHz. In both cases, the optimal loading condition for maximum output power was found close to the real axis. The optimal loading condition is found close to the real axis because the output capacitance C_{out} that normally consists of layout parasitics and the C_{ds} of the DUT is small. The C_{ds} of the device is expected to be low because of the device high speed (f_T = 280 GHz) and small dimensions (W/L = $500/20\ nm$). In addition, the capacitively loaded inverted CPW de-embedding technique [5] employed in the calibration / de-embedding kit brings the reference planes as close as



(a) Loading conditions and P_L at the -1 dB compression point for the benchmark at 5 GHz

(b) Loading conditions and P_L at the -1 dB compression point for the benchmark at 26 GHz



(c) Loading conditions and G_p at the -1 dB compression point for the benchmark at 5 GHz

(d) Loading conditions and G_p at the -1 dB compression point for the benchmark at 26 GHz

Figure 3.18: Contour plots of the output power P_L and power gain G_P at -1 dB compression for the benchmarks at 5 GHz and 26 GHz. The coloured loading conditions correspond to those shown in the EVM comparison plots (fig. 3.22 and fig. 3.24)

[3]A LRM calibration can be performed on the Global Foundries test die by using the matched and short structures of the SOLT kit, and the thru of the TRL kit

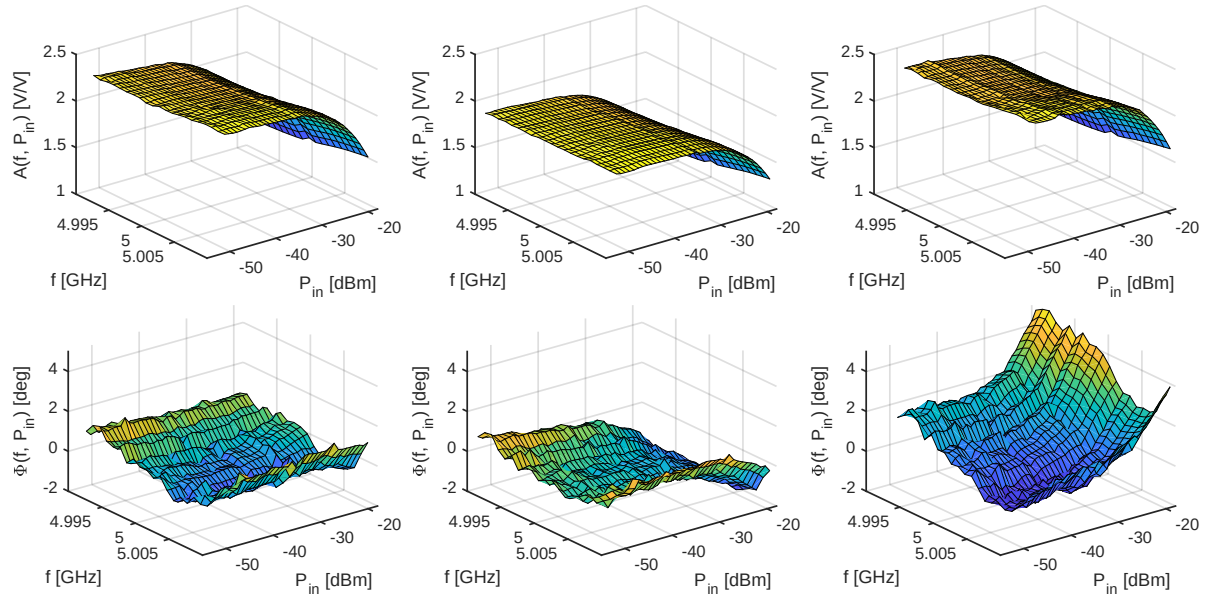
possible to the DUT (i.e. to the M1 device fixtures) while maintaining a high cut-off frequency allowing for accurate de-embedding up-to several hundreds of gigahertz. Therefore, at these low frequencies the output capacitance can be considered an open and the output impedance of the device is almost completely resistive. When operating at higher frequencies, as shown in section 3.5, it is necessary to resonate out C_{out} to achieve a positive gain.

Ideally, when using the vector gain based method for linearity evaluation of a device to be used as the last stage of a PA one would measure (closely) around the optimum loading condition. If that approach was taken here then roughly the same value for P_L at the -1 dB point would be seen for the two frequencies (and others) in fig. 3.18 as maximum output power under the same bias conditions at optimal loading is then only related to device size and not frequency, given a high enough drive level is used. In this set of experiments the constellations were moved somewhat away from the center of the Smith chart (i.e. away from $\Gamma_L = 0$) to ensure good convergence and stability during the measurements, and provide a diverse set of device behaviours to benchmark the approach.

3.4.3. Vector gain curves at 5 GHz

At 5 GHz the vector gain was measured using the Real Time (RT) operation mode of the MT2000 system. At this frequency the vector gain was sampled across a 20 MHz band using 1 MHz spacing with available source power levels from -30 dBm to 0 dBm with 1 dB spacing. The measurement was terminated once 4 dB of compression in the transducer gain G_T was measured or when the end of the sweep was reached. This was done to ensure that the effects of operating (far) into compression could be captured by the vector gain estimation approach.

The plots in fig. 3.19 show the measured vector gain versus frequency and drive level for three of the six loading conditions characterized at 5 GHz. Different AM-AM and AM-PM behaviour can be observed from the amplitude transfer (top plots) and the phase transfer (bottom plots) of the vector gain. The vector gain curves contain similar information as typical single tone AM-AM and AM-PM plots, they differ only from having a third dimension (frequency) and having the magnitude transfer plotted as an amplitude transfer instead of a power transfer. The phase in the vector gain transfer is normalized to the phase at the centre frequency for the lowest drive level, thereby removing the overall $180°$ that is seen from input to output.



(a) $\Gamma_L = 0.500$ (b) $\Gamma_L = 0.379 + 0.088j$ (c) $\Gamma_L = 0.546 - 0.143j$

Figure 3.19: Two dimensional vector gain transfers measured at three different loading conditions at 5 GHz of the slvt NMOS2 transistor on the Global Foundries 22nm FD-SOI test die.

Although the goal of this set of measurements is to benchmark the approach of vector gain based EVM estimation interesting behaviour can be seen in the vector gain of the device in fig. 3.19. Starting with the AM-AM, different amounts of compression are observed within the characterized power range depending on the maximum gain and compression point of the loading condition. For example, the loading condition in fig. 3.19c has a higher gain and lower compression point (due to the higher impedance value) than in fig. 3.19b and therefore a larger gain compression and phase variation is observed within the characterized range. There is little variation in the AM-AM across frequency and no explicit difference in the shape (i.e. hard or soft compression).

The AM-PM behaviour shows larger differences across loading conditions. As the CMOS transistor is biased in class-A a negative slope for the AM-PM is expected due to the non-linear nature of the C_{gs} of the transistor that is increasing with input swing [4]. However, in this case the output loading is deliberately not perfectly tuned, i.e. the small C_{ds} is not perfectly resonated out and some loading conditions are capacitive while some are inductive. This causes the slope of the AM-PM to flip depending on whether the loading condition is capacitive or inductive. For loading impedances with a positive imaginary part the phase variation is negative while for loading impedances with a negative imaginary part the phase variation is positive. When the loading condition is purely resistive as in fig. 3.19a $\Gamma_L = 0.5$ but also $\Gamma_L = 0.65$ (not shown in fig. 3.19) almost no phase variation with drive level is observed up until the maximum characterized drive level.

3.4.4. EVM estimation and comparison at 5 GHz

Following the post processing steps described in section 3.2 the vector gain curves are used to estimate the output spectrum of the modulated waveform for the different drive levels and loading conditions simulated. The output spectrum for the loading condition $\Gamma_L = 0.5$ is shown in fig. 3.20. As seen in the simulated cases of section 3.3 the output spectrum flattens when the input drive level is increased.

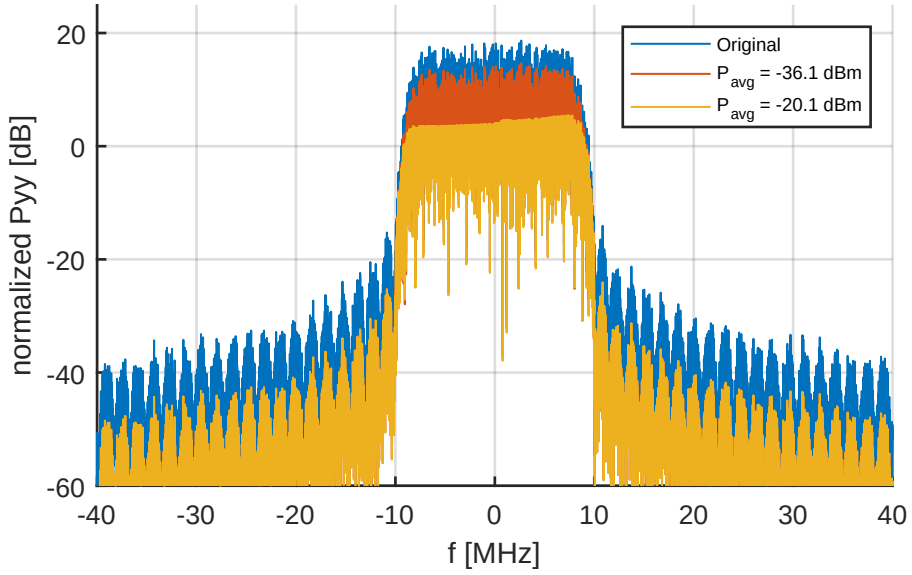


Figure 3.20: Output spectrum estimated by the vector gain based approach at 5 GHz for $\Gamma_L = 0.5$.

Figure 3.21 shows the IQ samples after transforming the distorted output spectra to time-domain and normalizing the result to remove the effects of the overall gain and phase shift from input to output. To show the error made because of the distortion of the transistor the difference between input and output I/Q samples is also shown in the figure. Here a similar increase in error can be seen when operating closer to the compression point as error in the IQ samples increases when the drive level is increased.

[4] A good example of the AM-PM behaviour changes for class-A to deep class-AB can be found in the paper by Kulkarni and Reynaert [48, fig. 5]

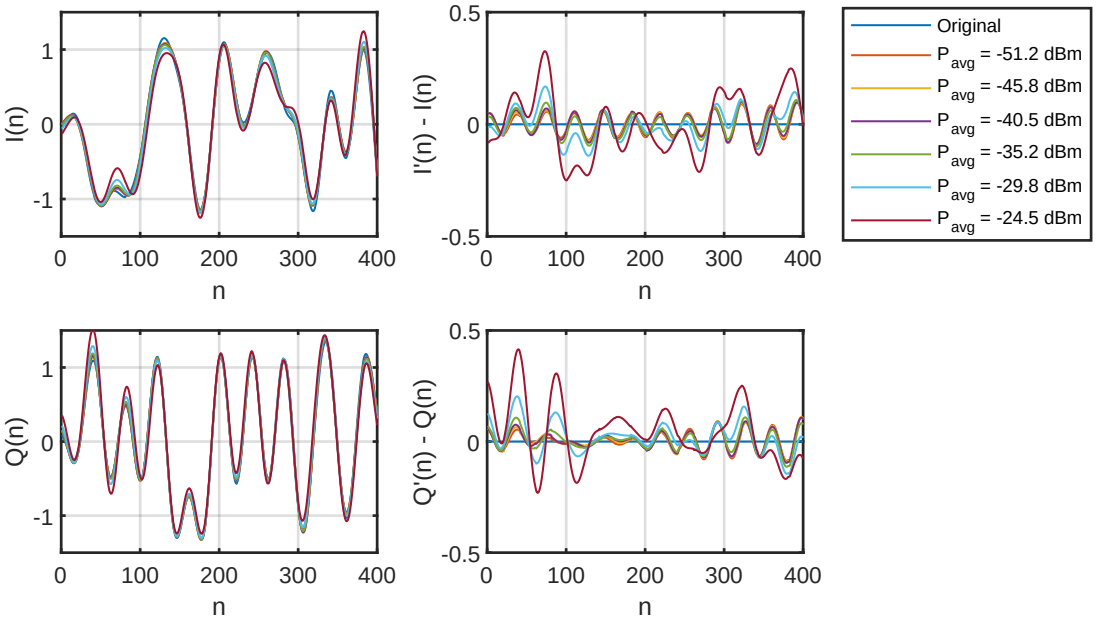


Figure 3.21: Output I and Q samples estimated by the vector gain based approach (left) and the difference with respect to the input I' and Q' samples (right) for increasing drive levels.

The estimated EVM using the vector gain based method and true measured EVM using setup 2 are shown in fig. 3.22 for matching drive levels and loading conditions. The dotted lines are the estimations of the EVM for this particular modulation scheme using the extracted vector gain data at each loading condition. The solid lines are the true measured EVM for the same modulation and loading conditions. The input -1 dB compression points for these loading conditions (from combination of fig. 3.18a and fig. 3.18c) lie around $P_{in,-1dB} = -26.5\ dBm$. As can be seen in fig. 3.22 the vector gain estimations are able to give predictions beyond the compression point without showing significant errors as opposed to what was seen in the simulations of section 3.3.

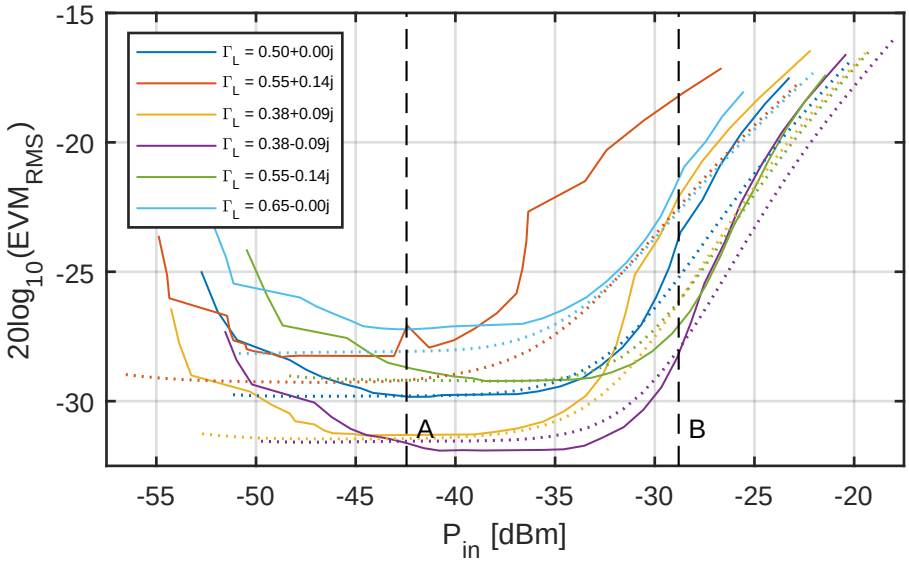


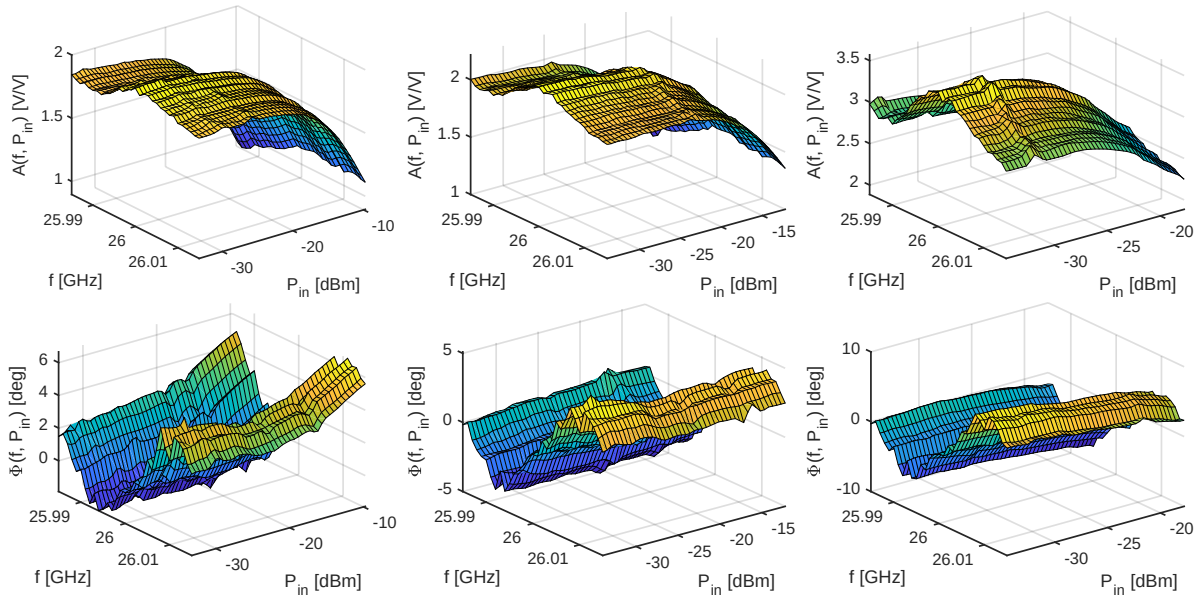
Figure 3.22: Correlation between estimated and measured EVM for a QAM16 signal at six loading conditions at 5 GHz. Solid lines are true modulated wave EVM measurements, dashed lines are estimations of the EVM using the vector gain.

There are three important observations to make from this graph. The first observation is that the estimated EVM is primarily lower than the measured EVM, which is as expected because the vector gain based approach does not take into account all the non-linearities. The second observation is that across a large range in drive level (indicated as the region between the lines A and B) the vector gain based estimation approach stays within 1 dB of the true value for all loading conditions with the exception of $\Gamma_L = 0.55 + 0.14j$ which is considered an outlier of the true EVM measurement performed with setup 2. Below point A the noise in the second measurement setup is limiting and the EVM starts to increase. The third observation is that although the vector gain based estimation method does not give an one-on-one absolute estimate of the EVM it is able to predict the best condition for lowest EVM among the set of loading conditions characterized. In this case the purple curve, of the loading condition $\Gamma_L = 0.38 - 0.09j$, gives the lowest absolute EVM which is confirmed by the true EVM measurement. In addition, the method is able to predict for each of the loading conditions at which drive level the EVM starts to increase if this modulation was used.

3.4.5. Vector gain curves at 26 GHz

The same benchmarking experiment was repeated at 26 GHz. At this frequency he vector gain was measured in CW mode as it was found that the device would quickly break down in RT or pulsed-RF modes. The vector gain was measured across a 25 MHz band using 1 MHz spacing with a source power drive levels from -25 dBm to 0 dBm with 1 dB spacing. A wider range of frequency points was used in the 26 GHz measurement as it was found after the 5 GHz measurement to be helpful in post-processing of the results and removing outliers (see section 3.6). Finally, the 26 GHz measurement was terminated once 2 dB of compression in the G_T was measured to further reduce the risk of breaking the device.

The plots in fig. 3.23 show the measured two dimensional vector gain against frequency and drive level for three of the sixteen loading conditions characterized at 26 GHz. At this frequency a larger variation in the amplitude transfer and phase transfer versus frequency was observed than at 5 GHz. This behaviour was observed for multiple devices on different dies. Therefore this effect is either device intrinsic or due to the calibration.



(a) $\Gamma_L = 0.282 - 0.103j$ (b) $\Gamma_L = 0.417$ (c) $\Gamma_L = 0.640 + 0.113j$

Figure 3.23: Two dimensional vector gain transfers measured at three different loading conditions at 26 GHz of the slvt NMOS2 transistor on the Global Foundries 22nm FD-SOI test die.

3.4.6. EVM estimation and comparison at 26 GHz

To benchmark the vector gain based method the estimated EVM and true measured EVM are compared for a QAM-16 modulated waveform. This comparison is shown in fig. 3.24. Here the results were less conclusive because of several reasons. As mentioned, the DUT would easily break down when real-time or pulsed-RF modes were used to extract the vector gain of the devices. In order to reduce the risk of more devices breaking down during the EVM measurements these measurements were terminated once 1 dB of compression in G_T was measured. The dynamic range in the EVM measurements was therefore constrained in the lower bound limited by the noise in the system (as seen in the 5 GHz measurement) and the upper bound limited by how far the devices could be pushed into compression. The results in fig. 3.24 therefore also do not show a clear EVM-'floor' being reached during the measurements as the EVM starts to increase due to the low signal to noise ratio at the lower power levels. In addition, two of the 16 loading conditions did not converge during the EVM measurement which is showing that the system had trouble operating in these ranges. This could also explain why the measured EVM curves show jagged edges and not smooth lines.

However, a more prominent reason why the results were inconclusive is that the EVM is estimated consistently higher with about 3 dB than the measured EVM for all loading conditions as can be seen by the dotted lines in fig. 3.24. This is an incorrect result as the vector gain based method is a simplified first order approach that does not take intermodulations into account. It therefore should give a lower bound to the EVM and should not be higher than the true measured EVM. The vector gain based approach is still able to estimate the best-to-worst loading condition for lowest EVM among the characterized set which is confirmed by the true EVM measurements. Albeit the difference being small in percentages (i.e. an EVM of -30 dB equals a 3.2% error while an EVM of -33 dB equals a 2.2% error) the method is clearly not able to give correct absolute estimations of the EVM in all cases.

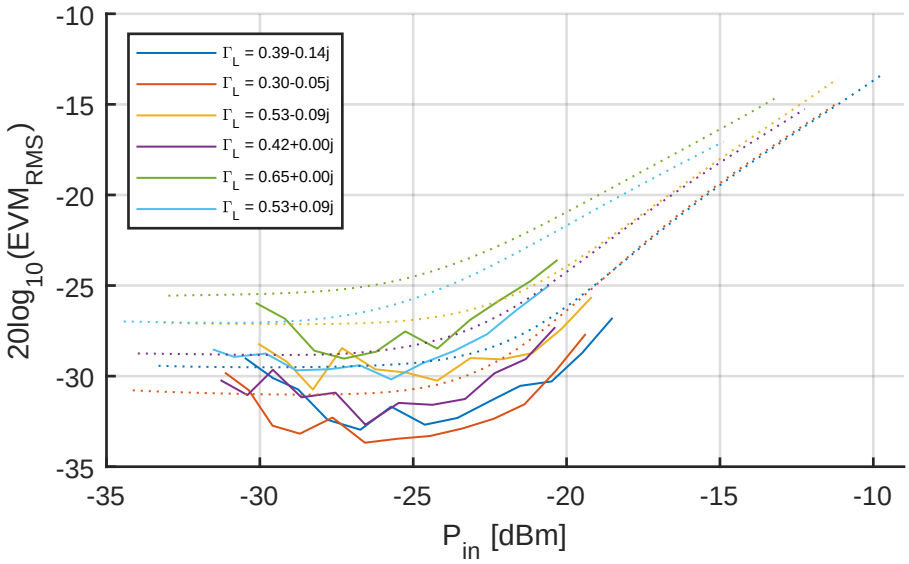
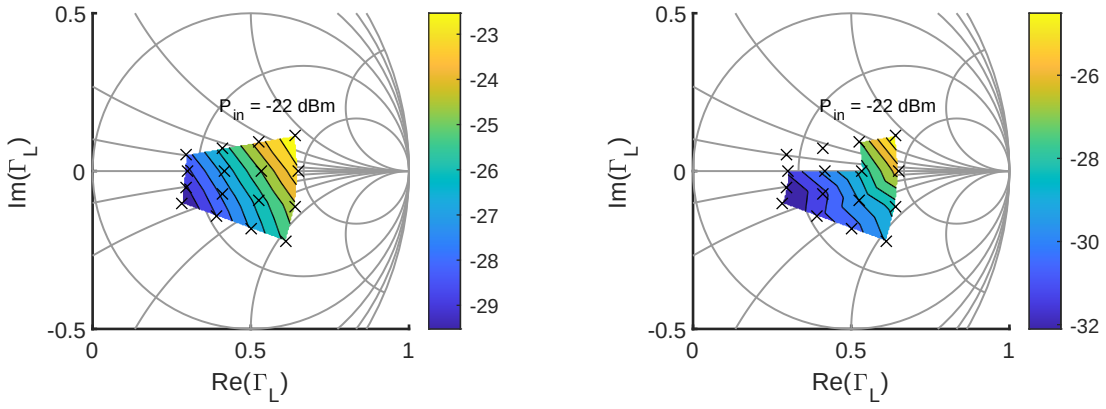


Figure 3.24: The vector gain based EVM estimation at 26 GHz shows the correct trends against loading conditions however is consistently higher than the true measured EVM

After a discussion with the designers of the Anteverta system the most likely cause of the difference was narrowed down to a difference in normalization used between the vector gain based EVM estimation method and the EVM of the measurements. During the 5 GHz experimental benchmarking both the estimation and measurement used a RMS based approach to normalize the signals after amplification by the PA. In the time between the 5 GHz and 26 GHz benchmark the Anteverta system changed to a more sophisticated approach for re-normalization as described in [49]. Implementation of this approach for the vector gain based estimation in order to see if a one-on-one estimation could be achieved was considered beyond the scope of this work. Therefore, no claims are made on the ability of the method to estimate absolute EVM performance as this could be influenced by the normalization method employed and would therefore always end up into an exercise of aligning the approach to a specific system

Instead, a focus is put on the methods prediction capabilities across the different characterized operating conditions. An example is shown in fig. 3.25 for the input drive level of $P_{in} = -22\ dBm$. Although the absolute prediction of the EVM is not correct, the estimation and measurement show a good correspondence in the prediction across the tested loading conditions.



(a) Estimated EVM at fixed drive level across loading conditions.

(b) Measured EVM at fixed drive level across loading conditions.

Figure 3.25: Comparison between estimated and measured EVM for a QAM16 signal at six loading conditions at 26 GHz. Solid lines are true modulated wave EVM measurements, dashed lines are estimations of the EVM using the vector gain.

3.5. Experimental results of D-band demonstration

Although the vector gain based method for EVM estimation could be used for frequencies below 30 GHz as was done during the benchmarking experiments. It is likely that existing transistor compact or behavioural models such as BSIM6 [50] or non-linear characterization techniques such as X-parameters are already better able to capture and predict non-linear device behaviour than then vector-gain based method at these frequencies. In addition, the current state-of-the-art active load-pull systems are able to perform modulated wave measurements up to 40 GHz [29]. The real use-case for the vector gain based method is estimating linearity performance far into the millimeter-wave regime where device behaviour is less well captured in models and non-linear characterization methods are challenging due to the limitations of the characterization equipment. To demonstrate the approach, the vector gain measurements were repeated at 165 GHz and used to estimate modulated wave performance for different loading conditions to see if an optimal loading condition for EVM performance could be identified using the approach.

3.5.1. Setup and method

To extract the vector gain behaviour of the DUT at 165 GHz a Vertigo millimeter-wave active load pull setup was used with WR-5.1 waveguide VDI VNA extension modules [51]. A center frequency of 165 GHz was chosen as at that frequency the extender module has the most power available within the band. Figure 3.26 shows the setup mounted on the Suss semi-automatic probe-station.

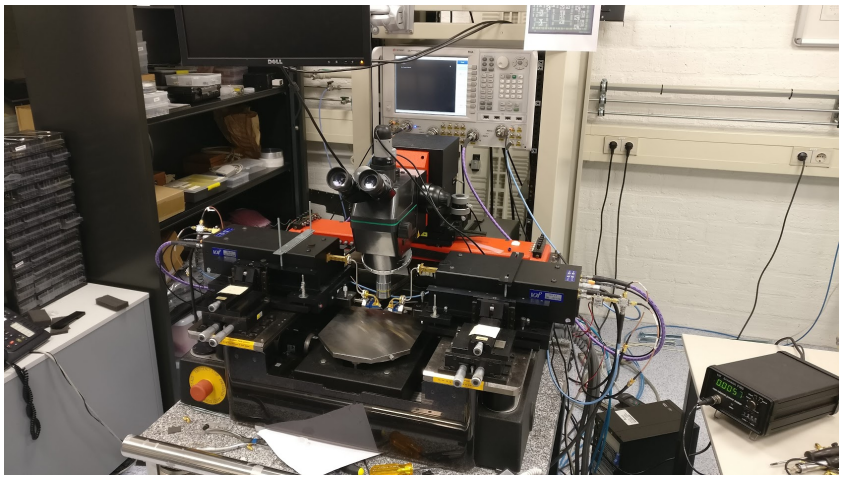


Figure 3.26: Vertigo millimeter-wave active load pull setup used for 165 GHz vector gain measurements

The first tier calibration was performed using the TRL VDI WR5.1 calibration kit. A short was used as the reflect and it is assumed that all standards in the kit are ideal, leaving only the length of the line ($1.07\ mm$) as the to be specified parameter in the calibration. A power calibration is performed on the source port (port-1) using the VDI PM5 calorimeter-style power meter [52]. Power levelling was performed using the matched loads of the calibration kit. The second tier calibration was performed on-wafer on the same die as the DUT using the metal-1 TRL calibration kit. The characteristic impedance (50Ω), the length of the thru ($0\ mm$) and the length of the line ($0.54102\ mm$) are specified using values previously found from electro magnetic (EM) simulation of the on-wafer calibration kit, the reflects (shorts) are assumed ideal. A zero-length for the thru may be used as the distance from the pads to the device intrinsic reference planes is equal to the that from the pads to the center of the thru.

The same measurement and post-processing procedure as in the benchmarking measurements is employed, as described in section 3.2. As now there is no validation case, the measurement setup and signal types are similar to that of setup 1 in fig. 3.17. Using CW excitation the vector gain is measured at several frequencies in the operating band of interest at several power levels. The bias levels and loading conditions could be varied to evaluate different operating conditions for the DUT however in these experiments only the loading condition is swept. Table 3.2 shows all parameters during the measurement.

Table 3.2: Overview of measurement parameters during 165 GHz vector gain measurements

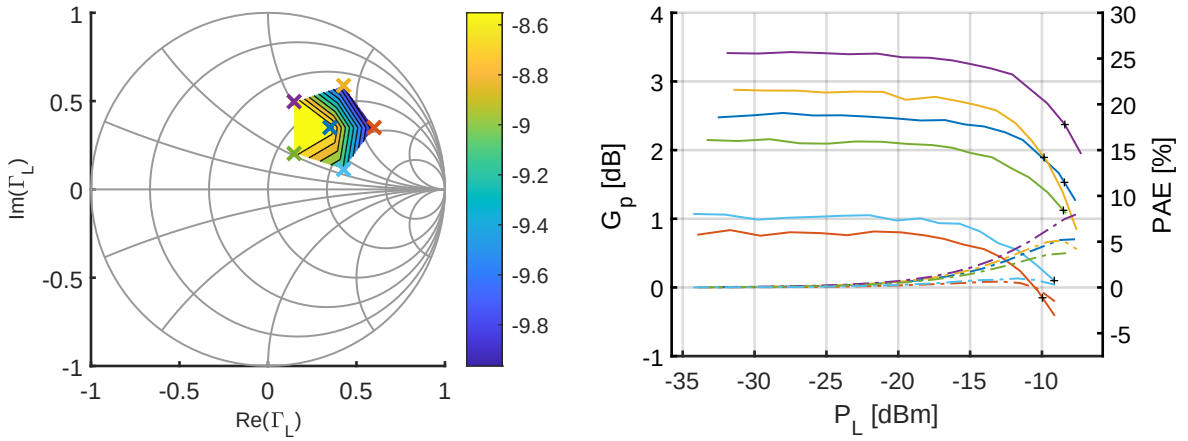
Parameter			Setup 1	Unit
			165 GHz	
Device Under Test				
Foundry			Global Foundries	
Technology			CMOS FD-SOI	
Device			slvt-nmos2	
Wafer #			a44de872	
Die #			5-1	
Bias / Supply				
Gate bias		V_{gs}	0.45 [1)]	V
Drain bias		V_{ds}	0.46 [1)]	V
Frequency				
Points		N_{freq}	11	
Sweep	start	f_1	164.99	GHz
	Δ	Δf	2	MHz
	stop	f_2	165.01	GHz
Drive level				
Points		$N_{P_{avs}}$	16	
Sweep	start	$P_{avs,1}$	-37.1	dBm
	Δ	ΔP_{avs}	2	dB
	stop	$P_{avs,2}$	-7.1	dBm
Max compr.			n.a.	dB
Loading condition				
Strategy			load-pull for max P_{L-1dB}	
Method			CW	
Points		N_{Γ_L}	6	
Sweep	type		circular	
	center	Γ_L	0.35 + 0.35j	
Calibration				
1st-tier			TRL-ideal	
2nd-tier			M1 TRL	

n.a. - Not applicable

[1)] - Bias was reduced in order to be able to compress the device

3.5.2. Loading conditions and characterization range

At these frequencies the output power of the VNA extender modules limits how far can be measured into compression. The available power at this frequency was not high enough to put the DUT into compression at the same bias conditions as the 5 GHz and 26 GHz benchmarks. The bias conditions were reduced to $V_{gs} = 0.45\ V$ and $V_{ds} = 0.46\ V$, ensuring that the device is quickly pushed into compression as the output swing on the drain gets high enough to push the device from saturation into the linear region ($V_{ds} < V_{gs} - V_{th}$). The vector gain was extracted at six loading conditions around the optimal loading condition for maximum output power as shown in fig. 3.27a. The optimal loading condition is now found in the inductive upper region of the Smith chart, as the capacitive reactance is now significant due to the higher frequency of operation. For reference, the gain compression curves and PAE at these loading conditions are show in fig. 3.27b. At these bias levels a compression of 1 dB could just be reached, further reduction of the bias levels would result into significantly lowering the f_T/f_{max} resulting into no gain at this frequency. The vector gain is measured at 11 tones (2 MHz spacing) across a 20 MHz operating band using 16 source power levels from -37.1 to -7.1 dBm (2 dB spacing) at six loading conditions. The number of power and frequency points during these measurements had to be reduced as during the long duration measurements probe contact losses would occur (see section 3.6).



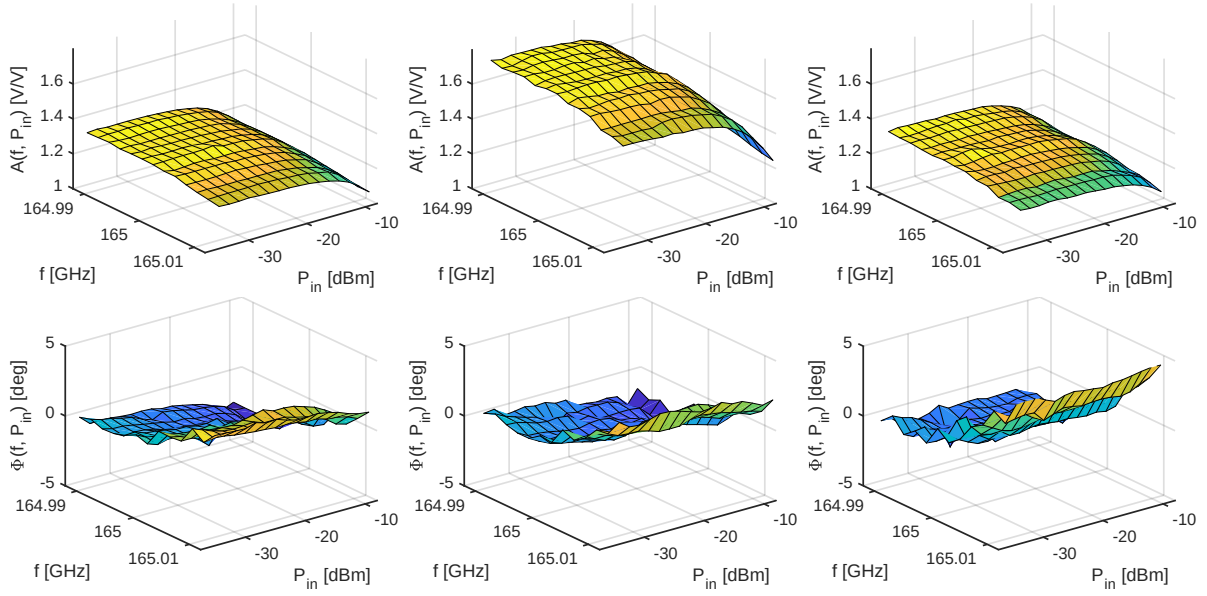
(a) Contour plot of P_L at the -1 dB compression point. Colour of loading conditions match with the figure on the right.

(b) Power gain (solid, left-axis) and PAE (dash-dot, right-axis) for the loading conditions. The '+' symbols show the -1 dB compression points.

Figure 3.27: The vector gain was extracted for six loading conditions at 165 GHz around the optimal loading condition for maximum output power. The colour of the loading conditions corresponds to those shown in the EVM estimation figure in fig. 3.29

3.5.3. Vector gain curves at 165 GHz

The vector gain curves for three of the six characterized loading conditions are shown in fig. 3.28. The phase of the vector gain transfer is normalized to the phase at the center frequency for the lowest drive level. Although it was not possible to measure far into compression, the amplitude transfer (top plots of fig. 3.28) of the vector gain show some AM-AM gain compression. The AM-PM phase variation (bottom plots of fig. 3.28) is less pronounced in the phase transfer of the vector gain. For both the amplitude and phase transfer some frequency variation across the band of operation is observed, varying among the loading conditions characterized, i.e. more variation across frequency is observed for the loading condition in fig. 3.28c than that of fig. 3.28a.



(a) $\Gamma_L = 0.148 + 0.497j$ (b) $\Gamma_L = 0.427 + 0.588j$ (c) $\Gamma_L = 0.600 + 0.350j$

Figure 3.28: Two dimensional vector gain transfers measured at three different loading conditions at 165 GHz of the slvt NMOS2 transistor on the Global Foundries 22nm FD-SOI test die.

3.5.4. EVM estimation at 165 GHz

Following the post-processing steps described in section 3.2 the measured vector gain transfers are used to estimate the EVM for a 16-QAM modulated signal. The properties of the modulated signal that is analysed are shown in table 3.3.

Table 3.3: Properties of 16-QAM modulated signal analysed at 165 GHz

Parameter			Unit
Modulation			
Constellation		16-QAM	
Symbol rate	R_s	16	$Msyms/s$
Sample rate	f_s	320	MHz
Number of symbols	N_{sym}	1000	
Pulse shaping			
Type		SRRC	
Filter span		8	$syms$
Roll-off	α	0.22	
Channel BW	BW_{ch}	$R_s \cdot (1+\alpha)$	

Figure 3.29 shows the EVM estimates for the different loading conditions against the input drive level. Since there is no validation case only the vector gain based estimations are shown. The figure shows that the different loading conditions result into a different EVM performance. Most notably the loading case $\Gamma_L = 0.43 + 0.59j$ has overall the worse EVM performance. The vector gain transfer for this loading condition is shown in fig. 3.28b which shows a large variation in the amplitude transfer across frequency which might explain the worse EVM performance. In addition, the gain-compression curve for this loading condition (yellow trace in fig. 3.27b) shows that this loading condition has a lower -1 dB compression point, but also a sharper gain-compression curve, which may also contribute to the overall higher EVM.

From the different EVM curves it is clear that a trade-off can be made between achievable EVM and power gain. The loading conditions $\Gamma_L = 0.15 + 0.20j$ and $\Gamma_L = 0.43 + 0.11j$ have the lowest EVM bound of the characterized set. However, $\Gamma_L = 0.15 + 0.20j$ has a higher gain as shown in fig. 3.27b. Alternatively, the loading condition $\Gamma_L = 0.15 + 0.50j$ achieves a reasonable EVM and has

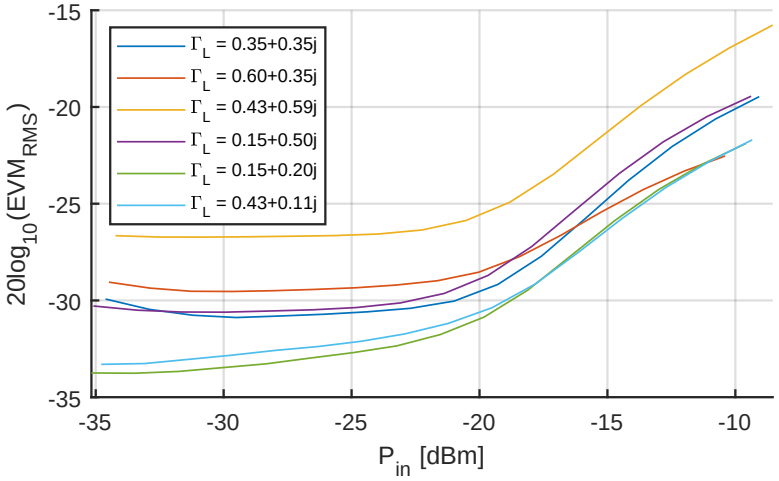


Figure 3.29: Vector gain based estimation of the EVM for a 16-QAM signal at six loading conditions at 165 GHz.

the highest gain, PAE and -1 dB compression point (see fig. 3.27b) among the set of characterized loading conditions. This could therefore be the optimum loading condition for a final stage of a PA.

On the lower range of drive levels the EVM starts to increase for some of the loading conditions. This may be explained by the extrapolation happening close to the ends of the characterized power range in the CW power to spectral power mapping process described in section 3.2.2. The offset

power ΔP for a spectral component may become negative for components with a lower than average spectral power. When the estimation is made at the lower power range characterized, the transfer for the spectral components with a negative ΔP are linearly extrapolated from the available data. Therefore an error is made as the linear extrapolation maintains the slope from the last measured points in the AM-AM and AM-PM curves while the actual curves flattens for this class-A biased transistor. Measuring at lower power levels becomes increasingly difficult because of the noise in the system affecting both the measurements readouts but also how good the active load-pull is able to converge.

3.6. Practical considerations for on-wafer vector gain measurements

The vector gain based approach relies on a large set of measurements to be made across frequency, drive-level and loading condition of the DUT. The state-of-the-art millimeter-wave active load-pull setups allow these measurements to be fully automated however some caution should be taken to ensure good results of the measurements. There are several factors that come into play during long duration on-wafer measurements. The drift in the setup limits the maximum duration time of a measurement, which depending on the stability of the setup, should no longer than a couple of hours. Depending on the frequency range and equipment it was found that in most cases the performance of a calibration would be deteriorate within 24 hours. It is therefore recommended to calibrate just before measuring and not use the calibration for more than 24 hours.

Careful consideration needs to be taken when selecting the number of measurement points in order to maximize the success of the measurement. Depending on the behaviour of interest, the number of frequency and drive-level points may be reduced. Less frequency points may be taken when a flat response in the band of interest is expected. In some cases, a finer sampling is required if for example the performance of a (wide-band) output matching network is to be tested that could show some variation across the band. The number of drive level points may also reduced, however in some cases it was found that this would reduce the convergence accuracy of the millimeter-wave active load-pull (see section 5.6).

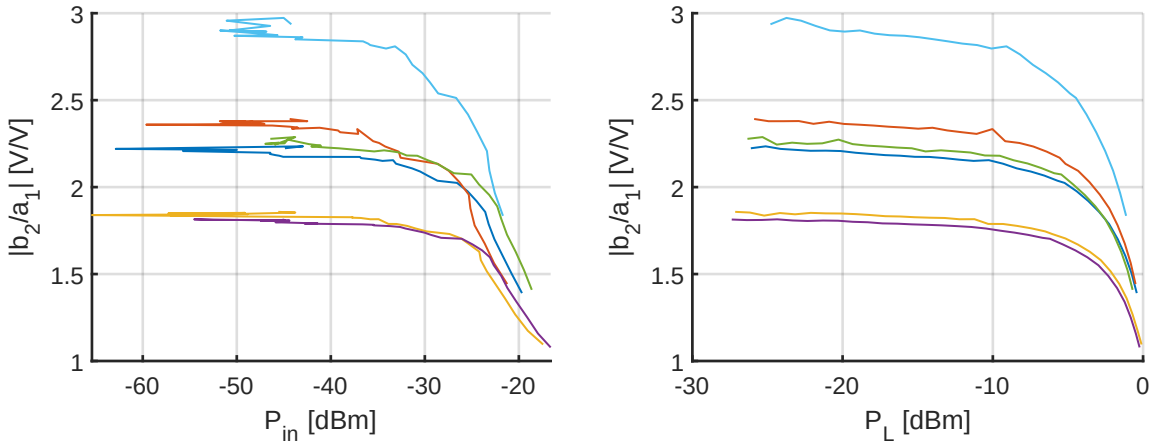
This section will discuss two other issues that occurred frequently during the vector-gain measurements, how they can be detected and possibly mitigated in post-processing. The first issue is related to measurement outliers caused by the low signal-to-noise levels at the lower end of the characterized power range. This was a primary issue in the 5 GHz measurements and that case will be discussed here, although the mitigation approach might be appropriate for other cases too. The second issue is related to probe loss of contact or change of contact during the measurement. This is primarily an issue during millimeter-wave measurements and some examples during in the D-band experiments are shown.

3.6.1. Input power correction

When there is an interest in large signal device behaviour it is common to investigate behaviour against the input power/drive level P_{in}. The controlled variable during an active load-pull measurement is however the available source power P_{avs}, while the P_{in} is calculated through the de-embedding / calibration parameters of the probes and the on-wafer structures. This makes the value of P_{in} more susceptible to noise and calibration errors, including its derived parameters such as power gain G_P. When displaying measurements against an uncontrolled quantity like P_{in} there is a risk that due to noise or other errors the 'x-axis values' are not strictly increasing. This may result into strange plots or problems in further post-processing and should be accounted for. This is particularly true for P_{in}, which depends on the measured Γ_{in} which may be itself a function of P_{in} due to non-linear effects. However, this may also happen when displaying data against the base-current while using voltage based biasing, i.e. during a bias plane characterization when displaying for example the metric of f_T against I_{be} for bipolar transistors.

A problematic case of this was found during the 5 GHz measurements using the Anteverta system. Here the input power reached well below -45 dBm levels and therefore became limited to the noise in the system. Figure 3.30 shows how the same measured value of the absolute vector gain at 5 GHz looks when plotted against the (uncorrected) input power P_{in} (fig. 3.30a) or output power P_L (fig. 3.30b). The output power is well above the noise level in the system because of the high gain of the device at this frequency. For the vector gain based approach it is important to refer the the signal power

levels to that of the input of the device. However, as can be seen in fig. 3.30a the input power is badly defined during the measurement. The cause of this can be traced back to the measured input

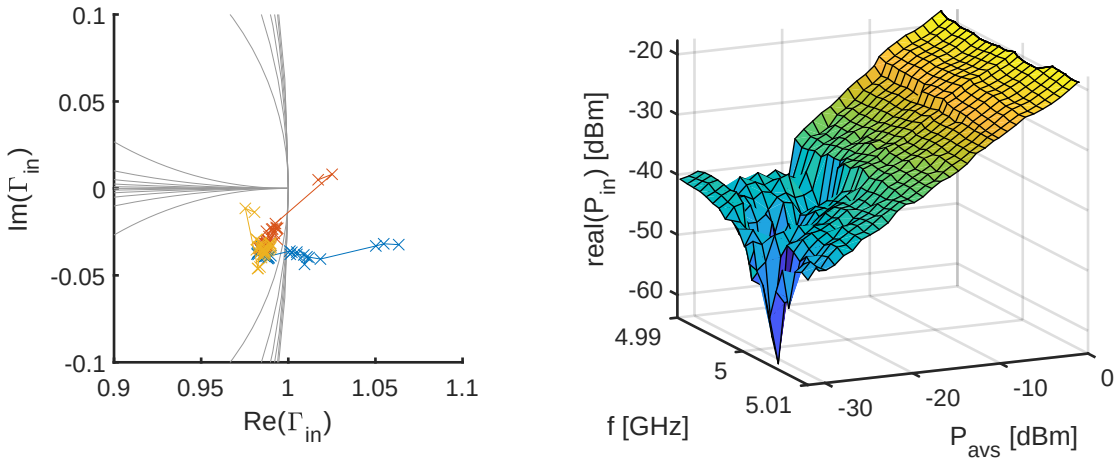


(a) Magnitude of vector gain at 5 GHz plotted against input power

(b) Magnitude of vector gain at 5 GHz plotted against output power

Figure 3.30: During load pull Pavs and PL are well defined as Pavs is controlled during the measurement and PL is often large enough. Pin can get low and eventually become difficult to measure close to the noise floor of the device. However Pin is more practical to use, as then known at which level the PA should be driven

reflection coefficient, which is needed in to calculate the input power from the available source power. The measured input reflection coefficient is shown in fig. 3.31a for the first three loading conditions. As can be seen the input reflection coefficient for this device at this frequency lies on the edge of the Smith chart. This is explained by the high speed and small size of the transistor, and the absence of parasitics by the de-embedding structures. Because the low power levels the input reflection coefficient is sometimes measured beyond the Smith Chart $|\Gamma_{in}| > 1$. Causing the input power calculated from to become negative (greater than the available source power and complex when expressed in dBm) resulting into the issues seen in fig. 3.30a. A later measurement using a VNA based load-pull system (as opposed to a ADC digitizer based system) confirmed that this issue was related to the noise in the system close to the edge of the Smith chart and not due to instabilities of the device.



(a) Measured input reflection coefficient of the slvt-nmos2 at 5 GHz for three loading conditions.

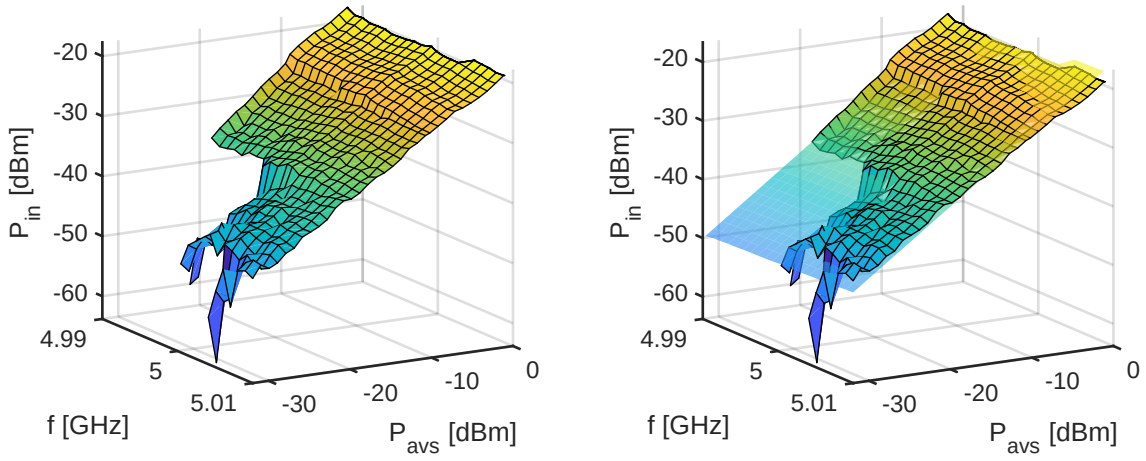
(b) Conversion of available source power to input power

Figure 3.31: When characterizing high speed devices at low frequencies and low input powers the noise in active load-pull setup might cause the Γ_{in} to be measured outside of the Smith chart ($|\Gamma_{in}| > 1$) causing issues in the calculation of the input power.

The errors in calculating the input power can be clearly seen from fig. 3.31b where the available source power P_{avs} to real input power $real(P_{in})$ is shown for one loading condition. Because the relation between P_{avs} and P_{in} is linear, determined by the mismatch between Z_{in} and Z_S, a linear trend

between the two variables is expected in fig. 3.31b. Taking this assumption, it is possible to correct P_{in} by removing outliers and interpolating across the P_{avs} and frequency plane. This can be done without severely affecting the measured vector gain as only the mapping from P_{avs} to P_{in} is corrected. The vector gain is well defined throughout the measurement because both power-waves b_2 and a_1 are well above the noise level.

The first step in this correction process is to remove the P_{in} data for the points in which $|\Gamma_{in}| > 1$ was measured so that they are not considered in the following processes. This puncturing step is shown in fig. 3.32a. Next a plane ($z = ax + by + c$) is fitted to the remaining data, shown in fig. 3.32b. At this point the plane could either be used as a replacement for the original P_{in} data or the process can be repeated to get a better fit by using the plane to find the outliers, removing those and fitting a new plane. During the experiments, the first planar fit was adequate for further processing of the data.

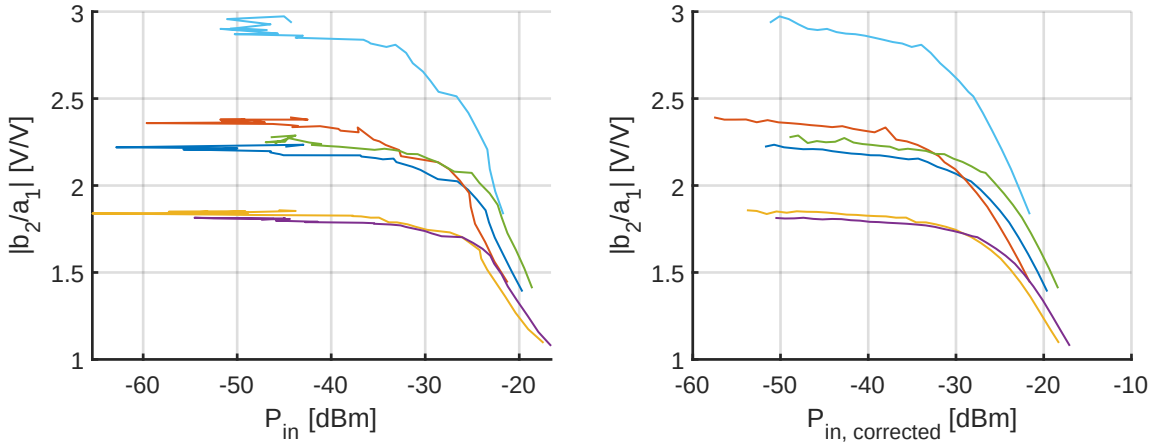


(a) Step 1: Points for which $|\Gamma_{in}| > 1$ are removed

(b) Step 2: A plane is fitted to the remaining data

Figure 3.32: When characterizing high speed devices at low frequencies and low input powers the noise in active load-pull setup might cause the Γ_{in} to be measured outside of the Smith chart ($\Gamma_{in} > 1$). As P_{in} is derived from Γ_{in} it might therefore become unphysical.

With the corrected P_{in} plane the vector gain can now be plotted against the input drive level as shown in fig. 3.33b. Similarly, a corrected power gain can be calculated: $G_p' = P_L - P_{in,corrected}$ however more caution should be taken to the accuracy in that case as the power gain will directly depend on the quality of the fit. The result before and after correction is shwon in fig. 3.33.



(a) Before correction.

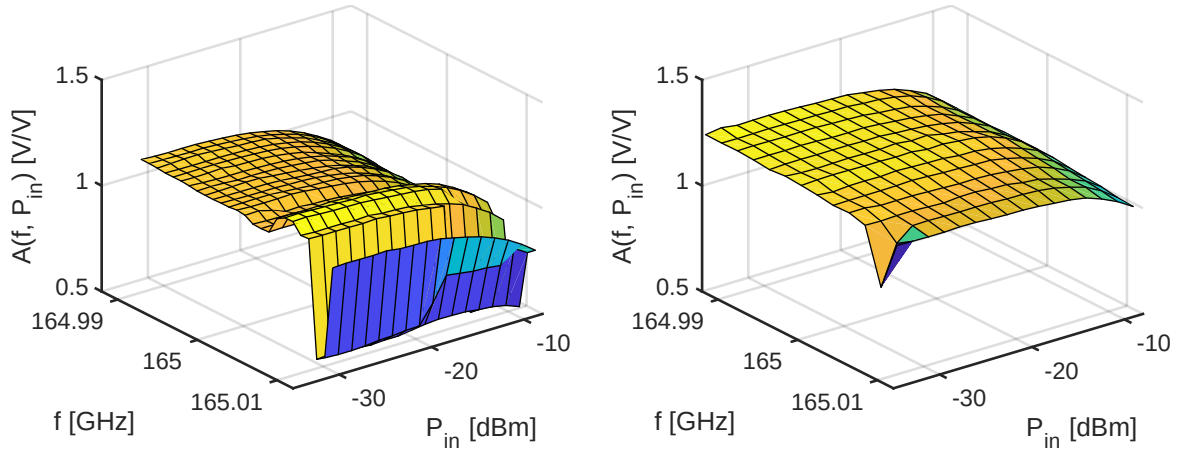
(b) After correction.

Figure 3.33: Magnitude of vector gain for the six loading conditions characterized at 5 GHz plotted before and after correction of the P_{avs} to P_{in} mapping.

3.6.2. Probe contact loss, bad convergence and outlier removal

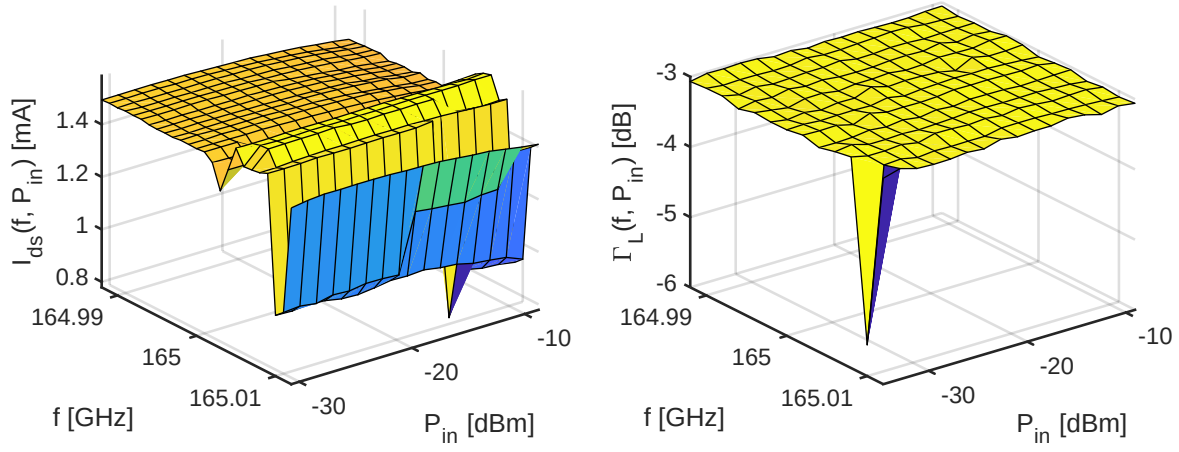
A second common problem occurring during long duration on-wafer measurements is the loss/change of contact in the probes. This can be primarily avoided by a vibration free environment and well adjusted probe manipulators. Unfortunately, this issue cannot always be fully mitigated and it is advised to use automated repeated measurements to allow for averaging or removal of bad measurement points. This problem is even more significant for millimeter-wave measurements as a slight change in probe contact might change the probe to wafer capacitance even-tough contact is still maintained.

One way to detect probe contact changes is to look at the device current after the measurement. Two examples are shown in fig. 3.34 appearing during the D-band 165 GHz measurements. In the example on the left (figs. 3.34a and 3.34c) it is clear from the drain current that the probe contact has changed during the last 5 frequency points, resulting into a drastic change in the magnitude of the vector gain. In fig. 3.34b only two points are outliers in the measurement. At first glance it may seem that this is also due to a probe contact change. However, in this case this is due to an issue in the load-pull convergence as can be seen in fig. 3.34d, as for those points the loading condition is significantly different.



(a) Measurement of vector gain showing probe-contact loss

(b) Measurement of vector gain showing convergence problem



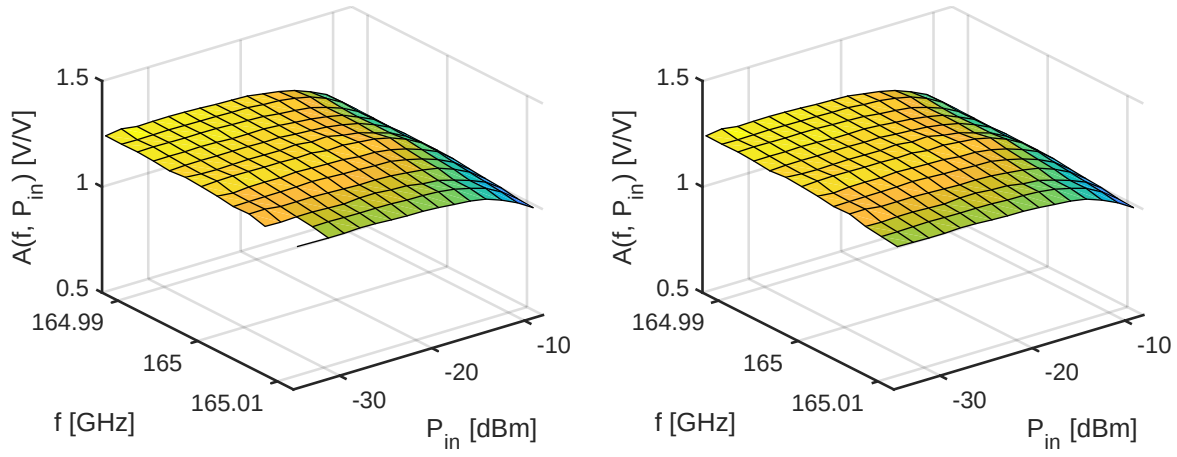
(c) Drain current during above measurement

(d) Loading condition during above measurement

Figure 3.34: Two common issues appearing during on-wafer vector gain measurements, probe contact changes (left) and bad convergence during some loading conditions (right)

Unfortunately, a loss of contact as bad as in fig. 3.34 cannot be corrected. To mitigate this issue the amount of measurement points was reduced, which in this case was possible as the DUT showed smooth behaviour across this range. The outlier in fig. 3.34b may be removed by an outlier detection and correction routine. For a complex valued quantity such as the vector gain this may be done on the real and imaginary parts separately. For the vector gain in fig. 3.34b this was done by outlier removal using a moving median across the frequency dimension. This removal process removes a point if it deviates more than three mean standard deviations from the moving medium. This is done

in fig. 3.35a, using a moving median window of 15 samples across the frequency dimension. This is where measuring additional frequency points below and above the band of interest becomes helpful as the data on the edges can still be repaired using neighbouring points. A 2D interpolation method works well for correcting small outliers as shown in fig. 3.35b.



(a) Outliers removed by moving median detection

(b) Outliers corrected by 2D interpolation

Figure 3.35: Small errors and outliers may be corrected using an outlier detection and correction routine

3.7. Summary & Recommendations

3.7.1. Summary

There is a growing interest in millimeter-wave communications at frequencies above 100 GHz. State-of-the-art millimeter-wave active load-pull characterization may enable development and reduction of design cycles for these communication systems. However, linearity characterization in these setups is limited to single tone AM-AM and AM-PM distortion due to the highly non-linear architecture of the VNA frequency extension modules. Modulated performance of a the DUT may be evaluated in the envelope domain using single tone AM-AM and AM-PM measurements however this is taking a narrow-band response and is therefore not fully utilizing the characterization capabilities of millimeter-wave active load-pull setups.

A new approach to first order linearity evaluation at millimeter-wave frequencies is proposed that uses vector gain measurements of the DUT. The vector gain is the complex large-signal voltage transfer function from the input to output of the DUT. It can be measured from the ratio of the down-converted b_2 over a_1 power waves and contains both AM-AM and AM-PM behaviour of the DUT. The proposed method uses two-dimensional vector gain measurements across the frequency band of operation and drive level. Thereby not taking a narrow-band assumption and allowing to observe in-band variations of the device response. The response fo the device to a complex modulated signal may then be estimated using a frequency domain approach. Where each frequency component in the signal is treated as a separate CW tone and is applied a magnitude and phase transfer according to the tone power level. The obtained output spectrum may be transformed back to time domain to calculate linearity metrics such as EVM.

The vector gain based method was tested in simulation and during on-wafer experiments for active devices. During these experiments the two-dimensional vector gain was measured at several loading conditions to investigate if differences in linearity performance could be determined using the method. The method was benchmarked using on-wafer measurements at two sub 30 GHz frequencies on an active load-pull capable setup that also supports complex modulated excitation. In both cases the vector gain based method showed correlation in the EVM across loading condition and drive levels for a QAM-16 signal. The method was then used in D-band at 165 GHz to determine the optimum loading condition for lowest EVM using a millimeter-wave active load-pull capable setup.

3.7.2. Recommendations

During the investigation into the vector gain based methods for linearity evaluation one of the primary goals was to show an on-wafer proof of concept demonstration at millimeter wave frequencies above 100 GHz. This made the development and validation of the method challenging as all measurements had to be performed on-wafer, using sensitive high speed transistors that would easily break down. This resulted into lengthy measurement campaigns that required equipment and time from external companies. Therefore extensive testing and validation of the method across different devices and modulation schemes was not available.

The first recommendation is therefore to extend the validation of the vector gain based method to a greater sample size. A quick way of doing so without having to go to an on-wafer environment would be to use a commercially of the shelf (wideband) connectorized amplifier as the DUT[5]. These amplifiers are available with a wide bandwidth up to 40 GHz and would allow to test the method for several modulation schemes and across a large frequency range, while still having the option to compare against true modulated wave measurements. For the millimeter-wave frequencies the same could be done with waveguide amplifier modules. For both cases this could be done in a 50 Ohm environment therefore relaxing the need to do active load-pull, making the measurements easier and quicker to perform. During this project such wideband amplifier modules were not available but the use of high speed single stage transistors and active load-pull allowed for testing across different frequencies.

It should also be investigated if and in which cases the wideband approach using the two-dimensional vector gain gives better results than a narrow band approach. Two situations may already be examined, first the vector gain measured at the center frequency may be used for all frequencies to see in which cases the inclusion of in-band variations makes a difference, and second the vector gain based method could be compared with a narrow-band time-domain envelope approach where the AM-AM and AM-PM is applied on the envelope of the signal. This investigation should be complemented with cases that have some frequency dependency in the design such as in the matching networks.

Investigation into the performance of the method for different modulation schemes could be interesting as well. During the project QAM modulations were examined, however the spectrum based approach would be suited for Orthogonal frequency-division multiplexing (OFDM) modulations as-well. It would therefore be interesting to see how the approach performs for a different class of modulation schemes.

During the on-wafer measurements a fixed loading condition across the frequency band of operation was used. Creating a wide-band output matching network with a flat response is difficult to achieve in practice. However, the two-dimensional vector gain should allow for a different loading condition at each of the frequency measured. This way a (simulated) loading condition with frequency variation could be tested. Allow for example to investigate how the frequency variation of for example an output-matching network, balun or output filter influences the behaviour of the device. Possibly even tuning of the output network would be possible to compensate for the effects of the transistor or other parts in the circuit. This ability was not tested as the current millimeter-wave active load-pull setups do not allow for automatic importing of a frequency dependent loading condition. This would however be a good addition to the software to increase automatic measurement capability.

Although the vector gain method is was designed for on-wafer transistor characterization it could offer some insights during the design phase of a (multi-stage) millimeter wave PA in simulation as well. The single stage SiGe PA that was used in simulation to test the approach did not show a significant difference between different loading conditions. In fact, based on the envelope simulation it showed great linearity up to the -1 dB compression point for all loading conditions and modulation schemes that were tested. The transistor model and simulation method used might play a role in the accuracy of the results and should be repeated using a transient simulation. However, the ADS Verification Test Bench does not support transient simulation. To do so a different implementation could be used that interfaces MATLAB and ADS to generate modulated signals, simulate them in ADS and post process them in MATLAB.

Some inspiration to investigate using the vector gain for linearity evaluation came from the paper by Qui et. al. [53]. In that paper a GaN HEMTs were characterized for single tone AM-AM/AM-PM, two-tone and modulated waveforms at 31.5 GHz. In the paper the results between the three characterization

[5] A great candidate would be to use a wide-band gain block amplifier as a DUT. These are commonly used at the input and output of the active load-pull setup. For example the ZVA-02443HP+ would offer a wide bandwidth to test the vector gain based method.

methods are compared. A linearity figure or merit is defined based on the single tone AM-AM that well correlates with the EVM measurements using the modulated waveforms across different bias condition / operation class (A, AB, B). A similar extension to the vector gain investigation was planned that included variation of bias (i.e. different class of operation) in the analysis. Thereby, possibly also determining an optimum bias condition (named ”balanced AB” in [53]) for a particular modulation. However, this idea was left for future work as there was only limited availability to measure using the modulated wave load-pull setup. It would not be possible to vary the bias much in the D-band measurements as here the bias already had to be reduced to ensure compression within the power ranges of the frequency extension modules.

4

16-term error calibration for sub-Terahertz measurements

When moving towards on-wafer measurements in the sub-Terahertz range (100 GHz to 300 GHz), the mechanical reduction of the waveguide to on-wafer co-planar environment, i.e. by the probes, does not scale equally with the smaller wavelengths. This makes the radiation effects of on-wafer measurements increasingly problematic when moving towards higher frequencies. In this chapter a structured approach is described towards quantifying and removing through calibration the impedance dependent cross-coupling occurring between probes.

4.1. Probe-to-probe cross-coupling

The quality of a calibration can be estimated by the worst case error bound metric [5] and its distance from the zero line. Figure 4.1 shows the worse case bound for three on-wafer calibration procedures: SOLT calibration using ideal standard values (SOLT-IDEAL), SOLT calibration using the S-parameters from the EM simulation (SOLT-EM) and TRL calibration. The test die that was used for this measurement is the TSMC 28nm die as discussed in section 2.4 that which uses a capacitively loaded inverted CPWs calibration kit [54].

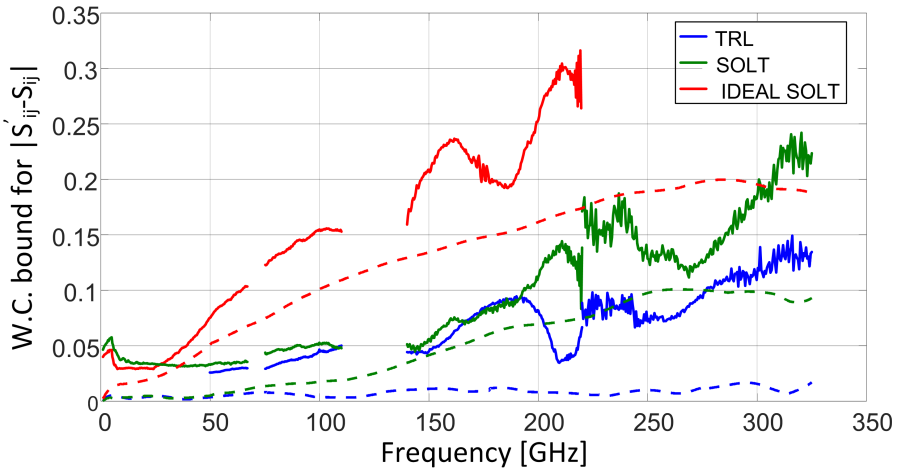


Figure 4.1: The worst case error bound: maximum absolute deviation between true and measured S-parameters (after calibration) on the TSMC 28nm calibration die. Dashed lines show simulated lower bounds while solid lines show measured values [55]

The solid lines in fig. 4.1 show the error bound of each calibration based on measured data. The differences in performance is be explained by the required information of the standards used and how

they can be extracted from EM simulation [8]. Aside from this difference the performance of the methods show a linear increase in error up to 120 GHz. After this point some interferometric error dominates and a standing wave pattern is observed. It is assumed that this pattern is dominated by the cross-coupling between the probes, an effect that has also been studied by Williams et. al. [11]. The focus of this chapter is on calibration algorithms that can correct for this type of error made in a probe environment where cross-coupling is present.

Probe-to-probe cross-coupling cannot be accurately captured using only the 8-term or 10/12-term error models [9], [31]. Of these models only the 12-term error model contains terms for leakage. However, these leakage terms do not accurately represent a coupling effect as they are static, while the coupling is dependent on the impedance of the DUT. An extended error model is required with additional error terms that capture this DUT impedance dependent cross-coupling. This is done in the 16-term error model, the 2-port variant of the generalized multi-port error definition defined by R. Speciale [56] and shown in fig. 4.2. The model contains eight conventional error terms, shown by the e_{xy} terms with solid lines in fig. 4.2, and eight cross error terms, shown by the e_{xy} terms with dashed lines in fig. 4.2.

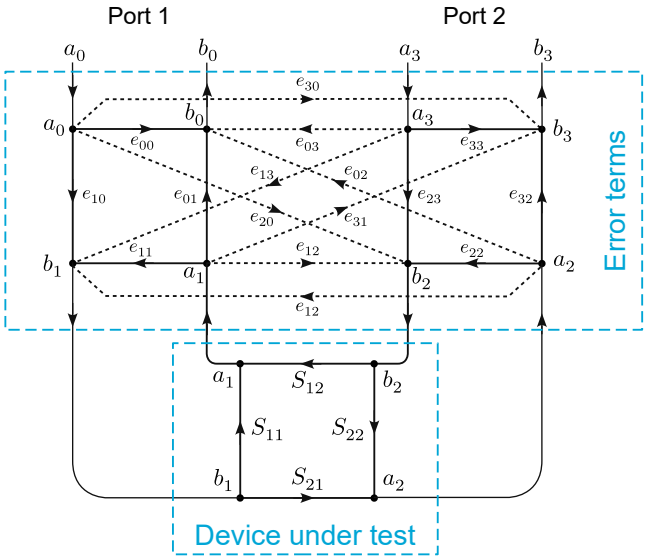


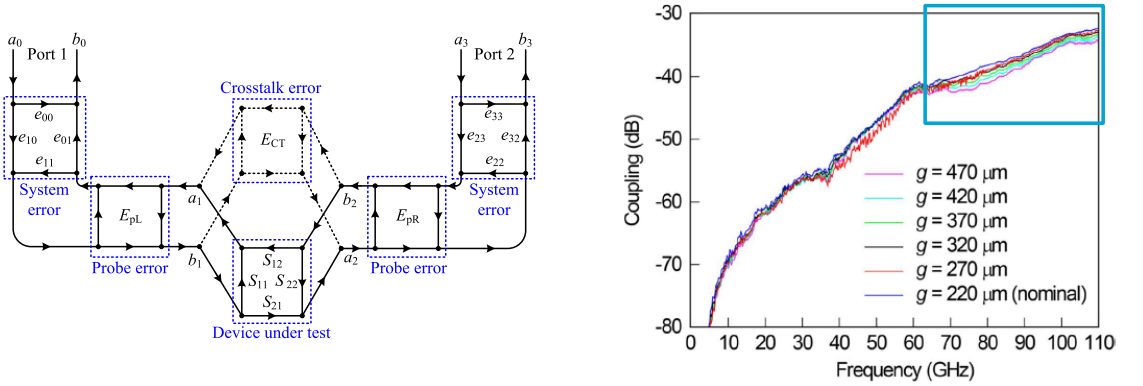
Figure 4.2: 16-term error model including all possible cross-terms (based of [9], [10], [31])

The terms e_{30} and e_{03} are the cross-terms that introduce leakage of energy independent of the impedance of the DUT measured, similar to the two cross-terms present in the 12-term model. The 16-error term model has additional leakage paths that couple energy from one-port through the DUT to the other port (i.e. $a_o \rightarrow e_{20} \rightarrow b_2 \rightarrow S_{22} \rightarrow a_2 \rightarrow e_{32} \rightarrow b_3$) and these leakage paths are not captured in the 12-term model and related calibration procedures. Work by Liu et al. [57] showed that the e_{21} and e_{12} leakage paths are primarily representing the over the air probe-to-probe cross-coupling.

The sixteen error terms can be found using several different calibration procedures requiring five or four fully known calibration standards to be measured [9], [10], [58], [59]. However, when using a conventional 16-term calibration the probe spacing must be kept constant during the calibration and following DUT measurements as the probe-to-probe cross-coupling is dependent on the spacing between the probes. This means that each differently spaced DUT should have its own calibration kit, which takes up costly space on an on-wafer or off-wafer calibration kit, i.e. an area of at least 4 times the DUT with landing pads is needed for the 16-term calibration. This also makes it difficult to commercialize a 16-term calibration kit as they could only be used if the spacing matches that of the customers DUT.

Several publications have addressed this issue, including the ”calibration on the fly method” [10] where the cross-coupling errors are described in a separate network parallel to the DUT, see fig. 4.3a. In this ”calibration on the fly method” the system and probe errors are found using one-port short-open-load (SOL) calibrations on both ports using an off-wafer calibration, then during a second on-wafer

calibration the cross-coupling errors are found using a calibration kit with the same pitch as the to-be-measured DUT. This method solves the problem of non-constant spacing during (multiple) DUT measurements. However, the method still requires an on-wafer calibration structure for each DUT with differently spaced landing pads. Ideally, the spacing dependent probe-to-probe cross-coupling terms should be parameterized. In this way, the cross-coupling may be determined for several spacings during an off-wafer calibration (preferably on a fused silica (FS) calibration kit to minimize the calibration transfer error [60]) and then be applied for arbitrary probe-spacing during a large wafer measurement. An publication by Williams et al. [11] showed a distance dependence in the cross-talk error terms, as shown in the blue box of fig. 4.3b.



(a) Calibration on the fly error model proposed by Wu et. al. [10] contains a parallel crosstalk error network that is obtained during a second-tier calibration

(b) Magnitude of cross-coupling term across probe spacing from Williams et. al. [11] showing a spacing dependence in the cross-coupling terms (see blue box)

Figure 4.3: Examples in literature showing efforts to create 16-term calibration algorithms that can cope with different probe spacing

The distance dependence shown by Williams et. al. could be used in a two-tier model having the cross-coupling parametrized for different spacing. However, in their approach a multi-line TRL calibration was used to find the scattering parameters of the two-port calibration structures. This calibration does not take into account cross coupling thereby losing the ability to resolve a spacing dependent cross-talk. EM based calibration as done in [8] may offer the solution here. The scattering parameters of the two-port calibration structures may be found in a 3D-EM simulation. After which they could be used to find the 16-error terms for several probe-to-probe spacing to give a conclusive answer on whether or not the 16-error term model may be parameterized for probe spacing.

4.2. Goal, research question & structure

The goal of the work described in this chapter is to investigate if a 16-term error model may be parameterized for different probe-to-probe spacing. To do so first a 16-term calibration method needs to be developed and validated its ability to resolve DUT dependent cross-coupling. Then it should be shown that the 16 error terms may be split in those that are strictly non-spacing dependent and those that are spacing dependent. Then a parameterized formulation of the 16-error terms may be developed and tested for its ability to correct un-calibrated measurements for a probe-to-probe spacing that is not included in the calibration. The research question associated to this chapter is formulated as followed:

How can a 16-term calibration algorithm be developed and validated to be used in a parameterized form for correction of spacing dependent probe-to-probe cross-coupling?

In section 4.3 the mathematical formulation of the 16-term error model and its calibration procedure is discussed. It is shown how the 16-terms may be found in a robust manner using singular value decomposition. In section 4.4 the implementation of the 16-term calibration algorithm in MATLAB is tested. For this tests calibration substrates are defined in Keysight ADS and are used in a simulated measurement including VNA non-idealities and probe like cross-coupling. In section 4.5 this validation is extended to a full 3D EM simulation of the standards in CST including probe models. In section 4.6 a parameterized version of the 16-term error model is presented and a procedure is defined that allows

for a probe-to-probe spacing interpolation of the error terms. This chapter concludes with recommendations for future work.

4.3. Mathematical model formulation

In the 16-term calibration model the errors in the measurement are modeled by a 4 port error box similar to that in an 8-term calibration however now including all possible leakage / cross-terms. By taking enough measurements of known DUTs one can solve the terms of the 4 port error box. After solving the 16 terms calibrated S-parameter measurements can be taken by correcting the measured S-parameters (often noted as S_m) to obtain the actual S-parameters at the calibration reference plane (simply noted as S or S_a) signals using (the inverse of) the error box. The S-parameter notation of the error model in fig. 4.2 takes the shape:

$$\begin{bmatrix} b_0 \\ b_3 \\ b_1 \\ b_2 \end{bmatrix} = S_E \begin{bmatrix} a_0 \\ a_3 \\ a_1 \\ a_2 \end{bmatrix} \tag{4.1}$$

$$S_E = \begin{bmatrix} e_{00} & e_{03} & e_{01} & e_{02} \\ e_{30} & e_{33} & e_{31} & e_{32} \\ e_{10} & e_{13} & e_{11} & e_{12} \\ e_{20} & e_{23} & e_{21} & e_{22} \end{bmatrix}. \tag{4.2}$$

Taking the measured and actual S-parameters as:

$$\begin{bmatrix} b_0 \\ b_3 \end{bmatrix} = S_m \begin{bmatrix} a_0 \\ a_3 \end{bmatrix} \qquad S_m = \begin{bmatrix} S_{11m} & S_{12m} \\ S_{21m} & S_{22m} \end{bmatrix} \tag{4.3}$$

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = S_a \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \qquad S_a = \begin{bmatrix} S_{11a} & S_{12a} \\ S_{21a} & S_{22a} \end{bmatrix} \tag{4.4}$$

and then applying linear algebraic operations on eq. (4.1), eq. (4.4) and eq. (4.2) would result into an equation combining S_a, E and S_m which is non-linear in the error terms making it difficult to solve for the error terms using typical linear algebra [9]. If however the system of equations is setup using transmission T-parameters[1] the resulting system of equations is linear in the T-parameters. That system of equations takes the shape [9]:

$$\begin{bmatrix} b_0 \\ b_3 \\ a_0 \\ a_3 \end{bmatrix} = T \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix}, \tag{4.5}$$

$$T = \left[\begin{array}{c|c} T_1 & T_2 \\ \hline T_3 & T_4 \end{array}\right] = \left[\begin{array}{cc|cc} t_0 & t_1 & t_4 & t_5 \\ t_2 & t_3 & t_6 & t_7 \\ \hline t_8 & t_9 & t_{12} & t_{13} \\ t_{10} & t_{11} & t_{14} & t_{15} \end{array}\right]. \tag{4.6}$$

One can see why the notation as in eq. (4.5) is beneficial to calibration because all the measured waves are on the left hand side and scattered and incident waves of the DUTs are on the right hand side. Making calibrated measurements of the DUT is therefore equal to applying the inverse of the T matrix on the measured S-parameters S_m.

Corruption equation: $$S_m = (T_1 S_a + T_2)(T_3 S_a + T_4)^{-1} \tag{4.7}$$

Correction equation: $$S_a = (T_1 - S_m T_3)^{-1}(S_m T_4 - T_2) \tag{4.8}$$

Calibration (linear-in-T) equation: $$T_1 S_a + T_2 - S_m T_3 S_a - S_m T_4 = 0 \tag{4.9}$$

[1] One should be cautious when working with T-parameter notations instead of S-parameter notations. Across literature different notation (orders) are used for the T-parameters. The notation in this work uses those used by D. Rytting [9], [31] and A. Ferrero [32], [37]. These however differ from those for example used by D. Frickey [61] or MATLABs s2t(...) function.

The linear-in-T equation eq. (4.9) is used to find the error terms of the 4-port transmission matrix. This equation takes the shape of $A\boldsymbol{t} = 0$ which is a homogeneous system of equations that can be solved for $\boldsymbol{t}$ using several algorithms. To show this, the linear-in-T equation can be expanded, resulting into four equations for each S_m, S_a measurement:

$$\begin{bmatrix} t_0 & t_1 \\ t_2 & t_3 \end{bmatrix}\begin{bmatrix} S_{11a} & S_{12a} \\ S_{21a} & S_{22a} \end{bmatrix} + \begin{bmatrix} t_4 & t_5 \\ t_6 & t_7 \end{bmatrix} - \begin{bmatrix} S_{11m} & S_{12m} \\ S_{21m} & S_{22m} \end{bmatrix}\begin{bmatrix} t_8 & t_9 \\ t_{10} & t_{11} \end{bmatrix}\begin{bmatrix} S_{11a} & S_{12a} \\ S_{21a} & S_{22a} \end{bmatrix} - \begin{bmatrix} S_{11m} & S_{12m} \\ S_{21m} & S_{22m} \end{bmatrix}\begin{bmatrix} t_{12} & t_{13} \\ t_{14} & t_{15} \end{bmatrix} = 0 \quad (4.10)$$

$$\begin{aligned}
&S_{11a}t_0 + S_{21a}t_1 + t_4 - S_{11a}S_{11m}t_8 - S_{21a}S_{11m}t_9 - S_{11a}S_{12m}t_{10} - S_{21a}S_{12m}t_{11} - S_{11m}t_{12} - S_{12m}t_{14} = 0 \\
&S_{11a}t_2 + S_{21a}t_3 + t_6 - S_{11a}S_{21m}t_8 - S_{21a}S_{21m}t_9 - S_{11a}S_{22m}t_{10} - S_{21a}S_{22m}t_{11} - S_{21m}t_{12} - S_{22m}t_{14} = 0 \\
&S_{12a}t_0 + S_{22a}t_1 + t_5 - S_{12a}S_{11m}t_8 - S_{22a}S_{11m}t_9 - S_{12a}S_{12m}t_{10} - S_{22a}S_{12m}t_{11} - S_{11m}t_{13} - S_{12m}t_{15} = 0 \\
&S_{12a}t_2 + S_{22a}t_3 + t_7 - S_{12a}S_{21m}t_8 - S_{22a}S_{21m}t_9 - S_{12a}S_{22m}t_{10} - S_{22a}S_{22m}t_{11} - S_{21m}t_{13} - S_{22m}t_{15} = 0
\end{aligned} \quad (4.11)$$

For each measurement a homogeneous system of equations can be setup with four equations and 16 unknowns:

$$A\boldsymbol{t} = 0 \quad (4.12)$$

$$A^{\{4\times16\}} = \begin{bmatrix}
S_{11a} & S_{21a} & 0 & 0 & 1 & 0 & 0 & 0 & -S_{11a}S_{11m} & -S_{21a}S_{11m} & -S_{11a}S_{12m} & -S_{21a}S_{12m} & -S_{11m} & 0 & -S_{12m} & 0 \\
0 & 0 & S_{11a} & S_{21a} & 0 & 0 & 1 & 0 & -S_{11a}S_{21m} & -S_{21a}S_{21m} & -S_{11a}S_{22m} & -S_{21a}S_{22m} & -S_{21m} & 0 & -S_{22m} & 0 \\
S_{12a} & S_{22a} & 0 & 0 & 0 & 1 & 0 & 0 & -S_{12a}S_{11m} & -S_{22a}S_{11m} & -S_{12a}S_{12m} & -S_{22a}S_{12m} & 0 & -S_{11m} & 0 & -S_{12m} \\
0 & 0 & S_{12a} & S_{22a} & 0 & 0 & 0 & 1 & -S_{12a}S_{21m} & -S_{22a}S_{21m} & -S_{12a}S_{22m} & -S_{22a}S_{22m} & 0 & -S_{21m} & 0 & -S_{22m}
\end{bmatrix} \quad (4.13)$$

$$\boldsymbol{t} = \begin{bmatrix} t_0 & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 & t_9 & t_{10} & t_{11} & t_{12} & t_{13} & t_{14} & t_{15} \end{bmatrix}^T \quad (4.14)$$

When K standards are measured the A matrices can simply be stacked:

$$A^{[K\cdot4\times16]} = \begin{bmatrix} A_1{}^{[4\times16]} \\ A_2{}^{[4\times16]} \\ \vdots \\ A_k{}^{[4\times16]} \end{bmatrix} \quad (4.15)$$

To calibrate for 16-error terms at least four standards need to be measured, resulting into 16 equations with 16 unknowns ($m = n$). Without further assumptions about symmetry in the error network at least five (fully-known) standards need to be used [10], [58] to solve the system of equations. Singularities may make the in the model unsolvable when less standards are used and the system is not symmetric. By using a fifth standard, the redundancy results into a solvable system of equations in which the number of equations is larger than the number of unknowns ($m > n$) [58]. At least one standard needs to have a finite and known transmission, a thru (T) and the other standards are a combination of double-one-port standards, load (L), short (S) or open (O). In case five standards are used the set of standard consists of a two-port standard T and four double-one-port standards L-L, X-X, L-X and X-L, where X is either a short or an open [58], [59].

4.3.1. Singular Value Decomposition

There are several methods to solve the homogeneous system of equations $A \cdot \boldsymbol{t} = 0$. The paper by Butler et. al. [9] mentions two; direct-inversion and singular value decompostion (SVD) to solve the system. According to the authors, direct-inversion did not provide satisfactory results on real measurements likely due to singularities and SVD being the better solution because it solves the system in a least squares sense providing the best approximation of the error model. Other least squares based methods may be used, such as a orthogonal distance regression algorithm used in [11]. Initial results with SVD were promising, therefore this method was used in this work.

SVD is applied here to find the null space of the system $A \cdot \boldsymbol{x} = 0$, i.e. the basis for x that makes A times x equal to zero. Since only ratio measurements are used in the VNA measurements t can be simply any vector in the null space, including the null space vector itself. The proof of finding the null space using SVD can be found in [62].

The principle behind SVD is that any $(m \times n)$ matrix can be decomposed into three matrices regardless of its singularities[2]:

$$A = U\Sigma V^T \tag{4.16}$$

where U is an orthogonal $m \times n$ matrix, $\Sigma = [diag(\sigma_j)]$ is a diagonal $n \times n$ matrix with singular values σ_j on the diagonal, and V^T is an orthogonal $m \times n$ matrix. The last columns j in Σ for which σ_j is zero correspond to the rows in V (or columns in V^T) that span the null space of A. In real measurements however none of the σ_j values will be exactly zero. According to [9] the value of the smallest σ_j corresponds approximately to the level of noise and systematic errors in the system, allowing for checking the quality of the solution after the calibration. As none of the values in Σ will be zero, simply the last column of V^T may be used as the null space for A, granted that if the last σ_j (the smallest) is much greater than zero the quality of the calibration will not be good. As mentioned, only ratio measurements are considered in a VNA measurement therefore the last column of $V^T = [\boldsymbol{v}_1^T, ..., \boldsymbol{v}_{16}^T]$ can simply be used as a solution for t:

$$\boldsymbol{t} = \boldsymbol{v}_{16} \tag{4.17}$$

4.3.2. Switch term correction

In a two-port measurement the VNA excites and measures the forward direction (port-1 to port-2) and reverse direction (port-2 to port-1) separately. The direction is determined by an internal (imperfect) switch, that terminates the non-excited port with an imperfect match. This imperfect match causes the waves a_3 and a_0' to be non-zero when exited in the forward and reverse direction respectively. The apostrophe ($'$) indicates the waves in the model of fig. 4.4 when excited in the reverse direction.

In the 10/12-term calibration method the correction for the effect of the non-perfect VNA switch are embedded inside the port-1/2 reverse/forward match terms (e_{11}' and e_{22}, also known as E_L' and E_L'') and transmission tracking terms ($e_{10}e_{32}$ and $e_{23}'e_{01}'$, also written as E_T' and E_T''). This is not the case for the 8-term or 16-term calibration method and therefore the raw measured S-parameters need to be pre-corrected prior to calibration (or DUT correction). This can be done by once measuring the reflection coefficients of the switch in the forward direction (Γ_2) and once in the reverse direction (Γ_1) given the switch is stable throughout the measurement [63]:

$$\Gamma_2 = \frac{a_3}{b_3} \qquad \Gamma_1 = \frac{a_0'}{b_0'}. \tag{4.18}$$

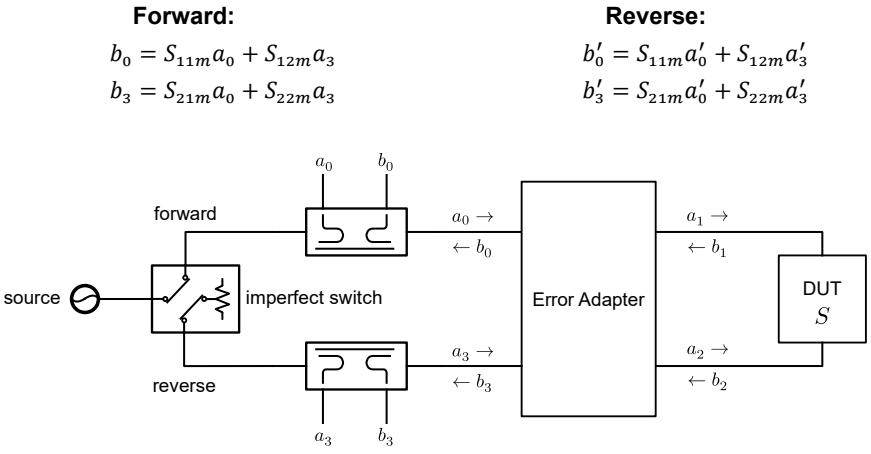


Figure 4.4: In a two-port VNA measurement the forward and reverse directions are measured separately while the opposite port is terminated by an imperfect switch (based of [31]).

[2]rows or columns in A that are (nearly) zero making A non-invertable

These two ratios can be measured once during calibration using a DUT with a non-zero transmission (i.e. a thru or a line) and then used to remove the effects of the switch during all succeeding measurements. Correcting for the switch terms may be done as followed:

$$S_{11m} = \frac{S_{11r} - S_{12r}S_{21r}\Gamma_2}{d} \qquad S_{12m} = \frac{S_{12r} - S_{11r}S_{12r}\Gamma_1}{d}$$
$$S_{21m} = \frac{S_{21r} - S_{22r}S_{21r}\Gamma_2}{d} \qquad S_{22m} = \frac{S_{22r} - S_{21r}S_{12r}\Gamma_1}{d} \tag{4.19}$$

$$d = 1 - S_{21r}S_{12r}\Gamma_1\Gamma_2$$

where $S_{11r} = \frac{b_0}{a_0}$, $S_{12r} = \frac{b_0'}{a_3'}$, $S_{21r} = \frac{b_3}{a_0}$ and $S_{22r} = \frac{b_3'}{a_3'}$ are 'raw' ratio measurements of the waves in forward and reverse direction.

4.4. Simulated VNA measurements in ADS

The 16-term calibration approach using SVD is validated using s-parameter simulations in ADS. Corrupted measurements of the standards are taken using a simulation model of a non-ideal VNA. Momentum EM simulated calibration standards and verification structures are used to test the calibration. The simulations are imported into MATLAB where the calibration and correction steps are performed.

4.4.1. Substrate standards

The two-port calibration structures for the 16-term calibration shown in fig. 4.5 were made in ADS on a $400\mu m$ Alumina substrate ($\epsilon_r = 9.9$) using $4.2\mu m$ thick gold conductors ($4 \cdot 10^7 S/m$). The structures are suitable for Ground-Signal-Ground (GSG) probes and are simulated individually for a $100\mu m$ pitch probes using six 'direct' S-parameter feed ports. The structures include standard SOLT structures: Short-Short (S-S), Open-Open (O-O), Match-Match (M-M) and Thru (T), two hybrid structures Short-Match (S-M) and Match-Short (M-S) and a $400\ \mu m$ verification line (LINE1). A Momentum Microwave simulation was done to generate the two-port S-parameters of the structures.

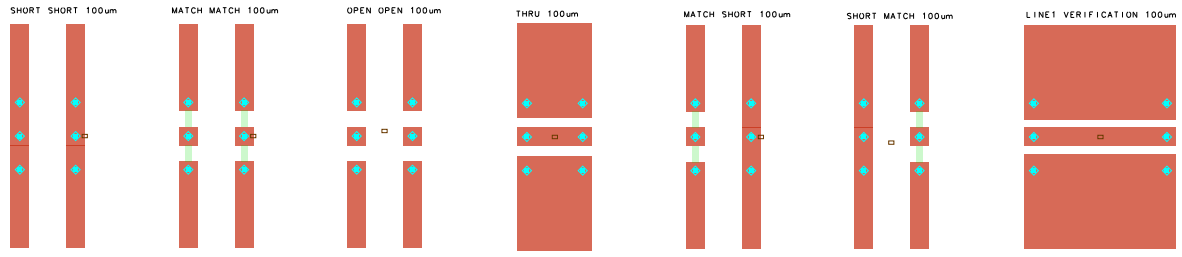
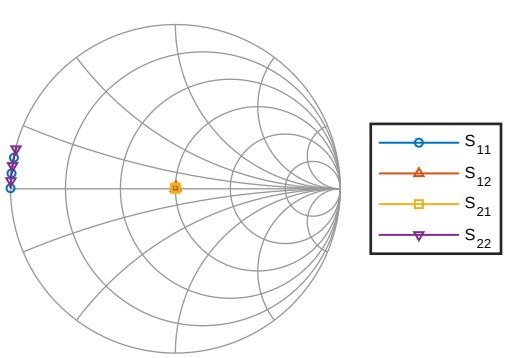
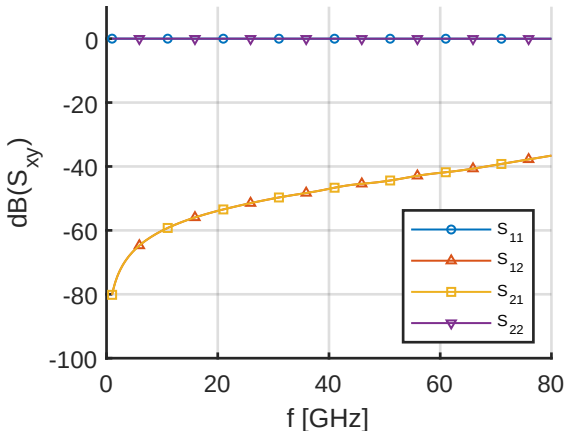


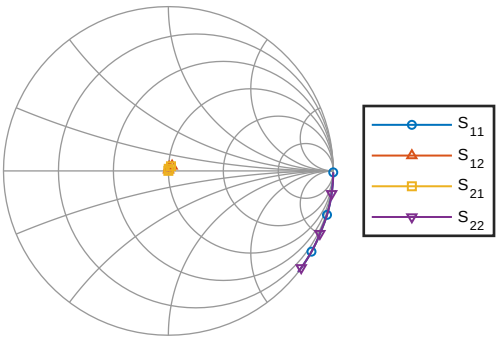
Figure 4.5: GSG $100\ \mu m$ pitch CPW calibration structures in ADS for validation the 16-term calibration algorithms. Structures include f.l.t.r. short-short (SS), load-load (LL), open-open (OO), thru (T) and load-short (LS), short-load (SL) and a $400\mu m$ verification line. Blue markers indicate probe/excitation locations, green areas in-between the probes is resistive layer ($50\ nm$ thick)

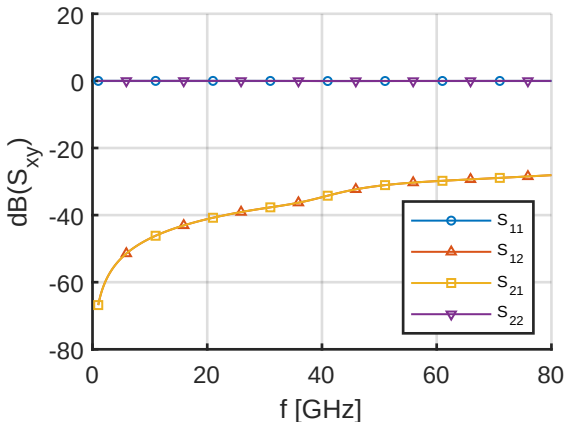
The S-parameters of the structures were simulated from 1 GHz to 100 GHz, however above 90 GHz the results are unphysical and non-passive. The simulations are therefore only used up to 80 GHz and are shown in fig. 4.6 and fig. 4.7. All standards apart from the thru have a non-zero S_{12} and S_{21} indicating that some cross-coupling between the ports is taking place, especially at the higher frequencies. The thru itself has a loss of less than 2 dB at 80 GHz.



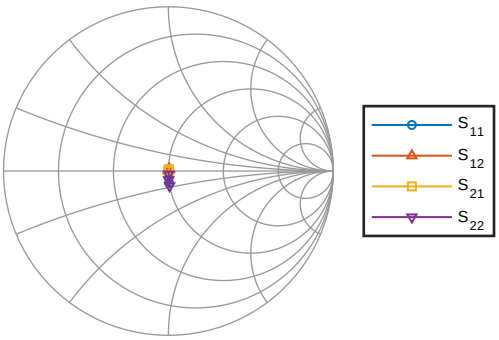


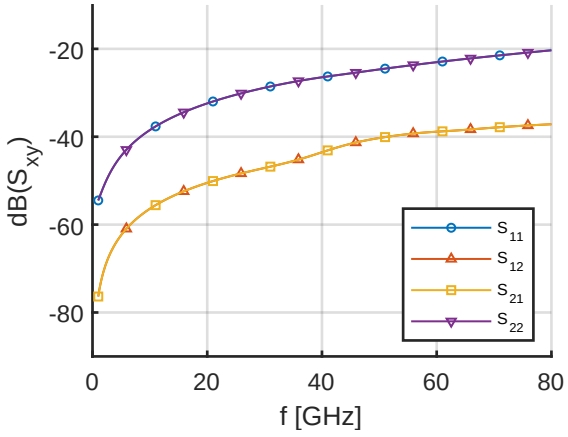
(a) Short-Short



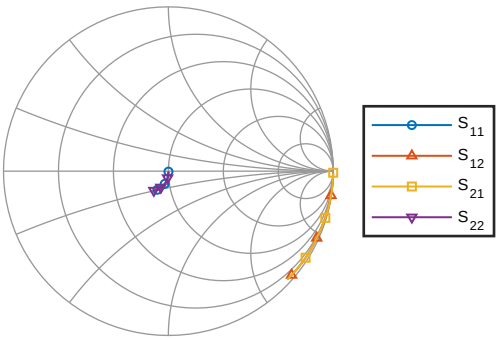


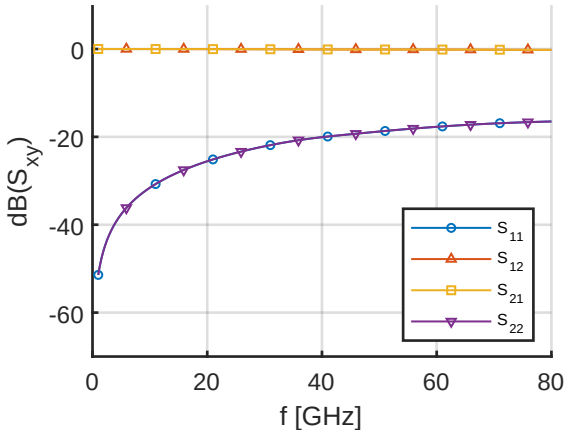
(b) Open-Open





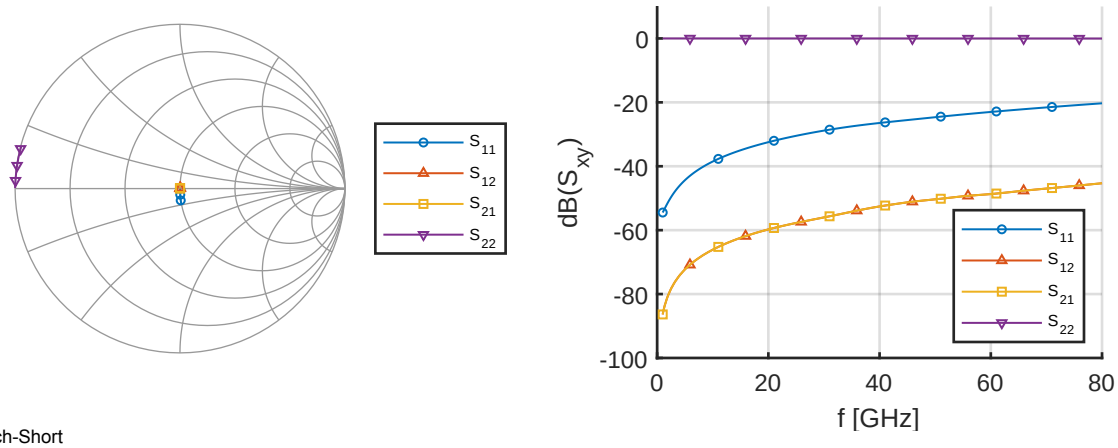
(c) Match-Match



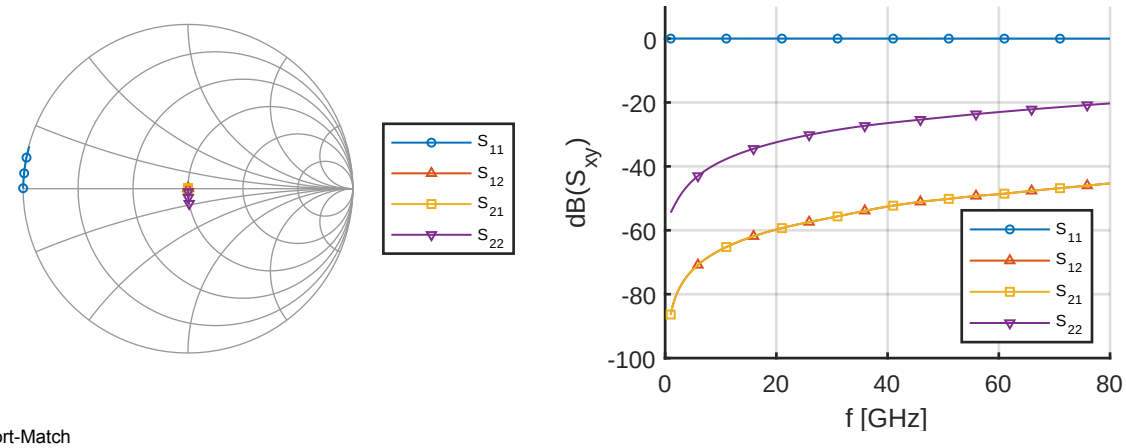


(d) Thru

Figure 4.6: EM simulated S-parameters of the two-port (SOLT) calibration structures

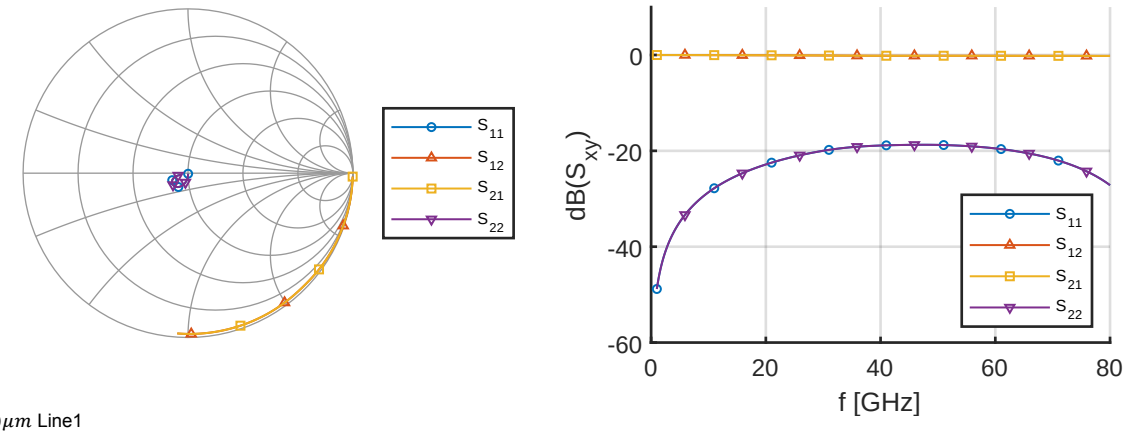


(a) Match-Short



(b) Short-Match

Figure 4.7: EM simulated S-parameters of the two hybrid two-port calibration structures (MS&SM)



(a) $400\mu m$ Line1

Figure 4.8: EM simulated S-parameters of the $400\ \mu m$ verification-line

4.4.2. Non-ideal VNA circuit

The ADS schematic of the non-ideal VNA is shown in fig. 4.9. The model includes different line lengths and attenuations for the connections between the ports and the couplers (ATTEN1/ATTEN2 and TL4/TL5), slightly different couplers (COUP1/COUP2) and an unbalance in the a-wave and b-wave paths (TL2/TL3/TL6/TL7). The effects of the imperfect switch are modelled in the impedance of the tone sources (PORT1 and PORT2). The model is simulated in a Harmonic Balance simulation operating at only the fundamental tone with $P = -5$ dBm.

To simulate the behaviour of a real VNA the power is first excited from port 1 and then port 2. During excitement from one port, the source impedance of that port is set to 50Ω while the other port has a non-matched condition, $16.35 + 13.4j\ \Omega$ for port-1 when measuring in reverse direction and $19.87 - 21.30j\ \Omega$ for port-2 when measuring in forward direction. These impedances were chosen arbitrarily and other values showed no difference in the results.

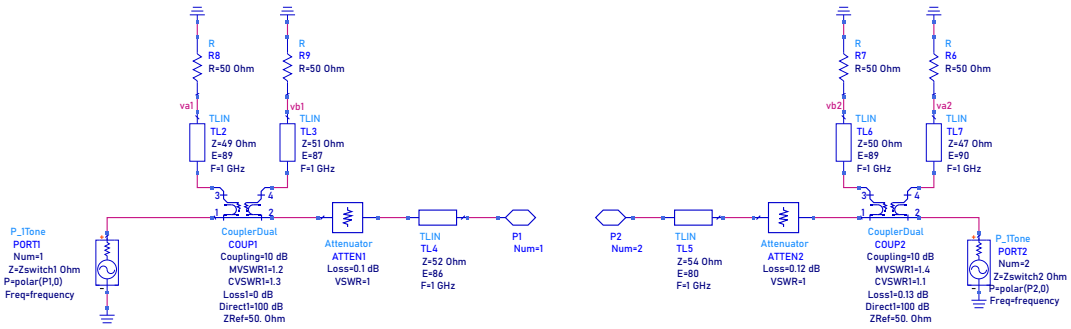


Figure 4.9: Schematic of the non-ideal VNA used for making corrupted measurements of the EM simulated standards. The model includes several non-idealities such as non-equal path lengths and attenuation for port-1 and port-2 and different length a-wave and b-wave paths.

4.4.3. Line verification

To validate the 16-term calibration algorithm, corrupted measurements of the standards and the verification line are taken using the non-ideal VNA. To emulate probe cross-coupling a coupled transmission line is added, between the non-ideal VNA and the DUTs. Figure 4.10 shows the setup including the CLIN element. The measurements of the five standards needed for calibration (M-M, S-S, M-S, S-M, T) are exported as Touchstone '.s2p' files and loaded into MATLAB. To correct for the switching terms the Γ_1 and Γ_2 are also exported using the Thru as the transmissive device. In MATLAB the values of the T-error matrix are found using SVD and then used to correct the measurement of the verification line using eq. (4.8).

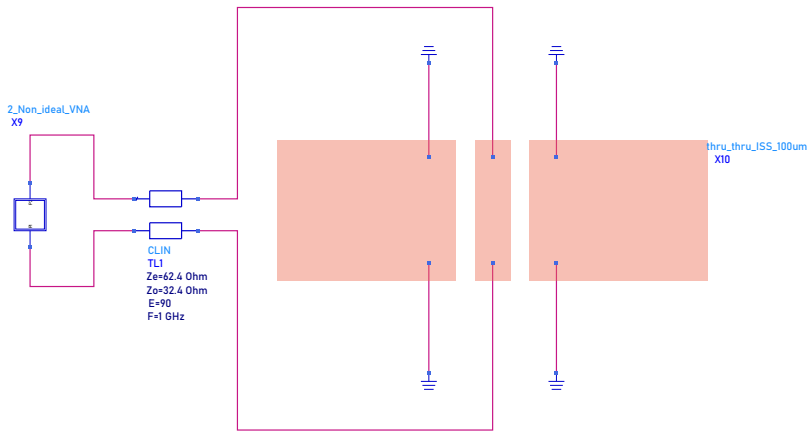
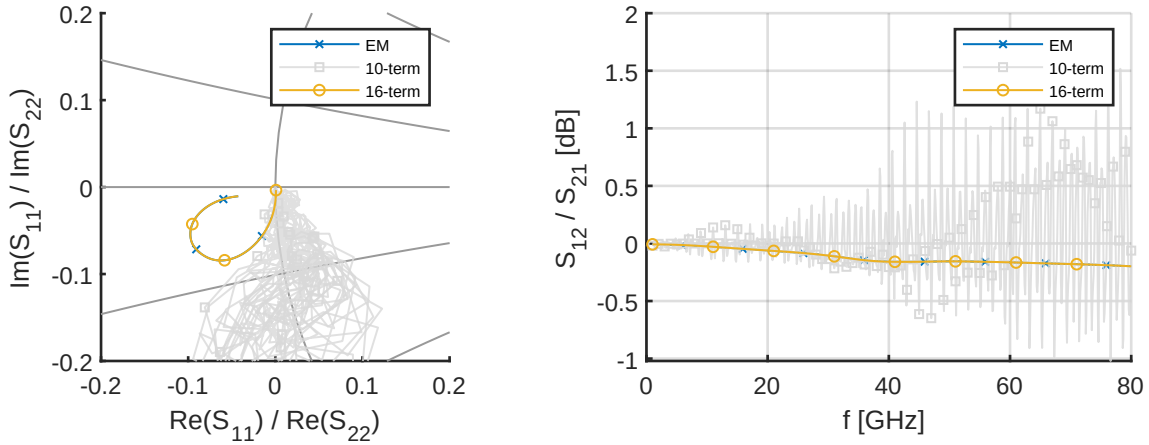


Figure 4.10: A coupled transmission line element is used in simulation to emulate probe-to-probe cross-coupling.

Figure 4.11 shows the S-parameters after first performing the switch term correction and then applying the 10-term and 16-term calibration. The line is perfectly reciprocal and therefore $S_{11} = S_{22}$ and $S_{21} = S_{12}$. As can be seen, the 16-term correction is able to provide an one-to-one correction of the line whereas the 10-term calibration is not.



(a) S_{11} / S_{22} corrected and EM simulated

(b) S_{12} / S_{21} corrected and EM simulated

Figure 4.11: Corrected S-parameters of the verification line using the 16-term error correction algorithm. Since the line is perfectly reciprocal $S_{11} = S_{22}$ and $S_{21} = S_{12}$

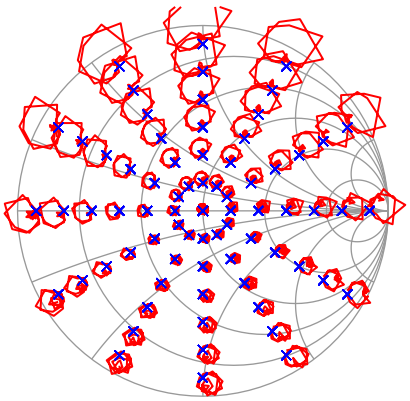
4.4.4. Comparison with 10/12-term SOLT calibration and error bound

The previous section showed that the 16-term calibration algorithm is able to correct measurements of a verification line taken with probe-like cross-coupling. However, the cross-coupling effects are dependent on the S-parameters of the DUT. Therefore it should be checked if the algorithm is able to accurately correct for S-parameters across the Smith chart. In this section the S-parameter depending performance of the algorithm is compared with that of an 10/12-term SOLT calibration. To test the performance of the three algorithms the whole Smith chart is sampled by sweeping the S-parameters of the DUT. Four sweeps were made, one for each of the four scattering parameters as shown in table 4.1 while the other variables were set to a non-zero value. The graphs in fig. 4.12 on the next page show the results of the 12-term SOLT calibration and the 16-term calibration from 40 GHz to 45 GHz.

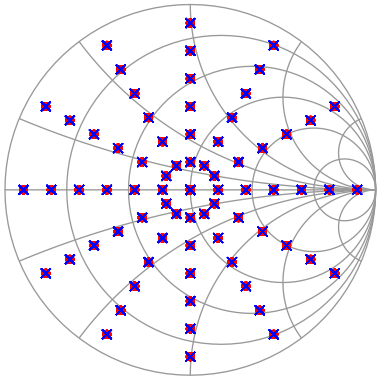
Table 4.1: To test the correction capability of the 16-term error calibration algorithm S-parameters of the DUT are swept across the Smith chart. Each S-parameter was individually swept, keeping the other four constant.

	Sweep 1	Sweep 2	Sweep 3	Sweep 4
S_{11}	$r \cdot e^{-j\phi}$	$0.1 + 0.1j$	$0.1 + 0.1j$	$0.1 + 0.1j$
S_{12}	$0.2 - 0.2j$	$r \cdot e^{-j\phi}$	$0.2 - 0.2j$	$0.2 - 0.2j$
S_{21}	$0.3 + 0.3j$	$0.3 + 0.3j$	$r \cdot e^{-j\phi}$	$0.3 + 0.3j$
S_{22}	$0.4 - 0.4j$	$0.4 - 0.4j$	$0.4 - 0.4j$	$r \cdot e^{-j\phi}$
Parameter sweep:				
r	$0, 0.15, 0.3, 0.45, 0.60, 0.75, 0.9$			
Φ	$0°, 30°, 60°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330°$			

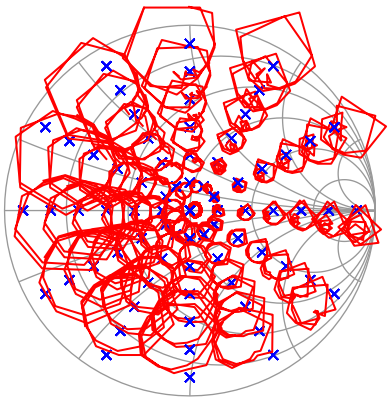
The graphs of fig. 4.12 show that the 12-term SOLT algorithm is not able to correctly resolve the S-parameters of the DUT in case probe-to-probe like cross-coupling is present. The inclusion of the two cross-terms in the 12-term model is therefore not enough to correct for probe-to-probe coupling. Because of the frequency dependency in the cross-coupling the impedances of the 12-term correction rotates around the correct values. Moreover it can be seen that the 12-term algorithm performs worse for S-parameters away from the center of the Smith chart and for the S_{21} and S_{12} parameters. The 16-term calibration algorithm is able to provide a one-to-one correction across the Smith chart.



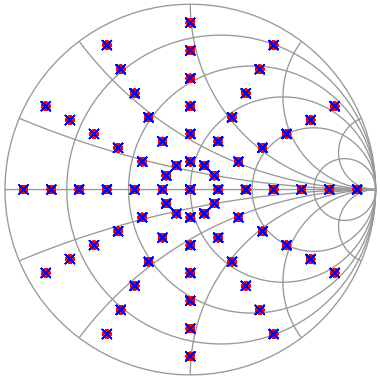
(a) Swept S_{11} corrected using 12-term SOLT calibration



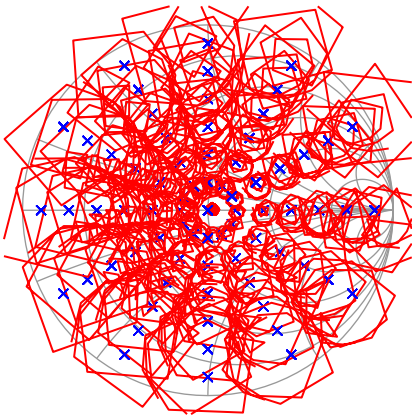
(b) Swept S_{11} corrected using 16-term SVD calibration



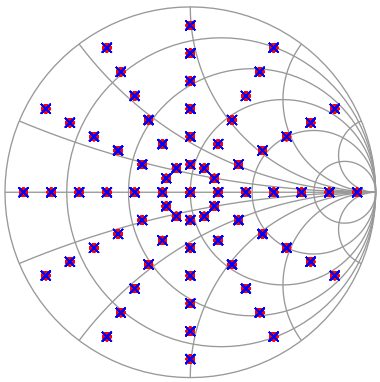
(c) Swept S_{12} corrected using 12-term SOLT calibration



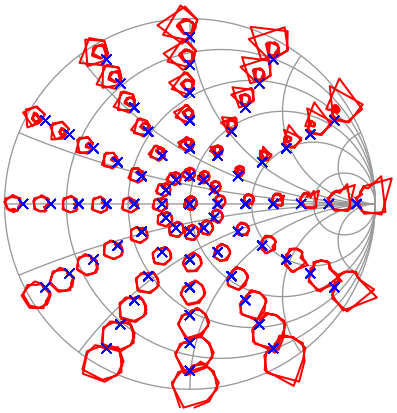
(d) Swept S_{12} corrected using 16-term SVD calibration



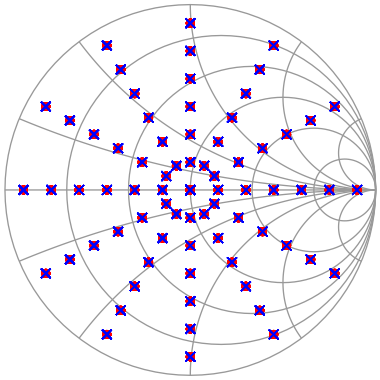
(e) Swept S_{21} corrected using 12-term SOLT calibration



(f) Swept S_{21} corrected using 16-term SVD calibration



(g) Swept S_{22} corrected using 12-term SOLT calibration



(h) Swept S_{22} corrected using 16-term SVD calibration

Figure 4.12: Swept two-port S-parameters corrected by 12-term and 16-term calibrations between 40 GHz and 45 GHz

To qualitatively compare the calibration algorithms the worst case error bound is calculated as followed [5]:

$$WCB(f) = max|S'_{ij}(f) - S_{ij}(f)| \tag{4.20}$$

where S' is the actual point on the Smith chart and S is the point after corruption and calibration. Figure 4.13 shows a comparison of the worst case bound of a 10-term SOLT, 12-term SOLT and the 16-term calibration. What can be seen is that the 10/12-term SOLT calibrations show a standing wave pattern while the 16-term calibration does not. The worst case error for the 16-term calibration is -60 dB or $1 \cdot 10^{-6}$ which is expected from the numerical precision in the SVD computation and the precision when exporting the Touchstone files from ADS to MATLAB. This is accurate enough to allow for use in a real on-wafer measurement.

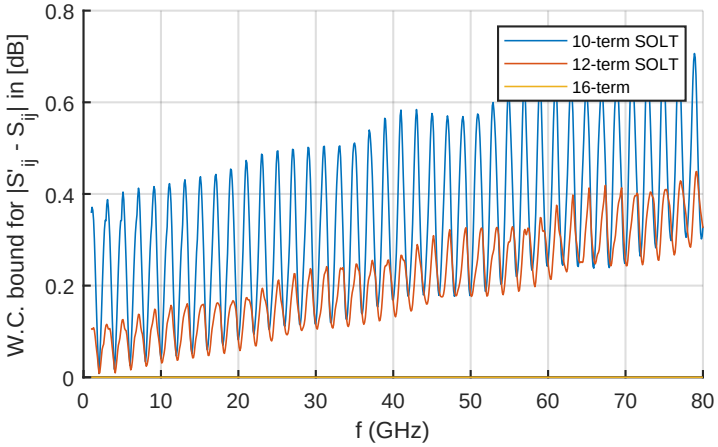
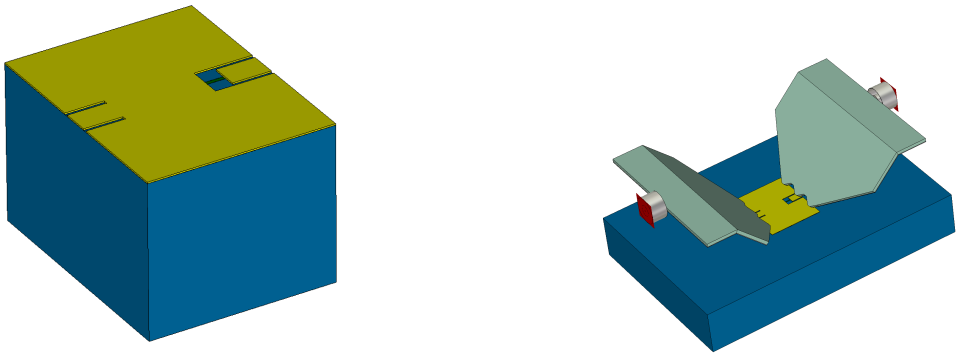


Figure 4.13: Worst case error bound across the Smith chart for 10-term SOLT, 12-term SOLT and the 16-term calibration.

4.5. Probe-to-probe cross-coupling 3D EM simulations

The previous section used coupled transmission lines to simulate probe-to-probe like cross-coupling in VNA measurements. It was shown that the 16-term error calibration is able to resolve this DUT impedance dependent cross-coupling. In this section the analysis is extended towards a full 3D EM simulation of the calibration and DUT structures in CST including models of the probe tips. Two sets of simulation models were used, one set with just the FS standards shown in fig. 4.14a and one set with the standards and probe tip models as shown in fig. 4.14a.



(a) CST model of FS short-load standard

(b) CST model of FS short-load standard with probe tips

Figure 4.14: CST models of FS standards previously made for EM based calibration [55]

The models of the standards and probes that were used are parametrized models that were previously designed and used for EM based calibration [55]. To investigate the probe-to-probe cross-coupling effect the probe spacing d is varied from $60\ um$ to $220\ um$ in the simulations. Figure 4.15 shows the definitions of the dimensions in the standards and verification DUTs used. Several simulation runs were performed with different standards, different spacing, with and without probes. The parameters and simulated standards in these the simulations are shown in table 4.2.

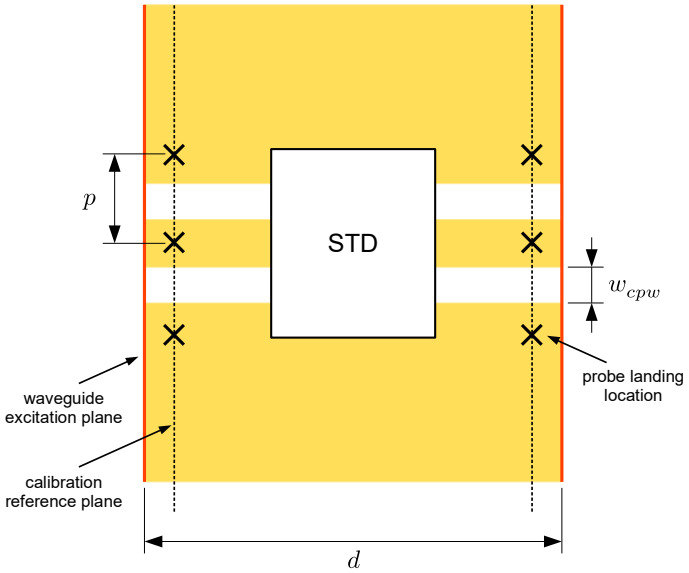


Figure 4.15: Dimension definition of the CST CPW structures with standards.

Table 4.2: Dimensions of the simulated calibration standards.

Property			*Unit*
Probe pitch	p	100	μm
Probe spacing	d	$60, 100, 140, 180, 220$	μm
CPW spacing	w_{cpw}	5.56	μm
Structures		SS, OO, LL, T, LS, SL	

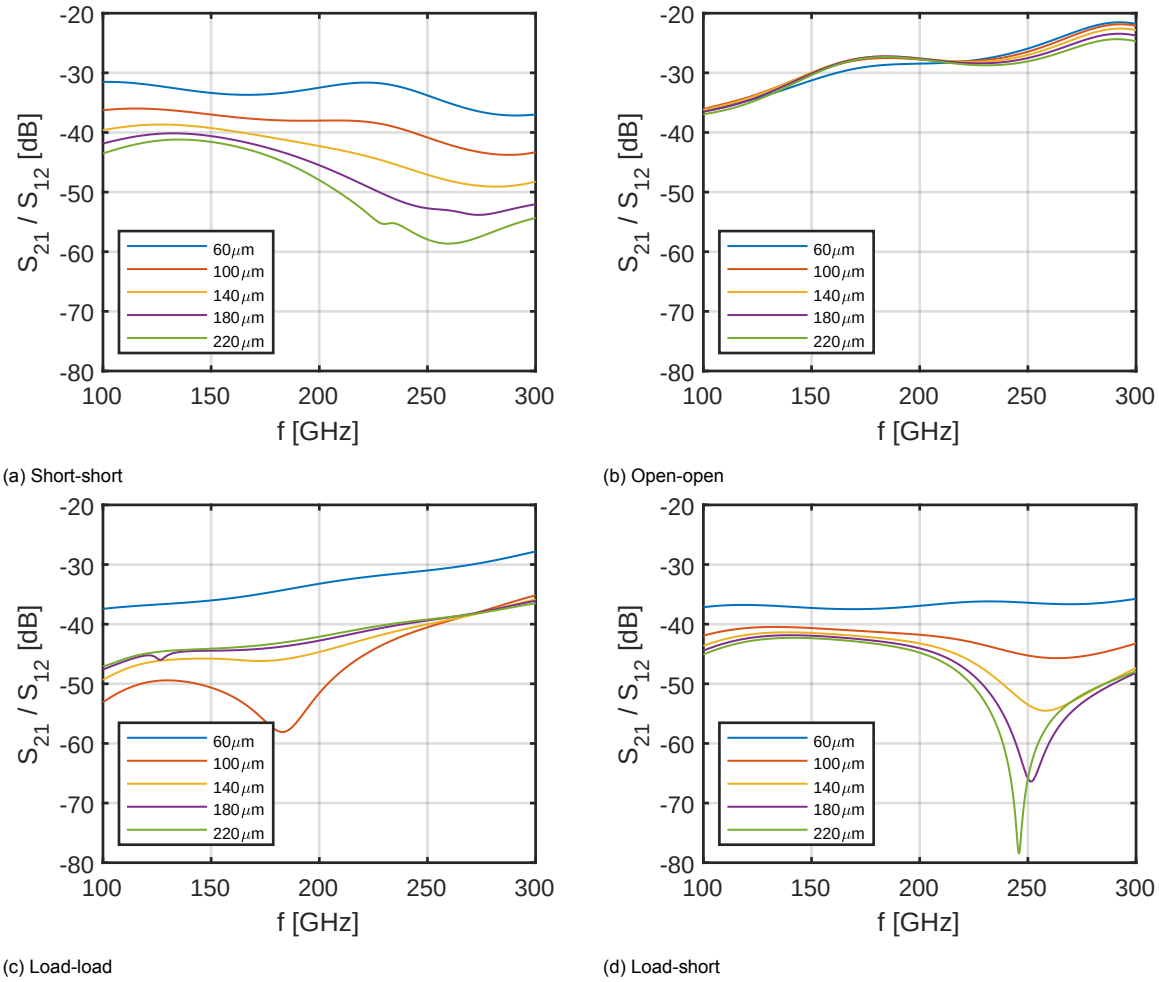
Two types of structures were used to verify the 16-term calibration algorithm and compare its performance with the 10/12-term algorithms. The first structure is a mismatched load-load (LL) with an impedance of 25Ω and the second is a line/thru. Both are simulated for different CPW spacings, resulting in a different characteristic impedances. table 4.3 shows an overview of the parameters of the verification structures that were used.

Table 4.3: Dimensions of the simulated verification structures.

Property			*Unit*
Probe pitch	p	100	μm
Probe spacing	d	$60, 100, 140, 180, 220$	μm
CPW spacing	w_{cpw}	$0.319, 1.308, 5.56, 15.45, 47$	μm
Characteristic impedance	Z_0	$15\Omega, 30\Omega, 50\Omega, 70\Omega, 100\Omega$	μm
Structures		LL (25Ω), T	

4.5.1. Probe-to-probe cross-coupling in 3D EM simulations.

The S_{12}/S_{21} parameters of the un-calibrated simulations of the standards with probes are compared across the different probe spacings to see if a spacing dependent probe-to-probe coupling can be observed. Figure 4.16 shows the un-calibrated simulations of the SS, OO, LL, and SL standards. As the 3D models are fully symmetric the S_{12} and S_{21} are equal and the short-load is identical to the short-load in S_{12} and S_{21}.



(a) Short-short

(b) Open-open

(c) Load-load

(d) Load-short

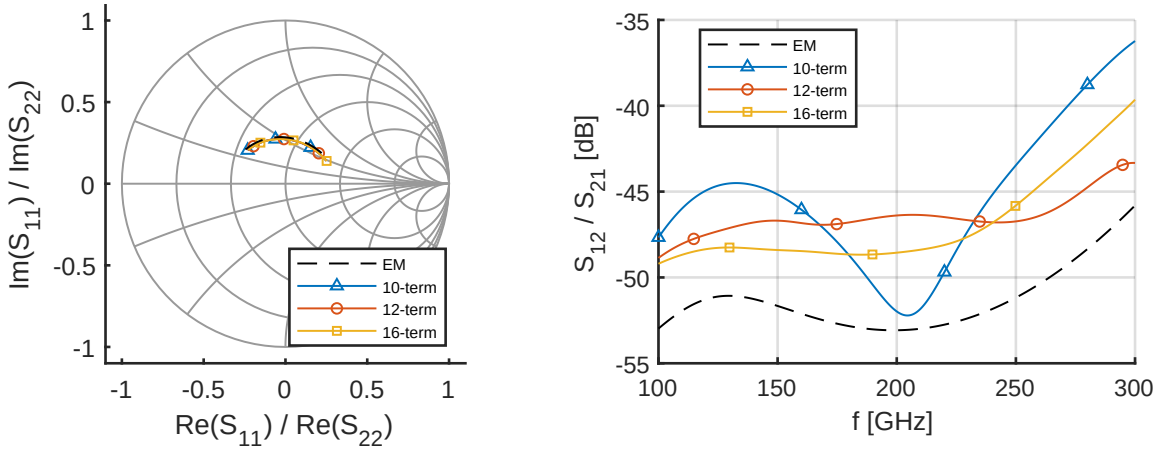
Figure 4.16: S_{21} and S_{12} of un-calibrated SS, OO, LL and LS standards simulated with probes for different spacings

The spacing dependent probe-to-probe coupling can be observed the graphs of fig. 4.16 as the S_{12} and S_{21} vary with the different probe spacings. This variation is however more pronounced in the SS and LS standards than in the LL and OO standards. The SS and LS show a clear trend where the S_{12} and S_{21} decreases with an increase in spacing. The LL and OO do vary for the different spacings, but are not consistent. In this work the combination of the standards SS-LL-LS-SL-T are used in the 16-term calibration to find the error terms. Future research may include an investigation if a difference in calibration quality can be observed if other standards were used (like the OO) that show less probe-to-probe cross-coupling.

4.5.2. Improvement over 10/12-term calibration using spacing dependent calibration

The 16-term calibration requires more standards to be measured during the calibration than the 10 or 12-term calibration algorithms. In addition, the standards also take up more space on an on-wafer or off-wafer calibration kit and could be more expensive to use. It is therefore investigated if in which cases the 16-term calibration algorithm shows significant improvement over the 10/12-term SOLT algorithm.

Figure 4.17 shows the 10, 12 and 16-term correction of the mismatched (25Ω) LL with a characteristic impedance of $Z_0 = 30\Omega$ for a probe spacing of $140\ \mu m$. As can be seen, all three calibration algorithms are able to get relatively close matching values on S_{11} and S_{22}. On the S_{12} and S_{21} parameters the 10 and 12-term algorithms show ripples that do not correspond to trend seen in the EM simulation of the structure. However, since the S_{12} and S_{21} have low values for this structure this does not significantly impact the performance. The 16-term algorithm follows the actual trend of the EM simulation of the structure in the S_{12} and S_{21}. However, it does not show an exact one-one-one match as was seen in ADS simulations of the previous section. This is due to the excitation of the CPW structure is slightly different when excited directly with the waveguide ports in the simulation without the probes as opposed to when the structure is excited with the probes.

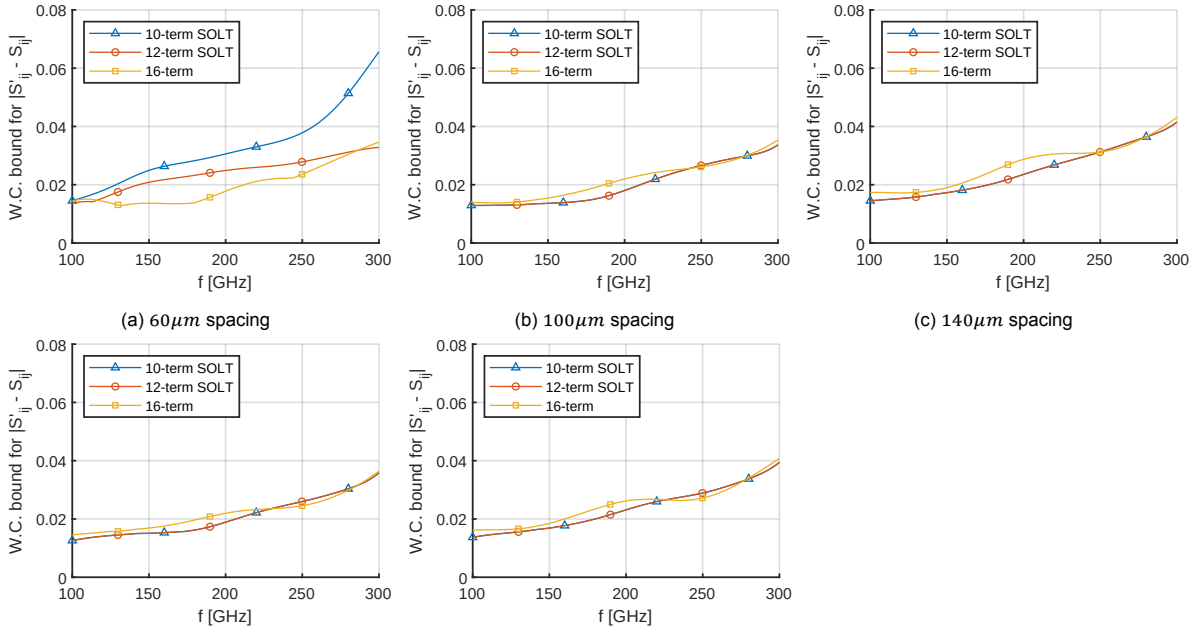


(a) S_{11} and S_{22} corrected and EM simulated

(b) S_{12} and S_{21} corrected and EM simulated

Figure 4.17: Corrected S-parameters of the (reciprocal) verification line using the 16-term error correction algorithm.

A better performance indicator is the worst case error bound, these are shown for the individual (spacing dependent) calibrations on all five spacings in fig. 4.18 for the mismatched (25Ω) LL with a characteristic impedance of $Z_0 = 30\Omega$. Here the same behaviour is seen; for this structure the 16-term



(a) $60\mu m$ spacing

(b) $100\mu m$ spacing

(c) $140\mu m$ spacing

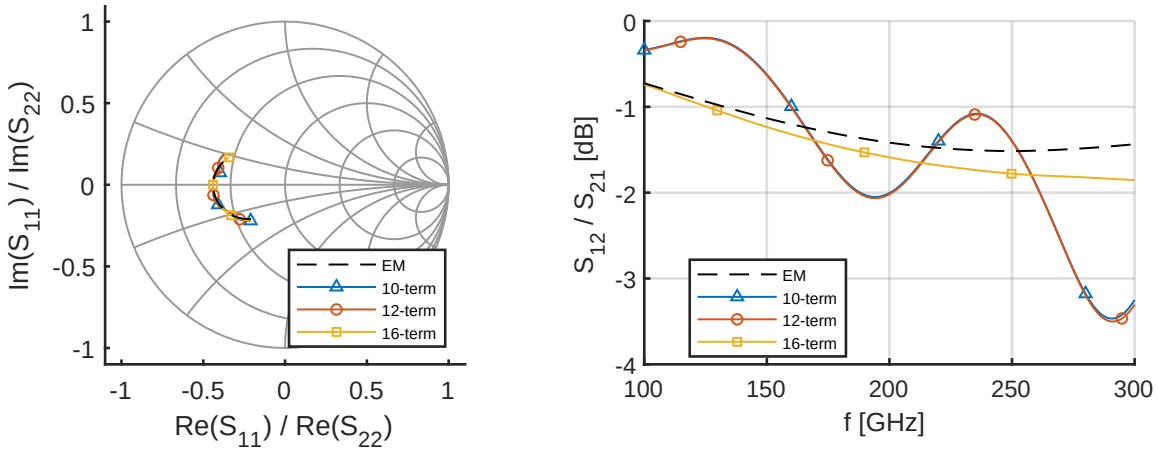
(d) $180\mu m$ spacing

(e) $220\mu m$ spacing

Figure 4.18: Worst case error bounds for the different probe-to-probe spacings for the mismatched (25Ω) load-load with $Z_0 = 30\Omega$. Individual calibrations are used for each different spacing.

calibration algorithm does not provide a significant improvement over the 10 or 12-term calibration algorithms.

Figure 4.19 shows the 10, 12 and 16-term correction of the through line with a characteristic impedance of $Z_0 = 30\Omega$ for a probe spacing of $140\ \mu m$. For this structure a significant improvement can be seen for from using the 16-term calibration algorithm as opposed to the 10 or 12-term algorithms. All three algorithms are able to closely correct the line for the S_{11} and S_{22} however the 10 and 12-term algorithms make a significant error on the S_{12} and S_{21} showing a ripple in the correction which is as expected from a frequency dependent coupling. This result is confirmed by the worst case error bounds shown in fig. 4.20. For each of different probe spacings the 10 or 12-term calibrations make unacceptable high errors, while error of the 16-term calibration algorithm remains constant.

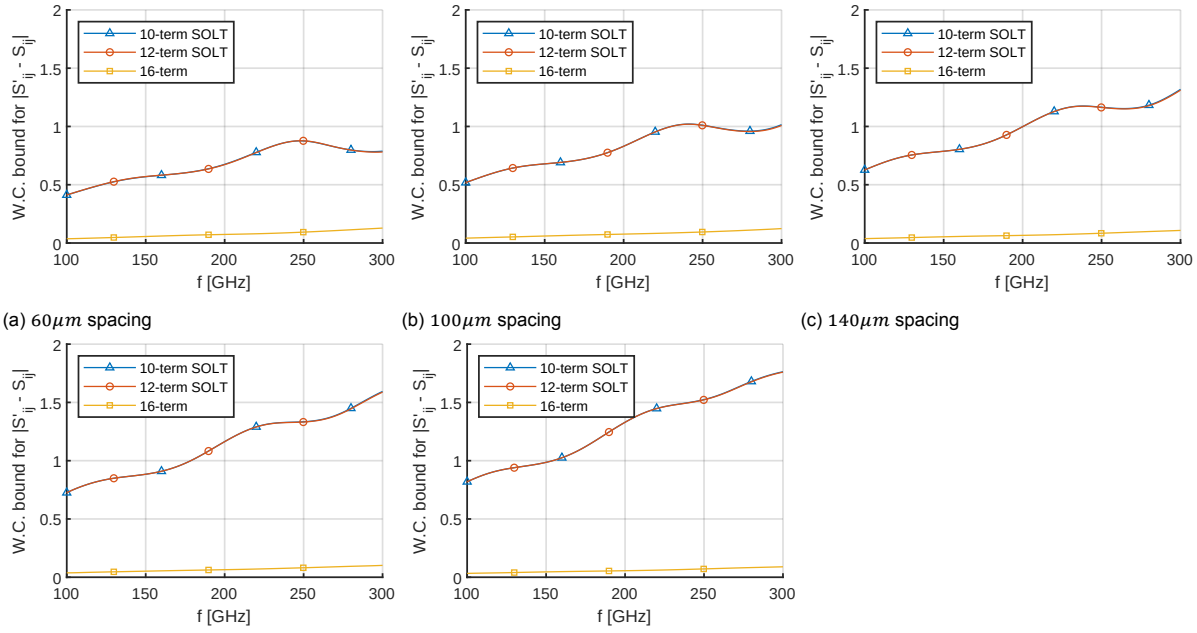


(a) S_{11} and S_{22} corrected and EM simulated

(b) S_{12} and S_{21} corrected and EM simulated

Figure 4.19: Corrected S-parameters of the verification line using the 16-term error correction algorithm

This result confirms what was previously seen in section 4.4.4 where the performance of the 10/12-term algorithm is depending on the S-parameters of the DUT when probe-to-probe coupling is present. For the 10/12-term algorithm the performance is better closer to the center of the Smith chart, and better



(a) $60\mu m$ spacing

(b) $100\mu m$ spacing

(c) $140\mu m$ spacing

(d) $180\mu m$ spacing

(e) $220\mu m$ spacing

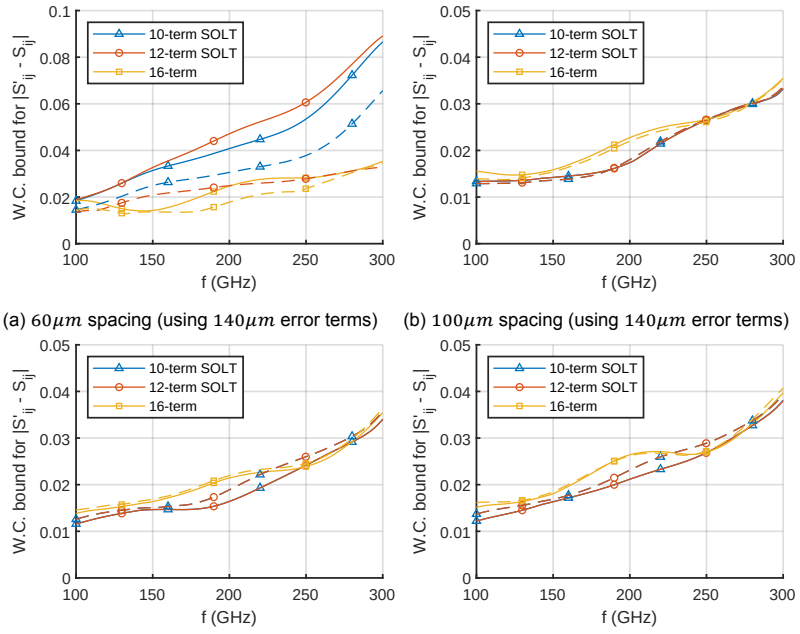
Figure 4.20: Worst case error bounds for the different probe-to-probe spacings for the mismatched (25Ω) LL with a characteristic impedance of $Z_0 = 30\Omega$

for the S_{11}/S_{22} than S_{12}/S_{21}. This explains why the 10/12-term algorithm still has a good performance for the mismatched LL but not for the line. The performance of the 16-term calibration algorithm is independent of the S-parameters of the DUT under probe-to-probe coupling.

4.5.3. Improvement over 10/12-term calibration using fixed spacing calibration

Another test was performed to see if the extra error terms in the 16-term algorithm offer an advantage when the probe-to-probe distance has to be changed during an on-wafer measurement. This could for example happen when the DUT has a (slightly) different spacing than the (off-wafer) calibration kit. The worst case bound is again evaluated on the mismatched 25Ω LL, however this time the error terms from the $140\mu m$ calibration are used for all the spacings. In other words, a fixed spacing calibration is performed that is used on a structures with different spacing. For this test the line verification structure is not used as the 16-term algorithm already showed significant improvement for this structure as opposed to the 10/12-term calibrations.

Figure 4.21 shows the result in the worst-case error bound when using the error terms of the $140\ \mu m$ calibration to correct the mismatched 25Ω LL with $Z_0 = 30\Omega$ for the $60\mu m$, $100\mu m$, $180\mu m$ and $220\mu m$ spacings. To illustrate the performance degradation, the error bound for the spacing dependent calibrations (those of fig. 4.18) are shown in dashed lines. A slight performance increase is seen for the closest spacing of $60\ \mu m$ when using the fixed 16-term calibration over a fixed 10/12-term calibration. For the other spacing, the difference in the error bound is similar when using either a spacing dependent 10-, 12- or 16-term calibration or the fixed spacing calibrations.



(a) $60\mu m$ spacing (using $140\mu m$ error terms) (b) $100\mu m$ spacing (using $140\mu m$ error terms)

(c) $180\mu m$ spacing (using $140\mu m$ error terms) (d) $220\mu m$ spacing (using $140\mu m$ error terms)

Figure 4.21: Worst case error bounds for the $60\mu m$, $100\mu m$, $180\mu m$ and $220\mu m$ probe-to-probe spacings on the mismatched (25Ω) LL with $Z_0 = 30\Omega$ when using the error terms of the $140\mu m$ calibration. The dashed lines show the performance when using a spacing dependent calibration (identical to fig. 4.18)

4.6. Parameterized 16-error terms

The previous sections showed that the 16-term calibration algorithm could offer better performance for on-wafer measurements that are influenced by probe-to-probe cross-coupling, especially for a transmissive DUT. In addition, the 16-term calibration algorithm showed better performance when the probe-to-probe spacing is changed (slightly) after calibration, for example deliberately when one of the DUTs has a different spacing than what was calibrated. This section describes the investigation into a form of the 16-term algorithm that allows for parameterization of (some of) the error terms with respect to probe-to-probe spacing.

In order to arrive at a parameterized form of the 16-term calibration algorithm a closer look needs to be taken into the actual error terms. Keeping in mind that for the 16-term calibration these still represent a (not strictly passive) 4-port S-parameter network that corrupts measurements taken of the DUT. Ideally, the 16-terms are split up into two parts; one non-changing with probe spacing (fixed) and one changing with probe spacing (dynamic).

4.6.1. Split 16-term error model

In the original, non-parameterized 16-term error model (shown in fig. 4.2) the 16 error terms are described in a single 4-port S-parameter matrix:

$$S_E = \begin{bmatrix} e_{00} & e_{03} & e_{01} & e_{02} \\ e_{30} & e_{33} & e_{31} & e_{32} \\ e_{10} & e_{13} & e_{11} & e_{12} \\ e_{20} & e_{23} & e_{21} & e_{22} \end{bmatrix} \tag{4.21}$$

which can be found by applying the four port S-parameter to T-parameter transformation described in [57]. For the parameterized 16-error terms it is assumed that the error model takes the shape of two cascaded four-port S-parameter networks as shown in fig. 4.22.

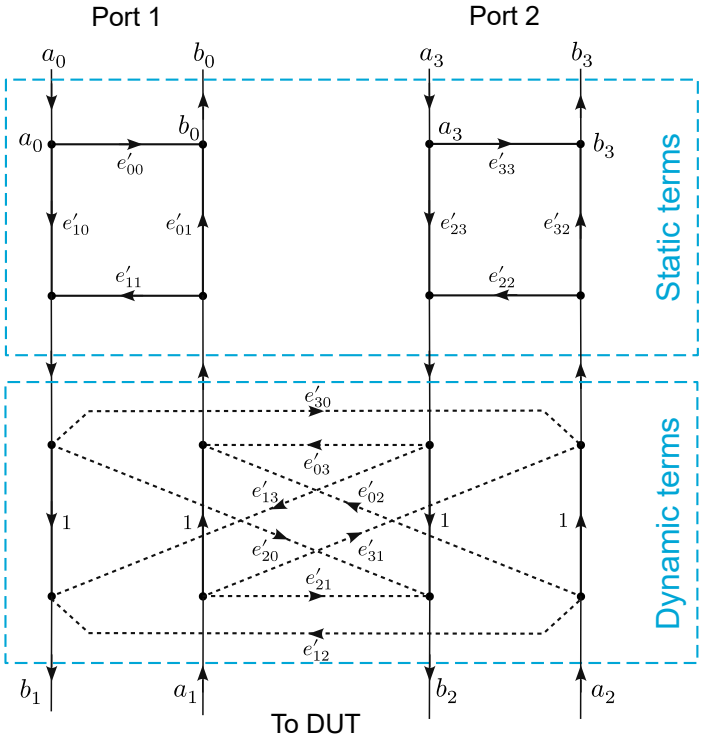


Figure 4.22: For the parameterized version of the 16-term error model the original four-port error matrix is split into a cascade of two four-port networks; one static with no cross/leakage terms and one with dynamic containing the cross/leakage terms.

The first four-port S-parameter network contains the static portion of the error terms, those that do not vary with different probe-to-probe spacing. This network does not include any cross-terms or leakage terms. The second four-port S-parameter network contains the dynamic portion of the error terms, those that do vary with different probe-to-probe spacing. This network includes the eight cross-terms and four terms equal to unity to achieve a transmissive behaviour of the network.

The static error terms may be considered similar to the terms in the 8-term error model. However, an 8-term calibration would not necessarily give the same 8-error terms as with this split 16-term method. When an 8-term calibration is performed on a system where energy is coupled or leaked between ports, some of this energy will contribute to the 8-terms even though the model does not accurately capture

the effects. In this split 16-term approach the 8-terms of the static model are assumed to be the remainder of the error terms in case there is no coupling or leakage in the system.

Mathematically the two four-port S-parameter networks take the shapes:

$$E_{static} = \begin{bmatrix} e'_{00} & 0 & e'_{01} & 0 \\ 0 & e'_{33} & 0 & e'_{32} \\ e'_{10} & 0 & e'_{11} & 0 \\ 0 & e'_{23} & 0 & e'_{22} \end{bmatrix} \qquad E_{dynamic} = \begin{bmatrix} 0 & e'_{03} & 1 & e'_{02} \\ e'_{30} & 0 & e'_{31} & 1 \\ 1 & e'_{13} & 0 & e'_{12} \\ e'_{31} & 1 & e'_{21} & 0 \end{bmatrix} \tag{4.22}$$

The calibration approach proposed here assumes that the static error terms may be extracted from one 16-term calibration, i.e. the nominal spacing calibration, and that these do not vary (significantly) for the other spacings. Then for each spacing there is a set of dynamic terms that represents the coupling and leakage present for that spacing. So when a correction is applied to a measurement, the static error terms are combined with the dynamic error terms according to the spacing used in the measurement.

The next step is to split the 16-error terms into two four-port networks; one with eight non-zero terms and 8 zero terms, and one with eight non-zero terms, four zero terms and four terms equal to unity. Figure 4.23 shows the magnitude of the 16-error terms for each of the different probe-to-probe spacing.

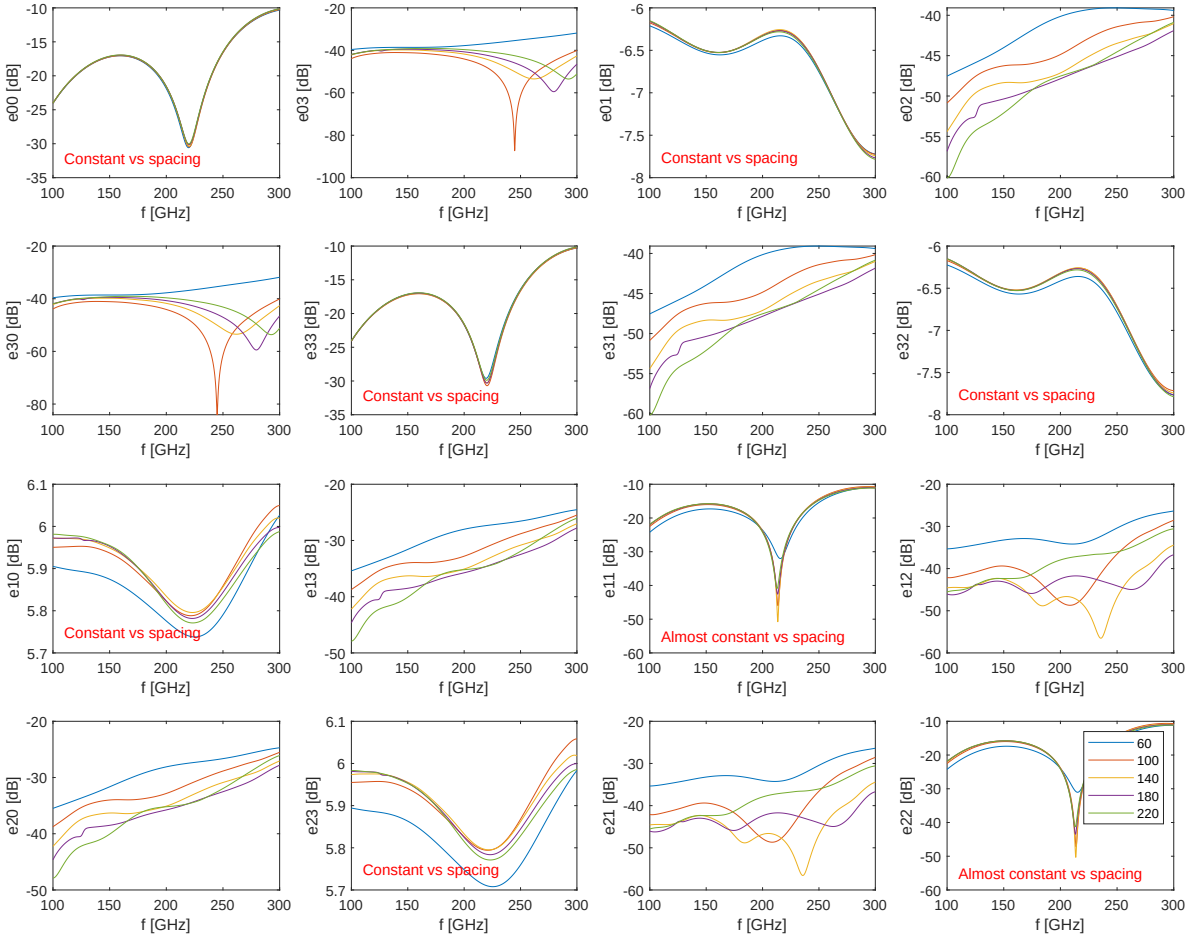


Figure 4.23: Magnitude of the 16-error terms for each of the different probe-to-probe spacing

As indicated in the fig. 4.23, eight of the sixteen error terms are (almost) constant against the different probe-to-probe spacing. As these do not seem to vary across the spacing, their values at the nominal spacing ($140\ \mu m$) are taken for the static error terms:

$$E_{static} = \begin{bmatrix} e'_{00} & 0 & e'_{01} & 0 \\ 0 & e'_{33} & 0 & e'_{32} \\ e'_{10} & 0 & e'_{11} & 0 \\ 0 & e'_{23} & 0 & e'_{22} \end{bmatrix} = \begin{bmatrix} e_{00} & 0 & e_{01} & 0 \\ 0 & e_{33} & 0 & e_{32} \\ e_{10} & 0 & e_{11} & 0 \\ 0 & e_{23} & 0 & e_{22} \end{bmatrix} \quad at\ 140\ \mu m \tag{4.23}$$

The dynamic error terms for each of the different spacing are found using the T-parameter notation. This is advantageous because in the T-parameter notation the cascade of the two four-port networks is a multiplication:

$$T_{16} = T_{static} \cdot T_{dynamic}. \tag{4.24}$$

To find the dynamic error terms first the static error terms are transformed to the T-parameter notation ($E_{static} \rightarrow T_{static}$). Then the 16 dynamic error terms are found by applying the (pseudo-)inverse:

$$T_{dynamic} = (T_{static})^{-1} \cdot T_{16}. \tag{4.25}$$

The resulting dynamic error terms in T-parameter notation may then be transformed to S-parameter notation ($T_{dynamic} \rightarrow E_{dynamic}$) to check the results. These are shown in fig. 4.24. As indicated in the figure, the resulting sixteen error terms of this four-port S-parameter matrix has the right shape as defined for $E_{dynamic}$ in eq. (4.22). The entries on the primary diagonal (top-left to bottom-right) are all close to zero, there are four values approximately equal to unity corresponding and there are eight non-zero values varying with spacing.

The entries on the secondary diagonal (bottom-left to top-right) are all decreasing with increasing probe-to-probe spacing as would be expected from a correct probe spacing parameterized model. The terms e_{03} and e_{30} are unchanged with respect to the full error term representation in fig. 4.23. These error terms represent the non-DUT dependent leakage paths. It is therefore expected that these are uncorrelated with the other error terms. The e_{12} and e_{21} do not follow a clear pattern with the changing of the probe-to-probe spacing. According to the paper by Liu et al. [57] these error terms should represent the air probe-to-probe cross-coupling. Their analysis however assumes all other cross-terms are negligible while in the result of fig. 4.24 shows that this is not necessarily the case and these terms have the same order of magnitude.

During correction of the measurements it is not necessary to transform the dynamic error terms to a S-parameter notation, as the 16-term correction may be fully performed in the T-parameter notation. In this case the T-parameters that are used for the correction is the product of the static T-parameters at the nominal case of $140\ \mu m$ and the spacing dependent dynamic T-parameters:

$$T_{16} = T_{static,nom} \cdot T_{dynamic}. \tag{4.26}$$

These 16-term T-parameters may then be used according to eq. (4.8) to correct DUT measurements.

4.6.2. Interpolation in the parameterized 16-term model

With the approach defined for splitting the 16-term error model in a static and dynamic part it is now possible to investigate if the dynamic parameters may be interpolated. This would allow for measurement of DUTs with different spacing than what is calibrated. To demonstrate this parameterized approach the dynamic error terms of the $60\ \mu m$, $140\ \mu m$ and $200\ \mu m$ are used to interpolate the error terms of the 'in-between' spacing $100\ \mu m$ and $180\ \mu m$. A linear relation is assumed between the value of the dynamic error terms and the probe-to-probe spacing. Therefore in the approach the dynamic error terms are linearly interpolated at each frequency. This interpolation is done in the T-parameter notation.

The dotted lines in fig. 4.24 show the result of interpolating the dynamic error terms for the $100\ \mu m$ and $180\ \mu m$ spacing using the $60\ \mu m$, $140\ \mu m$ and $200\ \mu m$ as the basis. For the e_{02} d_{31} e_{13} and e_{20} parameters a good correspondence with the original error terms is obtained. For the error terms

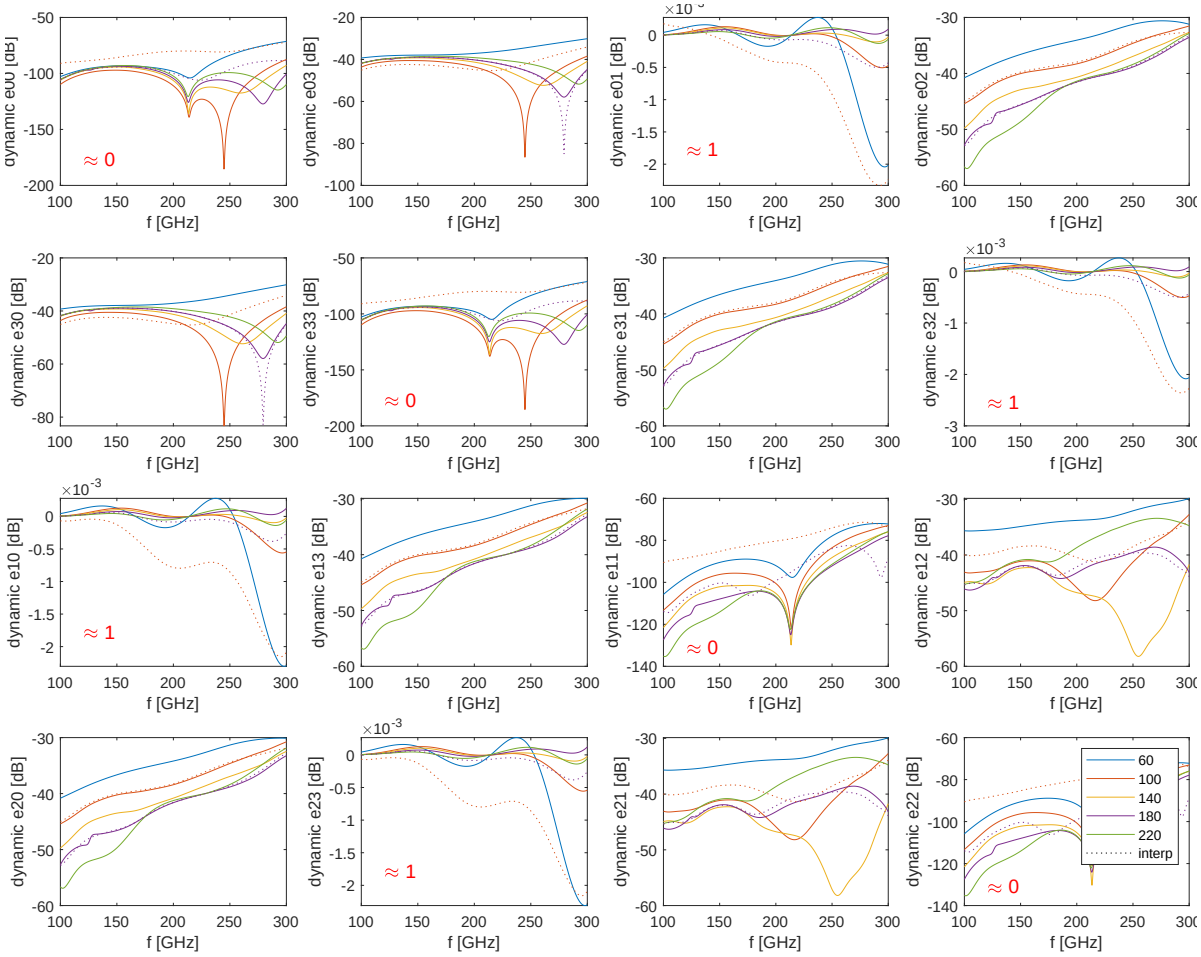
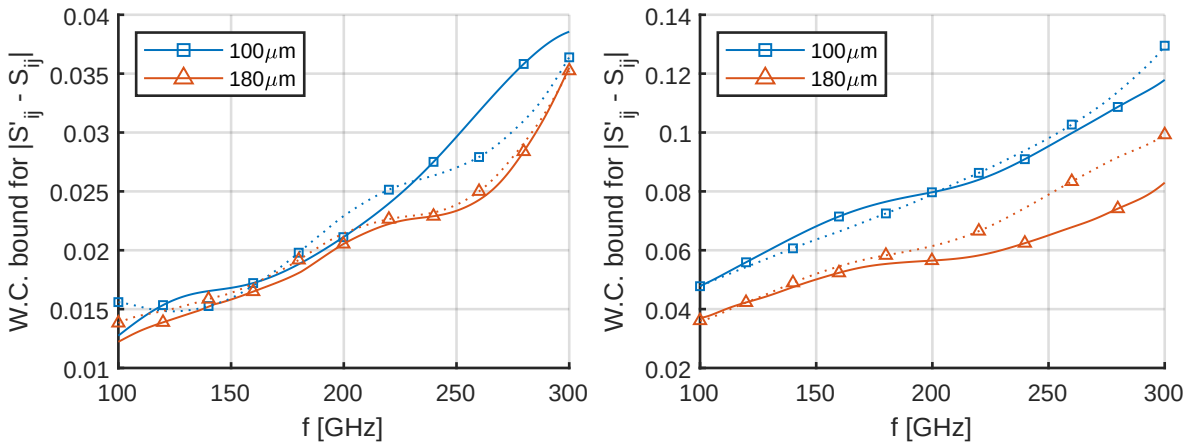


Figure 4.24: Magnitude of the dynamic 16-error terms for each of the different probe-to-probe spacings. The dotted lines show the result after interpolation of the parameterized 16-term model

among the primary diagonal (top-left to bottom-right) the results are not matching, however since the magnitude for these terms is quite low their impact is expected to be less significant. For the other error terms the interpolation has a better result for the $180\ \mu m$ spacing than for the $100\ \mu m$ spacing. The impact of this interpolation error is difficult to determine from just the error terms, as the actual error terms also have a phase component and their effects may combine or cancel out.

4.6.3. Performance of parameterized 16-term model

To asses the performance of the parameterized 16-term approach the worst case error bounds are compared with the original 16-term corrections are compared. The dynamic (interpolated) error terms are combined with the static error terms of the (nominal) $140\ \mu m$ calibration. The error bounds for the mismatched (25 Ω) load-load with $Z_0 = 30\Omega$ and the line with $Z_0 = 30\Omega$ are shown in fig. 4.25. The dotted lines show the performance of the parameterized approach. For both structures the error bound in the parameterized approach stay close to that of the original 16-term calibrations at these spacing. The small difference may be explained by the use of the $140\ \mu m$ static error terms and the interpolation of the dynamic error terms. The errors seen by interpolation of the dynamic error terms at the $100\ \ mum$ does not significantly affect the results. The error of the parameterized approach stays close to the error of the original 16-term approach for both structures meaning that the performance is still comparable. It can therefore be concluded that the parameterized approach, using a split 16-term error model and linear interpolation of the dynamic error terms, can be used to correct measurements that are made at non-calibrated probe-to-probe spacings in an environment where there is significant probe-to-probe coupling.



(a) Mismatched (25 Ω) load-load with $Z_0 = 30\Omega$

(b) Line with $Z_0 = 30\Omega$

Figure 4.25: Worst case error bounds for the $100\mu m$ and $180\mu m$ probe-to-probe spacings on the mismatched 25Ω load-load with $Z_0 = 30\Omega$ and the line with $Z_0 = 30\Omega$ using the original 16-term calibration algorithm (solid) and the parameterized/interpolated 16-term calibration algorithm (dotted).

4.7. Summary & Recommendations

4.7.1. Summary

Millimeter-wave on-wafer measurements may suffer from errors caused by probe-to-probe cross-coupling. It is expected that the error caused by this effect is dominating the overall worst-case bound of 8-term or 12-term calibration methods. Probe-to-probe cross-coupling is however not captured in traditional 8 or 10/12-term error models as these do not include leakage paths that depend on the measured DUT. The 16-term error model, originally proposed by Speciale [56] includes all possible cross-coupling terms and can capture probe-to-probe coupling [9]. A 16-term calibration however requires more standards to be measured and therefore more space on an off-wafer or on-wafer calibration kit. In addition, since the probe-to-probe cross-coupling is dependent on the spacing between the probes their distance cannot vary during or after the calibration and following measurements. Therefore the improvement of a 16-term calibration over the 8 or 10/12-term calibration needs to be justified. State-of-the-art modifications to the 16-term calibration approach allows for two tier calibration methods that include the effects of probe-to-probe coupling [10], these however still require calibration standards for each differently spaced DUT.

In this part of the thesis a 16-term calibration algorithm was implemented in MATLAB. The algorithm includes switch term corrections and uses singular value decomposition to solve the 16-error terms. The algorithm was extensively tested in Keysight ADS using a VNA model that includes the systematic errors made during a S-parameter measurements and used EM simulated two-port fused-silica calibration standards. The 16-term calibration algorithm was able to produce a one-on-one corresponding result when the DUT S-parameters were across the Smith chart. The performance of a traditional 12-term calibration algorithm was shown to be impacted by the DUT impedance dependent cross-coupling. This was also confirmed in the 3D EM simulations including models of the probe tips. The performance of the traditional 10/12-term calibration algorithm with the presence of probe-to-probe cross-coupling was shown to depend on the S-parameters of the DUT and is significantly worse for transmissive DUTs.

Using the validated 16-term calibration algorithm a new calibration approach was formulated using 3D EM simulations of the standards. In this calibration approach the 16-term error model is split after calibration for several different spacings into two independent parts, one that is non spacing dependent (static) and one that is spacing dependent (dynamic). This split 16-term error model allows for a parameterized approach where error terms for a non-calibrated spacing may be interpolated from the different spacings that are calibrated. Therefore not requiring calibration structures for each differently spaced DUT.

4.7.2. Recommendations

The developed 16-term calibration algorithm and parameterized calibration approach is ready to be tested in an on-wafer environment. A fused silica wafer that was produced prior to this work, shown in fig. 4.26, may be used for this purpose. This calibration kit includes mixed calibration standards and structures that allow measurements with different probe-to-probe spacing.

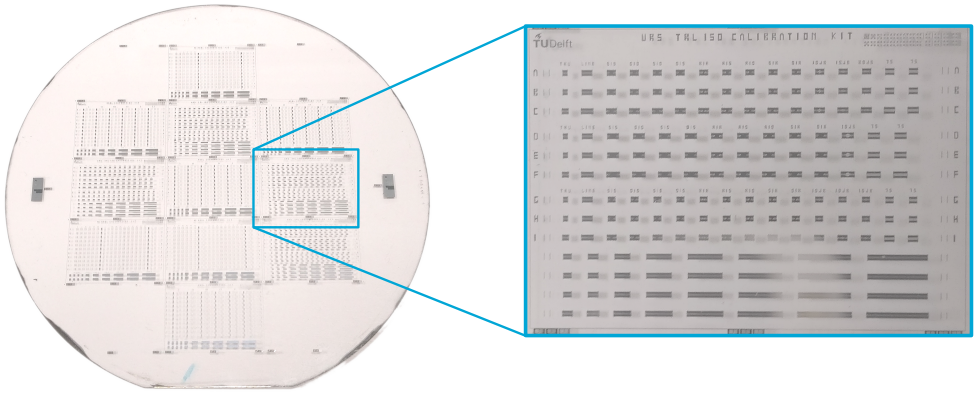


Figure 4.26: Fused silica wafer with mixed calibration structures at different probe-to-probe spacing to be used for (parameterized) 16-term calibration

It should however be confirmed if the calibration kit is suitable for measurement in the WR-5.1 band (140 - 220 GHz) as this is the band where the probes are suspected to have a significant cross-coupling. The MATLAB implementation of the algorithm is designed to work with Touchstone '.s2p' files and works with the RAW outputs of the VNA or WinCAL. The primary recommendation for future work is therefore to perform measurements on the fused silica wafer and confirm if the proposed calibration approach works in practice.

From the 3D EM simulation it was found that the open-open standard shows less probe-to-probe cross-coupling than the short-short standard. According to the papers by Heuermann [58] and Silvonen [59] either open or short combinations may be used. However, in this work it was not investigated if the use of open standards instead of short standards makes a difference in the performance of the calibration. This could therefore be included in future work or during an on-wafer measurement if theses structures are available.

Finally, the 16-term algorithm and this MATLAB implementation allows for more than the minimum of 4 or 5 standards to be used in the calibration. By including addition measurements of standards the redundancy could be used to improve the determination of the error terms. This could be particularly useful in low signal-to-noise cases, as more standards or repeated measurements allow for the removal of uncorrelated noise.

5

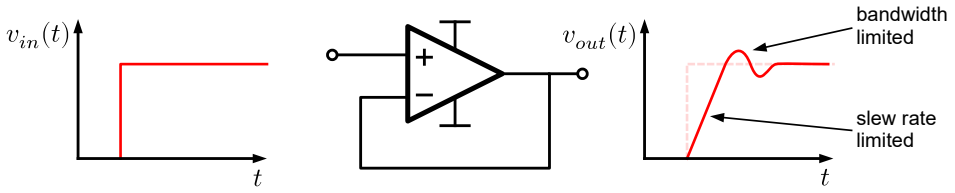
Preliminary investigation of device limiting effect when approaching f_T/f_{max}

In this chapter an investigation is conducted into the behaviour of active devices when operating at frequencies close to the transition frequency f_T or maximum frequency of oscillation f_{max}. An experiment is setup that could show changes across frequency in large signal operating modes of the active device.

5.1. Metrics for frequency limitations

In analog circuits, metrics such as the gain-bandwidth (GBW) product and slew-rate are used to describe the limits in speed and accuracy of (feedback) circuits. Both contribute differently in for example the speed to which the output, i.e. of an op-amp, settles to a step change in the input as illustrated in fig. 5.1. The maximum oscillation frequency f_{max} may be considered as the RF active device equivalent of the analog GBW product, relating achievable gain to operating frequency. The f_{max} is however a small signal metric and the question arises whether there exists a device limitation equivalent to slew-rate in analog circuits.

Bandwidth & slew rate limitations in **analog circuits:**



Device limitation operating close to f_T/f_{max} in **RF circuits:**

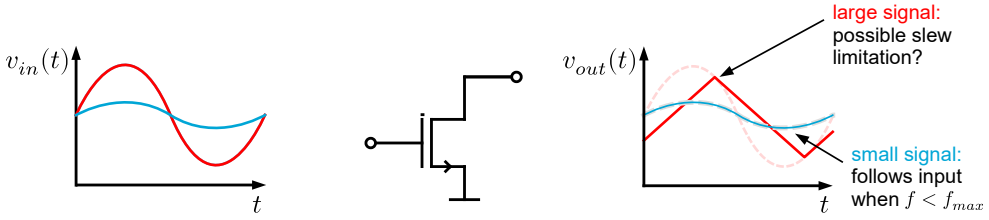


Figure 5.1: With power amplifiers operating close to the f_T/f_{max} of the active device the question arises whether a device speed limitation is present during the large signal operation that results in a slew-rate like operation similar to that in analog circuits.

A device limitation of active devices representing itself as a slew-rate effect could explain why it is increasingly difficult to generate large signal power at higher frequencies. The publication by Johnson [24] states that the ability for power amplification depends upon the volts/second capability of the semiconductor and can be directly related to the ”Johnson’s figure of merit” discussed in section 2.1. E. McCune relates slew-rate limited operation to the low achievable efficiencies for millimeter-wave PAs [64] and states that the theoretical transition to slew-rate limited operation should occur at $f_T/2\pi$. On a physical level a slew-rate like behaviour could possibly be explained by non-quasi static (NQS) effects. NQS effects originating from the finite response time of free carriers to changes in applied bias [65]. On this area active research is conducted to include these effects in device compact models (i.e. bi-polar [66], MOS [57]). However, this problem is mainly approached from a small-signal perspective and does not consider large-signal effects.

As it is now possible to design PAs at a fraction of the f_T/f_{max} of the technology, i.e. by minimizing layout parasitics [67, achieving $f_T/2.2$], it now becomes important to understand the large signal behaviour at these frequencies. This chapter describes a preliminary investigation into developing measurement methods that may allow for observation of a device limiting effects in large signal operation when operating close to the f_T/f_{max} of the device.

5.2. Goal, research question & structure

The goal of the work described in this chapter is to investigate how a device limiting effect, similar to slew-rate limiting in analog circuits, may be observed from large signal measurements of transistors operating close to their f_T/f_{max}. To do so a measurement approach is to be formulated that allows for identification of the device limiting behaviour using millimeter-wave active load-pull measurement setups. The research question associated to this chapter is formulated as followed:

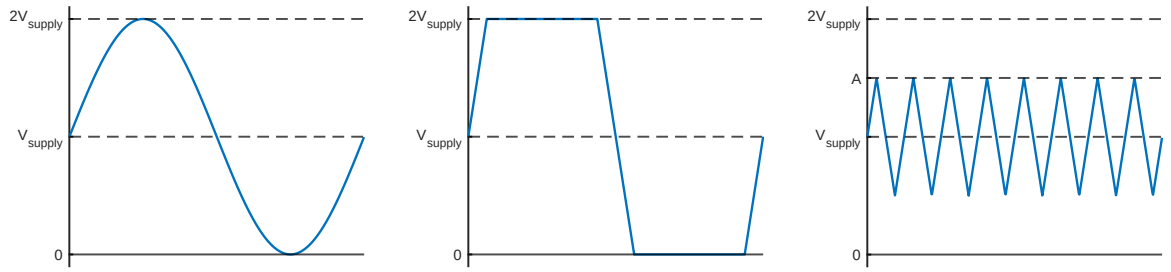
How can a millimeter-wave characterization method be formulated that allows for observation of large signal device limiting effects in transistors operating close to their f_T/f_{max}?

In section 5.3 an approach is formulated for an experiment that could show slew-rate like behaviour in transistors from trends across frequency in large signal characterization. A metric is defined as the change in output power by the change in supply voltage, which in case of slew-rate like behaviour should decrease when operating closer towards the f_T/f_{max}. In section 5.4 several device models are used in simulation to show how this metric should behave and to be able to later compare this behaviour with the measurement results. In section 5.5 experimental results of a large signal measurement in a 50Ω environment on the Global Foundries FD-SOI slvt-nmos2 transistor are presented. In section 5.6, similar experimental results are presented however then using a millimeter-wave active load-pull capable setup.

5.3. Power saturation when operating close to f_T/f_{max}

When a step response is applied to the input of a slew rate limited analog circuit the output requires a finite time to transition from the original value and settle to a new value as shown in fig. 5.1. In RF amplifiers the output never settles to a value as sinusoidal signals are used. However, a slew rate effect in RF amplifiers would manifests itself as a limiting effect in the achievable output power. This is illustrated in fig. 5.2, showing hypothetical output waveforms in three modes of operation of an RF amplifier stage: linear mode, switching mode and slew rate limited mode. In the linear mode the transistor is driven with an input power that just saturates the devices output power, i.e. no (soft) clipping occurs and the output waveform is rail-to-rail[1]. When the input power is increased and the transistor is effectively driven as a switch and a switching mode of operation is achieved (i.e. class-E operation) where the output of the transistor switches between an on-state and off-state with some finite transition between the states. When the operating frequency is significantly lower than the f_T of the device the transition times have little influence on the output waveform, which then still takes a square shape. However, when the frequency is increased, the transition time (limited by the slew rate of the device) becomes significant with respect to the period of the signal. The effect is that the amplitude of the waveform will not be able to reach the peak value of V_{supply} within the transition time and a triangular waveform is obtained. This effect may already happen even when the transistor is not

[1] Such a waveform would be difficult to create in a real waveform due to soft non-linearities and a finite r_{on} of the transistor.



(a) **Linear operation**
$f_0 << f_T$
No clipping
P_{out} relates to P_{in} and V_{supply}

(b) **Switch mode operation**
$f_0 << f_T$
Hard clipping
P_{out} relates to V_{supply} and f

(c) **Slew rate limited operation**
$f_0 > f_T/10$
Slew limiting
P_{out} relates to f

Figure 5.2: Theoretical voltage waveforms at the output terminal of the device assuming no (harmonic) filtering. When operating close to the f_T of the device a slew rate limiting mode may occur that limits the amplitude the waveform can reach, therefore limiting the device output power.

operated as a switch if simply the rate of change in the sinusoidal signal is higher than the apparent maximum slew rate of the device (i.e. by the response of the charge carriers). In this case the output power of the amplifier becomes independent of the supply voltage, as the output voltage never reaches the rail voltage.

If the device behaviour close to the f_T is described by a slew rate limit this would mean that as the frequency increases the output power would decrease. The slew rate in volts per second required for a sinusoidal waveform with peak amplitude V_p is equal to:

$$\begin{aligned} sr &\geq max\left|\frac{dv(t)}{dt}\right| \\ &\geq max\left|\frac{dV_p \cdot sin(2\pi ft)}{dt}\right| \\ &\geq 2\pi f V_p \cdot max\left|cos(2\pi ft)\right| = 2\pi f V_p. \end{aligned} \tag{5.1}$$

Equivalently, a sinusoidal signal that is being slew rate limited with a max rate of sr, results into a triangular output waveform that reaches a peak amplitude in a quarter of the period of:

$$V_p = sr \cdot \frac{T}{4} = sr \cdot \frac{1}{4f} \tag{5.2}$$

As can be seen in eq. (5.2), increasing the frequency reduces the peak amplitude of the output waveform. However, being able to identify behaviour as in fig. 5.2c is challenging above 100 GHz as direct observation of the waveform is difficult. Observation of the power in the harmonics with respect to the fundamental power would allow to differentiate between the output waveforms as shown from the Fourier series decompositions in table 5.1.

Table 5.1: Theoretical power observed at fundamental frequency and second/third harmonics for waveforms with amplitude $V_p = A$ from the Fourier series of the waveform

Waveform	b_n	$R_L \cdot P_{f_0}$	$R_L \cdot P_{2f_0}$	$R_L \cdot P_{3f_0}$
Sine	$\begin{cases} V_p & n = 1 \\ 0 & n \neq 1 \end{cases}$	$\frac{1}{2}V_p^2$	0	0
Square	$\frac{4V_p}{n\pi}\begin{cases} 1 & n\ even \\ 0 & n\ odd \end{cases}$	$\frac{8}{\pi^2}V_p^2$	0	$\frac{8}{9\pi^2}V_p^2$
Trapezoid ($t_{rise} = \pi/4$)	$\frac{8\sqrt{2}V_p}{\pi^2}\left(\frac{sin(n\pi/4)+sin(3n\pi/4)}{n^2}\right)$	$\frac{64}{\pi^2}V_p^2$	0	$\frac{64}{81\pi^2}V_p^2$
Triangle	$\frac{8V_p}{\pi^2 n^2}\begin{cases} (-1)^{(n-1)/2} & n\ even \\ 0 & n\ odd \end{cases}$	$\frac{32}{\pi^4}V_p^2$	0	$\frac{128}{81\pi^4}V_p^2$

Because of the octave-wide frequency extension modules used in the current millimeter-wave active load-pull setups it is only observe the power at the fundamental frequency. Therefore, a different method should be formulated to be able to observe slew rate like limiting behaviour as in fig. 5.2c.

5.3.1. Supply voltage dependency of the output power versus frequency

Under the hypothesis that when the frequency increases towards the f_T, the output waveform transitions from a sinusoidal (fig. 5.2a) waveform to a triangular waveform (fig. 5.2c), then the supply voltage dependency of the output power decreases. In other words, when a slew-rate like limiting behaviour is present, the $\delta P_{out}/\delta V_{supply}$ across frequency decreases towards higher frequencies. It should eventually reduce to zero when the output waveform is triangular and not increasing in amplitude with an increase in supply voltage. This results into a behaviour as illustrated in fig. 5.3.

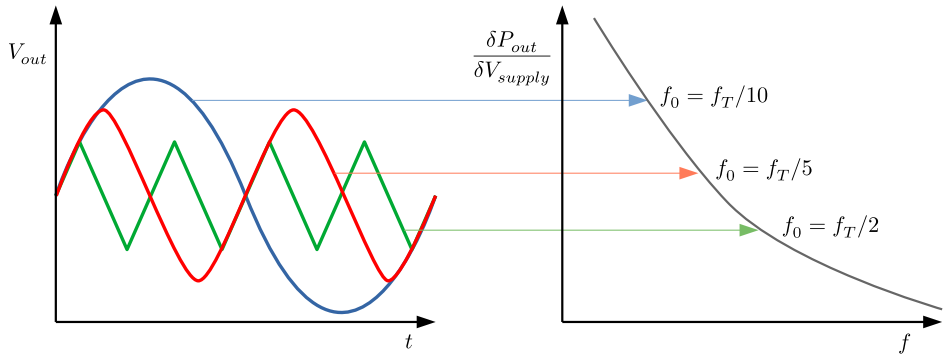


Figure 5.3: Concept behind the $\delta P_{out}/\delta V_{supply}$ metric across frequency. When the operating frequency is increase towards the f_T of the transistor the output power dependency on a change in supply voltage decreases, indicating a device limiting effect that shows a slew-rate like behaviour.

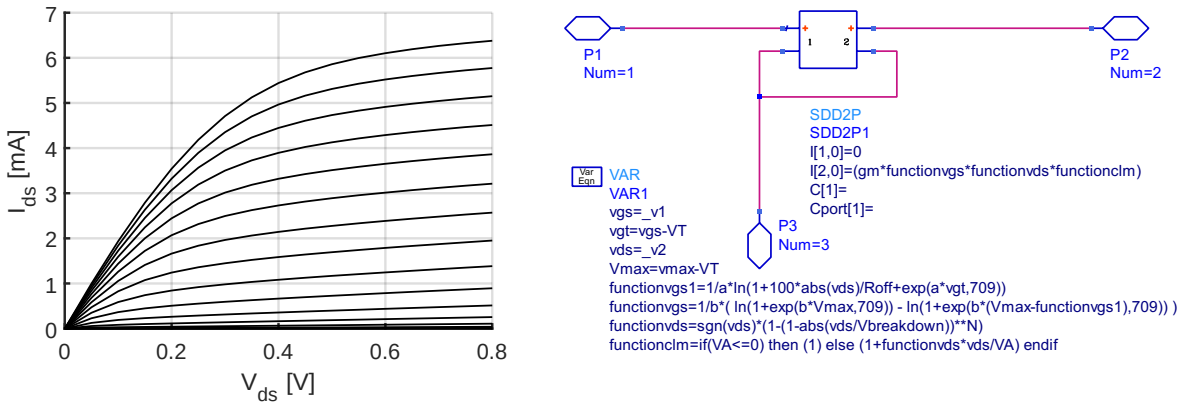
To measure this metric in a large signal measurement, the supply voltage is varied, i.e. the V_{ds} for a MOS or V_{ce} for a bipolar, while the change in output power of the DUT is observed. This procedure is then repeated at different frequencies up to the f_T/f_{max} of the device using frequency extenders in order to obtain the $\delta P_{out}/\delta V_{supply}$ metric across frequency. This comparison across frequency may be done regardless of the power gain of the transistor, which decreases as dictated by the f_{max} figure of merit, as only the change in output power for a change in supply voltage is used. Therefore, the absolute output power at a particular frequency is irrelevant. Ideally the slope of the ratio metric power gain $g_p = \delta P_{out}/\delta P_{in}$ is forced to zero by operating the device as a switch (as in fig. 5.2b) but this requires the device to be measured (far into) compression, up to saturation, which is difficult to do with the limited available power from frequency extension modules. To avoid different modes of operation of the transistor the supply voltage should be varied only a small amount during this measurement. The resulting relation between δP_{out} and V_{supply} may then be considered linear with a slope equal to $\delta P_{out}/\delta V_{supply}$.

5.4. Simulation results

To illustrate the concept, the $\delta P_{out}/\delta V_{supply}$ metric is simulated against frequency for simple MOS transistor models increasing in complexity with finally including non-quasi-static parameters. The models are not meant to be accurate representations of device behaviour close to f_T/f_{max}, they are only intended to show which elements may or may not influence the $\delta P_{out}/\delta V_{supply}$ metric against frequency.

5.4.1. IV-curve based

The first model that is used is a simple IV-curve based model. The IV-curves measured for the Global Foundries 22nm slvt-nmos2 transistor shown in fig. 5.4a are mapped on a transistor behavioural model proposed by S.C. Cripps [68] that includes a knee-function and channel length modulation. Figure 5.4b shows this model implemented in ADS by R. Bootsman.



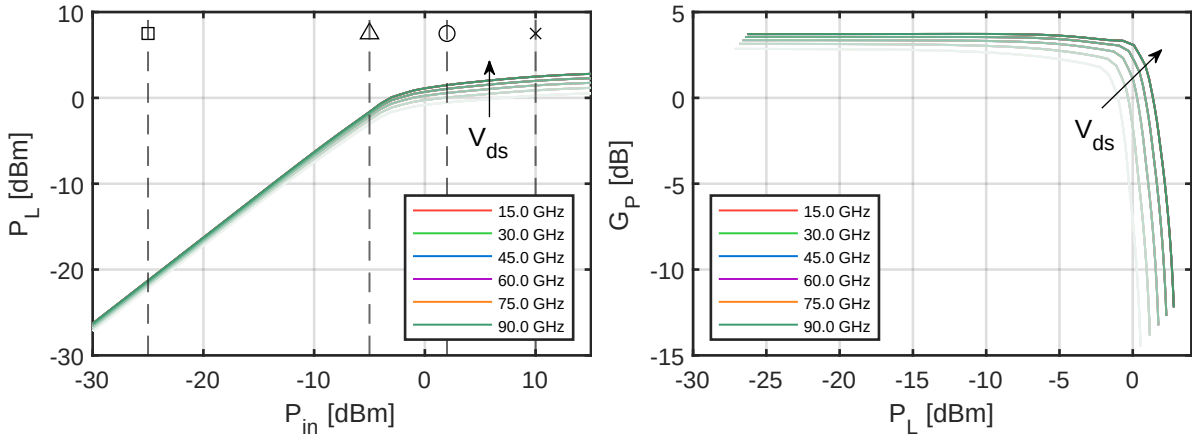
(a) Measured static IV-curve of the Global Foundries slvt-nmos2 transistor

(b) Implementation of transistor behavioural model proposed by S.C. Cripps [68] with added channel length modulation (created by R. Bootsman)

Figure 5.4: A measured IV-curve of a transistor is mapped to a behavioural model to show the $\delta P_{out}/\delta V_{supply}$ metric when assuming no frequency dependent behaviour.

A harmonic balance simulation was run for this static model using an output loading of 50Ω. The available source power was swept from -30 dBm to 15 dBm, the drain bias voltage V_{ds} from 0.6V to 0.8V, and the frequency from 25 GHz to 90 GHz. The gate bias voltage is fixed at $V_{gs} = 0.6V$. The input power to output power relations of the model are shown in fig. 5.5a. As there is no frequency behaviour in the model the curves overlap for each of the frequencies.

There are two distinct regions of operation for this transistor. Using the definitions defined by McCune in [69] two modes can be distinguished. Up until $P_{in} \approx 5dBm$ the transistor is operating in L-mode where the output power solely dependent on input power. In the L-mode region, the metric of interest is approximately zero $\delta P_{out}/\delta V_{ds} \approx 0$ and the gain is constant $\delta P_{out}/\delta P_{in} = C$. As the device is pushed further into compression eventually reaching output power saturation, the transistor starts to operate in C-mode where the output power is solely dependent on supply voltage. In the C-mode region, the metric of interest is non-zero $\delta P_{out}/\delta V_{ds} \neq 0$ while the power gain is approximately $\delta P_{out}/\delta P_{in} \approx 0$.



(a) Input power to output power relation. Darker colors show increase in supply voltage (direction of arrow).

(b) Power gain versus output power. Darker colors show increase in supply voltage (direction of arrow).

Figure 5.5: Input power to output power relation in the static IV based model. As no frequency behaviour is implemented the curves overlap for all frequencies. This also results into a constant $\delta P_{out}/\delta V_{ds}$ trend versus frequency.

The $\delta P_{out}/\delta V_{ds}$ metric is computed at four input power levels, -25 dBm, -5 dBm +2 dBm and +10 dBm, indicated by the dashed lines in fig. 5.5a. For each input power level the output power is plotted as function of the supply voltage, i.e. $P_L(V_{ds})$. From the resulting points the slope is calculated, resulting

into the $\delta P_{out}/\delta V_{ds}$ metric. This metric then has the unit of milliwatts per volts [mW/V]. The results are shown in fig. 5.6. At low drive levels (L-mode operation) the $\delta P_{out}/\delta V_{ds}$ is close to zero. At higher drive levels operating further into compression (C-mode operation) the $\delta P_{out}/\delta V_{ds}$ metric is higher. Because no frequency behaviour is implemented in this model the $\delta P_{out}/\delta V_{ds}$ is constant with frequency.

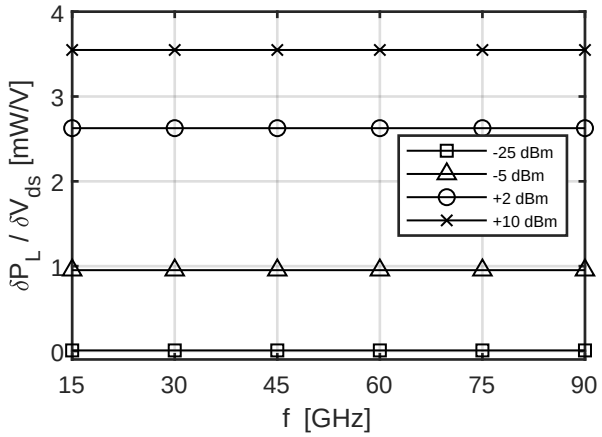
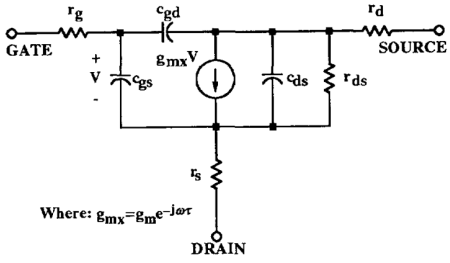


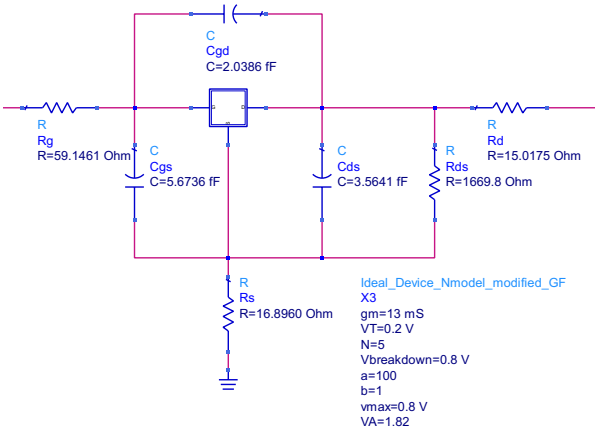
Figure 5.6: $\delta P_{out}/\delta V_{ds}$ metric at four drive levels for the IV-based model

5.4.2. IV-curve based with device extracted parameters

Device extracted parameters are added to the ideal IV-curve based model to get an idea of how the $\delta P_{out}/\delta V_{ds}$ metric changes when a frequency dependency is included. The model is extended with parameters extracted from an on-wafer small signal measurement of the Global Foundries slvt-nmos2 transistor in WR-10 band (67 GHz to 110 GHz). The small signal parameter extraction method for silicon metal oxide semi-conductor (MOS) transistors that is used, is described in a paper by Lovelace et. al. [70]. This method maps the small signal parameters on the model shown in fig. 5.7a. For a typical small signal model the device transconductance would also be extracted from the measurements. However, in this analysis large signal behaviour is investigated and therefore the IV-curve based transconductance is used as a small signal based transconductance would not show non-linear behaviour. This results into a quasi-large signal model, as the transconductance is large signal based but the capacitances used in the model are linear. The ADS implementation of this model is shown in fig. 5.7b.



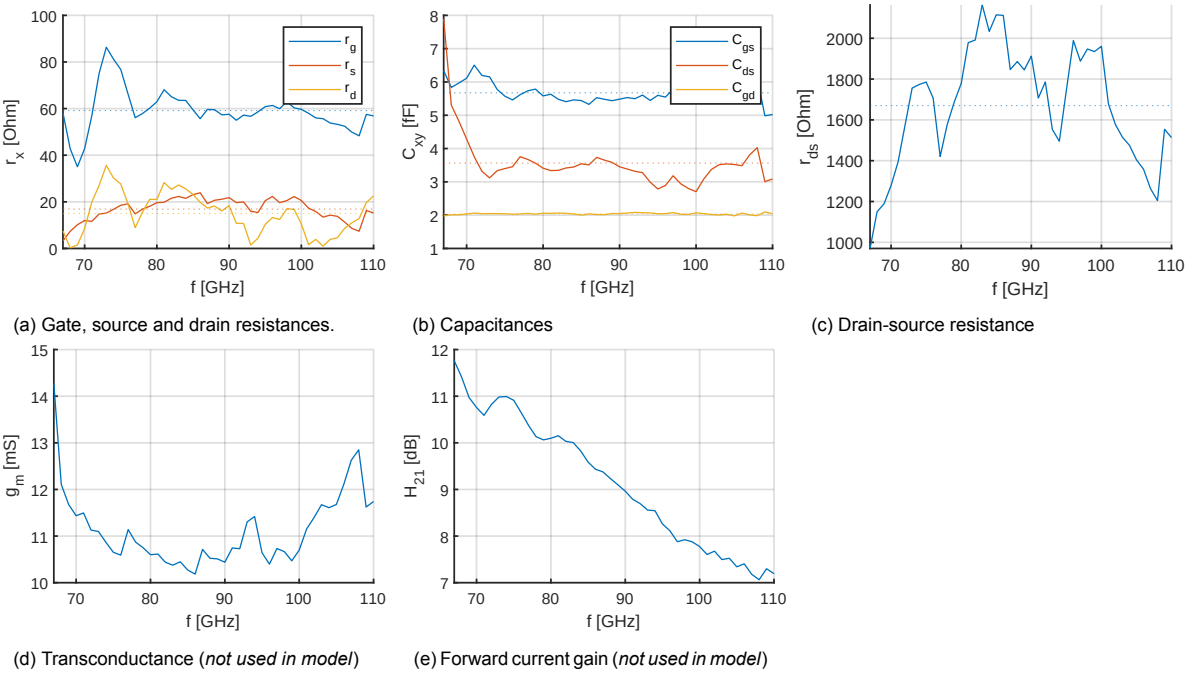
(a) Silicon MOSFET Small-Signal Equivalent model (from [70])



(b) $\delta P_{out}/\delta V_{ds}$ metric at two drive levels -20 dBm and 10 dBm versus frequency

Figure 5.7: A small signal MOSFET model is implemented in ADS using the IV model based trans conductance and device parameters extracted from small signal measurement

Figure 5.8 shows the device parameters extracted from the on-wafer small signal measurements of the Global Foundries slvt-nmos2 transistor at bias voltages of $V_{gs} = 0.6V$ and $V_{ds} = 0.8V$. The mean values of the resistances and capacitances (fig. 5.8a, fig. 5.8b and fig. 5.8c) across the band are used in the ADS model of fig. 5.7b. The small signal transconductance and forward current gain are shown for reference.



(a) Gate, source and drain resistances.

(b) Capacitances

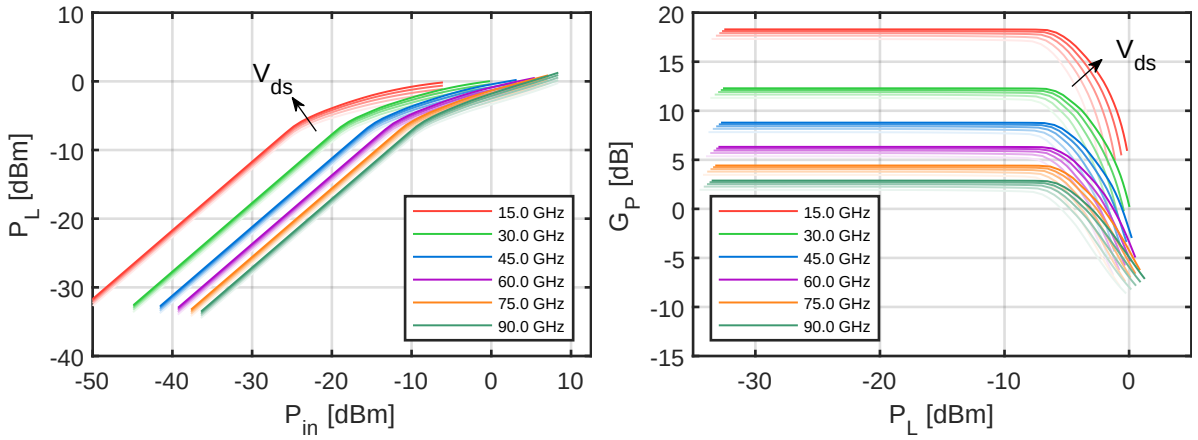
(c) Drain-source resistance

(d) Transconductance (*not used in model*)

(e) Forward current gain (*not used in model*)

Figure 5.8: Small signal parameters extracted from the Global Foundries slvt-nmos2 transistor at bias voltages of $V_{gs} = 0.6V$ and $V_{ds} = 0.8V$. Only the extracted capacitances and resistances are used in the model.

The same harmonic balance simulation was performed, sweeping the available source power from -30 dBm to 15 dBm, the drain bias voltage V_{ds} from 0.6V to 0.8V, and the frequency from 25 GHz to 150 GHz. The gate bias voltage is fixed at $V_{gs} = 0.6V$ and a 50Ω loading is used. The input power to output power relations and power gain from the simulation are shown in fig. 5.9 for this model.



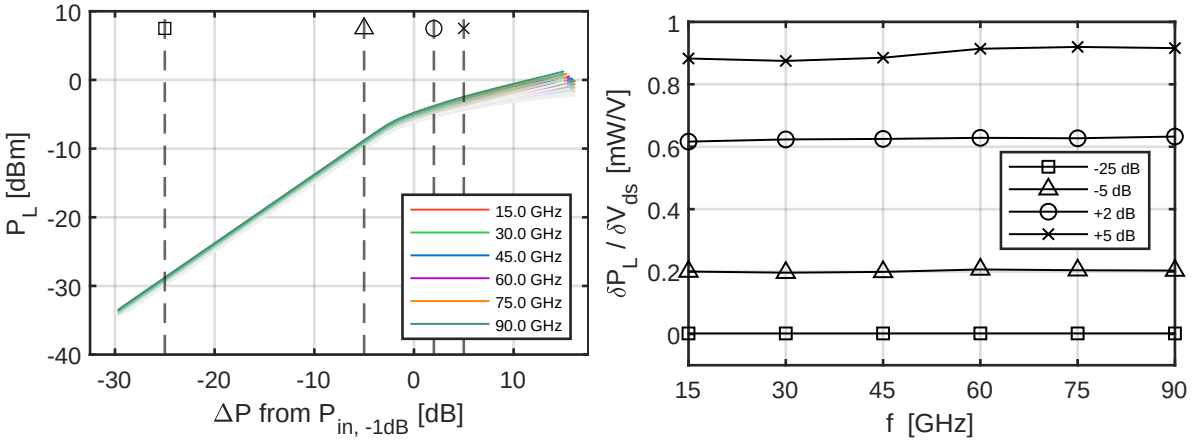
(a) Input power to output power relation. Darker colors show increase in supply voltage (direction of arrow).

(b) Power gain versus output power. Darker colors show increase in supply voltage (direction of arrow).

Figure 5.9: Simulated input power to output power relations and power gain of the IV-based model with device measured resistances and (linear) capacitances.

As now the power gain of the transistor is dependent on the frequency, the input power to output power curves do no longer overlap. Therefore, an additional step is required to evaluate the $\delta P_{out}/\delta V_{ds}$ metric at the different power levels. The input-to-output power curves versus frequency are aligned by the input power -1 dB compression point as shown in fig. 5.10a. This way the $\delta P_{out}/\delta V_{ds}$ metric can be evaluated at different input back-off levels across frequency even though the power gain is different at each of the frequencies.

The $\delta P_{out}/\delta V_{ds}$ metric for this model is shown for the four input back-off levels, -25, -5, +2 and +5 dB the $\delta P_{out}/\delta V_{ds}$ metric increases slightly with frequency at +5 dB into compression level. This seems counter intuitive but is explained by the trend in compression characteristic towards a higher frequency. As can be seen in fig. 5.9a and fig. 5.9b the maximum output power gets higher when the frequency is increased. This is because the gate drain capacitance has a lower impedance at higher frequencies and the maximum output power in saturation therefore increases with respect to the lower frequency when the gain is less than 0 dB. This results into more power for the same V_{ds} supply voltage and therefore a higher value of the $\delta P_{out}/\delta V_{ds}$ metric is observed. When the C_{gd} in circuit of fig. 5.7b is removed, the $\delta P_{out}/\delta V_{ds}$ metric flattens across the frequency band as was seen in fig. 5.6.



(a) Input power to output power relation aligned on the P_{in} -1 dB compression point

(b) $\delta P_{out}/\delta V_{ds}$ metric at four back-off levels

Figure 5.10: The $\delta P_{out}/\delta V_{ds}$ metric can be evaluated across the different frequencies by aligning the input to output power graphs on the -1 dB compression point.

It can be concluded that the inclusion of small signal device parameters does not does not effect a change in the $\delta P_{out}/\delta V_{ds}$ metric against frequency. Moreover, it does not show a deceasing trend with frequency as hypothesized. The next model includes some elements representing non-quasi-static effects to show that if these are included the $\delta P_{out}/\delta V_{ds}$ metric does decrease versus frequency.

5.4.3. IV-curve based with non-quasi-static parameters

The next model includes some NQS elements similar to those in the lumped-component NQS linear model in [71]. The model includes series resistances R_{gs}, R_{gd}, R_{sub} and a time-delay τ resulting into the circuit as shown in fig. 5.11. In order to show a decreasing trend is obtained in the trend the $\delta P_{out}/\delta V_{ds}$ metric against frequency, the time constants of the series RCs are set to a time constant corresponding to a cut-off frequency of 45 GHz. This results into fairly high and likely unrealistic resistance values, but here the model is solely used to illustrate how the inclusion of NQS time delays would influence the results. The time delay τ is set to match the propagation time of an electron through the channel, i.e. assuming a saturation velocity in silicon of $v = 1 \cdot 10^7\ cm/s$ [25, p. 461] and a channel length corresponding to that of the slvt-nmos2 $L = 20nm$.

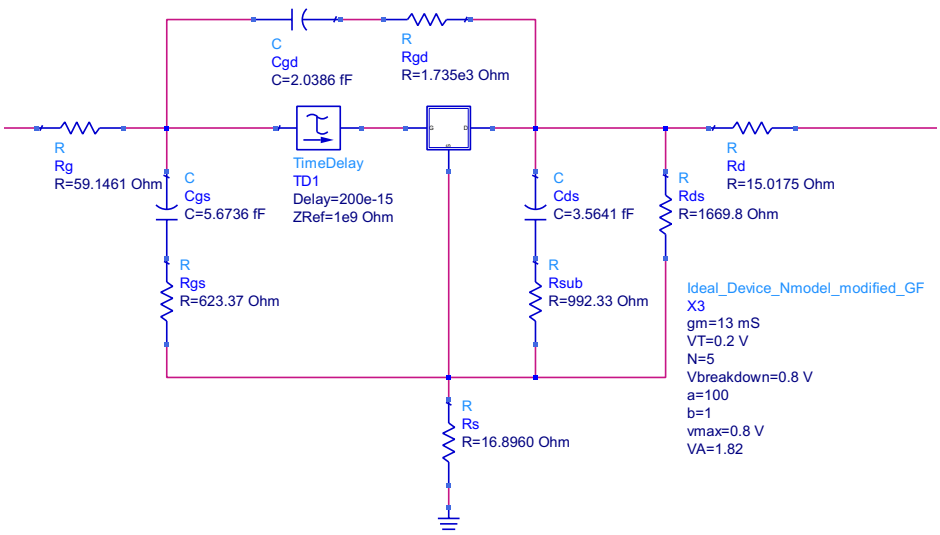
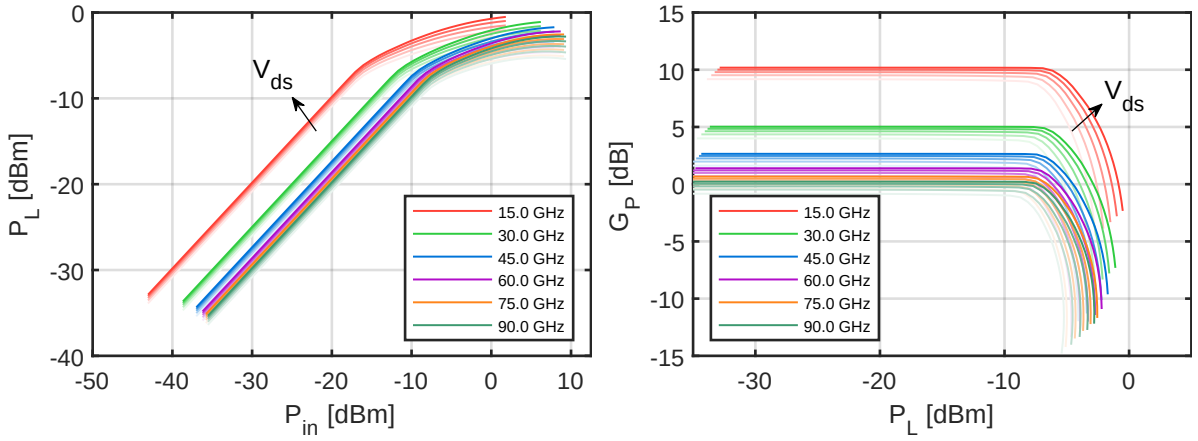


Figure 5.11: IV-curve based model with inclusion of linear non-quasi-static elements

The same harmonic balance simulation was performed and the input power to output power relations and power gain are shown in fig. 5.12. The inclusion of R_{gs} results into an overall lower gain for all the frequencies and the inclusion of R_{gd} results into a sharp roll-off for all the frequencies.



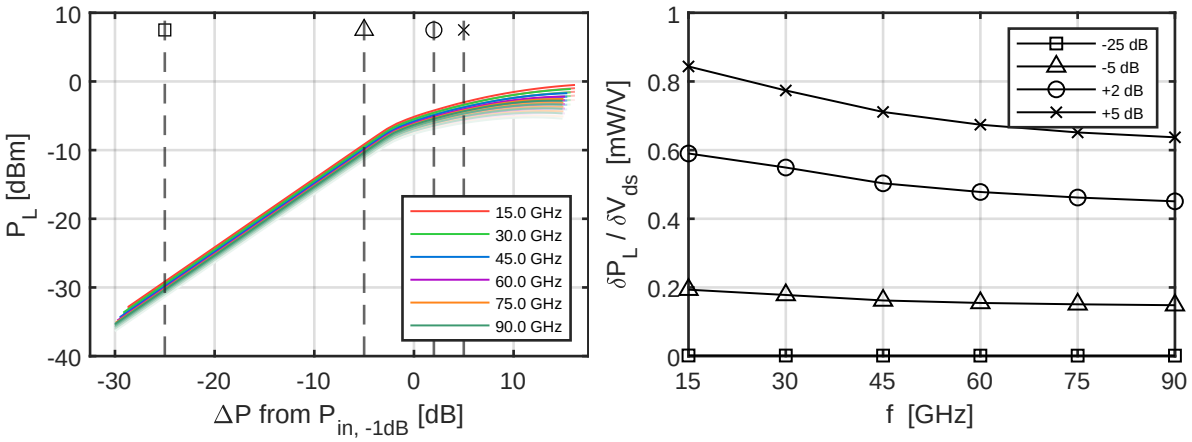
(a) Input power to output power relation. Darker colors show increase in supply voltage (direction of arrow).

(b) Power gain versus output power. Darker colors show increase in supply voltage (direction of arrow).

Figure 5.12: Simulated input power to output power relations and power gain of the IV-based model with non-quasi-static parameters.

The $\delta P_{out}/\delta V_{ds}$ metric for this model is shown for the four input back-off levels, -25, -5, +2 and +5 dB in fig. 5.13. As can be seen now, the metric is decreasing against the frequency. This effect is not seen when the speed of the transistor is simply reduced, i.e. by doubling the C_{gs}. The observation that can be made here is that the inclusion of (linear) NQS effects, commonly modelled by RC time constants and a delayed transconductance, may cause the $\delta P_{out}/\delta V_{ds}$ metric to decrease with frequency.

The important conclusion to make from the last two models is that a decreasing trend of the $\delta P_{out}/\delta V_{ds}$ metric against frequency cannot be explained by just the effects of the device capacitances, i.e. C_{gs}, C_{gd} and C_{ds}. The metric may be compared across frequency as it is not directly related to the power gain of the device. It may therefore also be used in large signal characterization in a 50Ω environment where the device capacitances are not resonated out. However, it is not excluded that device loading may change the value of the metric. Inclusion of delay like elements, such as those in a linear NQS may describe a decreasing trend of the $\delta P_{out}/\delta V_{ds}$ metric against frequency.



(a) Input power to output power relation aligned on the P_{in} -1 dB compression point

(b) $\delta P_{out}/\delta V_{ds}$ metric at two drive levels -20 dBm and 10 dBm versus frequency

Figure 5.13: The $\delta P_{out}/\delta V_{ds}$ metric for the model using elements representing non-quasi-static effects

5.5. Experimental results in a 50Ω environment <67 GHz

A large signal characterization was performed of slvt-nmos2 transistor on the Global Foundries 22nm test die to measure the $\delta P_{out}/\delta V_{ds}$ metric against frequency. This measurement was conducted in a 50Ω environment using a VNA based large signal characterization setup.

5.5.1. Setup and method

A large signal VNA based characterization setup as described in was used to conduct the measurement. The Keysight N5227A Power Network Analyser (PNA) was used that allows for measurements up to 67 GHz without the use of frequency extension modules. The setup was mounted on the Suss probe station and the $100\mu m$ Cascade Infinity probes were used for the on-wafer measurements.

The first tier LRM calibration was performed using a 1.85mm V-connector Anritsu calibration kit. A short was used as the reflect and it is assumed that the standards in the kit are ideal. The length of the line ($79.98\ ps$) was found prior to the calibration using a separate RSOL 'unknown thru' [32] calibration on the PNA. A power calibration was performed on the source port (port-1) using the Keysight U8488A power meter. Power levelling was performed using the matched loads of the calibration kit. The second tier calibration was performed on-wafer on the same die as the DUT using the metal-1 LRM calibration kit[2]. The short and thru were assumed ideal, for the matched loads an resistance of 53Ω was used that was measured beforehand in an IV-measurement. A zero-length for the thru may be used as the distance from the pads to the device intrinsic reference planes is equal to the that from the pads to the center of the thru.

During the measurement the available source power level was swept from $-30\ dBm$ to $6\ dBm$. However due to the limited power output of the PNA and the losses in the cables and probes the maximum available output power is lower at the higher frequencies. The gate bias level V_{gs} was fixed during the repeated measurements at 0.6V while the drain bias V_{ds} was altered from 0.6V to 0.8 V with a step size of 0.05 V. The measurement was repeated with around 10 GHz steps in between frequencies from $10\ GHz$ to $60\ GHz$. An overview of the important parameters during this measurement is shown in table 5.2.

[2]A LRM calibration can be performed on the Global Foundries test die by using the matched and short structures of the SOLT kit, and the thru of the TRL kit

Table 5.2: Overview of measurement parameters for measurement in a 50 Ohm environment

Parameter				Unit
Setup overview				
Type			Large signal PNA based	
PNA			Keysight N5227A	
Probes			Cascade Infinity 67 GHz $100\ \mu m$	
Station			Suss probe station	
Device Under Test				
Foundry			Global Foundries	
Technology			CMOS FD-SOI	
Device			slvt-nmos2	
Wafer #			a44de872	
Die #			7-1	
Bias / Supply				
Gate bias		V_{gs}	0.60	V
Drain bias	start	$V_{ds,1}$	0.60	V
	Δ	ΔV_{ds}	0.05	V
	stop	$V_{ds,2}$	0.80	V
Frequency				
Points		N_{freq}	11	
Sweep	start	f_1	10	GHz
	Δ	Δf	10	GHz
	stop	f_2	60	GHz
Drive level				
Points		$N_{P_{avs}}$	37	
Sweep	start	$P_{avs,1}$	-30	dBm
	Δ	ΔP_{avs}	1	dB
	stop	$P_{avs,2}$	6 [1)]	dBm
Max compr.		n.a.	n.a.	dB
Calibration				
1st-tier			LRM-ideal	
2nd-tier			M1 LRM-ideal	

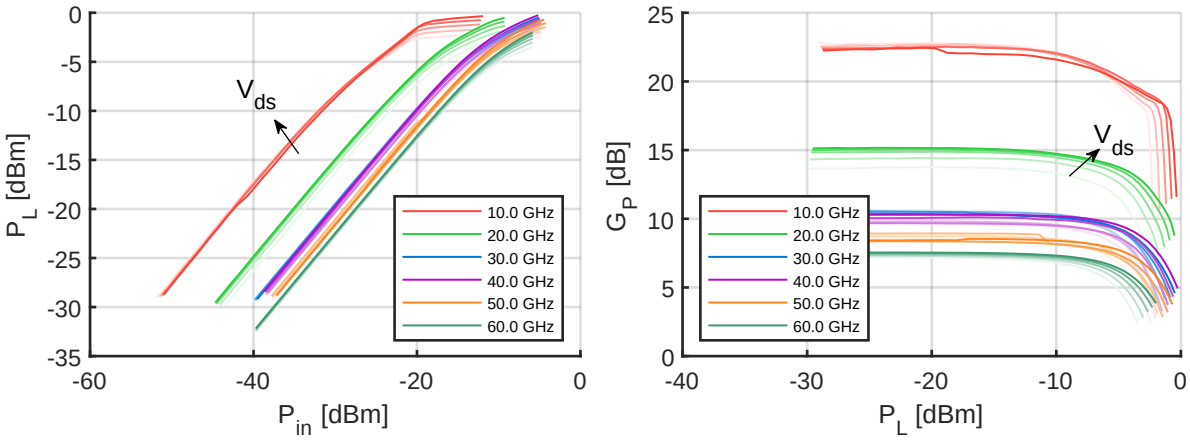
n.a. - Not applicable
[1)] - Lower at high frequencies

5.5.2. Results

Figure 5.14 shows the measured input to output power relations and power gain of the transistor at the different frequencies and bias conditions. There are some jumps in the measurement, visible in fig. 5.14b at $P_L = -20\ dBm$ for the 10 GHz measurement and at $P_L = -9\ dBm$ for the 50 GHz measurement. These are likely due to some probe change of contact during the measurement.

The gain compression curves have different shapes across the frequencies and are not as consistent as was seen for the simple model shown in fig. 5.9b. In particular, for the 10 GHz measurement there are two distinct regions visible, a weakly non-linear region with soft compression characteristic and highly non-linear region with hard compression characteristic. Up until $P_L = -5dBm$ the transistor has a soft compression characteristic, thereby giving it a high $P_{L,-1dB}$ compression point. Then with a sharp transition the device turns into a hard compression operation reaching a saturated output power of around $P_L = 0dBm$. The other frequencies do not show such an expressed difference in a soft and hard compression region.

The 10 GHz measurement shows almost no variation in the output power or gain for the different supply voltages in the soft compression region. Only in the hard compression region a variation appears



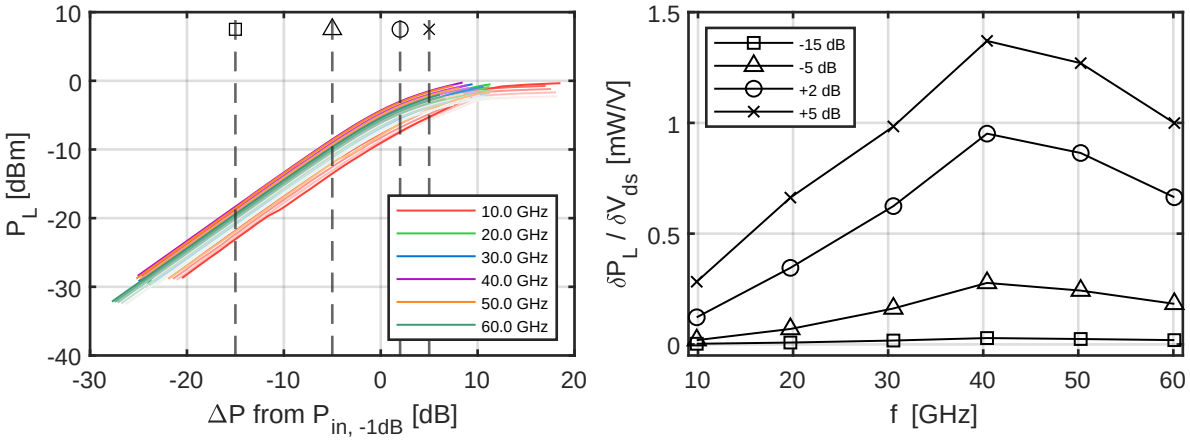
(a) Input power to output power relation. Darker colors show increase in supply voltage (direction of arrow).

(b) Power gain versus output power. Darker colors show increase in supply voltage (direction of arrow).

Figure 5.14: Input power to output power relations and power gain of the Global Foundries slvt-nmos2 transistor in the 50Ω environment <67 GHz.

in output power for the different supply voltages. Towards 40 GHz the variation against supply voltage increases, while at 60 GHz again only variation is seen in the hard compression region. This is also seen in the $\delta P_{out}/\delta V_{ds}$ metric against frequency shown in fig. 5.15b. The behaviour from 40 GHz onward is as expected under the hypothesis that a slew rate like behaviour is reached towards operation closer to the f_T/f_{max}. This would correspond with roughly an $f_T/8.75$ operation of the transistor. Interestingly this behaviour is seen at all four compression levels (-15 dB, -5 dB, +2 dB and +5 dB). Suggesting that the mechanism causing this behaviour is present during small signal and large signal operation.

The upward trend of the metric towards 40 GHz is unexpected and may be due to the spread in the alignment on the -1 dB compression points for the different frequencies. As can be seen in fig. 5.15a the lines do not overlap each other after normalization to the input -1 dB compression point and are much more spread out than was previously the case in fig. 5.10a and fig. 5.13b. This is due to the reliance of the method on using the input/output -1 dB compression points to align the curves. As was noted in fig. 5.14a the 10 GHz gain compression has a soft compression characteristic, with therefore a low -1 dB compression point. The measurement therefore ends up further on the right after normalization in fig. 5.15a. Therefore for the evaluation of the $\delta P_{out}/\delta V_{ds}$ metric +5 dB into compression the 10 GHz trend still has little dependence on the supply voltage, i.e. it still operates in L-mode.



(a) Input power to output power relation aligned on the P_{in} -1 dB compression point

(b) $\delta P_{out}/\delta V_{ds}$ metric at two drive levels -20 dBm and 10 dBm versus frequency

Figure 5.15: $\delta P_{out}/\delta V_{ds}$ metric for the Global Foundries slvt-nmos2 transistor in a 50Ω environment.

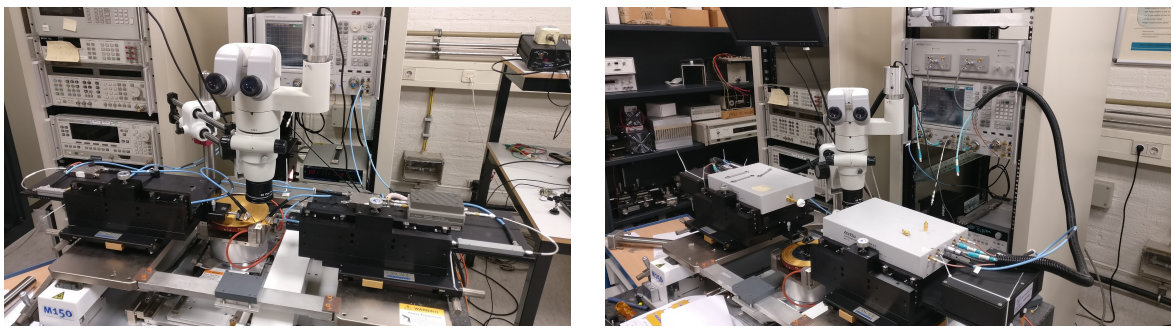
5.6. Experimental results millimeter-wave load-pull environment

An attempt was made to extend the analysis to large signal active load-pull measurements in order to investigate the change in the $\delta P_{out}/\delta V_{ds}$ metric when the transistor is tuned for a similar operation at each frequency. This investigation was to be conducted on all the available technologies described in section 2.4. However, due to the limitations of the available test equipment and inconclusive results this measurement campaign was ended pre-maturely. It was therefore decided not to continue this pursuit as upgraded equipment would have been required to come to meaningful conclusions.

The remainder of this section will describe the steps executed during this measurement campaign, present the limited results obtained and discuss why it was not considered worthwhile to continue this pursuit. The discussion section at the end of this chapter will elaborate on the limitations and make recommendations for future work in this direction.

5.6.1. Setups and methods

In order to be able to compare the $\delta P_{out}/\delta V_{ds}$ metric across frequency in an load-pull environment the same loading strategy should to be used for each of the frequencies. Tuning the loading condition for maximum output power of the device was chosen as the most suitable loading strategy. This would be the strategy used if the device were to be used as the last stage of a PA, with the preceding stages providing the required gain. In addition, with this loading the saturated output power of the device should be equal across the frequencies as in theory this primarily depends on device size. Other loading strategies such as tuning for maximum gain, would introduce a frequency dependency and might invalidate the evaluation of the $\delta P_{out}/\delta V_{ds}$ metric against frequency. This loading strategy however brings with it two issues in an active load-pull system: 1) since the gain will be lower when tuning for maximum saturated power, a high available source power is required in order to push the device far enough into compression, and 2) a high saturated output power also requires (at least) an equally high available power at the load to perform the active load-pull measurement. As was shown in the 165 GHz measurement of chapter 3 the available power of the frequency extension modules quickly gets less then desired for an active load-pull measurement. For this measurement Anritsu WR-10 (67 GHz - 110 GHz) frequency extender modules were used as they have the highest power of the available frequency extenders in the labs. In addition, a sub 20 GHz VNA based load-pull setup was build in order to get a baseline comparison with respect to 50Ω measurements for the change in the $\delta P_{out}/\delta V_{ds}$ metric when the loading condition is tuned. Figure 5.16 show the two active load-pull setups on the M150 probe station. An overview of the important parameters during the load-pull measurements is shown in table 5.3.



(a) Sub 20 GHz VNA based active load-pull setup

(b) WR-10 (75 GHz - 110 GHz) active load-pull setup

Figure 5.16: Two active load-pull setups were used. A VNA based one was developed for sub 20 GHz measurements and a WR-10 based setup for millimeter wave measurements.

Table 5.3: Overview of measurement parameters during the load-pull experiments to find device limiting effects

Parameter			<20 GHz	WR-10	Unit
Setup overview					
Type			Load-pull PNA based	Vertigo MT920	
PNA			Keysight N5242A	Keysight N5242A	
Extenders			n.a.	Anritsu WR10	
Probes			Formfactor Z-probe 40 GHz $100\ \mu m$	Cascade Infinity WR-10 $100\ \mu m$	
Station			M150 probe station	M150 probe station	
Device Under Test					
Foundry			Global Foundries	Global Foundries	
Technology			CMOS FD-SOI	CMOS FD-SOI	
Device			slvt-nmos2	slvt-nmos2	
Wafer #			a44de872	a44de872	
Die #			6-5	7-2	
Bias / Supply					
Gate bias		V_{gs}	0.60	0.60	V
Drain bias	start	$V_{ds,1}$	0.60	0.60	V
	Δ	ΔV_{ds}	0.05	0.05	V
	stop	$V_{ds,2}$	0.80	0.80	V
Frequency					
Points		N_{freq}	3	2	
Sweep	start	f_1	5	95	GHz
	Δ	Δf	5	10	GHz
	stop	f_2	19.5	105	GHz
Drive level					
Points		$N_{P_{avs}}$	37	61	
Sweep[1)]	start	$P_{avs,1}$	-30.5	-32	dBm
	Δ	ΔP_{avs}	1	0.5	dB
	stop	$P_{avs,2}$	6.5	-2	dBm
Max compr.			n.a.	n.a.	dB
Loading condition					
Strategy			load-pull for max P_{L-1dB}		
Points		N_{Γ_L}	1	1	
Calibration					
1st-tier			LRM-ideal	TRL-ideal	
2nd-tier			M1 LRM-ideal	M1 TRL-ideal	

n.a. - Not applicable

1) - Depending on available at frequency

5.6.2. Loading conditions

A load-pull measurement was performed to find the optimal loading condition for highest output power at the -1 dB compression point for each of the frequencies. An initial circular constellation of six loading conditions at the conjugate of the S_{22} of the DUT was moved around the Smith chart until the highest output power at the -1 dB compression point was found within its center. For the following measurements only this optimal loading condition was measured, i.e. a single loading condition per frequency was measured. The loading conditions at the measured frequencies are shown in fig. 5.17. On the sub-20 GHz setup the measurement was performed at 5 GHz, 10 GHz, 15 GHz and 19.5 GHz.

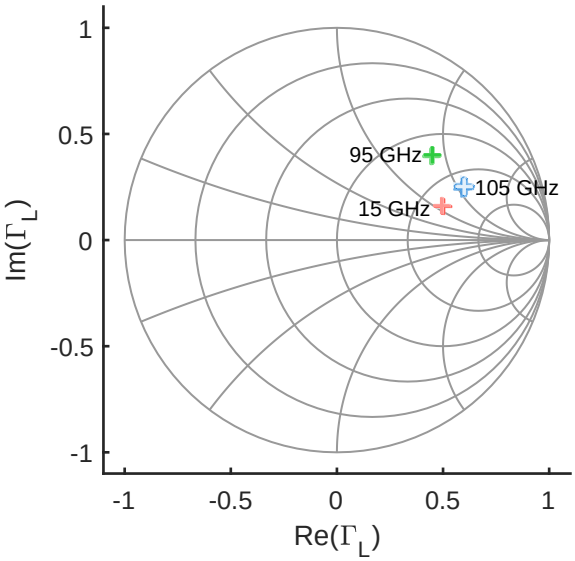
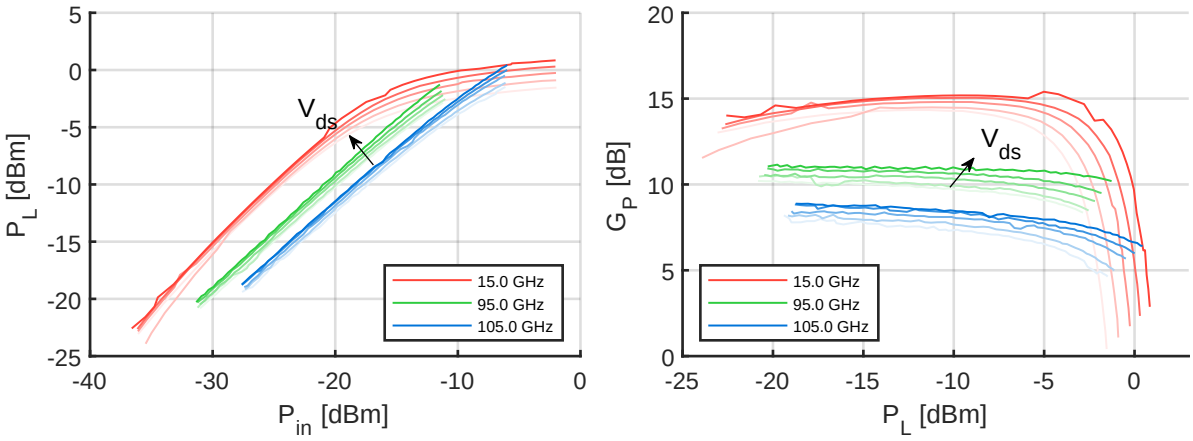


Figure 5.17: Location of optimum loading conditions at 15 GHz, 95 GHz and 105 GHz

However only the 15 GHz measurement was adequate, as the source power amplifier showed some instabilities at the other frequencies. This is likely due to the configuration of the setup for maximum available output power causing the input PA to see a non-constant load. For the WR-10 setup it was found that the available power of the modules was only adequate at 95 GHz and 105 GHz. Here however it was a challenge to find the optimum loading conditions because of two reasons. The first reason was that the algorithm had trouble converging when a constellation of loading conditions was tested. During the active load-pull sequence first a source available power level is set and then the software converges to each of the loading conditions one-by-one, makes the related measurements before moving to the next source available power level. It was found that when a large power step was used (more than 0.5 dB) the algorithm had a lot of trouble jumping between the loading conditions and lost accuracy. When the same loading conditions were measured individually with the same settings the algorithm had no problems converging. This however made the load-pull measurement much more time consuming. The second reason was that due to the limited available driving power it was not possible to measure far into compression and therefore see if the optimum loading condition for maximum output power was actually achieved.

5.6.3. Results

The input-to-output power relations and gain compression curves are shown in fig. 5.18a. The results of the load-pull measurements leave much to desire for the analysis approach described in this chapter. The limited dynamic range achieved in the WR-10 measurements does not allow for good observation of the compression characteristics of the DUT. By only being able to see the linear and weakly non-linear region of the gain compression curve it becomes difficult to determine with confidence the exact -1 dB compression point for alignment of the input-to-output power curves. As shown in fig. 5.19a, the measurements for 95 GHz are slightly below the -1 dB compression points. They could therefore only evaluated at input back-off levels of -12.5 dB and -5 dB. The 15 GHz and 105 GHz measurements could also be evaluated at +2.0 dB and +5.0 dB into compression.

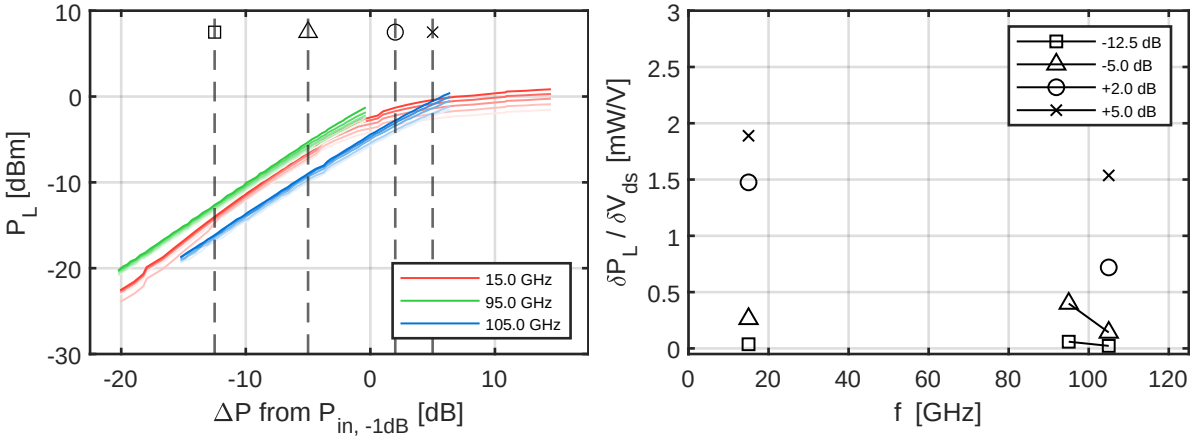


(a) Input power to output power relation. Darker colors show increase in supply voltage (direction of arrow).

(b) Power gain versus output power. Darker colors show increase in supply voltage (direction of arrow).

Figure 5.18: Input power to output power relations and power gain at the three frequencies measured under active load-pull

Figure 5.19b shows the evaluation of the $\delta P_{out}/\delta V_{ds}$ metric from the results obtained during the measurement. The value of $\delta P_{out}/\delta V_{ds}$ metric at 15 GHz has increased with respect to what was seen in the 50Ω measurements at 20 GHz in fig. 5.15b. Now at +5 dB compression the metric shows that the transistor is no longer operating in L-mode, hinting that the problem seen in the 50Ω measurements were due to a non-resonated output reactance effect. There is a slight downwards trend observable between the 15 GHz measurement and the measurements around 100 GHz but with the limited data this is not enough to be conclusive.



(a) Input power to output power relation aligned on the P_{in} -1 dB compression point

(b) $\delta P_{out}/\delta V_{ds}$ metric at two drive levels -20 dBm and 10 dBm versus frequency

Figure 5.19: $\delta P_{out}/\delta V_{ds}$ metric for the Global Foundries slvt-nmos2 transistor loaded for maximum output power at three frequencies.

The large gap in characterized frequencies creates a lack of knowledge that makes this investigation inconclusive. The slight downward trend seen is therefore not enough to conclude on if slew-rate like limiting behaviour is present and if there is a transition frequency at which this behaviour starts dominating.

5.7. Summary and recommendations

5.7.1. Summary

A device limiting effect that shows slew-rate like behaviour in the large signal operation of active devices at may explain the challenge to generate power at millimeter-wave frequencies. In this chapter a measurement method was formulated that could allow for observation of large signal device limiting effects in transistors. The method employs a simple metric that defines the sensitivity of the output power to the supply voltage variation. This metric can be extracted in single tone CW measurements in a 50Ω or active-load pull environment and allows for comparison in different modes across frequencies towards the f_T/f_{max}.

The method was employed in an on-wafer large signal measurement on a 22nm FD-SOI CMOS transistor <67 GHz in a 50Ω environment. The transistor showed a reduction in the sensitivity of the output power to supply voltage variation from 40 GHz onwards, corresponding roughly to an $f_T/8$ operation of the DUT. This hints towards an onset of device limiting behaviour, however this observation could not be confirmed in an active load-pull measurement up to 100 GHz due to limitations in the capabilities of the equipment available.

The investigation into liming effects when operating close to the f_T/f_{max} of the device was to be extended with active load-pull measurements across different technologies to evaluate if different technologies show different behaviour. Only the measurement on the Global Foundries 22nm FD-SOI were completed after which the efforts were terminated. The results showed a decrease in the output power to the supply voltage variation metric but the measurement needs to be repeated at different frequencies in order to be conclusive.

5.7.2. Recommendations

The investigation into device limiting effects presented in this chapter is preliminary and at this point inconclusive. A method was formulated that according to theory could show large signal slew-rate limiting effects however it proved to be challenging to use this method with the available equipment. The primary reason why this was challenging is because the method relies on tuning of the devices for highest output power, i.e. highest P_{sat} or highest P_L at the -1 dB compression point. Therefore the output power of the DUT will be maximal at the operating conditions of interested. This means that in order to do a successful active load-pull the setup should be able to deliver at least an equal amount of power at the output device reference plane. In case the source is highly mismatched a similar power may be needed at the input.

For the sub-20 GHz VNA based active load-pull setup the available power was maximized by careful minimization of the attenuators in the system. Unfortunately this resulted into slightly unstable behaviour of the input and output PAs. If this setup were to be again used it should be considered adding isolators after the input and output PAs to ensure these see a constant loading condition.

Nevertheless, a lot of measurement equipment is needed to do an broad study such as this one in active load-pull across frequencies from microwave to millimeter-wave. The current available measurement equipment leaves a large gap in the 26.5 GHz to 110 GHz range, roughly equal to the (interesting range) of $f_T/10$ to $f_T/3$ of the available transistors on the test dies. With the development in this project on the VNA based active load-pull setup it is possible to extend the range of the sub-20 GHz setup to 40 GHz by an investment into 40 GHz IQ mixers, amplifiers and couplers. Then the 50 GHz to 65 GHz range may be covered by converting the 50 GHz to 65 GHz wave-guide based active load-pull setup described in the work of L. Galatro [27] to a 67 GHz non-frequency multiplied VNA based setup. As the components in this use wave-guide transmission lines and PAs the driving power capabilities are high and losses are low. This would allow for characterizations roughly in the middle in the large frequency gap that exists now. Unfortunately, the driving capabilities and quality of the frequency extension modules in the WR-10 band (75 - 110 GHz) leave much to desire for large signal characterization.

Future efforts of the investigation into device limiting effects using the proposed method should include compact or behaviour models provided in the technologies development kits. A comparison of the metric extracted in simulation and on-wafer for the same technology could show interesting results and allow for the metric to be incorporated into design trade-offs.

6

Conclusions

Commercial interest for millimeter-wave wireless communication has increased significantly over the past years. Developments in this frequency range are enabled by the continuous improvements in silicon-based technologies over the past decades and are driven by frequency congestion in low-GHz bands. This thesis explored the current capabilities and challenges of device characterization at millimeter-wave frequencies. Three topics were addressed in this report related to: linearity characterization of active devices using the state-of-the-art millimeter-wave load-pull setups, calibration for cross-coupling correction during on-wafer characterization and investigations into device limiting effects when operating close to the f_T/f_{max} of the technology. The outcomes and conclusions of this thesis are summarized in section 6.1 and recommendations for future work are provided in section 6.2.

6.1. Outcome of the thesis

6.1.1. Vector gain based in-band linearity estimation for on-wafer characterization beyond 75 GHz

In the first part of this thesis a novel approach was presented for in-band linearity estimation that overcomes the limitation of harmonically generated spectral content by the frequency extension modules in the millimeter-wave active load-pull setups. These extension modules operate using non-linear $\times N$ frequency multipliers in the RF up-conversion path to synthesize frequencies beyond 75 GHz. The inclusion of this non-linear element makes conventional non-linear characterization methods, such as the two tone test challenging.

In this work it was shown that linearity metrics for modulated signals can be extracted by using vector gain measurements of the active device under test, across drive-level, frequency and loading condition by using a frequency-domain based approach. By using this method the AM-AM and AM-PM response of the device can be captured and used to evaluate linearity at the device level across a large bandwidth of operation, which is important as millimeter-wave systems are being intensively investigated for large bandwidth data transmission.

This vector gain based method was benchmarked in simulation and during on-wafer experiments in two sub-30 GHz bands on an active load-pull setup that also supports complex modulated excitation. The vector gain was used to estimate the EVM for a 16QAM signal at several drive levels and loading conditions. Although the method did not provide an one-to-one accurate estimate of the absolute value of the EVM, it was able to show correlation in the extracted trends for the EVM against drive-level and loading condition, as per its intended use for a first-order linearity estimation approach.

The method was then used during an on-wafer characterization to determine the optimum loading condition for a 22nm CMOS FD-SOI transistor excited by a 16QAM signal at 165 GHz. From this proof of concept it can be concluded that the approach is applicable for millimeter-wave linearity characterization beyond 75 GHz and can be used to determine optimum device operating conditions for drive-level and loading.

6.1.2. Cross-coupling correction for sub-Terahertz on-wafer measurements

The second part of this thesis has been dedicated to the development of a calibration algorithm that can correct for probe-to-probe cross coupling during on-wafer measurements. Probe-to-probe cross-coupling becomes an increasingly large problem as wavelengths reduce while mechanical transitions to the on-wafer environment are not scaled equally. Traditional 8-term or 10/12-term calibration algorithms cannot capture the effects of this cross-coupling as the amount of coupling is dependent on the impedance of the DUT. 16-term calibration algorithms can compensate the effects of cross-coupling and have existed for over three decades but are not commonly used because these are only effective if the probe spacing is not changed after calibration. This problem has gained interest in research to find a more practical solution to cross-coupling corrections for on-wafer measurements.

By taking a step-by-step approach in the development for a new 16-term calibration algorithm it was shown in which cases such a model offers an advantage with respect to 8-term or 10/12-term models. During this approach ADS/Momentum and 3D modelled CST simulations of realistic fused silica CPW calibration standards were used. Distinct cases were identified where a standard 16-term calibration algorithm would show significant improvement over 10/12-term calibration methods. The developed 16-term implementation in MATLAB showed a one-on-one correction capability across the Smith chart when impedance dependent cross-coupling was present.

This work led to the proposal of a new form of the 16-term error model that allows for parametrization of the probe-to-probe spacing. The new model splits the original 16 error terms into two parts, one static with spacing and one dynamic with spacing. The dynamic error terms may be interpolated for a probe spacing not included in the original calibration set. This allows for the creation of calibration kits with a set of standards having a different pad-to-pad spacing and maintaining the 16-term advantage for differently spaced DUTs. It was shown that the performance of this interpolated 16-term error model is equivalent to when the spacing is included in the calibration. The rigorous investigation performed into the parametrized 16-term algorithm could enable industry scale on-wafer coupling corrected measurements at millimeter-wave frequencies without requiring a dedicated calibration kit for each DUT with a different pad-to-pad spacing.

6.1.3. Preliminary investigation into device limiting effects

In the third part of this thesis a preliminary investigation was performed into device limiting effects when operating transistors at frequencies close to the f_T/f_{max} of the technology. Device liming behaviour has been investigated in the small-signal domain in for example NQS modelling, however as millimeter-wave PAs are reaching operating frequencies at fractions of the f_T/f_{max} of the technology it becomes important to understand large signal behaviour at these frequencies. A large signal slew-rate like behaviour in the transistor may explain the challenges in generating power at millimeter-wave frequencies.

In this work it was shown how such behaviour may be identified by measuring the sensitivity of the output power by a change in supply voltage at several frequencies. By observing the trend in this metric across frequency different operation modes may be identified. A slew-rate behaviour can be detected if the output power becomes in-dependent of the supply voltage for higher frequencies, i.e. when a downwards trend is observed in the metric against frequency.

This method was used in simulation and on-wafer experiments on a 22nm FD-SOI CMOS transistor. During a large signal measurement <67 GHz in a 50Ω environment the transistor showed a reduction in the sensitivity of the output power to supply voltage from 40 GHz onwards, corresponding roughly to an $f_T/8$ operation of this technology. This hints towards an onset of device limiting behaviour, however this observation could not be confirmed in an active load-pull measurement up to 100 GHz due to limitations in the capabilities of the equipment available.

6.2. Recommendations

6.2.1. Expanding vector gain based linearity characterization

At this moment the vector gain based linearity approach stands as a proof of concept demonstration to extract linearity metrics for active device characterization at millimeter-wave beyond single-tone AM-AM and AM-PM responses. This first step leaves opportunities for extending the validation and exploring the application of the approach. By distinguishing from envelope (narrow-band) based approaches the question rises on whether the advantage of the frequency-domain based approach can be quantified. From a device perspective this may be investigated by characterizing the amount of variation in the vector gain across frequency as these would not be accounted for in a narrow band approach. This investigation may also be evaluated on a design/application level by using the frequency-domain approach to investigate a frequency dependence in the applied loading condition, i.e. that of a simulated output matching network. In addition, the vector gain extraction and linearity estimation can be extended to include different bias levels to trade-off linearity and efficiency for different operating classes.

6.2.2. On-wafer validation of 16-term distance dependent cross-coupling correction approach

The step-by-step approach taken to develop the parametrized 16-term calibration algorithm showed interesting directions for future work. Different combinations of two-port calibration standards showed different amounts of cross-coupling in the simulations which raises the question which combination gives the best performance. In addition, publications of other researches in this field have shown to be able to correct cross-coupling in split error term models with only one spacing dependent standard. This might bring the possibility to further reduce the set of calibration standards, i.e. full set at a nominal spacing, and only one additional standard to parameterize the change in coupling for additional probe spacings. In its current state the developed parametrized 16-term calibration approach is ready to be tested in an on-wafer environment. To do so a calibration kit needs to be developed for a frequency band that shows significant probe-to-probe cross-coupling.

6.2.3. Continued investigation into device limiting effects

While the available measurement equipment limited obtaining conclusive results for the investigation into device limiting effects, the approach employed in this work could provide insight into device behaviour when operating close to the f_T/f_{max} of the technology. Two steps should be taken in future work to continue this investigation. First, using the same approach, additional frequency points should be observed in load-pull for the $f_T/20$ to $f_T/4$ range for the investigated technology. And second, these results should be compared with the same metric extracted from simulation using development kits of the technology. By comparing results, it could be highlighted if the current transistor models include the same behaviour and if not, how this would impact the performance of the device.

Acronyms

ACPR adjacent channel power ratio 18

ADC Analog-to-Digital converter 6, 37

ADS Advanced Design System 12, 13, 18, 45, 58, 72, 73

Al Aluminium 4

AM-AM amplitude to amplitude 11–13, 15, 17, 18, 21, 22, 26, 27, 34, 36, 40–42, 85, 87

AM-PM amplitude to phase 11–15, 17, 18, 21, 22, 26, 27, 34, 36, 40, 41, 85, 87

As Arsenide 4

ASK amplitude-shift-keying 3

B Boron 4

Bi Bismut 4

CMOS Complementary MOS 3, 9

CPW co-planar waveguide 8, 9, 25, 43, 49, 56, 58, 86

CW Continious Wave 14, 15, 23, 24, 29, 32, 33, 35, 40, 83

DAC Digital-to-Analog converter 6

DUT Device Under Test 5–7, 9, 11–14, 16, 23, 25, 26, 30, 32, 33, 36, 39–41, 44–46, 48, 49, 52, 53, 55, 56, 59–61, 63, 65, 70, 76, 81, 83, 86

ELCA Electronic Circuits and Architectures iii, 2, 3

EM electro magnetic 32, 43–45, 49–51, 56, 58, 65, 66

EVM Error Vector Magnitude iii, 12–14, 16–18, 21–23, 25, 27–32, 34, 35, 40, 42, 85, 95

f.l.t.r. from left to right 49

FD-SOI fully depleted silicon on insulator 9

FS fused silica 45, 55

Ga Gallium 4

GaAs Gallium Arsenide 4, 5

GaN Gallium Nitride 4, 5

GBW gain-bandwidth 67

GSG Ground-Signal-Ground 49

HBT heterojunction bipolar transistor 9

I in-phase 17

IF Intermediate Frequency 6

IMD inter modulation distortion 11

In Indium 4

InP Indium Phosphide 4, 5

IV current voltage (relation) 71

LL load-load 49, 57–60

LNA low noise amplifier 3

LO Local Oscillator 3, 6

LRM Line-Reflect-Match 7, 8, 24, 25, 76, 77, 80

LS load-short 49, 57

MOS metal oxide semi-conductor 9, 70, 72, 89

MW Modulated Wave 24

N Nitride 4

NQS non-quasi static 68, 74, 75, 86

OFDM Orthogonal frequency-division multiplexing 41

OO open-open 49, 57

OOK on-off-keying 3

P Phosphide 4

PA power amplifier 1–5, 15, 26, 30, 35, 41, 68, 79, 81, 83, 86

PNA Power Network Analyser 76, 77, 80

Q quadrature 17

QAM quadrature amplitude modulation 4, 17, 35, 41

RF Radio Frequency 1, 6, 11, 14, 67, 68, 85

RMS root mean square 16, 30

RSOL Reciprocal-Short-Open-Load 7, 76

RT Real Time 24, 26, 29

Si Silicon 4

SiGe Silicon Germanium 9, 18

SL short-load 49, 57

slvt super low voltage threshold NMOS2 9

SOL short-open-load 44

SOLT Short-Open-Load-Thru 7–9, 25, 43, 53–55, 57, 76

SRRC square root raised cosine 16, 24, 35

SS short-short 49, 57

SVD singular value decompostion 47–49, 52, 54, 55

T thru 49, 57

Ti Titanium 4

TRL Thru-Line-Reflect 7–9, 25, 32, 33, 43, 76, 80

VNA vector network analyser 6, 7, 11, 32, 33, 37, 40, 45, 47–49, 52, 55, 65, 66, 76, 79, 83

List of Symbols

Sign	Description	Unit	Ref.
G_P	Power gain	dB	
G_T	Transducer gain	dB	
$G_{TU,max}$	Maximum unilateral transducer gain	dB	[72, p. 230]
MAG	Maximum available gain	dB	
MSG	Maximum stable gain	dB	
PAE	Power added efficiency	%	[73, p. 756]
P_L	Output/load delivered power	dB or W	
P_{avs}	Available source power	dB or W	
P_{in}	Input power	dB or W	
P_{sat}	Saturated output power	W or dBm	
f_T	Transition/cut-off frequency *frequency where* $\lvert h_{21} \rvert = 0$	Hz	BJT [72, p. 78] FET [72, p. 84]
f_{max}	Maximum oscillation frequency b*frequency where* $\lvert G_{TU,max} \rvert = 0$	Hz	BJT [72, p. 78] FET [72, p. 84]

A

95th ARFTG paper - Vector gain based EVM estimation at mm-wave frequencies

The following pages include a paper that was written, accepted and presented at the 95th ARFTG conference on the topic of the vector gain based EVM estimation method that was developed during the works of this thesis.

Vector gain based EVM estimation at mm-wave frequencies

J. Van 't Hof[1], C. De Martino[1], S. Malotaux[1], M. Squillante[2], M. Marchetti[2], L. Galatro[3], and M. Spirito[1]

[1]Electronics Research Laboratory, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands
[2]Anteverta-mw B.V., High Tech Campus 41, 5656AE Eindhoven, The Netherlands
[3]Vertigo Technologies B.V., Mekelweg 4, 2628 CD Delft, The Netherlands

***Abstract* — In this contribution we present a method for estimating linearity performance of devices operating in the higher millimeter-wave region, under modulated signals and over different loading conditions. The proposed method uses the power dependent vector gain extracted during continuous-wave large signal (load pull) measurements. The EVM prediction capability of the method is benchmarked with experimental load pull data with realistic modulated signals (QAM16) in the 5 GHz (RF) and in the 26 GHz (5G) bands on a 22nm CMOS FD-SOI device. The EVM estimated by the model correlates to the load pull measurements under complex modulated stimulus and properly predicts the best loading condition for linearity. Finally, the proposed method is used to estimate the EVM performance (QAM16) and the optimal loading condition for a 22nm CMOS-SOI device operating in the higher millimeter-wave region, at 165 GHz.**

***Index Terms* — 5G, EVM, linearity, mm-wave characterization, vector gain.**

I. Introduction

Commercial interest for millimeter-wave communication has increased significantly over the past years, due to the appearance of large volume applications such as automotive radar and telecom applications (i.e., 5G systems operating in the FR2 band and E-band backhaul systems). These developments are enabled by the continuous improvements in silicon-based technologies over the past decades and driven by frequency congestion in low-GHz bands. One of the bands currently under consideration for future millimeter-wave communication (i.e., 6G) is the D-band with available frequency slots assigned from 140 GHz to 180 GHz where the low atmospheric adsorption window allows for medium to long distance communication [1].

To exploit the available frequencies and maximize the system throughput, modern communication standards use non-constant envelope modulation schemes and thus require linear operation of the power amplifier (PA) present in the transmitter chain. Due to the trade-off between linearity and power efficiency, proper transistors models, compact or behavioral [2, 3], are required to accurately predict all the non-linear contributions in order to properly tackle this trade-off early in the design cycle. These models should accurately capture the amplitude to amplitude (AM-AM) and amplitude to phase (AM-PM) distortion in order to reach the (standard defined) linearity target.

Reducing the number of design iterations of millimeter-wave PAs therefore requires accurate characterization and validation of these transistor models employed in the circuit design. Various developments have been made in and calibration/de-embedding techniques and advanced measurement setups in order to accurately characterize the large signal response of the transistors on wafer [4-6]. Non-linear characterization and modelling techniques such as the X-parameters and Poly-Harmonic Distortion model [7, 8], have been applied in the characterization and design phase to support the development of power amplifiers. However, when frequencies approach the upper mm-wave region, i.e., in D-band, characterization of linearity requires complex and rarely available setups [9].

In order to make accurate predictions of the device linearity there is a need to extract figure of merits (FOM) on accessible millimeter-wave characterization test benches. A linearity FOM proposed in [10] based on the AM-AM distortion in single-tone measurement was shown to closely correlate with the linearity performance of a GaN HEMT measured under two-tone as well as modulated excitation at 31.5 GHz. In this work a similar approach is proposed to predict the device linearity performance, including both the AM-AM and AM-PM distortion mechanisms. Moreover, the proposed approach also includes the frequency dependency of the device non-linear transfer function across the operating bandwidth, which plays an important role in the large modulation bandwidths of mm-wave systems.

The method proposed is based on a power swept, single tone vector gain measurements of the transistor. These characterization routines are employed to setup a complex, power dependent, transfer function in the frequency domain. By measuring the load dependent, vector gain over the (signal) operation band at several power levels an estimation can be made of the error vector magnitude (EVM) response of the device under modulated signals. In this contribution the proposed vector gain based EVM estimation method is benchmarked for its prediction capability versus load pull measurements under true modulated signals in the 5 GHz and 26 GHz bands on a 22 nm CMOS FD-SOI.

Finally, vector gain measurement under CW load pull conditions at 165GHz are used to predict the EVM performance of the device for a QAM16 stimulus signal in this band.

II. The Vector Gain Model

The vector gain of a two port device can be measured from the ratio of the down-converted waves b_2/a_1 in a large signal characterization test bench as was shown in [11]. Following this definition, the vector gain represents a complex transfer

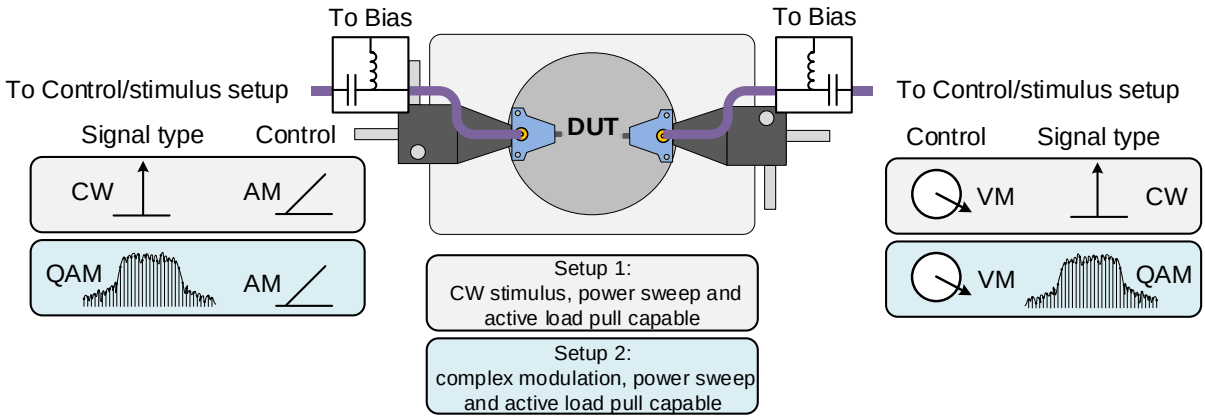


Fig. 1. Two measurement techniques are used in this contribution. A Continuous Wave (CW) load pull measurement is used to measure the vector gain of the DUT at the intrinsic device reference planes and estimate EVM performance using a behavioral model. This approach is benchmarked with a second true modulated wave EVM measurement at 5 GHz and 26 GHz.

function describing the magnitude and phase response from the input to the output of the device. The non-linear nature of this transfer function imposes dependency versus the bias, frequency, drive level as well as loading condition.

This input to output behavior can be described by the complex function in eq. (1), for every bias and load condition. Here both the amplitude and phase response are frequency and power dependent. This transfer function is defined by eq. (1) acquired measuring the vector gain at various power levels, frequency points and for several load conditions (during active load-pull).

$$H(\omega,\ P_{in}) = A(\omega, P_{in}) \cdot\ e^{j\Phi(\omega,\ P_{in})} \tag{1}$$

where $A(\omega,\ P_{in})$ describes the power depending magnitude transfer $\phi(\omega,\ P_{in})$ the power depending phase transfer versus frequency.

While the non-linear device response is acquired under continuous wave (CW) excitation at different frequencies, the input to output behavior for modulated signals is computed (numerically) for a multi-tone signal, with a frequency grid compliant to a user selectable standard definition.

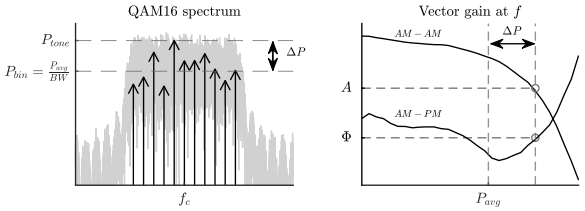


Fig. 2. Each spectral component is scaled in amplitude (A) and phase shifted (Φ) as determined by the measured vector gain at that frequency depending on the offset (ΔP) of the power in the frequency component from the average power in the signal.

For this reason, a relation between the CW power level and bin related power level in the modulated signal case needs to be defined. The proposed method employs the constant power (CW case) to the average bin power (modulated case) mapping described in Fig. 2. Each spectral bin component is then compared to the bin average power (i.e., P_{avg}/BW) and the relative power offset in the vector gain transfer function is used to apply the effective distortion (AM-AM and AM-PM) to that frequency bin.

After obtaining the now distorted spectrum at the output of the DUT, in the frequency domain, the resulting signal is transformed back to time domain. The EVM can then be calculated from eq. 2.

$$EVM_{rms} = \sqrt{\frac{\sum_n |y_n - x_n|^2}{\sum_n |x_n|^2}}. \tag{2}$$

III. Benchmarking At 5 GHz and 26 GHz

A. Setup Description

In order to validate the proposed vector gain based EVM estimation, two different test benches are employed, as sketched in Fig. 1.

Setup 1, allows the extraction of CW vector gain versus power and at different frequencies, under varying loading conditions.

Setup 2, allows to provide a complex modulated signal to the device under varying loading conditions and acquire the true EVM response of the device. For this experiment the Anteverta MT-2000 Series Mixed-Signal Active Load Pull system is employed [12]. The validation is done at 5 GHz and 26 GHz for a QAM16 modulated signal over a 20 MHz band.

B. Vector Gain Measurements At 5 GHz and 26 GHz

At 5 GHz the vector gain was measured over 21 tones (1 MHz spacing) in the operating band, at 31 source power levels from -30 dBm to 0 dBm (1 dB spacing) and for six different loading conditions around the maximum -1 dB compression point.

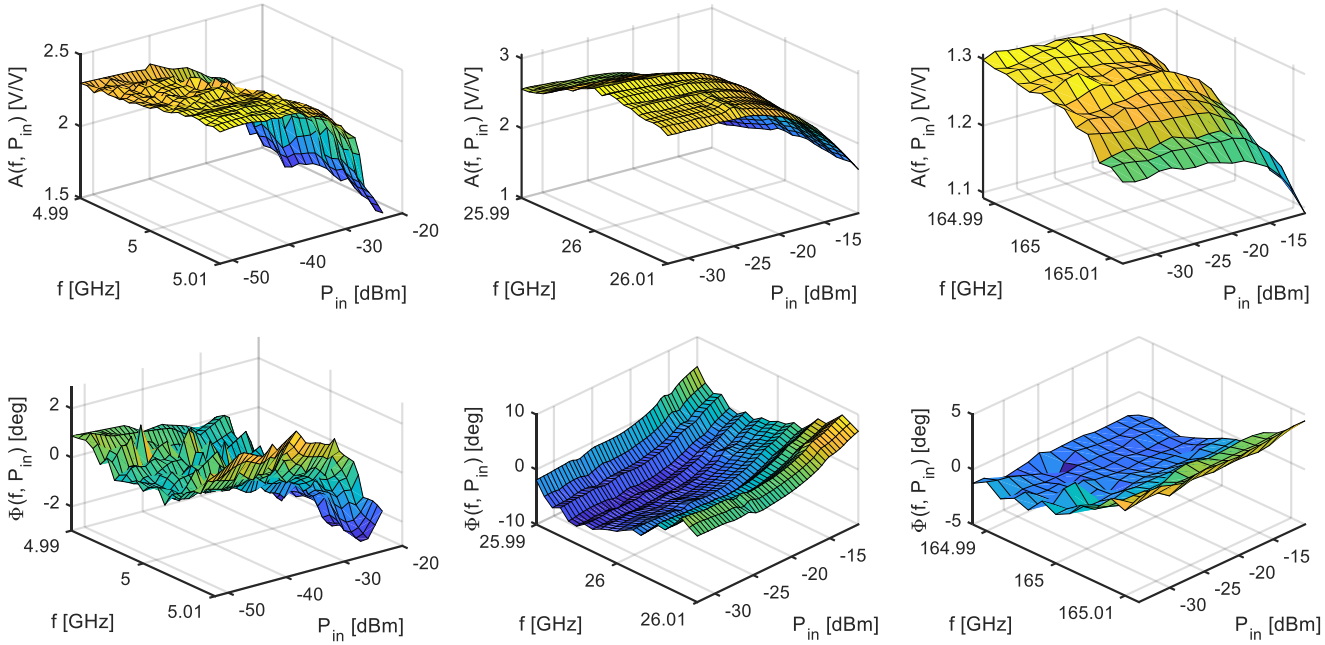


Fig. 3. Magnitude (top) and phase (bottom) of vector gain measured against frequency centered at 5 GHz and output power at the load for $\Gamma_L = 0.5$

Fig. 4. Magnitude (top) and phase (bottom) of vector gain measured against frequency centered at 26 GHz and output power at the load for $\Gamma_L = 0.64 - 0.11\mathrm{j}$

Fig. 5. Magnitude (top) and phase (bottom) of vector gain measured against frequency centered at 165 GHz and output power at the load for $\Gamma_L = 0.60 + 0.35j$

Fig. 3 shows the AM-AM gain compression characteristic and the AM-PM phase variation across the frequency band and input power for one of the six loading conditions.

At 26 GHz the vector gain was measured at 21 tones (1 MHz spacing) in the operating band, at 26 source power levels from -25 dBm to 0 dBm (1 dB spacing) and at sixteen loading conditions around the maximum -1 dB compression point.

Fig. 4 shows the AM-AM gain compression characteristic and the AM-PM phase variation across the frequency band and input power for one of the six loading conditions.

C. EVM Comparison At 5 GHz

Fig. 6 shows estimated EVM using the vector gain proposed approach versus the direct measured EVM using setup 2 at 5 GHz for a QAM16 signal for four of the six measured loading conditions. Between the lines marked A and B in the figure the difference in the estimated and measured EVM is less than 1 dB for all four loading conditions, however it is important to note that the absolute value of the EVM can also depend on renormalization choices made at firmware level that are difficult to replicate. For this reason, the authors are more inclined to evaluate the agreement across loading conditions between the proposed approach and the direct measured EVM values. The prediction of the constellation points for the loading condition $\Gamma_L = 0.5$ at an input power of -37.85 dBm are shown in Fig. 7In the region between A and B region the estimate is mostly lower than the measured EVM which is explained by the fact that the estimation does only capture the frequency bin self-non-linearities while the cross bin non-linearities are neglected. Left of point A, the measured EVM using setup 2 starts to increase due to the low signal-to-noise ratio in the measurement. To the right of point B, the difference between estimated and measured EVM increases due to limited non-linearities captured by the proposed approach as mentioned before. Across the whole characterized range, the vector gain EVM estimation can predict which of the loading conditions would achieve the lowest EVM. This shows that the approach is useful in providing a first-order estimation of the best loading condition for EVM, directly from measurements without the need of a model/circuit simulation framework.

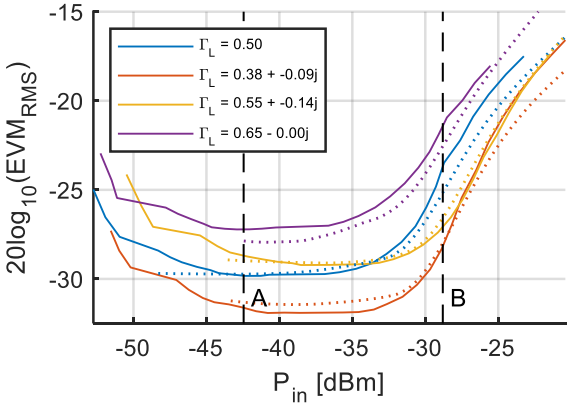


Fig. 6. Correlation between estimated and measured EVM for a QAM16 signal at four loading conditions at 5 GHz. Solid lines are true modulated wave EVM measurements, dashed lines are estimations of the EVM using the vector gain.

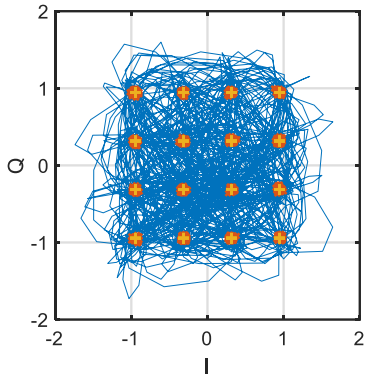


Fig. 7. Constellation points estimated during using the vector gain at 5 GHz for a QAM16 waveform ($P_{in} = -37.85\ dBm$, $\Gamma_L = 0.5$) using 1000 random symbols.

D. *EVM Comparison at 26 GHz*

A similar experiment as the one conducted at 5 GHz was carried out in the at 26 GHz with CW tones as described in section III-A. In this measurement the vector gain based EVM estimation was consistently higher than the direct measured EVM using setup 2 however the agreement across the loading conditions and input power was consistent. This is for example shown in the two contour plots of Fig. 8 showing the EVM trend for one power condition $P_{in} = -22\ dBm$. The two contours show a good correlation between the best loading condition predicted and measured. Similar contour plots can be made using only large signal CW measurement to evaluate bias conditions, frequencies, drive levels and loading conditions allowing the optimum point of operation to be estimated.

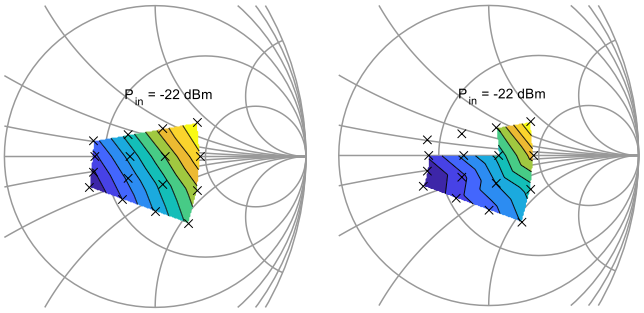


Fig. 8. Comparison between the contour plots of the EVM estimated by the vector gain (left) and measured using setup 2 (right) for one power condition $P_{in} = -22\ dBm$. Two loading conditions in the direct the EVM measurement did not converge (across the whole power range) and have therefore missing values in the right figure.

IV. Optimal Loading Condition Estimation At 165 GHz

To demonstrate the use of the EVM estimation approach, the vector gain of the CMOS devices is measured in a 20 MHz band centered at 165 GHz using a Vertigo mmWave Studio active load pull setup. The vector gain is measured using single tone CW load pull measurements and is used to estimate the EVM for a QAM16 signal.

The vector gain is measured at 11 tones (2 MHz spacing) across the operating band, at 16 source power levels from -37.1 to -7.1 dBm (2 dB spacing) and at six loading conditions. Fig. 5 shows the AM-AM gain compression characteristic and the AM-PM phase variation across the frequency band and input power for one of the six loading conditions at 165 GHz. The EVM for a QAM16 modulated signal is estimated for each of the loading conditions and the result is shown in Fig. 9.

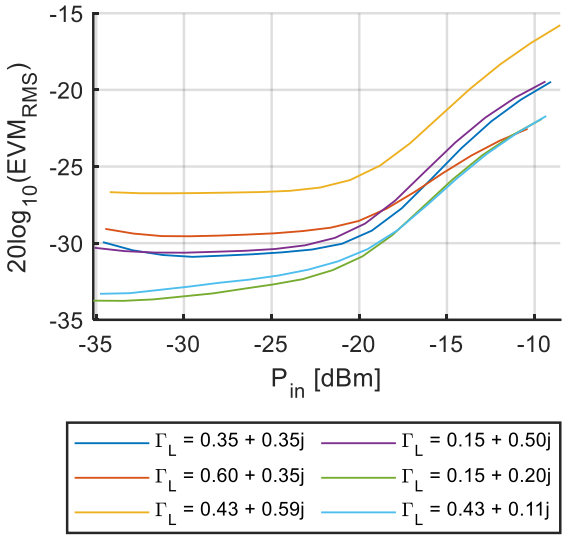


Fig. 9. Vector gain based EVM estimation at 165 GHz for a QAM16 signal at six loading conditions.

With the proposed approach it is possible to predict the optimal loading condition for lowest EVM, $\Gamma_L = 0.15 + 0.20j$ in this case. With the proposed approach of using the vector gain to estimate EVM performance it is possible to evaluate bias conditions, frequencies, drive levels and loading conditions during a large signal CW measurement in the upper millimeter-wave regime without requiring a complex modulated wave setup or the need of a circuit simulator.

V. Conclusion

In this contribution a method is presented for the evaluation of linearity performance under modulated signals of devices operating far into the millimeter-wave regime. The method uses the power dependent vector gain that is extracted during a continuous wave large signal (load pull) measurement of the device under test to directly estimate the EVM performance of the device. The behavioral model was benchmarked at 5 GHz and 26 GHz with modulated wave measurements on a 22nm FD-SOI device transistor. The vector gain approach shows good correlation with measured EVM versus power and applied loading conditions for a QAM16 waveform however the absolute accuracy in predicting the EVM is limited.

As a demonstration the method is used to estimate EVM performance at 165 GHz where other non-linear

characterization methods such as a two-tone test or modulated signal measurements are difficult to perform. The vector gain based EVM estimation could therefore be used for prediction of optimal bias conditions, frequencies, drive levels and loading conditions during a large signal CW measurement in the upper millimeter-wave regime.

ACKNOWLEDGEMENT

The authors would like to thank M. Alavi for his support in the formulation of the vector gain approach for EVM estimation.

REFERENCES

[1] M. Marcus and B. Pattan, "Millimeter wave propagation: spectrum management implications," *IEEE Microwave Magazine*, vol. 6, no. 2, pp. 54-62, 2005.

[2] D. Schreurs, M. O'Droma, A. A. Goacher, and M. Gadringer, *RF power amplifier behavioral modeling*. Cambridge university press New York, NY, USA, 2008.

[3] M. Schröter and B. Ardouin, "The HiCuM bipolar transistor model," in *Compact Modeling*: Springer, 2010, pp. 231-267.

[4] M. Marchetti *et al.*, "Active harmonic load–pull with realistic wideband communications signals," vol. 56, no. 12, pp. 2979-2988, 2008.

[5] L. Galatro, R. Romano, C. D. Martino, and M. Spirito, "Frequency Scalable Power Control and Active Tuning for Sub-THz Large-Signal Measurements," *Microwave Journal, test and measurement channel*, Feb. 2020.

[6] L. Galatro, A. Pawlak, M. Schroter, and M. Spirito, "Capacitively Loaded Inverted CPWs for Distributed TRL-Based De-Embedding at (Sub) mm-Waves," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 12, pp. 4914-4924, 2017.

[7] D. E. Root, J. Verspecht, D. Sharrit, J. Wood, A. J. I. T. o. M. T. Cognata, and Techniques, "Broad-band poly-harmonic distortion (PHD) behavioral models from fast automated simulations and large-signal vectorial network measurements," vol. 53, no. 11, pp. 3656-3664, 2005.

[8] D. E. Root, J. Verspecht, J. Horn, and M. Marcu, *X-Parameters: Characterization, Modeling, and Design of Nonlinear RF and Microwave Components*. Cambridge University Press, 2013.

[9] *Dual Source Module from Virginia Diodes*. Available: https://www.vadiodes.com/en/products/vector-network-analyzer-extension-modules

[10] J. X. Qiu, A. M. Darwish, E. A. Viveiros, K. Kingkeo, and H. A. Hung, "Linearity Characterization and Optimization of Millimeter-Wave GaN HEMTs," (in English), *Ieee Transactions on Microwave Theory and Techniques*, Article; Proceedings Paper vol. 59, no. 12, pp. 3419-3427, Dec 2011.

[11] A. Ferrero and U. Pisani, "An improved calibration technique for on-wafer large-signal transistor characterization," *IEEE Transactions on Instrumentation and Measurement*, vol. 42, no. 2, pp. 360-364, 1993.

[12] M. Marchetti, G. Avolio, M. Squillante, A. K. Doggalli, and B. V. Anteverta-Mw, "Wideband Load Pull Measurement Techniques: Architecture, Accuracy, and Applications," in *2019 92nd ARFTG Microwave Measurement Conference (ARFTG)*, 2019, pp. 1-7.

Bibliography

[1] J. Edstam, J. Hansryd, S. Carpenter, T. Emanuelsson, Y. Li, and H. Zirath, "'Microwave backhaul beyond 100 GHz,'" Ericsson Technology Review, vol. 2, pp. 1–15, 2017.

[2] M. Marcus and B. Pattan, "'Millimeter wave propagation: spectrum management implications,'" IEEE Microwave Magazine, vol. 6, no. 2, pp. 54–62, 2005, ISSN: 1557-9581. DOI: 10.1109/MMW.2005.1491267. [Online]. Available: https://ieeexplore.ieee.org/document/1491267/.

[3] M. Schröter and B. Ardouin, "'The HiCuM bipolar transistor model,'" in *Compact Modeling*. Springer, 2010, pp. 231–267.

[4] D. Schreurs, M. O'Droma, A. A. Goacher, and M. Gadringer, RF power amplifier behavioral modeling. Cambridge university press New York, NY, USA, 2008, ISBN: 0521881730.

[5] L. Galatro, A. Pawlak, M. Schroter, and M. Spirito, "'Capacitively Loaded Inverted CPWs for Distributed TRL-Based De-Embedding at (Sub) mm-Waves,'" IEEE Transactions on Microwave Theory and Techniques, vol. 65, no. 12, pp. 4914–4924, 2017, ISSN: 1557-9670. DOI: 10.1109/TMTT.2017.2727498. [Online]. Available: https://ieeexplore.ieee.org/ielx7/22/8187762/08000339.pdf?tp=&arnumber=8000339&isnumber=8187762&ref=.

[6] C. D. Martino, L. Galatro, R. Romano, G. Parisi, and M. Spirito, "'Hardware and Software Solutions for Active Frequency Scalable (Sub)mm-Wave Load-Pull,'" IEEE Transactions on Microwave Theory and Techniques, pp. 1–1, 2020, ISSN: 1557-9670. DOI: 10.1109/TMTT.2020.3005178.

[7] M. Marchetti, M. J. Pelk, K. Buisman, W. E. Neo, M. Spirito, and L. C. de Vreede, "'Active harmonic load–pull with realistic wideband communications signals,'" IEEE Transactions on Microwave Theory and Techniques, vol. 56, no. 12, pp. 2979–2988, 2008, ISSN: 0018-9480.

[8] L. Galatro and M. Spirito, "'Millimeter-Wave On-Wafer TRL Calibration Employing 3-D EM Simulation-Based Characteristic Impedance Extraction,'" IEEE Transactions on Microwave Theory and Techniques, vol. 65, no. 4, pp. 1315–1323, 2017, ISSN: 1557-9670. DOI: 10.1109/TMTT.2016.2609413.

[9] J. V. Butler, D. K. Rytting, M. F. Iskander, R. D. Pollard, and M. V. Bossche, "'16-term error model and calibration procedure for on-wafer network analysis measurements,'" IEEE Transactions on Microwave Theory and Techniques, vol. 39, no. 12, pp. 2211–2217, 1991, ISSN: 1557-9670. DOI: 10.1109/22.106567. [Online]. Available: https://ieeexplore.ieee.org/ielx1/22/3252/00106567.pdf?tp=&arnumber=106567&isnumber=3252&ref=.

[10] A. Wu, C. Liu, F. Liang, X. Zou, Y. Wang, P. Luan, C. Li, and N. Ridler, "'Calibration on the Fly–A Novel Two-Port S-Parameter Measurement Method for On-Wafer Leaky Systems,'" IEEE Transactions on Microwave Theory and Techniques, pp. 1–1, 2020, ISSN: 1557-9670. DOI: 10.1109/TMTT.2020.2988461.

[11] D. F. Williams, F. Schmückle, R. Doerner, G. N. Phung, U. Arz, and W. Heinrich, "'Crosstalk Corrections for Coplanar-Waveguide Scattering-Parameter Calibrations,'" IEEE Transactions on Microwave Theory and Techniques, vol. 62, no. 8, pp. 1748–1761, 2014, ISSN: 1557-9670. DOI: 10.1109/TMTT.2014.2331623.

[12] Z. Wang, P. Chiang, P. Nazari, C. Wang, Z. Chen, and P. Heydari, "'A CMOS 210-GHz Fundamental Transceiver With OOK Modulation,'" IEEE Journal of Solid-State Circuits, vol. 49, no. 3, pp. 564–580, 2014, ISSN: 1558-173X. DOI: 10.1109/JSSC.2013.2297415.

[13] K. Katayama, M. Motoyoshi, K. Takano, L. C. Yang, and M. Fujishima, "'209mW 11Gbps 130GHz CMOS transceiver for indoor wireless communication,'" in *2013 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, pp. 409–412. DOI: 10.1109/ASSCC.2013.6691069.

[14] B. Suh, H. Lee, S. Kim, and S. Jeon, "'A D-Band Multiplier-Based OOK Transceiver With Supplementary Transistor Modeling in 65-nm Bulk CMOS Technology,'" IEEE Access, vol. 7, pp. 7783–7793, 2019, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2889165.

[15] M. Fujishima, M. Motoyoshi, K. Katayama, K. Takano, N. Ono, and R. Fujimoto, "'98 mW 10 Gbps Wireless Transceiver Chipset With D-Band CMOS Circuits,'" IEEE Journal of Solid-State Circuits, vol. 48, no. 10, pp. 2273–2284, 2013, ISSN: 1558-173X. DOI: 10.1109/JSSC.2013.2261192.

[16] S. Carpenter, D. Nopchinda, M. Abbasi, Z. S. He, M. Bao, T. Eriksson, and H. Zirath, "'A D-Band 48-Gbit/s 64-QAM/QPSK Direct-Conversion I/Q Transceiver Chipset,'" IEEE Transactions on Microwave Theory and Techniques, vol. 64, no. 4, pp. 1285–1296, 2016, ISSN: 1557-9670. DOI: 10.1109/TMTT.2016.2533491.

[17] M. Ito, T. Okawa, and T. Marumoto, "'D-band Transceiver Utilizing 70-nm GaAs-mHEMT Technology for FDD System,'" in *2019 IEEE BiCMOS and Compound semiconductor Integrated Circuits and Technology Symposium (BCICTS)*, pp. 1–4. DOI: 10.1109/BCICTS45179.2019.8972754.

[18] I. Ando, M. Tanio, M. Ito, T. Kuwabara, T. Marumoto, and K. Kunihiro, "'Wireless D-band communication up to 60 Gbit/s with 64QAM using GaAs HEMT technology,'" in *2016 IEEE Radio and Wireless Symposium (RWS)*, pp. 193–195, ISBN: 2164-2974. DOI: 10.1109/RWS.2016.7444402.

[19] F. F. Manzillo, A. Clemente, and J. L. González-Jiménez, "'High-Gain D -Band Transmitarrays in Standard PCB Technology for Beyond-5G Communications,'" IEEE Transactions on Antennas and Propagation, vol. 68, no. 1, pp. 587–592, 2020, ISSN: 1558-2221. DOI: 10.1109/TAP.2019.2938630.

[20] H. Wang, F. Wang, S. Li, T.-Y. Huang, A. S. Ahmed, N. S. Mannem, J. Lee, E. Garay, D. Munzer, C. Snyder, S. Lee, H. T. Nguyen, and M. E. D. Smith, "'Power Amplifiers Performance Survey 2000-Present,'" [Online]. Available: https://gems.ece.gatech.edu/PA_survey.html.

[21] T. Razzak, H. Xue, Z. Xia, S. Hwang, A. Khan, W. Lu, and S. Rajan, "'Ultra-wide band gap materials for high frequency applications,'" in *2018 IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP)*, pp. 1–3. DOI: 10.1109/IMWS-AMP.2018.8457144.

[22] D. Garrido-Diez and I. Baraia, "'Review of wide bandgap materials and their impact in new power devices,'" in *2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)*, pp. 1–6. DOI: 10.1109/ECMSM.2017.7945876.

[23] M. Rais-Zadeh, V. J. Gokhale, A. Ansari, M. Faucher, D. Théron, Y. Cordier, and L. Buchaillot, "'Gallium Nitride as an Electromechanical Material,'" Journal of Microelectromechanical Systems, vol. 23, no. 6, pp. 1252–1271, 2014, ISSN: 1941-0158. DOI: 10.1109/JMEMS.2014.2352617.

[24] E. Johnson, "'Physical limitations on frequency and power parameters of transistors,'" in *1958 IRE International Convention Record*, vol. 13, pp. 27–34. DOI: 10.1109/IRECON.1965.1147520.

[25] D. A. Neamen, Semiconductor physics and devices: basic principles, 4th ed. New York, NY: McGraw-Hill, 2012, p. 758, ISBN: 9780073529585.

[26] V. Teppati, A. Ferrero, and M. Sayed, Modern RF and microwave measurement techniques. Cambridge University Press, 2013, ISBN: 1107245184.

[27] L. Galatro, "'Advanced calibration and measurement techniques for (sub)millimeter wave devices characterization,'" Thesis, 2019. DOI: 10.4233/uuid:245192d2-34c4-44a8-81da-61f12fda5c33.

[28] M. Marchetti, M. J. Pelk, K. Buisman, W. C. E. Neo, M. Spirito, and L. C. N. de Vreede, "'Active Harmonic Load-Pull With Realistic Wideband Communications Signals,'" IEEE Transactions on Microwave Theory and Techniques, vol. 56, no. 12, pp. 2979–2988, 2008, ISSN: 0018-9480. DOI: 10.1109/Tmtt.2008.2007330. [Online]. Available: %3CGo%20to%20ISI%3E://WOS:000261895200035.

[29] M. Marchetti, G. Avolio, M. Squillante, A. K. Doggalli, and B. V. Anteverta-Mw, "'Wideband Load Pull Measurement Techniques: Architecture, Accuracy, and Applications,'" in *2019 92nd ARFTG Microwave Measurement Conference (ARFTG)*, pp. 1–7. DOI: 10.1109/ARFTG.2019.8637240. [Online]. Available: https://ieeexplore.ieee.org/ielx7/8632668/8637213/08637240.pdf?tp=&arnumber=8637240&isnumber=8637213&ref=.

[30] L. Galatro, M. Marchetti, and M. Spirito, "'60 GHz mixed signal active load-pull system for millimeter wave devices characterization,'" in *80th ARFTG Microwave Measurement Conference*, pp. 1–6. DOI: 10.1109/ARFTG.2012.6422434.

[31] D. Rytting, "'Network Analyser Error Models and Calibration Methods,'" White Paper,, 1998.

[32] A. Ferrero and U. Pisani, "'Two-port network analyzer calibration using an unknown 'thru','" IEEE Microwave and Guided Wave Letters, vol. 2, no. 12, pp. 505–507, 1992, ISSN: 1558-2329. DOI: 10.1109/75.173410. [Online]. Available: https://ieeexplore.ieee.org/ielx4/75/4451/00173410.pdf?tp=&arnumber=173410&isnumber=4451&ref=.

[33] H. Eul and B. Schiek, "'Thru-Match-Reflect: One Result of a Rigorous Theory for De-Embedding and Network Analyzer Calibration,'" in *1988 18th European Microwave Conference*, pp. 909–914. DOI: 10.1109/EUMA.1988.333924.

[34] G. F. Engen and C. A. Hoer, "'Thru-Reflect-Line: An Improved Technique for Calibrating the Dual Six-Port Automatic Network Analyzer,'" IEEE Transactions on Microwave Theory and Techniques, vol. 27, no. 12, pp. 987–993, 1979, ISSN: 1557-9670. DOI: 10.1109/TMTT.1979.1129778.

[35] L. Galatro, S. Galbano, A. Santaniello, and M. Spirito, "'Power control for S-parameters and large signal characterization at (sub)-mmWave frequencies,'" in *2015 85th Microwave Measurement Conference (ARFTG)*, IEEE, pp. 1–4, ISBN: 1479988871.

[36] C. D. Martino, Z. Hu, L. Galatro, G. Sarris, and M. Spirito, "'Power level control of mm-wave test benches for accurate small and large-signal DUT measurements,'" in *2016 88th ARFTG Microwave Measurement Conference (ARFTG)*, pp. 1–4. DOI: 10.1109/ARFTG.2016.7839723.

[37] A. Ferrero and U. Pisani, "'An improved calibration technique for on-wafer large-signal transistor characterization,'" IEEE Transactions on Instrumentation and Measurement, vol. 42, no. 2, pp. 360–364, 1993, ISSN: 1557-9662. DOI: 10.1109/19.278582. [Online]. Available: https://ieeexplore.ieee.org/document/278582/.

[38] S. N. Ong, S. Lehmann, W. H. Chow, C. Zhang, C. Schippel, L. H. K. Chan, Y. Andee, M. Hauschildt, K. K. S. Tan, J. Watts, C. K. Lim, A. Divay, J. S. Wong, Z. Zhao, M. Govindarajan, C. Schwan, A. Huschka, A. Bcllaouar, W. LOo, J. Mazurier, C. Grass, R. Taylor, K. W. J. Chew, S. Embabi, G. Workman, A. Pakfar, S. Morvan, K. Sundaram, M. T. Lau, B. Rice, and D. Harame, "'A 22nm FDSOI Technology Optimized for RF/mmWave Applications,'" in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 72–75, ISBN: 2375-0995. DOI: 10.1109/RFIC.2018.8429035.

[39] H. Rücker and B. Heinemann, "'SiGe BiCMOS technology for mm-wave systems,'" in *2012 International SoC Design Conference (ISOCC)*, pp. 266–268. DOI: 10.1109/ISOCC.2012.6407091.

[40] J. Böck, K. Aufinger, S. Boguth, C. Dahl, H. Knapp, W. Liebl, D. Manger, T. F. Meister, A. Pribil, J. Wursthorn, R. Lachner, B. Heinemann, H. Rücker, A. Fox, R. Barth, G. Fischer, S. Marschmeyer, D. Schmidt, A. Trusch, and C. Wipf, "'SiGe HBT and BiCMOS process integration optimization within the DOTSEVEN project,'" in *2015 IEEE Bipolar/BiCMOS Circuits and Technology Meeting - BCTM*, pp. 121–124. DOI: 10.1109/BCTM.2015.7340549. [Online]. Available: https://ieeexplore.ieee.org/ielx7/7329075/7340543/07340549.pdf?tp=&arnumber=7340549&isnumber=7340543&ref=.

[41] C. D. Martino, E. S. Malotaux, and M. Spirito, "'A TRL Error-box Split Procedure to Compensate for the Bias Dependency Effects in Device Test-Fixtures,'" in *2019 93rd ARFTG Microwave Measurement Conference (ARFTG)*, pp. 1–6. DOI: 10.1109/ARFTG.2019.8739197.

[42] Virginia Diodes, Inc., "'Dual Source Module from Virginia Diodes,'" [Online]. Available: https://www.vadiodes.com/en/products/vector-network-analyzer-extension-modules.

[43] F. Vanaverbeke, L. Zhang, D. Holmes, M. Marchetti, and K. Kim, "'De-mystifying AM-PM characteristics through the definition of the complex Transducer Gain ĜT,'" in *2020 94th ARFTG Microwave Measurement Symposium (ARFTG)*, pp. 1–8. DOI: 10.1109/ARFTG47584.2020.9071718. [Online]. Available: https://ieeexplore.ieee.org/document/9071718/.

[44] Keysight Technologies, "'Circuit Envelope Simulation,'" 2011. [Online]. Available: http://edadownload.software.keysight.com/eedl/ads/2011/pdf/cktsimenv.pdf.

[45] D. E. Root, J. Verspecht, J. Horn, and M. Marcu, X-Parameters: Characterization, Modeling, and Design of Nonlinear RF and Microwave Components. Cambridge University Press, 2013, ISBN: 0521193230.

[46] M. Vigilante, E. McCune, and P. Reynaert, "'To EVM or Two EVMs?: An Answer to the Question,'" IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 36–39, 2017, ISSN: 1943-0590. DOI: 10.1109/MSSC.2017.2714398. [Online]. Available: https://ieeexplore.ieee.org/ielx7/4563670/8016454/08016487.pdf?tp=&arnumber=8016487&isnumber=8016454&ref=.

[47] Keysigth Technologies, "'ADS VTB Documentation,'" [Online]. Available: http://edadownload.software.keysight.com/eedl/ads/2017update1/Add-ons/vtb/pdf/ADS_Verification_Test_Bench.pdf.

[48] S. Kulkarni and P. Reynaert, "'A 60-GHz Power Amplifier With AM–PM Distortion Cancellation in 40-nm CMOS,'" IEEE Transactions on Microwave Theory and Techniques, vol. 64, no. 7, pp. 2284–2291, 2016, ISSN: 1557-9670. DOI: 10.1109/TMTT.2016.2574866. [Online]. Available: https://ieeexplore.ieee.org/ielx7/22/7506364/07494673.pdf?tp=&arnumber=7494673&isnumber=7506364&ref=.

[49] K. M. Gharaibeh, K. G. Gard, and M. B. Steer, "'In-band distortion of multisines,'" IEEE Transactions on Microwave Theory and Techniques, vol. 54, no. 8, pp. 3227–3236, 2006, ISSN: 1557-9670. DOI: 10.1109/TMTT.2006.879170.

[50] Y. S. Chauhan, S. Venugopalan, M. Chalkiadaki, M. A. U. Karim, H. Agarwal, S. Khandelwal, N. Paydavosi, J. P. Duarte, C. C. Enz, A. M. Niknejad, and C. Hu, "'BSIM6: Analog and RF Compact Model for Bulk MOSFET,'" IEEE Transactions on Electron Devices, vol. 61, no. 2, pp. 234–244, 2014, ISSN: 1557-9646. DOI: 10.1109/TED.2013.2283084.

[51] Virginia Diodes, Inc., "'Vector Network Analyzer Extension Modules (VNAX),'" [Online]. Available: https://www.vadiodes.com/en/products/vector-network-analyzer-extension-modules.

[52] ——, "'Power Meter (PM5),'" [Online]. Available: https://www.vadiodes.com/en/power-meter-pm5.

[53] J. X. Qiu, A. M. Darwish, E. A. Viveiros, K. Kingkeo, and H. A. Hung, "'Linearity Characterization and Optimization of Millimeter-Wave GaN HEMTs,'" Ieee Transactions on Microwave Theory and Techniques, vol. 59, no. 12, pp. 3419–3427, 2011, ISSN: 0018-9480. DOI: 10.1109/tmtt.2011.2171986. [Online]. Available: %3CGo%20to%20ISI%3E://WOS:000298052000018%20https://ieeexplore.ieee.org/document/6082421/.

[54] D. F. Williams and R. B. Marks, "'Calibrating On-Wafer Probes to the Probe Tips,'" in *40th ARFTG Conference Digest*, vol. 22, pp. 136–143. DOI: 10.1109/ARFTG.1992.327008.

[55] C. D. Martino, "'PHd. Dissertation (unpublished work),'" Unpublished Work, 2019.

[56] R. A. Speciale, "'A Generalization of the TSD Network-Analyzer Calibration Procedure, Covering n-Port Scattering-Parameter Measurements, Affected by Leakage Errors,'" IEEE Transactions on Microwave Theory and Techniques, vol. 25, no. 12, pp. 1100–1115, 1977, ISSN: 1557-9670. DOI: 10.1109/TMTT.1977.1129282.

[57] N. Mohamed, H. Agarwal, C. Gupta, and Y. S. Chauhan, "'Compact modeling of nonquasi-static effect in bulk MOSFETs for RF circuit design in sub-THz regime,'" in *2016 3rd International Conference on Emerging Electronics (ICEE)*, pp. 1–4. DOI: 10.1109/ICEmElec.2016.8074573.

[58] H. Heuermann and B. Schiek, "'Results of network analyzer measurements with leakage errors corrected with the TMS-15-term procedure,'" in *1994 IEEE MTT-S International Microwave Symposium Digest (Cat. No.94CH3389-4)*, 1361–1364 vol.3. DOI: 10.1109/MWSYM.1994.335297.

[59] K. J. Silvonen, "'Calibration of 16-term error model (microwave measurement),'" Electronics Letters, vol. 29, no. 17, pp. 1544–1545, 1993, ISSN: 0013-5194. DOI: 10.1049/el:19931029.

[60] L. Galatro and M. Spirito, "'Analysis of residual errors due to calibration transfer in on-wafer measurements at mm-wave frequencies,'" in *2015 IEEE Bipolar/BiCMOS Circuits and Technology Meeting - BCTM*, pp. 141–144. DOI: 10.1109/BCTM.2015.7340569.

[61] D. A. Frickey, "'Conversions between S, Z, Y, H, ABCD, and T parameters which are valid for complex source and load impedances,'" IEEE Transactions on Microwave Theory and Techniques, vol. 42, no. 2, pp. 205–211, 1994, ISSN: 1557-9670. DOI: 10.1109/22.275248. [Online]. Available: https://ieeexplore.ieee.org/ielx1/22/6818/00275248.pdf?tp=&arnumber=275248&isnumber=6818&ref=.

[62] J. Pillow, "'SVD applications: rank, column, row, and null spaces,'" Statistical Modeling and Analysis of Neural Data (NEU 560), 2018. [Online]. Available: http://pillowlab.princeton.edu/teaching/statneuro2018/slides/notes03a_SVDandLinSys.pdf.

[63] J. T. Valliarampath and S. Sinha, "'Designing Linear PAs at Millimeter-Wave Frequencies Using Volterra Series Analysis,'" Canadian Journal of Electrical and Computer Engineering-Revue Canadienne De Genie Electrique Et Informatique, vol. 38, no. 3, pp. 232–237, 2015, ISSN: 0840-8688. DOI: 10.1109/cjece.2015.2434941. [Online]. Available: %3CGo%20to%20ISI%3E://WOS:000362408000004%20https://ieeexplore.ieee.org/ielx7/9754/7230329/07230330.pdf?tp=&arnumber=7230330&isnumber=7230329&ref=.

[64] E. McCune, "'5G-NR Bandwidth Efficient Modulation Options for Efficient Link Operation that are Compatible with mmW Transistor Nonlinearities,'" in *2018 IEEE 5G World Forum (5GWF)*, pp. 289–293. DOI: 10.1109/5GWF.2018.8516944. [Online]. Available: https://ieeexplore.ieee.org/ielx7/8489870/8516707/08516944.pdf?tp=&arnumber=8516944&isnumber=8516707&ref=.

[65] E. P. Vandamme and G. Badenes, "'Impact of non-quasi-static effects on the high frequency small-signal behaviour of MOSFETs,'" in *ICM 2000. Proceedings of the 12th International Conference on Microelectronics. (IEEE Cat. No.00EX453)*, pp. 371–374. DOI: 10.1109/ICM.2000.916480.

[66] S. Yadav, A. Chakravorty, and M. Schröter, "'Small-signal modeling of the lateral NQS effect in SiGe HBTs,'" in *2014 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 203–206. DOI: 10.1109/BCTM.2014.6981315.

[67] Z. Hu, G. Sarris, C. D. Martino, M. Spirito, and E. McCune, "'Design and linearity analysis of a D-band power amplifier in 0.13 µm SiGe BiCMOS technology,'" in *2017 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, pp. 1–4, ISBN: 2374-8443. DOI: 10.1109/CSICS.2017.8240469. [Online]. Available: https://ieeexplore.ieee.org/ielx7/8211425/8240418/08240469.pdf?tp=&arnumber=8240469&isnumber=8240418&ref=.

[68] S. Cripps, "'Waveform Engineering in RFPAs - Facts and Fallacies,'" European Microwave Week, 2019. [Online]. Available: https://vimeo.com/411406377/15e573c615.

[69] E. McCune, "'Operating Modes of Dynamic Power Supply Transmitter Amplifiers,'" IEEE Transactions on Microwave Theory and Techniques, vol. 62, no. 11, pp. 2511–2517, 2014, ISSN: 1557-9670. DOI: 10.1109/TMTT.2014.2360858.

[70] D. Lovelace, J. Costa, and N. Camilleri, "'Extracting small-signal model parameters of silicon MOSFET transistors,'" in *1994 IEEE MTT-S International Microwave Symposium Digest (Cat. No.94CH3389-4)*, 865–868 vol.2. DOI: 10.1109/MWSYM.1994.335220.

[71] S. M. Homayouni, D. M. M. Schreurs, G. Crupi, and B. K. J. C. Nauwelaers, "'Technology-Independent Non-Quasi-Static Table-Based Nonlinear Model Generation,'" IEEE Transactions on Microwave Theory and Techniques, vol. 57, no. 12, pp. 2845–2852, 2009, ISSN: 1557-9670. DOI: 10.1109/TMTT.2009.2033840.

[72] G. Gonzalez, Microwave transistor amplifiers analysis and design. Prentice-Hall, Inc., 1996, ISBN: 0132543354.

[73] B. Razavi and R. Behzad, RF microelectronics. Prentice Hall New Jersey, 1998, vol. 2.