RELIABILITY ENGINEERING IN RF CMOS

Guido T. Sasse

Samenstelling promotiecommissie:

Voorzitter:	prof.dr.ir. J. van Amerongen	Universiteit Twente
Secretaris:	prof.dr.ir. J. van Amerongen	Universiteit Twente
Promotoren:	prof.dr. J. Schmitz	Universiteit Twente, NXP Semiconductors
	prof.dr.ir. F.G. Kuper	Universiteit Twente
Referenten:	dr.ing. L.C.N. de Vreede	Technische Universiteit Delft
	dr. P.J. van der Wel	NXP Semiconductors
Leden:	prof.dr.ir. W. van Etten	Universiteit Twente
	prof.dr. G. Groeseneken	Katholieke Universiteit Leuven, IMEC
	prof.dr.ir. A.J. Mouthaan	Universiteit Twente

The work described in this thesis was supported by the Dutch Technology Foundation STW (Reliable RF, TCS.6015) and carried out in the Semiconductor Components Group, MESA$^+$ Institute for Nanotechnology, University of Twente, The Netherlands.

G.T. Sasse
Reliability Engineering in RF CMOS
Ph.D. thesis, University of Twente, The Netherlands

ISBN: 978-90-365-2690-6

Cover design: J. Warnaar

Printed by PrintPartners Ipskamp, Enschede

© G.T. Sasse, Enschede, 2008

RELIABILITY ENGINEERING IN RF CMOS

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof.dr. W.H.M. Zijm,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op 4 juli 2008 om 15.00 uur

door

Guido Theodor Sasse
geboren op 29 september 1977
te Lochem

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. J. Schmitz
prof.dr.ir. F.G. Kuper

Contents

1 Introduction **1**
- 1.1 RF CMOS . . . 1
- 1.2 Reliability engineering . . . 2
- 1.3 MOSFET degradation mechanisms . . . 4
  - 1.3.1 Hot carrier degradation . . . 4
  - 1.3.2 Gate-oxide breakdown . . . 5
  - 1.3.3 NBTI . . . 6
- 1.4 Outline of the thesis . . . 7

2 RF MOS measurements **9**
- 2.1 Introduction . . . 9
- 2.2 RF vs. AC/DC measurements . . . 9
- 2.3 Small signal two-port characterization . . . 12
  - 2.3.1 *s*-parameters . . . 12
  - 2.3.2 Calibration . . . 13
  - 2.3.3 De-embedding . . . 14
- 2.4 Linear one-port RF voltage generation . . . 16
  - 2.4.1 Setting the amplitude . . . 16
  - 2.4.2 Verifying linearity . . . 19
- 2.5 Conclusions . . . 23

3 MOSFET degradation under RF stress **25**
- 3.1 Introduction . . . 25
- 3.2 RF hot carrier degradation . . . 26
  - 3.2.1 DC model . . . 26
  - 3.2.2 AC effects . . . 27
  - 3.2.3 Measurement setup . . . 29
  - 3.2.4 Measurement results . . . 32
  - 3.2.5 Discussion . . . 36
- 3.3 RF NBTI degradation . . . 37
  - 3.3.1 DC model . . . 37
  - 3.3.2 AC effects . . . 39
  - 3.3.3 Measurement setup . . . 39

3.3.4 Measurement results 40
3.3.5 Discussion 42
3.4 Gate-oxide breakdown under RF stress 43
3.4.1 DC model 43
3.4.2 AC effects 44
3.4.3 Measurement setup 45
3.4.4 Measurement results 46
3.4.5 Discussion 48
3.5 Conclusions 49

4 RF PA lifetime prediction 51
4.1 Introduction 51
4.2 Degraded MOSFET model 52
4.2.1 Breakdown paths 52
4.2.2 Hot carrier degradation 55
4.3 Design of the simulator 59
4.3.1 Probability function for multiple breakdown events 60
4.3.2 Implementation 61
4.4 Simulation results 63
4.4.1 Description of the evaluated circuits 63
4.4.2 Stress conditions 64
4.4.3 Circuit lifetime prediction 69
4.5 Discussion 73
4.6 Conclusions 75

5 RF reliability characterization 77
5.1 Introduction 77
5.2 Capacitance extraction from RF C-V measurements 78
5.2.1 Basics of the RF C-V technique 78
5.2.2 Two-port analysis of the test structure 79
5.2.3 Capacitance extraction methodologies 82
5.2.4 Comparison of the extraction methodologies 84
5.2.5 Discussion 88
5.3 Charge pumping at radio frequencies 88
5.3.1 Introduction 88
5.3.2 Measurement setup and methodology 90
5.3.3 Measurement results 93
5.3.4 Trap Response 96
5.3.5 Applications of RF CP measurements 100
5.3.6 Discussion 104
5.4 Conclusions 105

6 Conclusions and recommendations 107
6.1 General conclusions 107
6.2 Recommendations for future work 108

A Calibration equations **121**
A.1 Error terms 121
A.2 Calculating $V_{\mathrm{DUT,pp}}$ 123

Summary **125**

Samenvatting **127**

List of publications **129**

Dankwoord **131**

Chapter 1

Introduction

1.1 RF CMOS

The use of mobile wireless communication systems is rapidly increasing. Electronics involved in these systems are typically referred to as RF electronics. RF stands for radio frequencies, a term covering all frequencies of the electromagnetic spectrum used for transmitting radio signals. It spans from a few Hz to 100's of GHz. When dealing with present-day mobile applications the frequencies involved typically range between ∼500 MHz and ∼ 5 GHz. For example the four frequency bands used for GSM communication throughout the world are located around 850, 900, 1800 and 1900 MHz. In electronics design the term RF is often used to describe only those frequencies at which design and measurement issues start to arise that are typical for the high frequencies used in mobile wireless communication systems. Throughout this thesis the term RF also reflects only these high frequencies; in chapter 2 the precise definition of RF as used in this thesis will be explained.

The IC technologies involved in wireless systems operating at such high frequencies include GaAs, SiGe, Bipolar, BiCMOS and CMOS. In earlier years CMOS was not applicable for RF electronics, but through the efforts made in scaling CMOS technology, the RF performance of CMOS has increased significantly [1]. As a consequence CMOS has become the dominating technology used for mobile wireless applications, like it has already dominated logic applications, such as CPU's, for decades. When used in RF electronics, CMOS is often referred to as RF CMOS.

The only part of mobile applications in which CMOS still does not dominate is the Power Amplifier (PA). This module is used in the transmitting end of a wireless communication system. In the PA an electromagnetic wave is generated carrying all information, such as speech or data, which is fed to an antenna. The information this wave carries must be read at the receiver end, such as the base-station of a GSM network, of the wireless communication system. For the information to be available at the receiver end, the power of this electromagnetic

wave must be sufficiently large in order to prevent loss of information between the transmitter and the receiver.

The reason for the absence of CMOS in a PA is the fact that present-day CMOS technologies operate at relatively low voltage levels. While this may be desired for logic applications, it is easy to verify that when the maximum output voltage of a circuit is relatively low, its maximum output power will also be relatively low. This is the main bottleneck of implementing all electronics of a wireless mobile application into a single chip.

Operating the MOSFETs in an RF CMOS circuit at voltage levels exceeding nominal supply voltage may possibly be allowed. This requires a good understanding on the limitations of RF CMOS in terms of stress conditions. Lifetime specifications of CMOS are at present only well-investigated for DC and low frequency conditions. In general the reliability performance of RF CMOS is not well understood at present; designers make use of design specifications made for digital CMOS. Developing such specifications for RF CMOS may relax these design guidelines and as a consequence performance of RF CMOS electronics may be boosted. In the field of reliability engineering the topic of RF CMOS has only been addressed marginally. In this thesis new developments are described for a better incorporation of RF CMOS in the field of reliability engineering. As PA's are a critical block in terms of reliability, the developments described in this thesis were performed with a special attention to the design of reliable RF PA's.

1.2 Reliability engineering

Reliability engineering is an engineering field that deals with the reliability of products. *Reliability*, as defined in [2] is the probability of operating a product for a given time under specified conditions without failure. Reliability engineering makes it possible to set specifications on the number of products that still operate under normal use conditions years ahead in the future, thereby predicting its physical lifetime. Typically, the electronics in mobile applications have an economic lifetime of a couple of years. If it is found that the physical lifetime is far beyond what is required, designs may be adjusted in such a way that the performance of the product can be boosted. In this context the *physical lifetime* of a product is defined as the moment in time at which the reliability of the product has decreased below a given value. Throughout this thesis, the *lifetime* always refers to this physical lifetime.

A term related to reliability is *degradation*; this is used to describe the change in performance of a product or one of the components the product is composed of. If the degradation has reached a given level, the product is said to have reached its lifetime. The *degradation rate* is the speed at which this process of degradation takes place.

Physical lifetime prediction requires a good description of the reliability of the product as a function of time. For this purpose a reliability engineer has several tools, ranging from statistical algorithms to failure analysis tools. Describing the

complete set of tools used by reliability engineers lies outside the scope of this thesis. In this thesis new developments are described in three important topics of reliability engineering, with a focus on RF CMOS:

1. *Understanding the physical processes involved in the degradation of a product.* If the physical mechanisms are well understood, their effect can accurately be extrapolated to the future for making lifetime predictions. In the field of micro-electronics this so-called physics-of-failure approach is the custom way of dealing with reliability [3]. Different failure mechanisms have been found and in the next section the three typically encountered failure mechanisms in MOSFETs will briefly be discussed. In this thesis the manifestation of these degradation mechanisms under RF conditions is investigated.

2. *Characterization of the level of degradation of a product and its components.* As the degradation of a product, and the components it is composed of, under normal use conditions takes place on very long timescales, information on the degradation rate is typically obtained from *accelerated stress* experiments. In such experiments a product is operated under stress conditions much more severe than it would under normal operation conditions and as a consequence the degradation takes place on a timescale much smaller than the expected physical lifetime under normal operation. Assessing the amount of degradation involves the use of specifically designed characterization techniques. The development of characterization methods is an essential part of reliability engineering. In this thesis this topic will be addressed: new characterization tools are developed for the reliability evaluation of CMOS, making use of RF measurement techniques. These characterization tools may be applied in the reliability evaluation of both RF CMOS as well as digital CMOS.

3. *Translating degradation rates of individual components to product lifetime.* Even with accurate knowledge on the physical processes involved with degradation and experimentally obtained degradation rates of the different components of a product, it is not straightforward to predict product lifetime under use conditions. For this to be available use can be made of several tools. An important tool that is used, especially in the field of micro-electronics, is reliability simulation. In a reliability simulator the degradation of a product can be predicted as a function of time, based on the degradation of the individual components of the product. In this thesis the development of a new reliability simulator is described, intended to model the lifetime of PA's in RF CMOS.

While reliability engineering spans an even broader set of topics, this is a good starting point for the development of reliability engineering tools for RF CMOS. This development is what will be discussed in the following chapters.

1.3 MOSFET degradation mechanisms

Several degradation mechanisms can affect the performance of CMOS circuits. In this section the three well-accepted degradation mechanisms found in the active devices in a CMOS circuit, i.e. the MOSFETs, will be described. Other degradation mechanisms may also affect the performance of CMOS circuits, but in this thesis only those mechanisms affecting the actual MOSFETs are put under investigation. All experimental verification as described in this thesis is performed on-wafer. Only individual MOSFETs are considered, thereby not investigating reliability issues related to interconnects, such as electromigration.

The degradation mechanisms discussed all affect either the gate-oxide or the interface between the gate-oxide and the silicon substrate. Important elements in all three degradation mechanisms are defects in the crystal structure of the gate-oxide. These defects are often referred to as *traps*; traps near the interface between the oxide and the silicon substrate are called *interface traps* and traps further away from the silicon substrate are called *oxide traps*. Interface traps are sometimes also called *interface states*. In the three mechanisms discussed below these traps are either being formed or they are filled with electrons or holes originating from the silicon substrate, thereby causing device degradation. A filled oxide trap causes a fixed oxide charge in a MOSFET.

1.3.1 Hot carrier degradation

Hot carrier degradation is the effect caused by high energetic charge carriers flowing in the channel of a MOSFET. It affects both nMOSFETs and pMOSFETs. When a MOSFET is biased in inversion and a drain-source voltage V_{DS} is applied, charge carriers flow from the source region towards the drain. These charge carriers gain energy from the electric field induced by the applied V_{DS}. The most energy is gained in the region alongside the channel where the lateral electric field is highest; this will be near the drain region. If one looks at the distribution of the kinetic energy of the carriers near the drain, and compares it to a population without external bias it appears that the carrier population resembles a population at a temperature higher than the temperature of the silicon. The carriers are said to have become hot. When these hot carriers have an energy sufficiently high, they may cause damage to the device. This damage can be the result of trapping of charge carriers in oxide traps or the formation of new oxide traps or interface states.

Some of the high energetic carriers do not directly cause damage to the device, but they are the origin of a substrate current I_{sub} or gate current I_{G}. Both these currents can be used to monitor the device degradation due to hot carrier stress. I_{G} originates from Channel Hot Carriers (CHC). In CHC mode some hot carriers have gained sufficient energy to surmount the Si-SiO_2 barrier. These carriers may reach the gate and thereby contribute to a measurable I_{G}. I_{sub} results from Drain Avalanche Hot Carriers (DAHC). In this mode some hot carriers in the channel cause impact ionization, thereby generating a new electron hole pair.

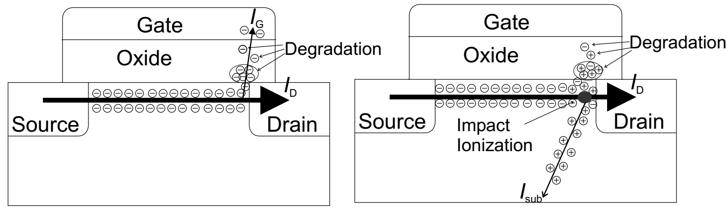

Figure 1.1: Illustration of CHC (left) and DAHC (right) in an nMOSFET. Electrons flow from the source towards the drain, thereby gaining kinetic energy from the applied V_{DS}. Some of these carriers cause device degradation, while other contribute to a measurable I_{G} (CHC mode) or I_{sub} (DAHC mode).

As a consequence carriers may flow towards the substrate contact (holes in an nMOSFET an electrons in a pMOSFET), thereby contributing to a measurable I_{sub}.

Both I_{G} and I_{sub} are indicators for the kinetic energy distribution of the charge carriers in the channel. They can be used for estimating the amount of degradation of a device, without measuring the device parameters. Accelerated stress experiments for characterizing hot-carrier degradation in MOSFETs are often performed at bias conditions for maximum I_{sub} or I_{G}. Also in reliability simulators use is made of I_{sub} and I_{G}, obtained from a circuit simulation; device degradation due to hot carriers is directly linked to these currents.

In CHC mode only carriers of one polarity are involved, while in DAHC both electrons and holes contribute to device degradation. CHC and DAHC are illustrated in figure 1.1 for an nMOSFET.

1.3.2 Gate-oxide breakdown

Gate-oxide breakdown is the sudden formation of a conductive path in the gate-oxide of a MOS device. Oxide breakdown is generally considered a two-step process [4]: in the first phase a gradual build-up of damage occurs in the oxide and the second step is the sudden formation of a breakdown path. Different models for describing the degradation phase have been proposed, such as the anode hole injection [5], the anode hydrogen release [6] or the thermochemical model [7].

The formation of the breakdown path can be explained using percolation theory [8], as illustrated in figure 1.2. Due to a stress signal at the gate, traps are being formed inside the oxide with a certain trap generation rate. The exact position where individual traps are formed are randomly distributed; if two traps are situated close to each other, charge can easily flow from one trap to another. In this way clusters of conducting paths will be formed throughout the oxide. Once a critical defect density has been reached these clusters will be distributed in such a way that charge can flow from the anode of the gate dielectric to the cathode. At this moment a large discharge will take place, thereby generating the conductive path in the oxide. This is the moment of gate-oxide breakdown.

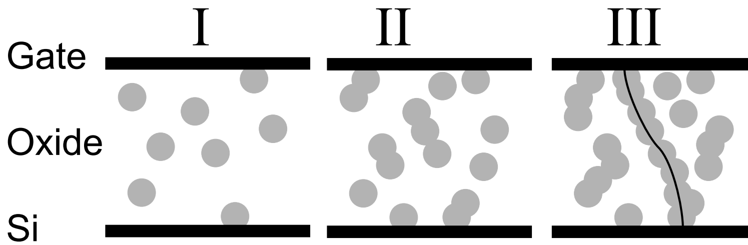

Figure 1.2: Illustration of the formation of a breakdown path as it can be modeled using percolation theory. Three different phases of degradation are shown. Traps are represented as spheres, randomly distributed throughout the oxide. In stage I a few traps have been formed, in stage II clusters of traps can be recognized and in stage III a sufficient number of traps has been formed to have a continuous link of traps between the gate and the substrate. This is the onset of oxide breakdown.

After a breakdown path has been formed it can manifest itself in different ways:

- In hard breakdown, a large gate current increase can be observed. This type of breakdown is typically found in thick oxide devices and high voltage stress.
- In soft breakdown [9] a small sudden gate current increase and a sudden gate current noise increase is observed.
- Progressive breakdown [10] is the term used for the non-instantaneous formation of a hard breakdown path.

Hard breakdown has the most disastrous effect on device performance. The effect of progressive breakdown is similar, but only after the degradation has progressed significantly.

1.3.3 NBTI

Negative bias temperature instability (NBTI) is the name for the degradation mechanism mainly seen in pMOSFETs, where a relatively small negative gate voltage can cause device degradation. It cannot be explained using either high energetic charge carriers or degradation mechanisms associated with a tunneling current through the gate-oxide. A similar mechanism can be observed in nMOSFETs, labeled PBTI. NBTI and PBTI combined are referred to as BTI. The first report on NBTI was published in 1966 [11]. The effect has not been considered as an important reliability hazard for a long time. However due to the scaling of the supply voltage and the threshold voltage V_T in CMOS technology, the effect has

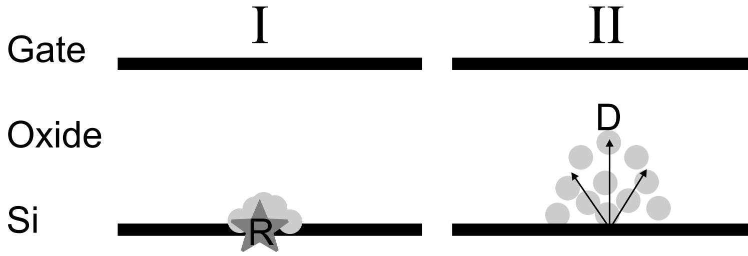

Figure 1.3: Schematic illustration of the R-D process for explaining NBTI. In the first phase an electrochemical reaction takes place at the interface between the silicon substrate and the oxide. In this reaction a diffusion species is released. In the second phase this species diffuses into the oxide.

become more and more important since late 1990's [12, 13, 14]. A lot of questions answering the mechanism underlying this effect have since then be answered while other questions still remain open.

It is commonly believed that during NBTI stress both interface traps and positive oxide charge are being formed. Furthermore it is well known that NBTI may recover after removal of the stress signal. This recovery effect may severely affect the obtained degradation rate under NBTI stress if characterization is performed using conventional stress-measure-stress sequences.

The exact nature of the physical mechanism underlying NBTI degradation is still a matter of debate, but the reaction-diffusion (R-D) model is generally believed to appropriately model interface generation under NBTI stress [15, 16, 17, 18]. In this model, illustrated in figure 1.3 it is believed that a two-step process is involved. First a reaction takes place at the Si-SiO_2 interface, thereby breaking Si-H bonds, this is followed by diffusion of some hydrogen species into the oxide. Only for very short stress times the reaction rate controls the degradation rate under NBTI stress. For stress times $>$ 0.02-0.03 s [17], the degradation rate becomes limited by the diffusion process.

While it is generally accepted that the nature of interface state degradation can be very well described using the R-D model, controversy still exists on the exact nature of the both the positive oxide charge [17] and the recovery mechanism. Some authors claim that the recovery effect can be completely explained from the R-D framework [15, 16], while others attribute this completely or partly to the detrapping of holes in deep oxide traps [17, 18].

1.4 Outline of the thesis

After this introductory chapter, the first topic that is discussed in this thesis is how to perform accurate RF measurements for the purpose of the experiments described later in the thesis. This will be done in chapter 2. Next, in chap-

ter 3, wafer-level MOSFET degradation under RF stress is investigated. Here RF hot carrier, RF NBTI and RF gate-oxide breakdown effects are experimentally investigated and the results are compared to existing models on DC and AC degradation. In chapter 4 a new simulation methodology is presented that allows for the lifetime prediction of RF CMOS PA's. Then, in chapter 5 the use of RF measurement techniques for the reliability characterization of leaky MOS devices will be discussed. The thesis ends with conclusions in chapter 6.

Chapter 2

RF MOS measurements

2.1 Introduction

Present day state-of-the-art DC measurement equipment is capable of measuring voltage and current signals in the nV respectively fA range [19]. Provided that proper cabling and connections are used, it is relatively easy for a user to obtain these high levels of accuracy. Although highly sophisticated equipment is needed, the user does not need a deep understanding of the underlying measurement principles used, the measurement equipment takes care of this. For RF measurements the situation is a bit more complicated: some aspects of RF measurements require that the user understands the basic measurement principles in order to obtain accurate results. Especially for on-wafer characterization this can be very intricate.

In this chapter some important aspects of RF measurements will be explained, focusing on those measurements relevant for the work described later in this thesis. First in section 2.2 a short overview will be given on how RF measurements differ from DC or AC measurements. Then in section 2.3 it will explained how accurate small-signal characterization of two-port networks can be performed. The work described in section 5.2 of this thesis makes use of these measurement techniques. Finally, in section 2.4 a method is described for generating sinusoidal RF voltage signals with a well-defined amplitude for experiments performed in a one-port on-wafer measurement setup. This method is used for the experiments discussed in chapter 3 and section 5.3 of this thesis.

2.2 RF vs. AC/DC measurements

Typical measurement issues that are not encountered in DC and low-frequency measurements, but which are important for RF measurements are the following:

1. Cables and connections become very lossy when used to transmit high frequency signals.

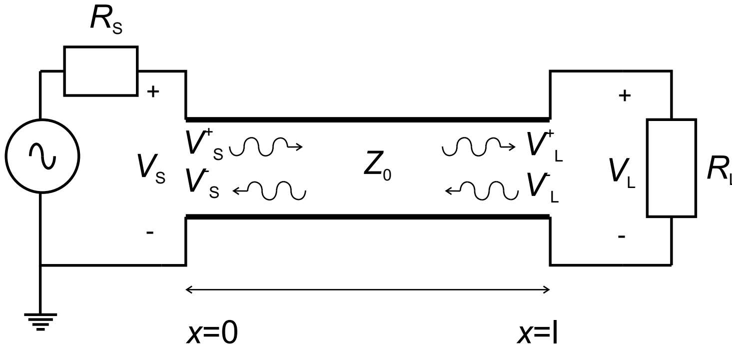

Figure 2.1: Illustration of a transmission line with voltage waves travelling in two directions. R_{S} is the source series resistance and R_{L} is the load resistance. In this example $R_{\mathrm{L}} \neq Z_0$, causing partial reflection of the voltage wave V_{L}^{+}.

2. Parasitic capacitances and inductances present in the measurement setup can no longer be considered as ideal opens respectively shorts for the high frequencies used.

3. The wavelengths of the voltage and current waves of interest are in the same order of magnitude as, or smaller than, the length of the measurement cables.

The first issue can be minimized by making use of dedicated cabling and connections, such as semi-rigid cables and SMA connectors. Furthermore the use of ground-signal-ground probe pad configurations can greatly reduce signal losses for on-wafer RF measurements. Loss free connections are however not available.

The second issue is the cause for large errors in the measured impedance levels of devices if they are not accurately accounted for. Both lossy connections as well as parasitic capacitances and inductances require proper modeling of all parasitic elements in the measurement setup for making accurate measurements. This can be realized by making use of calibration and de-embedding techniques as will be discussed in sections 2.3.2, 2.3.3 and 2.4.

The third issue can best be understood by looking at figure 2.1. In this figure a voltage source is shown with an internal series resistance R_{s}. It is connected to a transmission line with length l which is terminated with a load resistance R_{L}. Z_0 represents the characteristic impedance of the transmission line. Throughout this thesis this is always 50 Ω, the characteristic impedance of the coaxial cables used in the measurements. Voltage signals V_{S} and V_{L} are the voltage signals at the source respectively load side. Within the transmission line 4 different travelling voltage waves are indicated. V_{S}^{+} and V_{S}^{-} represent the voltage waves at the source side, travelling in the positive x-direction respectively negative x-direction. V_{L}^{+} and V_{L}^{-} are the voltage waves at the load side, travelling in the positive x-direction respectively negative x-direction. The occurrence of voltage waves travelling in

both directions stems from an impedance mismatch between the characteristic cable impedance Z_0 and the load resistance R_L.

The voltage source causes the voltage wave V_S^+, to propagate along the transmission line in the positive x-direction. This results in V_L^+ flowing at the load side of the transmission line. These two voltage waves are related through [20]:

$$V_\mathrm{L}^+ = V_\mathrm{S}^+ e^{-\gamma l} \tag{2.1}$$

In this expression γ is the complex propagation constant of the transmission line. It comprises of a real part, describing power loss of the travelling signal and an imaginary part describing the phase shift of the signal. In the given example in figure 2.1 $R_\mathrm{L} \neq Z_0$; this is called an impedance mismatch. As a result of this mismatch, part of the travelling voltage wave V_L^+ is reflected, causing the voltage wave travelling in the negative x-direction, V_L^-. This wave also propagates along the transmission line resulting in V_S^-. V_S^- and V_L^- are related through [20]:

$$V_\mathrm{S}^- = V_\mathrm{L}^- e^{-\gamma l} \tag{2.2}$$

If $R_\mathrm{S} \neq Z_0$, again this voltage wave will be partly reflected; V_S^+ is composed of this reflected component and the component directly generated by the source. The two voltage signals V_S and V_L are given by [20]:

$$\begin{aligned} V_\mathrm{S} &= V_\mathrm{S}^+ + V_\mathrm{S}^- \\ V_\mathrm{L} &= V_\mathrm{L}^+ + V_\mathrm{L}^- \end{aligned} \tag{2.3}$$

Clearly for a nonzero value of γ these two voltage signals are not equal. Due to the impedance mismatch and the relatively short wavelengths both standing waves and travelling waves occur on the transmission line; this is the cause for this difference. Even if signal loss in the transmission line is negligible (i.e. a zero-valued real part of γ), the phase shift between the two ends of the transmission line can not be neglected.

This effect gets worse with increasing frequencies as the phase shift is proportional to the frequency of the signal. As a result it is not possible to determine V_L by performing a measurement at the source side of the transmission line. This is what signifies the difficulties in performing RF measurements with respect to DC and AC measurements. In RF measurements the effect of reflections due to impedance mismatch should be carefully taken into account.

An important remark should be made here: the term RF stands for radio frequencies. In this sense it can refer to the frequency of any electromagnetic wave that is used to transmit radio signals. In this thesis use is made of a different definition of RF: it is any frequency high enough for the issues described in this section to appear using measurement cables (transmission lines) with lengths in the order of 1 m. It is common practice to use such a kind of definition; RF measurements are typically considered as being different from AC measurements. The issues described in this section are the problems that make this distinction necessary.

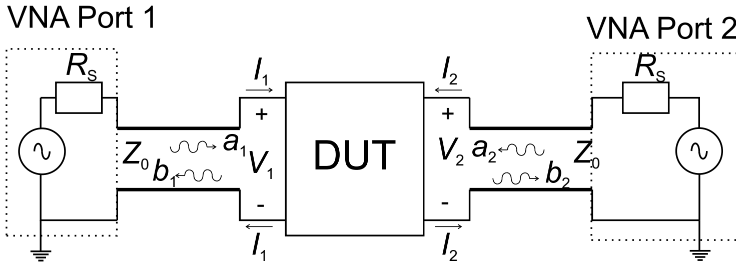

Figure 2.2: Two-port s-parameter measurement setup.

As a rule of thumb the lower frequency for which the term RF is used is chosen to be 10 MHz in this thesis. This corresponds to a wavelength of 20 m assuming a propagation speed of $2 \cdot 10^8$ m/s. This is a typical propagation velocity, it is the propagation velocity of TEM waves in coaxial cables that use poly-ethylene as a dielectric material. Frequencies used in this thesis do not exceed the few GHz range, which is still RF in any commonly used definition.

2.3 Small signal two-port characterization

The RF measurement issues discussed in the previous section should be taken care of in order to make accurate measurements of devices operating at radio frequencies. In this section small-signal measurements will be discussed. In small signal characterization the measurement setup is assumed to be a linear system. Although typically the device that are characterized are nonlinear by nature, this assumption is allowed if very small amplitudes of the voltage signals are used. In section 2.4.2 a method will be described to determine whether this assumption is allowed or not for a given device and voltage amplitude. Typical voltage amplitudes used in the small signal characterization of MOS devices lie in the order of 100 mV (~-15dBm for $Z_0 = 50\ \Omega$) or lower.

2.3.1 s-parameters

Due to the occurrence of both travelling and standing waves, it is not straightforward to characterize a DUT in terms of voltages and currents at RF conditions. Furthermore if one wants to characterize a MOS device, it should be well realized that the DUT has more than two terminals. For this purpose MOS devices are typically characterized in a two-port s parameter measurement setup as illustrated in figure 2.2. Such type of measurements can be performed using a Vector Network Analyzer (VNA). The DUT has two ports, which are connected to the signal ports of the VNA. The VNA is capable of generating well defined travelling waves as well as measure them. The measurement cables and contacts are represented as transmission lines. In the transmission lines travelling waves a_1, b_1, a_2

and b_2 are indicated. These quantities are typically used when performing VNA measurements. They are defined by [20]:

$$a_{\mathrm{i}} = \frac{V_{\mathrm{i}}^{+}}{\sqrt{Z_0}}$$
$$b_{\mathrm{i}} = \frac{V_{\mathrm{i}}^{-}}{\sqrt{Z_0}} \tag{2.4}$$

Using the measured wave quantities, the s-parameters of the DUT can be found. A VNA is capable of obtaining the s-parameters of the DUT. The s-parameters are defined by [21]:

$$s_{11} = \frac{b_1}{a_1}|_{a_2=0}$$
$$s_{12} = \frac{b_1}{a_2}|_{a_1=0}$$
$$s_{21} = \frac{b_2}{a_1}|_{a_2=0}$$
$$s_{22} = \frac{b_2}{a_2}|_{a_1=0} \tag{2.5}$$

These four s-parameters completely describe the two-port behavior of the DUT. They can easily be transformed into other types of two-port parameters, (e.g. z- or y-parameters), describing the DUT in terms of voltages and currents. So, in short, by measuring the wave quantities a_1, a_2, b_1 and b_2 and deriving the s-parameters as in 2.5, VNA measurements provide the complete small-signal two-port behavior of a DUT in terms of voltages and currents. When performed over a broad frequency range, these measurements can be used to determine important device parameters such as the cut-off frequency f_{T} and the maximum frequency of oscillation f_{max}. They can also be used for obtaining C-V curves as will be explained in section 5.2.

2.3.2 Calibration

A VNA is capable of measuring s-parameters; it cannot, however, measure the power waves directly at the DUT, but only at the VNA side of the transmission lines. In order to make s-parameter measurement at the DUT possible, the measurement setup needs to be calibrated first. The calibration process takes into account any error due to nonzero propagation constants of the transmission lines as well as errors introduced by the VNA itself. Several calibration procedures have been developed, these will not all be discussed here; a good explanation on different calibration procedures can be found in [22]. Any commercially available VNA provides the necessary tools for performing calibration for s-parameter measurements.

The key idea of calibration is to characterize the errors in the measurement setup by replacing the DUT in figure 2.2 with well characterized calibration stan-

dards. A commonly used calibration procedure for two-port networks is SHORT-OPEN-LOAD-THROUGH (SOLT) calibration. This is the calibration procedure used in this thesis when two-port s-parameter measurements are performed. In SOLT calibration the calibration standards consisting of a SHORT, OPEN and LOAD standard are consecutively connected at both ports of the DUT followed by a THROUGH standard that links both ports of the DUT. The LOAD standard is typically a 50 Ω resistive structure, thereby providing perfect matching to Z_0. For all these connections the a and b waves are measured at the VNA side of the transmission lines in figure 2.2. In this way the error terms present in the measurement setup can be found. This type of calibration can be performed on-wafer, thereby providing accurate s-parameter measurements at the tip of the probe needles.

The one-port equivalent of SOLT calibration is SOL calibration. It only differs form SOLT calibration in the sense that no measurements are made at port 2, hence the THROUGH connection cannot be used. SOL calibration consists of three measurements, a SHORT, OPEN, and a LOAD measurement. For this purpose the measurement setup is typically described using a three-term error model [22]. This model is a simplification of a four-term error model [22], but it is adequate for calibration a measurement for obtaining the reflection coefficient at the DUT, Γ_{DUT}. As with SOLT, SOL calibration provides no information on the absolute values of the power and voltage waves flowing in the measurement setup, only the ratio between the incoming and reflected waves at the DUT is known. In section 2.4 it will be explained how the use of a 4-term error model provides the possibility of also obtaining these absolute levels.

2.3.3 De-embedding

Even though the measurement setup can be calibrated to the tip of the probe needles for on-wafer measurements, some errors still remain unaccounted for. These can be attributed to the bond pads and interconnect lines, that connect the probe needles to the intrinsic DUT. The bond pads can cause a parasitic capacitance parallel to the DUT. Typical values of this bond pad capacitance are in the range of 100 fF. This results in a parasitic admittance of $\sim j{\cdot}0.6$ mS for a measurement at 1 GHz. The interconnect lines cause a parasitic series line inductance, typical values of this inductance are in the range of 100 pH. This coincides with a series parasitic impedance of $\sim j\cdot\, 0.6\ \Omega$ for a 1 GHz signal.

These parasitic impedances and admittances can seriously affect the accuracy of the obtained measurement results, even after calibration to the tip of the probe needle. Especially for frequencies exceeding 1 GHz this becomes more and more important. These test structure parasitics must therefore be accounted for as well. The technique for correcting for these parasitics is called de-embedding. For frequencies up to ∼20 GHz, the most commonly used de-embedding method is OPEN-SHORT de-embedding.

OPEN-SHORT de-embedding can be understood by modeling the test structure parasitics as done in figure 2.3 [23]. The parasitic admittances originating

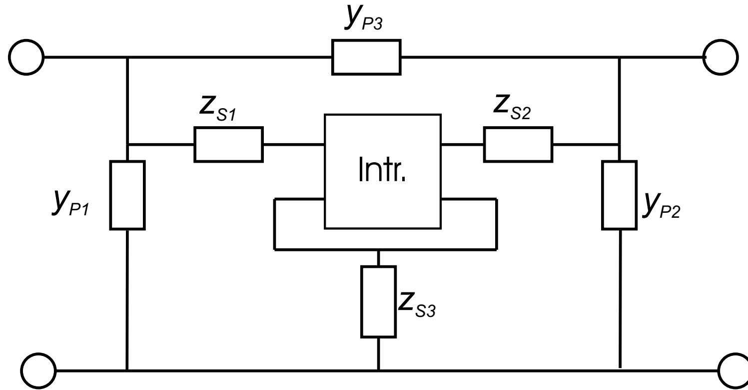

Figure 2.3: Two-port model of a test structure with parasitics as it is used in OPEN-SHORT de-embedding.

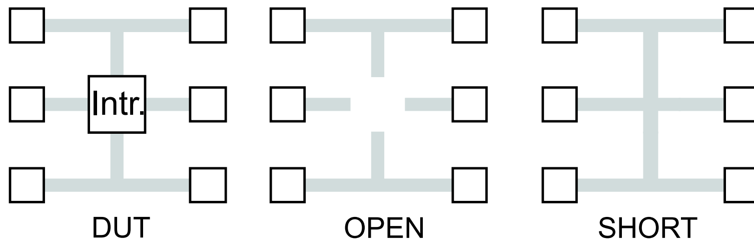

Figure 2.4: Layout of test structures for OPEN-SHORT de-embedding in a ground-signal-ground configuration. From left to right: the actual DUT, the OPEN de-embedding structure and the SHORT de-embedding structure.

from the bond pad capacitances are represented by y_{P1}, y_{P2} and y_{P3}. The parasitic impedances originating from the line impedances are represented by z_{S1}, z_{S2} and z_{S3}. The goal of de-embedding is to account for these parasitics so that the two-port behavior of the intrinsic DUT can be found.

In OPEN-SHORT de-embedding the two-port parameters of the three structures of figure 2.4 are measured. The OPEN and SHORT de-embedding structures are different from the OPEN and SHORT calibration structures used in SOLT calibration. The de-embedding structures are two-port structures whereas the OPEN and SHORT calibration structures are one-port structures. Furthermore de-embedding structures are not ideal structures as opposed to calibration structures. The de-embedding structures have the same parasitics as the DUT structure. De-embedding measurements are performed on-wafer, while calibration measurements are performed on a separate calibration substrate.

If the two-port parameters of the three structures of figure 2.4 are determined

it can be derived that [23]:

$$\overline{y_{\mathrm{I}}} = \left((\overline{y_{\mathrm{D}}} - \overline{y_{\mathrm{O}}})^{-1} - (\overline{y_{\mathrm{S}}} - \overline{y_{\mathrm{O}}})^{-1} \right)^{-1} \tag{2.6}$$

In this expression $\overline{y_{\mathrm{I}}}$, $\overline{y_{\mathrm{D}}}$, $\overline{y_{\mathrm{O}}}$ and $\overline{y_{\mathrm{S}}}$ represent two-by-two y-parameter of the intrinsic device, the DUT, the OPEN structure respectively SHORT de-embedding structure. Using this expression the two-port behavior of the intrinsic device can be found. Expression 2.6 only provides accurate results as long as $\overline{y_{\mathrm{I}}}$ is significant compared to $\overline{y_{\mathrm{O}}}$ and $\overline{y_{\mathrm{S}}}$, due to limitations in the resolution of the measurement setup. This can be assured by careful design of the test structure.

De-embedding is not always automated in VNA equipment. Typically VNA measurements are performed on the DUT followed by the two de-embedding structures. Working out equation 2.6 is part of the data analysis performed by the user. In the work described in this thesis use is made of a MATLAB routine for this purpose.

2.4 Linear one-port RF voltage generation

Whereas measurements in terms of small signal s-parameters are adequate for characterizing a vast amount of device parameters, this may not always be applicable. In chapter 3 and section 5.3 experiments will be discussed that require well-defined voltage signals with frequencies up to 4 GHz. Assessing voltage waveforms at the DUT is not straightforward, especially for on-wafer measurements. Nowadays commercially available equipment exists that allows for obtaining the time-domain voltage and current waveforms on nonlinear devices [24]. The experiments discussed in this thesis, however, do not require all the complexity involved in such equipment: only sinusoidal voltage signals are considered, where its phase is irrelevant in its analysis. This makes large-signal measurement equipment overly complex- and expensive- for the purpose of the experiments described in this thesis. In this section the methodology will be described that is used throughout this thesis for generating sinusoidal voltage signals with a well-defined amplitude in a one-port on-wafer experiment.

As only purely sinusoidal voltage signals are used, a linear VNA is suitable for generating the required signal for the experiments of chapter 3 and section 5.3. Since a VNA is not designed for performing these experiments, they typically do not provide information on the voltage signals during measurements. Based on well-accepted models it is, however, possible, to obtain this information without the need for any additional complexity to the measurement setup.

2.4.1 Setting the amplitude

For generating sinusoidal voltage signals, use is made from the measurement setup shown in figure 2.5. In this figure a VNA is represented as consisting of a power source, two directional couplers, local oscillator (LO) circuits and A/D and IF

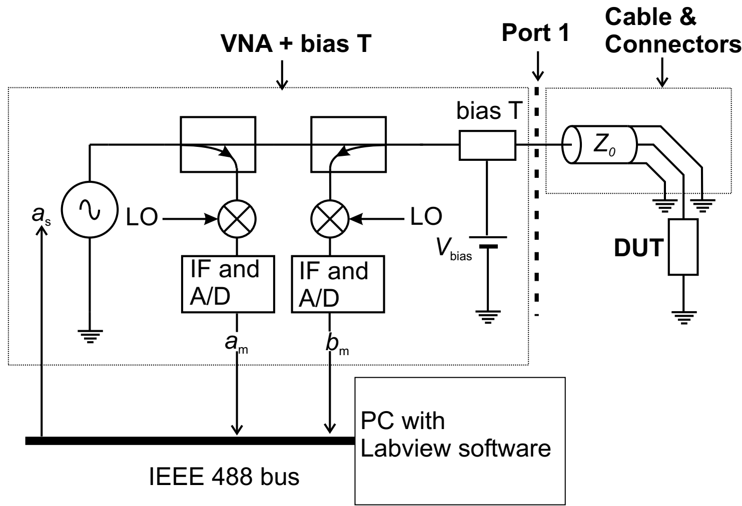

Figure 2.5: Schematic drawing of the setup used for the RF voltage generation procedure used in this thesis. A VNA is used to generate large amplitude sinusoidal voltage signals at the DUT. The DUT is laid out in a ground-signal-ground configuration and is connected on-wafer. The VNA is controlled via Labview software on a PC that is connected through an IEEE 488 bus.

processing circuitry. It is a very simple description of a VNA, but it is suitable for understanding the voltage generation procedure discussed in this section. The VNA is connected to a PC with Labview software, through an IEEE 488 bus. An externally generated DC voltage V_{bias} is superimposed on the RF signal by means of the bias T.

The VNA in this setup is connected in a one-port arrangement, the VNA operates in continuous wave mode. The power source generates a_{s}; this wave will flow towards the DUT and it will be partly reflected, causing a wave flowing in the opposite direction. With the use of the two directional couplers it is possible to measure both the a and b waves separately. The coupled signal is multiplied with an LO signal and after IF processing and A/D conversion the complex measured waves a_{m} and b_{m} are known. These variables differ from the actual waves at the DUT; calibration is needed to find the amplitude of the voltage signal at the DUT by recording a_{m} and b_{m}.

The amplitude and phase of a_{m} and b_{m} are read by the PC, and from these the peak-to-peak value of the voltage signal applied at the DUT, $V_{\mathrm{DUT,pp}}$ is determined in a Labview routine. Setting the desired value of $V_{\mathrm{DUT,pp}}$ is realized by increasing a_{s} in small steps and monitoring $V_{\mathrm{DUT,pp}}$ until its desired value is reached.

In order to determine $V_{\mathrm{DUT,pp}}$ from the obtained complex parameters a_{m} and b_{m} the measurement setup needs to be calibrated. As this setup is custom built,

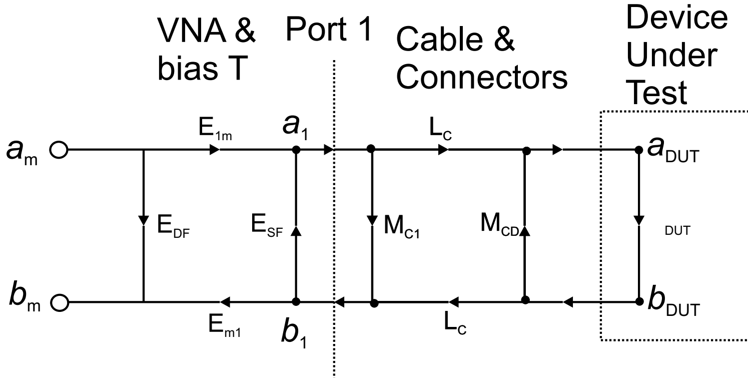

Figure 2.6: Signal flow graph depicting the 7 error terms of the setup of figure 2.5.

only for the purpose of the measurements discussed in this thesis, the necessary equations for the calibrations must be derived; calibration software inside the VNA used for the experiments discussed in this thesis does not provide the appropriate control over $V_{\mathrm{DUT,pp}}$.

Similar to the calibration procedures discussed in section 2.3.2 a VNA can be modeled in a Signal Flow Graph for the purpose of characterizing the error terms. The SFG used is shown in figure 2.6. This SFG has 7 error terms. It is a combination of a four-term error model and a three-term error model. These error models are commonly used models in one-port calibration of VNA's [22]. The four-term error model describes the behavior of the VNA in combination with the used bias Tees and some cabling. The three-term error model consists only of cabling, connectors and the probes for on-wafer characterization. This cable & connector network is a reciprocal network, reflected by the two equal L_C terms shown in figure 2.6.

The reason for splitting the measurement setup into a "VNA & bias Tee" and a "Cable & Connectors" network stems from the fact that an absolute power measurement is needed in order to calculate $V_{\mathrm{DUT,pp}}$ from a_m and b_m, as will be made clear below. This power meter measurement cannot be performed on-wafer, therefore this is done at Port 1. Making use of the reciprocity of the "Cable & Connector" network this allows for determining the absolute value of the waves at the DUT, in an on-wafer experiment.

The error terms are determined by making use of calibration standards. As the SFG of figure 2.6 has 7 error terms, 7 well-chosen calibration measurements are needed for characterizing the complete measurement setup. The error terms can be dependent on both frequency and power, therefore these calibration measurements are performed over a broad range of frequencies and power levels. The result of the calibration procedure is a large database containing information on the 7 error terms for all frequencies and power levels used in the experiments,

stored on the PC. Using this database, $V_{\mathrm{DUT,pp}}$ can be calculated from any a_{m}, b_{m} combination at any given frequency or power level. This calculation is done within a Labview routine.

The calibration measurements start with a SHORT, OPEN and LOAD measurement at Port 1. From these three measurements error terms E_{DF}, E_{SF} and $E_{\mathrm{1m}} \cdot E_{\mathrm{m1}}$ can be found. The next step is to measure the absolute power level at Port 1 using an external power meter for all frequencies and power levels of interest. In the work described in this thesis an HP 437B power meter is used for this purpose; it is also connected to the PC, through an IEEE 488 bus. Using this power measurement a value for $|E_{\mathrm{1m}}|$ can be found. Next the "Cable & Connector" network is connected to Port 1. The calibration procedure completes with a SHORT, OPEN and a LOAD measurement using standards on a calibration substrate for on-wafer calibration. These three measurements provide values for M_{C1}, M_{CD} and L_{C}.

The equations underlying this calibration procedure follow straightforwardly from the SFG of figure 2.6. They are worked out in appendix A, together with the equations needed to relate $V_{\mathrm{DUT,pp}}$ to a_{m} and b_{m}.

2.4.2 Verifying linearity

The calibration technique discussed above is only suitable if the voltage signal at the DUT is purely sinusoidal; this can only be realized using a DUT with linear input impedance. The input impedance of the devices discussed in chapter 3 and section 5.3 are, however, not perfectly linear. This can be overcome by choosing a voltage amplitude sufficiently low, so that the device is operating in small-signal regime. Choosing an amplitude too low, on the other hand, counteracts the purpose of the experiments of chapter 3 and section 5.3. In these experiments, the amplitude is necessarily chosen as large as possible. Setting up an experiment that fulfills both the demand of linearity and of having an amplitude sufficiently large for making the experiments relevant, is a crucial step in the methodology used in this thesis for generating RF voltage signals.

Various methods exist that allow for checking for nonlinear behavior of devices [25]. These methods can be used to determine the maximum voltage amplitude of an RF signal that can be superimposed on a DC voltage V_{bias} for which linearity can be guaranteed. For the experiments in chapter 3 however, it is needed to know the maximum voltage amplitude that can be generated for which linearity can be guaranteed, with a fixed value of the maximum voltage level, V_{max}. The existing methods could be used for this purpose, by measuring the maximum tolerable power level as a function of V_{bias}. Using the calibration procedure discussed in the previous section, the maximum tolerable voltage amplitudes could then be calculated from the obtained power levels. This would result in a large database from which the information of interest (i.e. the largest tolerable voltage amplitude with a given V_{max}) could be extracted. For the work described in this thesis a different approach was adopted in which considerably less measurements are required for obtaining the information of interest.

The key idea of the applied method is to calculate the voltage signal at the DUT on the basis of a measured small-signal input impedance of the DUT and check whether the resulting waveform is indeed sinusoidal. For this purpose the input impedance of the DUT is measured as a function of V_{bias} using a VNA connected in one-port setup with the power level set to -15 dBm. The voltage waveform is determined by solving transmission line equations within a MATLAB routine. The small-signal characterization is performed on one device, the actual experiments as discussed in chapter 3 and section 5.3 are performed on the same type of device, but on different dies on the same wafer.

In order to perform the necessary calculations, first the appropriate equations will be derived. When an RF power signal is applied to a DUT, a part of the signal can be reflected, due to an impedance mismatch between the characteristic impedance of the cable and the input impedance of the DUT. The resulting voltage waveform at DUT level is a direct function of this amount of mismatch. The reflection coefficient, and hence the resulting voltage waveform at the DUT, can be determined by solving transmission line equations. These equations are based on the fact that the current flowing in the measurement cable must equal the current flowing into the DUT. When dealing with a linear input impedance of the DUT it is given by [20]:

$$\frac{V^{+} - V^{-}}{Z_0} = \frac{V_{\text{DUT}}}{Z_{\text{DUT}}} \tag{2.7}$$

In this expression V^{+} and V^{-} represent the incoming respectively reflected voltage wave, as defined in [20] and V_{DUT} is the resulting voltage signal at the DUT. All three voltage signals in this expression are represented in a complex phasor notation. In this expression perfect linear behavior of the DUT is assumed. This equation can be solved by realizing that [20]:

$$V_{\text{DUT}} = V^{+} + V^{-} \tag{2.8}$$

The idea is now that equation 2.7 is solved using the actual measured input impedance of the DUT and verify whether the outcome indeed represents a perfectly sinusoidal signal (i.e. higher harmonics are negligible). If this is true, the assumption of linear behavior is allowed, thereby proving the validity of equation 2.7 and that of the resulting calculated waveform.

Expressions are needed for the following parameters: V^{+}, Z_0 and Z_{DUT}. V^{-} and V_{DUT} are solutions to equation 2.7. V^{+} is the incoming voltage wave, its amplitude should be chosen such that it leads to the desired value of V_{DUT}. In practice this means that the solution to equation 2.7 will be obtained with various values of V^{+}; only the value of V^{+} that leads to the desired value of V_{DUT} is considered.

More difficult is choosing the appropriate value of Z_{DUT}, it can be a complex variable, that is dependent on V_{DUT}. The latter is typical for the devices used in this thesis, consider, e.g., the voltage dependent gate capacitance of a MOSFET structure. Due to the fact that Z_{DUT} can be voltage dependent, equation 2.7 cannot be solved in the frequency domain, thereby prohibiting the use of complex

phasor notations. Another way of describing this equation is needed. The equation is rewritten in time-domain, with the imaginary component of Z_{DUT} described in a differential equation.

For the devices used in this thesis it will turn out that the imaginary part of Z_{DUT} is either negligible or negative, under the measurement conditions of interest. This corresponds to a capacitive input. When assuming sinusoidal steady-state and a negative imaginary component of Z_{DUT}, it can be derived that equation 2.7 can be written in time-domain as:

$$\begin{aligned}\frac{V^+(t) - V^-(t)}{Z_0} &= \\ & V_{\text{DUT}}(t)\left[G_{\text{DUT}}(V_{\text{DUT}}(t)) + \frac{dC_{\text{DUT}}(V_{\text{DUT}}(t))}{dt}\right] + \\ & \frac{dV_{\text{DUT}}(t)}{dt} C_{\text{DUT}}(V_{\text{DUT}}(t))\end{aligned} \tag{2.9}$$

In this expression all voltage signals are expressed in time domain. $G_{\text{DUT}}(V_{\text{DUT}}(t))$ and $C_{\text{DUT}}(V_{\text{DUT}}(t))$ are the input admittance respectively input capacitance of the DUT at time t, if the input voltage is equal to $V_{\text{DUT}}(t)$. The relation between parameters G_{DUT}, C_{DUT} and V_{DUT} can be obtained from the measured Γ_{i}, obtained in a one-port setup and after SOL calibration of the VNA. This should be done over the entire voltage and frequency range of interest. Having a value of Γ_{DUT} for the appropriate value of V_{DUT} and the correct frequency, parameter z_{in} can be found from:

$$z_{\text{in}}(V_{\text{DUT}}) = Z_0 \frac{1 + \Gamma_{\text{DUT}}(V_{\text{DUT}})}{1 - \Gamma_{\text{DUT}}(V_{\text{DUT}})} \tag{2.10}$$

Now if $V_{\text{DUT}}(t)$ is known, $z_{\text{in}}(V_{\text{DUT}}(t))$ can be found from the measured data and $G_{\text{DUT}}(V_{\text{DUT}}(t))$ and $C_{\text{DUT}}(V_{\text{DUT}}t))$ can be derived using:

$$\begin{aligned} G_{\text{DUT}}(V_{\text{DUT}}(t)) &= \Re\left(\frac{1}{z_{\text{in}}(V_{\text{DUT}}(t))}\right) \\ C_{\text{DUT}}(V_{\text{DUT}}(t)) &= \frac{1}{2\pi f}\Im\left(\frac{1}{z_{\text{in}}(V_{\text{DUT}}(t))}\right) \end{aligned} \tag{2.11}$$

The solution to equation 2.9 can now be found by making use of the measured values of G_{DUT} and C_{DUT} as a function of V_{DUT}.

No general analytical solution can be found for equation 2.9 and therefore a numerical solution technique is used. This means that equation 2.9 is solved in discrete time domain. For this purpose voltage signals $V^+(t)$, $V^-(t)$ and $V_{\text{DUT}}(t)$ are discretized in 200 time steps per period. Now for every time step an optimization algorithm embedded in a MATLAB routine finds the solution to equation 2.9. The time derivatives in equation 2.9 are replaced by their discrete-time equivalent for this purpose. In this way the time domain representation of $V_{\text{DUT}}(t)$ can be found. The peak-peak voltage of the sinusoidal voltage signal $V^+(t)$, V^+_{pp}, can be tuned to result in the desired amplitude of $V_{\text{DUT}}(t)$ by repeatedly performing this solution procedure for different values of V^+_{pp}.

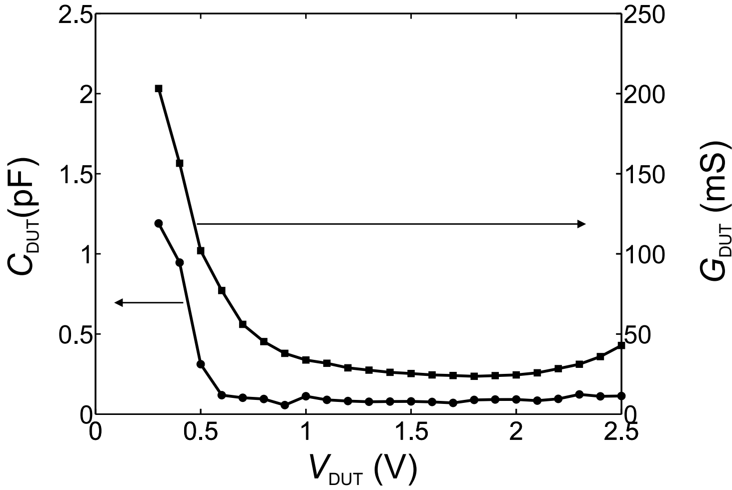

Figure 2.7: Example of the DUT impedance measured as a function of V_{DUT}. Small-signal s-parameters were performed in a one-port measurement setup and G_{DUT} and C_{DUT} were extracted using the equations in 2.11 at a frequency of 3.2 GHz. The DUT is an nMOSFET connected at its drain terminal with the gate voltage set to 1.5 V. For voltages above $\sim$ 1 V the nMOSFET operates in the saturation regime. This DUT is useful for performing RF hot carrier experiments as will be explained in section 3.2.

The resulting waveforms indicate whether sinusoidal voltage signals at the input of the DUT can be guaranteed, for the desired amplitude of $V_{\mathrm{DUT}}(t)$ and the measured values of $G_{\mathrm{DUT}}(V_{\mathrm{DUT}})$ and $C_{\mathrm{DUT}}(V_{\mathrm{DUT}})$. In figure 2.7 and 2.8 the importance of this analysis is illustrated. Figure 2.7 shows an example of the DUT impedance measured as a function of V_{DUT}. The DUT is an actually used DUT in this thesis; it is used for RF hot carrier experiments. The DUT consists of an nMOSFET, connected at the drain side, with the gate voltage set to 1.5 V. This means that the observed G_{DUT} and C_{DUT} are the conductance and capacitance are the impedance seen at the drain, as a function of drain voltage. For voltage levels above $\sim$ 1 V, the nMOSFET operates in the saturation regime; at lower gate voltages the DUT impedance has a stronger dependency on the applied voltage V_{DUT}.

The signal integrity analysis is applied to this DUT at a frequency of 3.2 GHz, the frequency for which this DUT is used in the RF hot carrier experiment. The DC offset voltage was set to 1.25 V and the waveform was calculated for an incoming voltage wave with an amplitude of 1.77 V (16.5 dBm). The resulting waveform is shown in figure 2.8. Clearly the resulting waveform suffers from harmonic distortion. It is not a sinus and therefore the amplitude should be reduced in order to use the DUT in an RF hot carrier experiment. In section 3.2

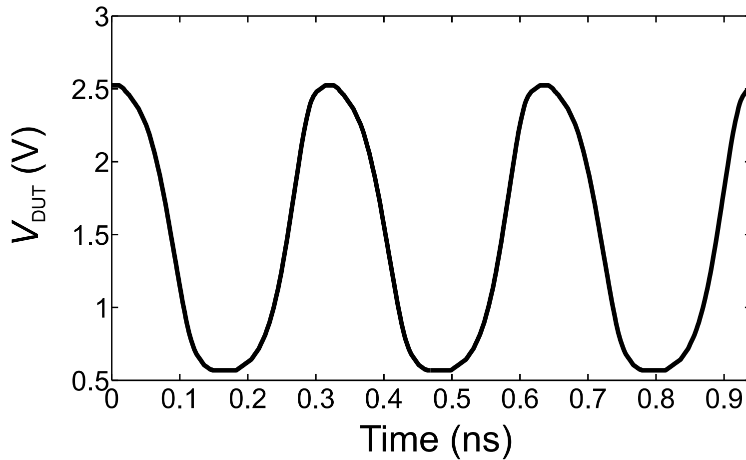

Figure 2.8: Calculated waveform at 3.2 GHz using the measured input impedance of figure 2.7. In this example the amplitude of the DUT voltage is chosen too high for obtaining a sinusoidal voltage signal at the DUT. The amplitude should be reduced in order to guarantee linearity. In figure 3.3 b) the waveform as used in RF hot carrier experiments on this device is shown.

it is explained which voltage level is chosen for this DUT, the result can be seen in figure 3.3 b).

The signal integrity analysis discussed here was performed prior to any measurement discussed in this thesis. If the resulting waveform indicated that nonlinearities were not negligible, different measurement conditions were chosen (e.g. lower amplitudes of $V_{\mathrm{DUT}}(t)$), such that this could be guaranteed. The effect of the introduction of higher harmonics due to a nonlinear input impedance is one of the causes that limits the use of higher frequencies than the ones discussed in chapter 3 and section 5.3, i.e. up to 4 GHz. It should be noted however that the measurements discussed in these chapters are already performed at frequencies far beyond conventional measurements and suffice to answer the research questions on RF reliability in CMOS for contemporary wireless applications.

2.5 Conclusions

RF measurements differ from AC and DC measurements in the fact that voltage and current signals occurring at the DUT may significantly differ from those at the source. In this chapter an overview is given of the RF measurement techniques used for the experiments discussed later in this thesis. These techniques include widely used small-signal characterization tools as well as a methodology for generating well-defined large-amplitude voltage signals for on-wafer experiments.

Small-signal s-parameter characterization is a well-established technique. Ac-

companied with calibration and de-embedding it provides a very accurate measurement of a two-port network connected on-wafer. Various calibration and de-embedding techniques exist, but for the purpose of the experiments discussed in this thesis it suffices to make use of SOLT calibration and OPEN-SHORT de-embedding.

In this chapter the voltage generation procedure as used for the RF stress and RF charge pumping measurements, discussed later in this thesis, is presented. The key idea of the technique is to generate sinusoidal voltage signals with a voltage amplitude as large as possible, but low enough for nonlinearities to be negligible. Sinusoidal voltage signals are useful for both RF stress and RF charge pumping measurements. A consequence of using only sinusoidal signals, is that it suffices to use only a VNA in the measurement setup, thereby allowing to omit the use of complex large-signal measurement equipment for determining the exact voltage waveform at DUT level. This methodology is very effective for performing the experiments discussed later in this thesis.

Chapter 3

MOSFET degradation under RF stress

3.1 Introduction

A key aspect of reliability engineering is getting a good understanding of physical mechanisms underlying device failure. For MOSFET degradation mechanisms this issue has been widely discussed for the DC case and various models have been developed for explaining degradation rates of the devices under different biasing conditions. These models can be used for making lifetime predictions of different circuits. For use in RF circuits it is of crucial importance to understand how applicable these DC models are under RF conditions. Some authors have compared DC stress conditions to AC conditions (see e.g. [26]), but stress under RF conditions is only marginally addressed in literature. One of the main reasons for this is that it is not straightforward to perform accurate reliability experiments at frequencies exceeding 10 MHz. Some authors did measure device degradation when operated in RF circuits [27, 28], but an accurate comparison with low frequency results could not be made. To do this, experimental results are needed that reveal whether or not degradation mechanisms are frequency dependent between the MHz and the GHz range.

Given the fact that all models describing the degradation mechanisms discussed in section 1.3, are all either voltage or field based, the generation of well-defined RF voltage signals is a critical issue. In this chapter the frequency response of the degradation mechanisms discussed in section 1.3 will be investigated using the voltage generation procedure as explained in section 2.4. This will be preceded, for every degradation mechanism, by a short overview on what is known from literature concerning the applicability of DC models for AC stress conditions. The combination of these DC to AC and AC to RF comparisons can be used to shed insight on the applicability of DC models for RF stress conditions. In section 3.2 this will be done for the hot carrier effect. This is followed by the

NBTI effect in section 3.3 and gate-oxide breakdown in section 3.4.

3.2 RF hot carrier degradation

3.2.1 DC model

As explained in section 1.3, the hot carrier effect is the degradation effect caused by high energetic charge carriers flowing in the channel of a MOSFET. This effect has been widely discussed in literature (see e.g. [29, 30, 31, 32, 33]). The effect can be observed in both nMOSFETs and pMOSFETs; hot carrier degradation in nMOSFETs however has received much more attention in literature than pMOSFET degradation. This is due to the fact that the effect of nMOSFET hot carrier degradation on digital circuit performance is more severe than pMOSFET hot carrier degradation. Similarly, when considering hot carrier degradation in RF circuits, investigating nMOSFET degradation is much more relevant than pMOSFET degradation, as will also be explained in subsection 3.2.2. Therefore in this section only hot carrier degradation in nMOSFETs will be considered.

From literature it is known that the hot carrier degradation rate is dependent on both the drain and the gate voltage. An increase in drain voltage leads to a higher lateral electric field in the channel, thereby generating more hot carriers. With a gate voltage biased at maximum degradation rate conditions, the hot carrier lifetime is usually related to the drain voltage as in [29]:

$$\tau_{\mathrm{HC}} = A \cdot e^{\frac{B}{V_{\mathrm{D}}}} \tag{3.1}$$

In this expression τ_{HC} is the hot carrier lifetime of the device. It can be defined as the time at which any device parameter has degraded by a given quantity (such as a 50 mV shift in the threshold voltage V_{T}) or fraction (typically 10%). Parameters A and B are dependent on the device parameter it concerns, the technology used and the stress conditions. V_{D} is the applied drain voltage level.

For a given drain voltage level, the rate of hot carrier degradation is strongly dependent on the gate voltage level applied to the device. For devices with channel lengths $> 0.25\ \mu$m three different regions of gate bias voltage can be distinguished under which hot carrier degradation is most severe. These three regions are listed in table 3.1 for nMOSFETs [32]. In this table it can be seen that hole trapping and interface state generation dominate at low gate voltages and electron trapping at high gate voltage levels. At medium gate voltage levels, the generation of interface states is dominant. For such long channel length devices, the maximum degradation rate was found to occur at medium gate voltage levels. This condition coincides with the condition for maximum substrate current. For this reason, the worst-case hot carrier lifetime of conventional nMOSFETs was typically characterized at maximum substrate current condition.

With the newer generations of MOSFETs, with channel lengths below 0.25 μm the gate voltage for maximum degradation rate was found to be shifted to high gate voltage levels ($V_{\mathrm{G}} \cong V_{\mathrm{D}}$), experimentally demonstrated in [34].

Table 3.1: Dominant nMOSFET hot carrier degradation mechanisms at different gate bias for long channel devices [32].

V_G **range**	**Degradation mechanism**
$V_G \cong V_T$	Interface state generation
	Hole trapping
$V_G \cong \frac{V_D}{2}$	Interface state generation
$V_G \cong V_D$	Electron trapping

The occurrence of different hot carrier degradation mechanisms at different gate voltage signals makes the lifetime prediction of AC and RF circuits challenging. For DC stress signals it is obvious which hot carrier degradation mechanism is dominant for a given MOSFET. For AC and RF circuits on the other hand, the total hot carrier degradation may consist of the contribution of different hot carrier degradation mechanisms. This must carefully be taken into account.

3.2.2 AC effects

Early reports on AC hot carrier degradation reported on an enhanced AC effect with increasing frequency, which was later explained by Bellens et al. as a measurement artifact [35]. They showed that the self inductance of the wiring in the measurement cables can cause large voltage overshoots. This conclusion made the interpretation of earlier reported results questionable and it signifies the importance of generating well-defined voltage levels for performing AC reliability experiments. For AC hot carrier experiments up to 10 MHz this signal distortion due the cable inductance can be overcome by adding a large parallel capacitor to the DUT at the drain side [35]. For higher frequencies additional effects such as signal distortion due to impedance mismatch come forward, as explained in chapter 2.

Since in digital circuitry clock frequencies exceeding 100 MHz have been readily available in the early 1990's, the desire for evaluating device degradation under AC stress signals with frequencies far beyond 10 MHz has emerged years before CMOS was introduced into RF circuits. In order to solve the measurement issues that arise at such high frequencies, self stressing devices were introduced [36, 37, 38, 39]. These self-stressing structures consist of oscillator and inverter circuits in which high-frequency stressing signals are generated on-chip. Using these structures, hot carrier experiments with inverter switching frequencies up to 369 MHz were already reported as early as 1994 [39]. None of these experiments revealed any unexpected effects occurring during AC stress.

When comparing AC hot carrier degradation to DC degradation Mistry et al. showed that use can be made of quasi-static assumptions [40]. Care should be taken that the contributions of all hot carrier degradation mechanisms as depicted in table 3.1 are taken into account. This may be done using a Matthiesen-like

formula [40]:

$$\frac{1}{\tau_{\mathrm{AC}}} = \frac{1}{\tau_{N_{\mathrm{it}}}} + \frac{1}{\tau_{\mathrm{e}}} + \frac{1}{\tau_{\mathrm{h}}} \tag{3.2}$$

In this expression τ_{AC} is the actual device lifetime under AC stress conditions, $\tau_{N_{\mathrm{it}}}$ is the device lifetime taking into account only the generation of interface states, τ_{e}, device lifetime based on electron trapping and τ_{h}, lifetime concerning only hole trapping.

With reducing supply voltages in CMOS technologies since the late 1990's, hot carrier degradation has become less of an issue for digital circuitry. Therefore the need for investigating AC hot carrier degradation at even higher frequencies became less evident. In RF circuits, on the other hand, voltage signals exceeding nominal supply voltage are no exception, especially in PA's. Therefore in RF circuits, the AC hot carrier effect may become a very important degradation mechanism if this is not accurately taken care of. This was acknowledged by Presti at al. in [28], where hot carrier degradation was measured for an nMOSFET operating at 1.9 GHz. Their approach however prevented an accurate comparison with lower frequency signals, thereby not shedding insight on any frequency dependence of the hot carrier effect.

When designing PA's in CMOS technology, it is most desirable to use nMOSFETs as the power driving transistors. This is because of their superior current -and hence power- driving capabilities over pMOSFETs. One of the main issues with RF PA design is guaranteeing a sufficiently high output power while keeping the drain voltage at an acceptably low level. This may not always be feasible with voltage signals below nominal supply voltage. It is for this reason that a proper investigation of hot carrier degradation under RF conditions is very relevant. As typically no pMOSFETs are used for this purpose, the work in this section only focuses on RF hot carrier degradation in nMOSFETs.

When performing RF hot carrier experiments it is important to realize that the highest voltage signals occur at the drain side of the devices, if the devices are used in PA's. It is therefore most relevant to investigate RF hot carrier degradation with an AC drain voltage. Such a kind of stress is used in this section with the gate voltage kept constant. This is different from earlier AC hot carrier experiments where the behavior of digital circuits was mimicked. For these experiments an AC gate voltage with constant drain voltage was more appropriate for describing actual circuit behavior.

In [41] a theoretical analysis is given of RF to DC lifetime ratios under such stressing conditions. The model assumes quasi-static behavior while it has not been experimentally verified. This will be done in this section for sinusoidal drain voltage signals by investigating the frequency dependence from the MHz towards the GHz range, and knowing that quasi-static assumptions are allowed in the MHz range, as discussed above.

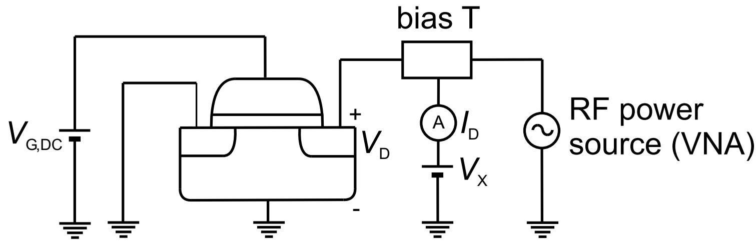

Figure 3.1: Schematic drawing of the RF hot carrier measurement setup.

3.2.3 Measurement setup

RF hot carrier experiments were performed on nMOSFETs in two different technologies. The type A devices were 90 nm process devices and the type B devices were processed in a 0.13 μm process. The type A devices have a gate channel length of 0.10 μm and the total gate width is 120 μm, consisting of 4 identical cells each having 6 gate fingers of 5 μm wide. The type B devices have a channel length of 0.13 μm; the total gate width is 192 μm, consisting of 8 identical cells each having 8 gate fingers of 3 μm wide. All devices were laid out in a two-port ground-signal-ground configuration, optimized for RF measurements. The gate was connected to one signal pad and the drain to the other. The source was tied to the substrate and connected to the ground plane. The devices were connected on-wafer using Suss |Z| probes. For each individual measurement a new device was connected.

In figure 3.1 a schematic illustration of the measurement setup is given. The equipment used in the setup consisted of a Rohde & Schwarz ZVB 20 Vector Network Analyzer for generating the RF voltage and an HP4156A semiconductor parameter analyzer for the DC biasing and measurements. The equipment was connected, through an IEEE 488 bus, to a PC. The complete measurement procedure was controlled using Labview software.

In the experiments a DC voltage was applied to the gate, while a sinusoidal RF voltage was generated at the drain port. Such kind of stress best mimics operation in an RF PA. The gate voltage must be held constant in order to ensure similar stressing conditions for different frequencies.

The gate and drain voltage signals have to be chosen such that a considerable amount of hot carrier degradation takes place that can be measured on a timescale suitable for experiments (i.e. 1,000 to 10,000 s). The devices need to be biased in inversion for the hot carrier effect to take place. This biasing poses an extra difficulty in generating well-defined RF voltage signals: the impedance seen by the RF power source can become very low and has a strong dependency on the drain voltage. The result of this is that the reflection coefficient during an entire cycle of the drain voltage signal may vary a lot. This may lead to severe signal distortion.

As stated before, the goal of the RF hot carrier experiments discussed in

this section is to investigate whether any frequency dependence can be observed between the MHz and GHz range. In this way it can be determined whether quasi-static models can be used for estimating hot carrier degradation under RF stress conditions, just as it is allowed for low frequency stress. Considering this purpose, it is important to guarantee that the hot carrier degradation during stress can not be attributed to the DC component of the stress signal during these experiments. This can be made sure by using an RF drain voltage signal with sufficiently large peak-peak value. This can be understood from figure 3.2, where device degradation was plotted as a function of DC drain voltage: if for instance a drain voltage signal would be chosen ranging between 2 and 2.5 V, a considerable portion of the measured degradation could be attributed to the DC degradation at $V_{\mathrm{D}} = 2$ V. These measurement results would not be suitable for investigating the frequency dependence of hot carrier degradation. If the drain voltage is varied between 1.5 V and 2.5 V on the other hand, the degradation at 1.5 V is negligible compared to degradation at 2.5 V. This means that hot carrier degradation for such a drain voltage signal is indeed dominated by RF stress rather than the DC component.

For this reason, the RF hot carrier degradation as discussed in this section is analyzed with an amplitude of the drain voltage signal as high as possible. The value of this amplitude is limited by the limitation posed by the voltage generation technique as discussed in section 2.4. The impedance seen at the drain is strongly dependent on the drain voltage. This may cause a harmonic distortion that is unwanted for the experiments discussed here. Therefore a signal integrity analysis was done before any RF hot carrier experiment was performed. The amplitude of the drain signal was chosen such that it had the maximum value for which the harmonic distortion of the drain voltage signal is negligible. As this harmonic distortion is most severe at the highest frequencies, this analysis was performed at the highest frequency used in the experiments, i.e. 3.2 GHz. The result of this is shown in figure 3.3 for both devices. For the type A devices the gate voltage was set to 2.5 V and for the type B devices this was 1.5 V. The figure shows that the resulting waveforms do not have a considerable harmonic distortion. These voltage conditions were used to perform the RF hot carrier experiments. The frequency of the drain voltage signal was varied from 10 MHz to 3.2 GHz.

Another difficulty that is encountered in RF hot carrier experiments is the fact that a considerable DC drain current can be observed to flow. Figure 3.1 shows that the DC component of the drain voltage is set using V_{X}, which is connected to a bias T. The bias T's used during the measurements were the internal bias T's of the VNA, connected at its DC biasing ports. This connection caused an extra parasitic series resistance of approximately 7 Ω. With DC drain currents flowing in the order of 100 mA to 300 mA, this may cause a DC voltage drop between V_{X} and V_{D} of 0.7 to 2.1 V. If this is not properly accounted for, this may lead to obscurities in the applied drain voltage signals and as a result large inaccuracies in the measurement results may arise.

This difficulty can be solved by characterizing the DC performance of the

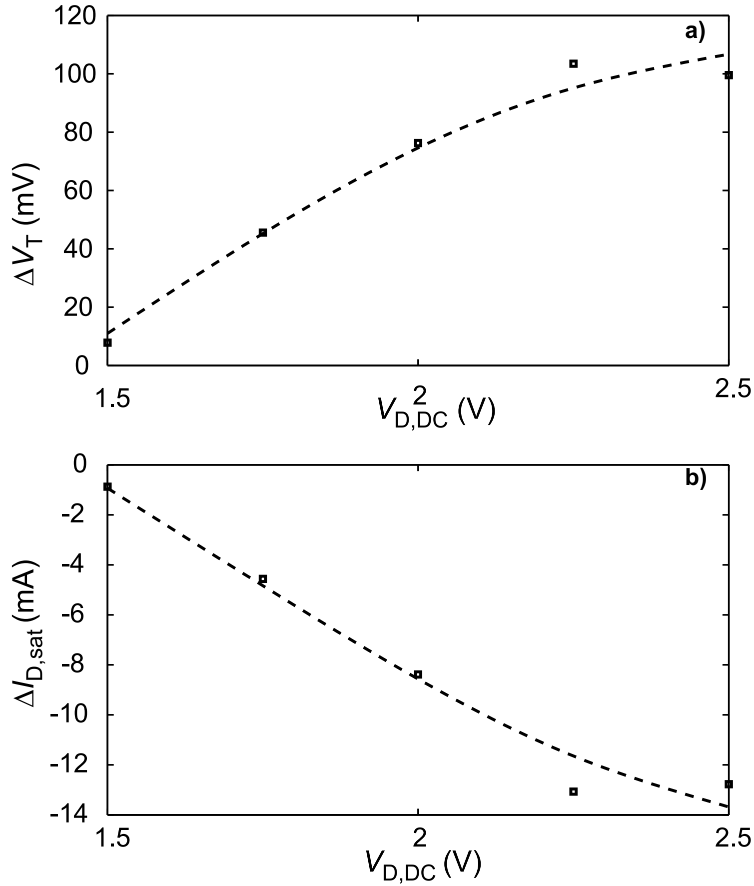

Figure 3.2: DC hot carrier degradation as a function of drain voltage for the type A devices after 10,000 s of stress and a gate voltage of 2.5 V. In a) V_T degradation is shown and in b) this is done for degradation in $I_{D,sat}$. The dashed lines are included as a guide to the eye.

connection between V_X and V_D prior to any measurement. If the parasitic DC series resistance R_s is accurately known, the DC component of V_D, $V_{D,DC}$ can be directly found from V_X, R_s and the measured DC drain current I_D:

$$V_{D,DC} = V_X - I_D \cdot R_s \qquad (3.3)$$

Any desired value of $V_{D,DC}$ can now carefully be set by a routine that gradually increases V_X and monitors I_D. $V_{D,DC}$ is calculated using expression 3.3. The routine is automatically stopped when the desired value of $V_{D,DC}$ is reached. The RF component of the drain voltage signal can subsequently be set by making use of the RF voltage generation approach of 2.4.

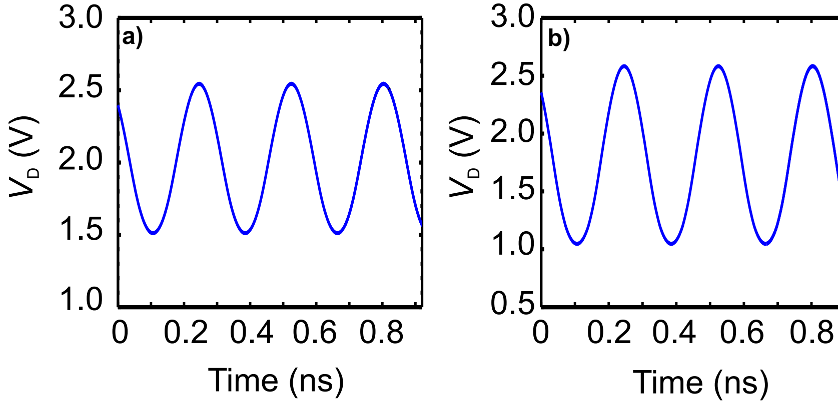

Figure 3.3: Drain voltage waveforms as they can be obtained at 3.2 GHz for a) the type A devices and b) the type B devices. The signals are calculated using the technique described in section 2.4, using impedance levels measured with a gate voltage of 2.5 V for the type A devices and 1.5 V for the type B devices. The shown voltage signals have the maximum realizable amplitude for which nonlinearities are negligible.

3.2.4 Measurement results

Using the measurement setup of figure 3.1, hot carrier degradation under RF stress was analyzed. This was done using a stress-measure-stress procedure in which the hot carrier stress was periodically interrupted. Both V_{T} and $I_{\mathrm{D,sat}}$ were monitored during this procedure. V_{T} was obtained by measuring I_{D} as a function of V_{G} with V_{X} set to 0.1 V and using the extraction method of [42]. In this extraction method V_{T} is found by extrapolating the tangent line in the g_{m} vs. V_{G} curve from the point of maximum derivative. V_{T} is defined as the value of V_{G} where this line intersects with the x-axis. $I_{\mathrm{D,sat}}$ was defined as the measured DC drain current with V_{G} set to 1.2 V and V_{X} set to a value corresponding to a V_{D} of 1.2 V.

In figure 3.4 device parameter degradation is plotted against the frequency of the stress signal for the type A devices. Degradation is plotted after 100 s, 1,000 s and 10,000 s of stress. Devices were stressed with a V_{G} of 2.5 V and a drain voltage signal as shown in figure 3.3 a). The frequency of the drain voltage signal was varied between 10 MHz and 3.2 GHz. The results indicate that the frequency of the stress signal has no influence on device parameter degradation for the type A devices. Some spread can be found on the data, but no significant frequency dependence can be observed.

The results presented in figure 3.5 show the device parameter degradation as a function of stress time for the type A devices. Again no clear frequency dependence on device parameter degradation can be observed . This, and the results of figure 3.4 indicate that for these devices and stress conditions, the hot

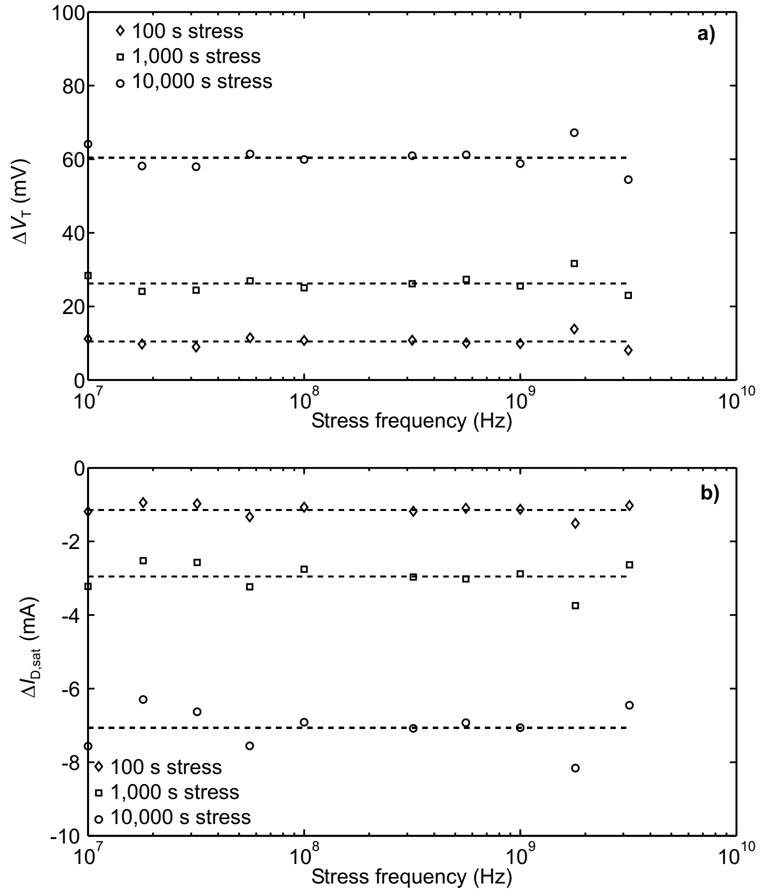

Figure 3.4: Hot carrier degradation as a function of frequency for the type A devices and a drain voltage signal as shown in figure 3.3 a) as a function of frequency. The gate voltage was 2.5 V during all stress measurements. In a) V_{T} degradation is shown and in b) this is done for $I_{\mathrm{D,sat}}$ degradation. The dashed lines are meant as a guide to the eye.

carrier effect may be assumed to be frequency independent. The solid lines in figure 3.5 show the best fit of the average value of degradation if it is assumed to follow a power law in time as in:

$$\Delta Par(t) = A \cdot t^n \tag{3.4}$$

In this expression $\Delta Par(t)$ represents device parameter degradation as a function of time, t is the stress time, A is a parameter dependent on technology and stress conditions and n is the time exponent. In figure 3.5 this n was found to be 0.41 respectively 0.39 for the average degradation in V_{T} and $I_{\mathrm{D,sat}}$. The average was

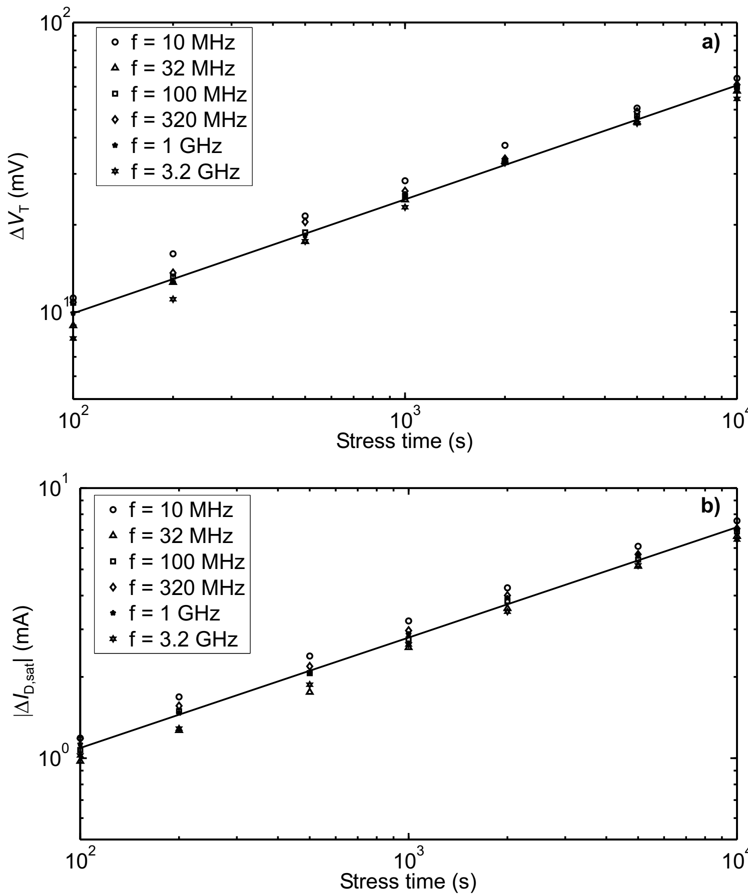

Figure 3.5: Hot carrier degradation as a function of time for the type A devices. The stress consisted of a constant gate voltage of 2.5 V and a drain voltage signal as shown in figure 3.3 b). The frequency of the drain voltage signal was varied. In a) V_T degradation is shown and in b) $I_\mathrm{D,sat}$ degradation. The solid lines show the degradation as a function of time using the average time exponent of 0.41 respectively 0.39 for figures a) and b).

taken over the measurement results obtained for all frequencies. These values are typical values for the time dependence of hot carrier degradation.

Similar experiments were performed on the type B devices where the stress signal consisted of a gate voltage of 1.5 V and the drain voltage signal as illustrated in figure 3.3 b). The frequency was varied from 10 MHz to 3.2 GHz. In figure 3.6 device parameter degradation is plotted against frequency and device degradation as a function of stress time is shown in figure 3.7. Similar to figure 3.5 the solid lines in figure 3.7 represent the best fit of the average degradation to expression 3.4.

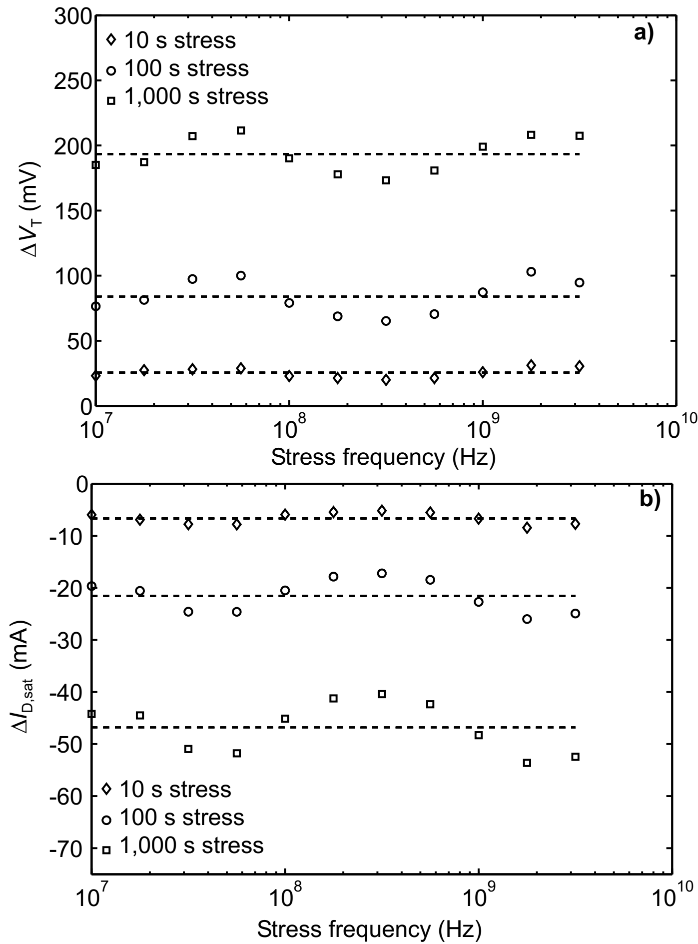

Figure 3.6: Hot carrier degradation as a function of frequency for the type B devices and a drain voltage signal as shown in figure 3.3 b). The gate voltage was 1.5 V during all stress measurements. Degradation levels are plotted after 10 s, 100 s and 1,000 s s of stress. In a) V_T degradation is shown and in b) this is done for $I_{D,sat}$ degradation. The dashed lines are meant as a guide to the eye.

The value of the average time exponent was found to be 0.45 respectively 0.42 for V_T degradation and $I_{D,sat}$ degradation. As with the type A device, the results on the type B devices do not show any frequency dependence. The results on the type B devices seem to exhibit a sinusoidal-shaped frequency dependence. This can be attributed to die-to-die variation: as an example, the measurements performed at 56 MHz are performed on a device located far away from the device used for the 316 MHz stress experiments, but near the device for the 1.8 GHz experiments.

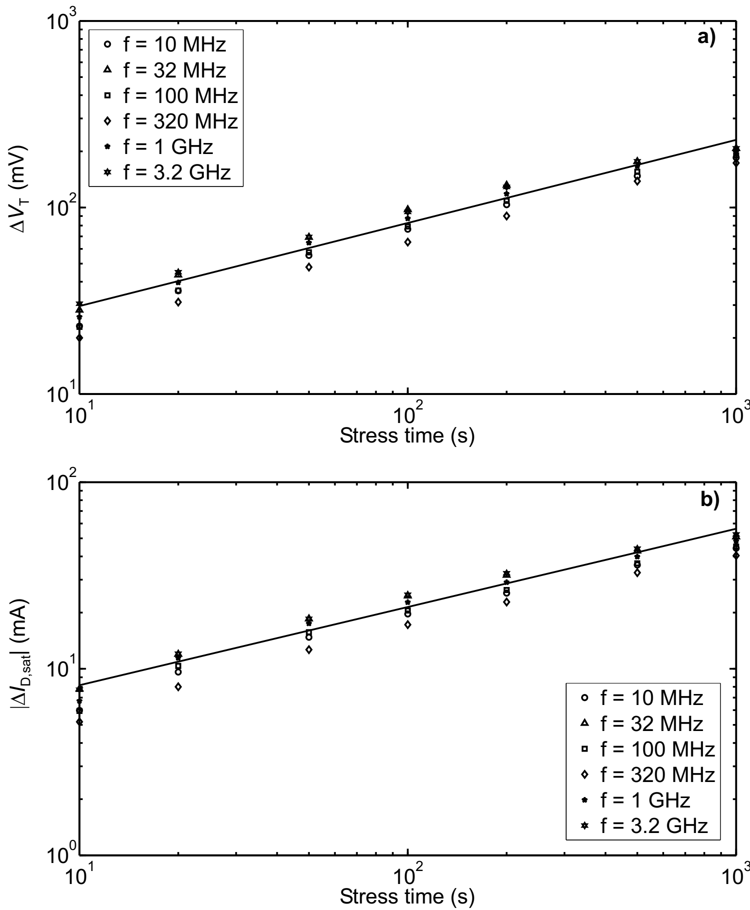

Figure 3.7: Hot carrier degradation as a function of time for the type B devices. The stress consisted of a constant gate voltage of 1.5 V and a drain voltage signal as shown in figure 3.3 b). The frequency of the drain voltage signal was varied. In a) V_{T} degradation is shown and in b) $I_{\mathrm{D,sat}}$ degradation. The solid lines show the degradation as a function of time using the average time exponent of 0.45 respectively 0.42 for figures a) and b).

Measuring device characteristics on an unstressed device show a similar trend. No consistent frequency dependency can be observed from these results.

3.2.5 Discussion

The RF hot carrier experiments discussed in this section were performed for investigating the frequency dependence of hot carrier degradation between the MHz and the GHz range. The results shown in figures 3.4, 3.5, 3.6 and 3.7 indicate that

hot carrier degradation is frequency independent between 10 MHz and 3.2 GHz. Furthermore, it is well-accepted that AC hot carrier degradation may be described using quasi-static models for frequencies into the MHz range, as long as the contribution of all degradation mechanisms is accurately taken into account as in equation 3.2.2. Combining both these statements, one can conclude that hot carrier degradation can be accurately described using quasi-static models up to 3.2 GHz. This has never been verified before; the highest frequency for which this was experimentally verified was 369 MHz [39]. Any hot carrier experiment described in literature with higher stress frequencies provided no information on the applicability of these quasi-static models. This verification is very important for predicting circuit lifetime based on known voltage waveforms, such as the work presented in [41]. Also for use in reliability simulators this observation is very useful, as will explained in chapter 4

Contrary to most of the AC hot carrier experiments discussed in literature, the work presented in this section made use of AC drain voltage signals rather than AC gate voltage signals. Such kind of voltage signals are of more interest for evaluating RF hot carrier degradation as the hot carrier effect can be expected to be mainly affecting the circuit performance of PA's. These PA's typically have drain voltage signal exceeding nominal supply voltage, while the gate voltage is kept below supply voltage. While it is of course useful to also investigate other types of stress signals, this would not be feasible using the voltage generation procedure of section 2.4 due to distortion of the voltage signals. The approach discussed in this section however, is sufficiently accurate for the evaluation of hot carrier degradation in PA's.

For other RF building blocks the relatively low voltage signals in these circuits make the hot carrier effect less of an issue compared to other degradation mechanisms. Furthermore as nMOSFETs have superior performance for use in PA's, the hot carrier degradation in pMOSFETs is less of an issue. This means that work discussed in this section can be used as evidence for the applicability of quasi-static models for hot carrier degradation for the most critical block in RF circuits.

3.3 RF NBTI degradation

3.3.1 DC model

NBTI is a mechanism that causes device degradation at relatively low voltage levels compared to hot carrier degradation or oxide breakdown. For this reason it has received a lot of attention in recent years. In present understanding of the phenomenon was recently reviewed [43]. In section 1.3 it was already explained that NBTI is a degradation mechanism, mainly affecting pMOSFETs where both interface and as positive oxide charge are being formed when a negative voltage is applied to the gate of a pMOSFET. It is generally accepted that the generation of interface states can be described using the reaction-diffusion (R-D) model, while the exact nature of the positive oxide charge is still a matter of debate.

Furthermore recovery effects were introduced, which is the effect that after removal of the stress signal, partial recovery of device degradation occurs.

Similar to hot carrier degradation NBTI degradation as a function of time can be described using a power law expression as in:

$$\Delta Par(t) = B \cdot t^n \tag{3.5}$$

In this expression ΔPar is the degradation of any device parameter; often V_T is used. Parameter B is dependent on process technology and stress conditions and t represents the time duration of the stress signal. Parameter n is the time exponent and for evaluating NBTI measurements it is a very important parameter. Recent literature has revealed a typical value for n of 0.16-0.17. This value can be obtained if parameter degradation is measured without removal of the stress signal, using the on-the-fly technique of [18]. This value coincides with a theoretical value that can be found if the diffusing species in the R-D model is molecular hydrogen [15, 16]. If recovery effects are not properly taken care of, higher values of n may be found. Its value is strongly dependent on the time between removal of the stress signal and the actual measurement of the device parameter of interest. If NBTI degradation is monitored in a conventional stress-measures-stress sequence, typical reported values of n are in the range of 0.25. After longer stress times this value may decrease towards 0.16-0.17, which can be explained within the framework of the R-D model [15].

When performing reliability measurements in an accelerated stress measurement, knowledge on accelerating factors is needed. NBTI is known to be strongly dependent on temperature; therefore NBTI experiments are typically performed at temperatures of 100°C and higher. Besides this temperature dependence it is known that lowering the gate voltage increases degradation in pMOSFETs. There has been some debate in literature whether the gate voltage or the oxide field is the accelerating parameter. Huard et al. have shown experimental evidence for an oxide field dependency rather than voltage dependency [18]. Based on this observation the following expression may be used for modeling V_T degradation under NBTI stress [17]:

$$\Delta V_\mathrm{T}(t) = C \cdot e^{\gamma \cdot E_\mathrm{ox}} \cdot e^{-E_\mathrm{A}/kT} \cdot t^n \tag{3.6}$$

In this expression E_ox is the field across the oxide, E_A the activation energy for the temperature dependence of degradation and k and T are Boltzmann's constant and the absolute temperature respectively. Parameter n is the time exponent and parameters C and γ are dependent on process technology and the stress conditions used. This expression reveals that NBTI degradation may be accelerated by increasing the temperature and increasing the absolute value of the (negative) gate voltage in pMOSFETs. This can be used for setting up NBTI experiments.

3.3.2 AC effects

NBTI degradation under AC stressing conditions has been discussed in various papers. As can be expected from the known recovery effects found in NBTI degradation results, AC NBTI degradation can not be modeled using quasi-static models that do not take these effects into account. In this section the purpose is not to solve this issue, but to shed insight on the question on whether AC NBTI degradation is a frequency dependent effect. In literature this issue has been discussed but controversy remains on whether or not a frequency dependent component is present in NBTI degradation. Alam et al. claim that on the basis of the R-D model no frequency dependence should be found [15]. This model however makes use of assumptions on the recovery mechanism that are not generally agreed upon. Some authors claim that the recovery mechanism can be completely or partly be attributed to the detrapping of holes in deep oxide traps [17, 18]. Under this assumption some frequency dependence may be expected.

Experimental results concerning AC NBTI degradation reflect a similar disagreement on the frequency dependence of AC NBTI degradation. Different papers show a clear frequency dependence of NBTI degradation [44, 45, 46, 47, 48, 49]. It has also been claimed that only the formation of positive oxide charge contributes to any frequency dependency, while interface state generation is frequency independent [50]. Others have experienced no frequency dependence at all for AC NBTI degradation [51, 52, 53].

The frequency range discussed in these papers generally does not exceed the MHz range. Only in [49, 53] NBTI degradation under RF stress conditions is discussed. In [49] use is made of a ring oscillator circuit and device parameter degradation is monitored indirectly using degradation in the oscillation frequency. The approach in [53] on the other hand allows direct extraction of V_{T} shifts. In [53] no frequency dependence of NBTI degradation was reported in the entire frequency range from 1 Hz to 2 GHz. On the contrast the results in [49] do show a small frequency dependence by comparing the degradation observed at 100 MHz and 3 GHz. In this section additional experimental results will be presented that may shed more insight on any frequency dependence of NBTI under RF stress conditions.

3.3.3 Measurement setup

RF NBTI measurements were performed using the measurement setup as shown in figure 3.8. The equipment used in the setup consisted of a Rohde & Schwarz ZVB 20 Vector Network Analyzer for generating the RF voltage and an HP4156A semiconductor parameter analyzer for the DC biasing and measurements. Measurements were performed using a Suss PA200 semi-automatic probe station. The chuck temperature was set using an ATT P40 cooling unit. The equipment was connected, through an IEEE 488 bus, to a PC. The complete measurement procedure was controlled using Labview software. RF voltage signals were generated and these were superimposed on the DC voltage V_{G} through the use of a bias T.

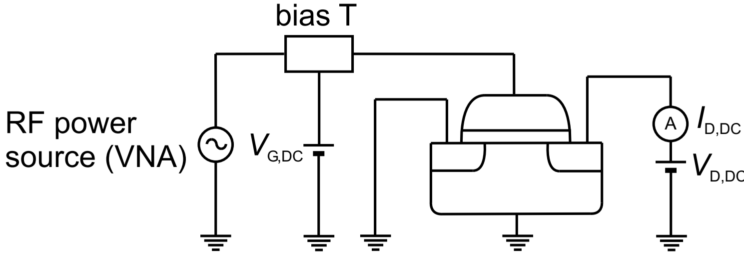

Figure 3.8: Schematic drawing of the RF NBTI measurement setup.

During stress $V_{D,DC}$ was set to 0 V.

Measurements were performed on pMOSFETs in two different technologies. The type C devices were processed in a 90 nm node process and the type D devices were pMOSFETs processes in a 0.13 μm process. The type C devices have a gate channel length of 0.10μm and the total gate width is 120 μm, consisting of 12 identical cells each having 1 gate finger of 10 μm wide. The type D devices, have a gate channel length of 0.13 μm and the total gate width is 192 μm, consisting of 8 identical cells each having 8 gate fingers of 3 μm wide. All devices were laid out in a two-port ground-signal-ground configuration, optimized for RF measurements. The gate was connected to one signal pad and the drain to the other. The source was tied to the substrate and connected to the ground plane. The devices were connected on-wafer using Suss |Z| probes. For each individual measurement a separate device was connected.

RF NBTI measurements were performed using a stress-measure-stress procedure in which the devices were stressed with a sinusoidal voltage signal and periodically interrupted. V_T was obtained following [42] from a DC I_D-V_G curve with V_D set to 0.1 V. The time delay between removal of the stress signal and the start of the V_T measurement was 16 s. In order to prevent any pre-stress from occuring the appropriate power level for generating the desired V_{pp} was performed on a separate device. The chuck temperature used in the experiments was 125°C. The gate voltage during stress can be expressed as:

$$V_G(t) = V_{G,DC} + \frac{V_{pp}}{2} \cdot \sin(2\pi f t) \tag{3.7}$$

In this expression $V_G(t)$ represents the voltage applied to the gate as a function of time and $V_{G,DC}$ is the DC component of this signal. V_{pp} was chosen to be 3 V for the type C devices and 2.5 V for the type D devices. $V_{G,DC}$ was set to -1.5 V for the type C devices and -1.25 V for the type D devices.

3.3.4 Measurement results

The above discussed measurement procedure was applied for investigating the frequency dependence of NBTI degradation in the frequency range from 10 MHz

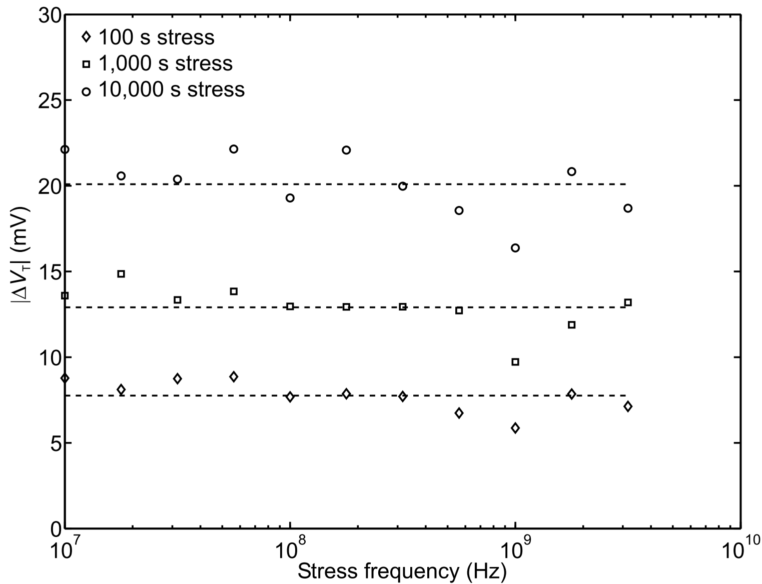

Figure 3.9: V_{T} degradation plotted against frequency for RF NBTI stress on the type C devices. The stress signal consisted of a constant gate voltage of -1.5 V superimposed on a sinusoidal voltage signal with V_{pp} set to 3 V. The measurements were performed at 125°C. The dashed lines are meant as a guide to the eye.

to 3.2 GHz. In figures 3.9 and 3.10 V_{T} degradation was plotted against stress frequency for the type C and type D devices respectively. Although both figures show some spread on the data, a clear trend is visible: for both the type C and type D devices NBTI degradation has no observable frequency dependence. This observation is in agreement with the theory of [15] and the experimental results presented in [51, 52, 53]. The time dependence of NBTI degradation under RF stress is depicted in figure 3.11 for the type C devices. The exponential time dependence reported before can be clearly recognized. The solid line shown in this figure is the average value of $|V_{\mathrm{T}}|$ degradation over all frequencies. The time exponent for the measurements performed for the various frequencies varies randomly between 0.19 and 0.27. The average value of this time exponent was found to be 0.22. As a comparison NBTI degradation after a DC stress was also measured using the same stress-measure-stress procedure, but without any RF component superimposed on this DC voltage. The gate voltage used for this experiment was set to -1.5 V. The time exponent found in this DC experiment was 0.24, with $|\Delta V_{\mathrm{T}}|$ ranging from 1.1 mV to 5.3 mV after 10 respectively 10,000 s of stress. From these results no significant difference between DC NBTI and RF NBTI degradation can be observed.

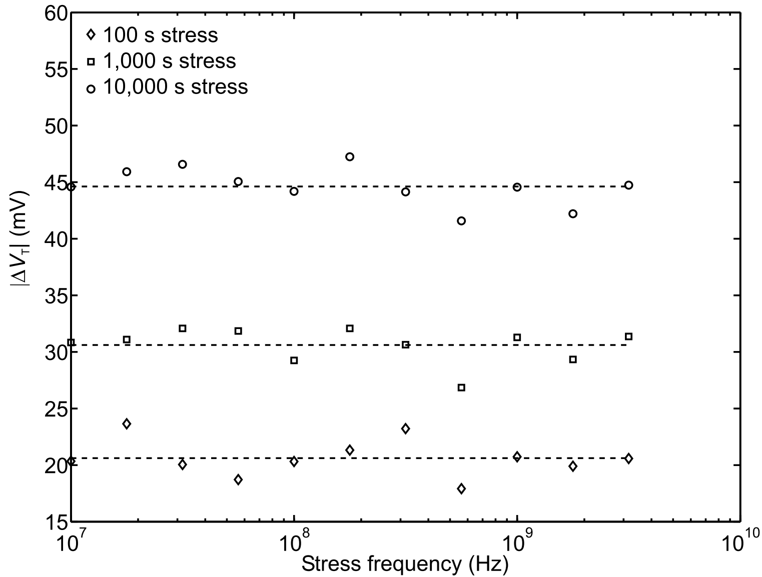

Figure 3.10: V_{T} degradation plotted against frequency for RF NBTI stress on the type D devices. The stress signal consisted of a constant gate voltage of -1.25 V superimposed on a sinusoidal voltage signal with V_{pp} set to 2.5 V. The measurements were performed at 125°C. The dashed lines are meant as a guide to the eye.

3.3.5 Discussion

The RF NBTI degradation experiments presented in this section provide an important contribution to getting a proper understanding of NBTI degradation, especially for AC conditions. NBTI degradation is not an issue specifically hazardous for RF circuits, as was the case with the hot carrier effects in the previous section or oxide breakdown in the next. The results of this section are however more generally applicable than for RF circuits only. In the presented experiments a pMOSFET was periodically switched on and off. The only difference with operation in digital circuitry is the sinusoidal voltage waveform rather than trapezoidal.

From the data presented in this section an important observation can be made: RF NBTI degradation appears to be frequency independent. This observation is a very useful contribution in determining the exact nature of the physical mechanisms underlying NBTI degradation. At this moment however the full picture is too premature to draw any generalized conclusions. All that can be stated is that the presented results support the theory of Alam et al. [15] and are in accordance with the experimental results presented in [51, 52, 53].

For developing accurate models describing RF NBTI degradation, recovery effects should be carefully taken into account. For this purpose the use of various

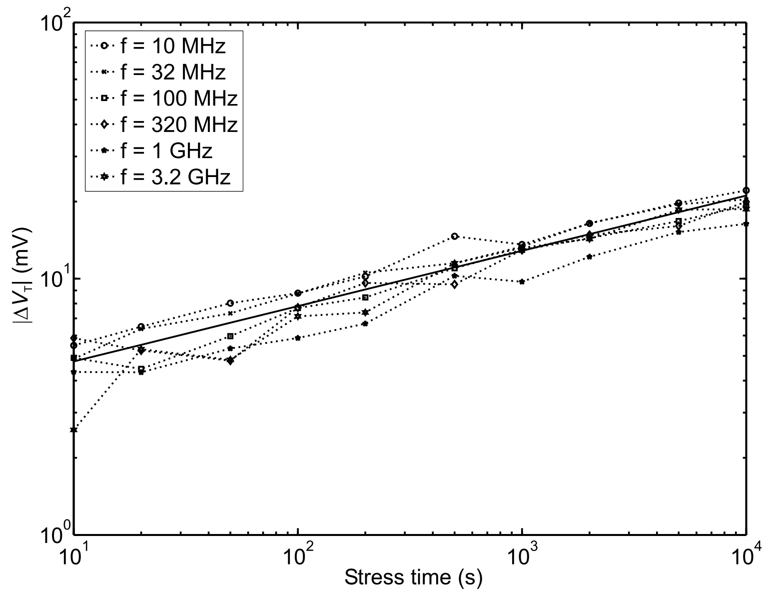

Figure 3.11: V_T degradation as a function of stress time for the type C devices after an RF NBTI stress. The RF stress consisted of a DC gate voltage of -1.5 V superimposed on a sinusoidal voltage signal with V_{pp} set to 3 V. This was done for various frequencies ranging from 10 MHz to 3.2 GHz. The measurements were performed at 125°C. The solid line is fitted to the average value of the RF stress data; this line has a time exponent of 0.22.

kinds of the voltage waveforms might prove to be very insightful. Generating these waveforms at frequencies exceeding 10 MHz might be cumbersome. Based on the observed frequency independence on the other hand, it can be concluded that it suffices to perform these experiments at AC frequencies (i.e. $\leq$ 10 MHz). These low frequency data can then also be applied in describing RF NBTI degradation.

3.4 Gate-oxide breakdown under RF stress

3.4.1 DC model

Gate-oxide breakdown is the event of a sudden increase in the gate current, while a device is under stress. It is the manifestation of a conducting channel inside the gate dielectric. In section 1.3 it was explained how gate-oxide breakdown can be modeled using percolation theory. Also the different mechanisms for trap generation were discussed there. In this section oxide breakdown under RF stress conditions will be discussed. First an overview will be given on the known acceleration models for gate-oxide breakdown. These are important to understand for performing RF breakdown experiments. Next a brief summary will be given of

the known AC effects of oxide breakdown and finally RF breakdown experiments will be discussed that make use of the voltage generation approach of section 2.4.

Gate-oxide breakdown is typically characterized using the time-to-breakdown t_{BD}. This is the time it takes for oxide breakdown to take place while it is under stress. For devices with an oxide thickness greater than 5 nm, field driven models were used to relate t_{BD} to stress conditions: the E-model [54, 55] and the $1/E$-model [4].

With oxide thicknesses below 5 nm neither of the two field based models is applicable. It has been shown [56] that a gate voltage driven model is more accurate for ultra-thin dielectrics. These voltage driven models can be described using an exponential dependence, similar to the field driven models, but Wu et al. have experimentally shown that a power-law model is better applicable [57]. For ultra-thin oxide devices t_{BD} can therefore be written as:

$$t_{\mathrm{BD}} \propto V_{\mathrm{G}}^{-m} \tag{3.8}$$

In this expression V_{G} is the value of the gate voltage and m is the power-law exponent. Typical values of m range between 30 and 40. This expression reveals that a minor increase in the gate voltage may dramatically decrease device lifetime.

3.4.2 AC effects

Lifetime enhancement under AC stress has been observed by several authors [58, 59, 60, 61, 62, 63]. This effect is most pronounced under bipolar stress conditions. It has been attributed to frequency dependent hole trapping kinetics and the detrapping of holes during zero or negative bias stress [60]. In this model it is assumed that hole generation and trapping is a precursor to oxide breakdown. With higher stress frequencies, the time at which the actual stress voltage is present at the gate per cycle reduces. The trapped hole density therefore has less time to spread out in the oxide leading to fewer trapped holes in the oxide and hence an increased lifetime with increasing frequency. For explaining the large difference between unipolar and bipolar stress however, it was stated that detrapping of holes during the off-state of the stress is the dominating effect causing lifetime enhancement [60].

These experiments have only been performed for frequencies of up to 10 MHz. For RF circuits it is important to know whether any unexpected behavior occurs at stress voltage signals with frequencies in the GHz range. In [27] oxide breakdown under RF stress conditions was investigated by designing RF PA circuits and recording t_{BD} values. The presented results revealed no difference in t_{BD} values between PA's operating at 80 MHz or 1.8 GHz. This approach very accurately mimics device operation under RF circuit conditions, but it is a very tedious task to properly investigate any frequency dependence.

The RF voltage generation approach discussed in section 2.4 enables the generation of equivalent voltage signals with different frequencies. Therefore it is perfectly suitable for evaluating the frequency dependence of gate-oxide break-

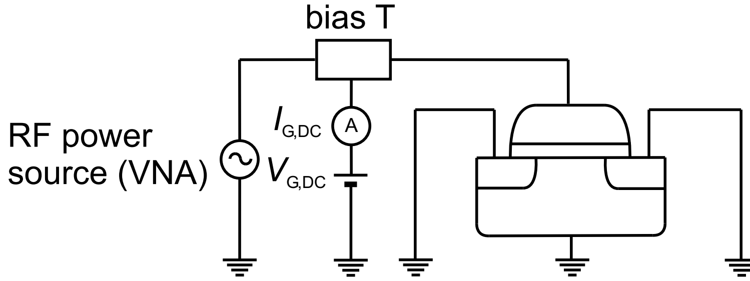

Figure 3.12: Schematic drawing of the measurement setup used for RF gate-oxide breakdown experiments.

down. In this section experimental results will be demonstrated that show the frequency dependence of t_{BD} under unipolar sinusoidal gate voltage signals.

3.4.3 Measurement setup

RF breakdown experiments were performed using the measurement setup as shown in figure 3.12. The equipment used was a Rohde & Schwarz ZVB 20 Vector Network Analyzer for setting the RF gate voltage and an HP 4156A semiconductor parameter analyzer for setting the DC bias voltage and measurement of the DC currents. The devices were connected on-wafer using Suss |Z| probes. An experiment was performed using n-type MOS transistors laid out in a two-port ground-signal-ground configuration. The gate was connected to one signal pad while the drain was connected to the other. The source was tied to the substrate and connected to the ground plane. The devices were processed in a 0.13 μm process flow and have a gate channel length of 0.13 μm. The total gate width of the devices is 128 μm, consisting of 2 identical cells, each having 64 gate fingers of 2 μm wide.

During the experiments the gate of the devices was stressed with a sinusoidal voltage signal:

$$V_{\mathrm{G}}(t) = V_{\mathrm{G,DC}} + \frac{V_{\mathrm{pp}}}{2} \cdot \sin(2\pi f t) \tag{3.9}$$

Here $V_{\mathrm{G}}(t)$ is the gate voltage signal and f the frequency of the stress signal. V_{pp} was set to 3.35 V and $V_{\mathrm{G,DC}}$ to 1.675 V. During the experiments the DC gate current $I_{\mathrm{G,DC}}$ was monitored and used for the detection of breakdown events. Breakdown was defined as the moment where between two measured samples a relative increase of 5% in $I_{\mathrm{G,DC}}$ was found. The result of this definition is illustrated in figure 3.13.

In this figure the measured I_{Gate} was plotted against stress time for different measurement examples. The examples used were typical examples for every measurement frequency used in the experiment. The figure also shows the moments at which breakdown was found using the 5 % definition used. It is clear that this

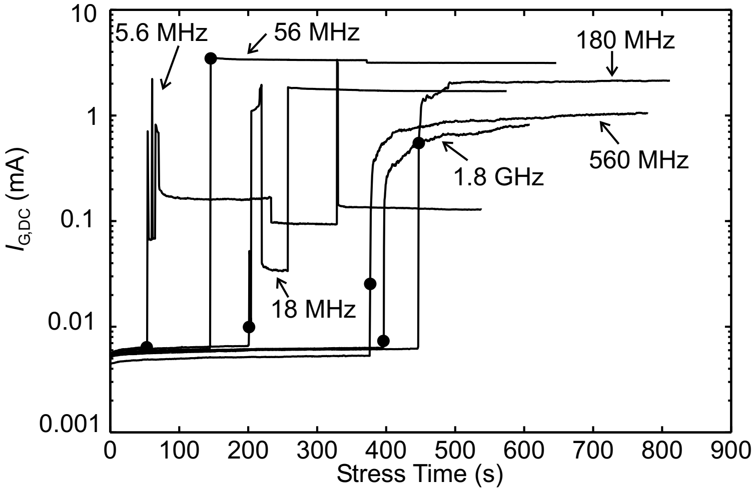

Figure 3.13: Recorded $I_{\mathrm{G,DC}}$ plotted against time. For each stress frequency one typical example is shown. The big dots indicate the moment in time where breakdown is detected using the 5 % definition.

definition is accurate for defining hard breakdown events on the given samples. Soft breakdown events may possibly be missed in this definition, but the use of the bias T limits a more accurate measurement of $I_{\mathrm{G,DC}}$. In the remainder of this chapter t_{BD} will always refer to the first hard breakdown event.

3.4.4 Measurement results

Measurements were performed with stress frequencies ranging from 5.6 MHz to 1.8 GHz. The measurements performed at 5.6 MHz were done using a measurement setup in which the VNA was replaced by an Agilent 33250A signal generator. For every frequency 20 samples were stressed and the t_{BD} values were recorded. The samples used for the different frequencies were randomly distributed over the wafer. In figure 3.14 the t_{BD} values are plotted in a Weibull plot for all these stress frequencies. The figure reveals an increase in the recorded t_{BD} values with increasing stress frequency. To be able to better quantify this effect an accurate description of the failure probability function is helpful. The results in figure 3.14 indicate that for all stress frequencies the probability function of the time for the first hard breakdown event can be described using the Weibull distribution:

$$F_{\mathrm{BD}}(t) = 1 - e^{-(\frac{t}{t_{63}})^{\beta}} \tag{3.10}$$

In this expression $F_{\mathrm{BD}}(t)$ is the probability that a device has suffered hard breakdown at time t; t_{63} and β are the two characteristic parameters of the Weibull distribution. These two parameters can be determined for every frequency making use of the recorded t_{BD} values shown in figure 3.14. This was done using

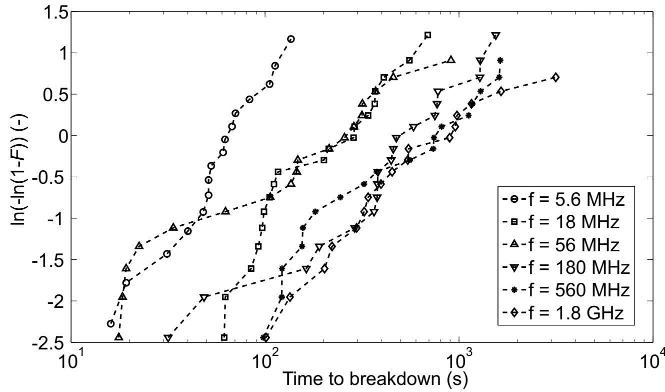

Figure 3.14: Weibull plots showing the distribution of t_{BD} values for all measurement frequencies used.

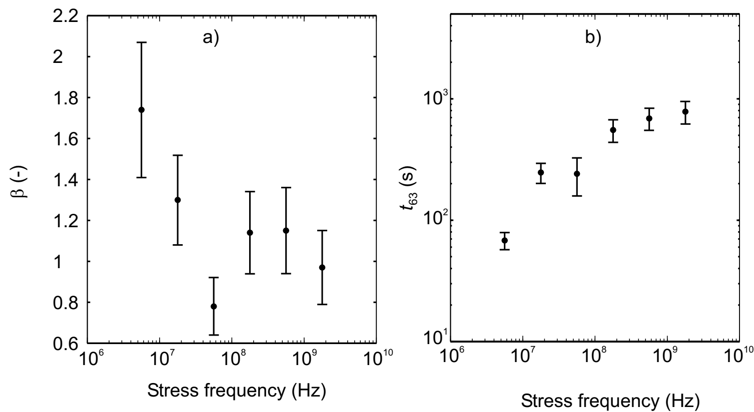

Figure 3.15: Weibull parameters as obtained using a maximum likelihood estimation. The parameters are plotted against stress frequency. The error bars indicate the standard deviation of the likelihood function included. A clear increase in t_{63} can be observed as a function of frequency; β appears to remain constant.

a maximum likelihood estimator embedded in a Matlab routine. The resulting most probable values of t_{63} and β are plotted in figure 3.15. In this figure a clear increase in t_{63} can be observed with increasing frequency. β appears to remain constant. As no frequency dependence of β is reported in literature and it is not to be expected to be frequency dependent, it is desirable to also extract t_{63} with an assumed constant β for all frequencies. This was done using a maximum likelihood estimator with β set to 1.11, a value that can be expected for the di-

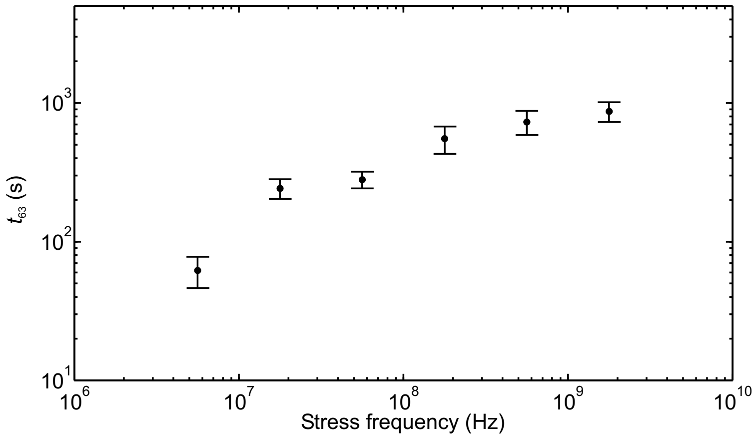

Figure 3.16: Extracted values of t_{63} using a maximum likelihood estimator with β set to 1.11. The error bars represent the standard deviation of the likelihood function.

electric thickness of the devices used (EOT = 1.5 nm) [64]. The result is shown in figure 3.16, the figure reveals a similar trend as in figure 3.15: with increasing frequency, t_{63} also increases. The exact values of t_{63} differ between the two figures, but the relative increase of t_{63} with increasing frequency is clearly visible.

3.4.5 Discussion

The results presented in this section show a clear frequency dependence of t_{63} for unipolar, sinusoidal stress voltage signals between 5.6 MHz and 1.8 GHz. In figure 3.15 b) extracted values of t_{63} were plotted that were determined for the maximum likelihood of both β and t_{63} while in figure 3.16 this was done for the maximum likelihood of only t_{63} with β assumed frequency independent. In figure 3.15 b) t_{63} shifts from 68 s for the 5.6 MHz signal to 783 s for the 1.8 GHz signal. These values are 62 s and 869 s respectively for the 5.6 MHz and 1.8 GHz signals of figure 3.16. This means that some discrepancy exists between the two methods of extracting t_{63}; extracting quantitative information from these results might be inaccurate, but the results do show that a considerable lifetime enhancement may be observed under unipolar AC stress signals.

This lifetime enhancement is in accordance with earlier reported results at lower frequencies [58, 59, 60, 61, 62, 63]. A striking difference with these earlier reported results is the relative large increase in lifetime for unipolar stress signals. This has not been shown before. A possible explanation for this effect could be the increased importance of hole trapping kinetics for frequencies above 10 MHz rather than hole detrapping. This explanation is in line with the model presented in [60] where hole detrapping was considered to be the dominating lifetime enhancement

effect for bipolar stress signals. Hole trapping kinetics was considered a possible mechanism for lifetime enhancement for AC stress signals, but it could not explain the large difference between unipolar and bipolar stress signals. The fact that lifetime enhancement for unipolar stress signals can be considerable for frequencies between 5.6 MHz and 1.8 GHz, is an indicator that this effect of reduced trapping of holes is becoming more dominant at these frequencies.

In [27] it was proposed that a dielectric relaxation time of 2-9 ns is needed for a trapped hole to create permanent damage. This means that if a stress signal is applied for a time-period shorter than this dielectric relaxation time, the stress becomes ineffective. With the use of sinusoidal stress voltages it is to be expected that the maximum stress level (in terms of gate voltage) that is effective for creating damage, decreases with increasing frequency.

Considering the results of figure 3.15 b) a lifetime increase of a factor of 14 can be observed between 5.6 MHz and 1.8 GHz. In order to relate this to the allowable stress voltage level at different frequencies, use should be made of equation 3.8 with parameter m ranging between 30 and 40. As an example one can determine the amplitude of a unipolar sinusoidal stress voltage signal, that has the same value of t_{63} as a lower frequency signal, knowing that a 14 times increase in lifetime under equivalent voltage signals exists between the low frequency and high frequency stress signal. Taking m to be 35 this would results in an increase in V_{pp} from, for instance, 1.2 V to 1.28 V. This is an increase of 80 mV or 10 % headroom for an analog designer.

Furthermore, as stated before, it is risky to draw quantitative conclusions from the measurement results presented in this section. An important feature of these results is the fact that quasi-static breakdown models for predicting RF circuit lifetime, will results in conservative estimates for the actual lifetime. Although conservative, the use of quasi-static models does not result in overly optimistic lifetime predictions. This knowledge can be very useful for the lifetime prediction of RF circuits, an example of this will be shown in chapter 4.

3.5 Conclusions

In this chapter it was investigated how MOSFETs degrade under RF stress conditions. Based on literature a comparison was made between well-known DC degradation models and AC stress conditions. This was combined with a new experimental approach for comparing low frequency ($\leq$ 10 MHz) stress to RF stress. This was done for the hot carrier effect, NBTI degradation and gate-oxide breakdown. Investigating hot carrier degradation and oxide breakdown under RF stress is of specific interest in designing RF circuits: Especially in PA's voltage levels may easily exceed nominal supply voltage. While RF NBTI degradation is of less specific interest in RF circuits, it may on the other hand be especially useful in getting a better understanding of the exact mechanism underlying NBTI degradation.

For hot carrier degradation it is important to realize that both the applied

drain and gate voltage levels affect device degradation rates. During an AC stress cycle, the device degrades under different modes of hot carrier degradation. When modeling AC hot carrier degradation quasi-static assumptions are allowed as long as the contribution of all modes of hot carrier degradation is included. When comparing low frequency hot carrier stress to RF hot carrier stress no frequency dependence can be observed. This was experimentally verified on devices in two different technologies. This means that for hot carrier lifetime assessment of RF circuits use can be made of quasi-static models and DC degradation data.

Due to recovery effects in NBTI degradation, it is not allowed to adopt quasi-static assumptions for accurate lifetime evaluation of devices stress under AC NBTI conditions. How to properly tackle this problem remains a matter of discussion. When accurate AC modeling of NBTI degradation is available, the extension to RF stress can easily be made, based on the experimental results shown in this chapter. The results reveal no frequency dependence of NBTI degradation from 10 MHz to 3.2 GHz.

Gate-oxide breakdown is known to be frequency dependent. However in literature this effect was most pronounced for bipolar stress signals. In this chapter a considerable frequency dependence has been observed for unipolar stress signals ranging between 5.6 MHz and 1.8 GHz; this has never been shown before in literature. A possible explanation for this lifetime increase is the increasing importance of limited hole trapping during the peak-level of the stress besides hole detrapping during low gate voltage signals. It has been shown that this lifetime increase results in a 10 % gain in voltage headroom. If this headroom can be fully utilized, RF circuits may be designed more aggressively in terms of voltage levels, compared to DC reliability specifications. In this way the performance of RF circuits may be increased while reliability specifications can be maintained at the same level as for DC and AC circuits.

Chapter 4

RF PA lifetime prediction

4.1 Introduction

The occurrence of a breakdown event in the gate-oxide of a MOSFET may not necessarily lead to circuit failure, as shown in [65] for digital circuits. Later Pantisano et al. showed that the impact of oxide breakdown on AC circuit performance diminishes at increasing frequencies thereby predicting RF circuits to be very robust against oxide breakdown [66]. In [67, 68, 69] the impact of breakdown paths on RF PA performance was investigated and the results indeed showed a very high robustness, even to multiple gate-oxide breakdown events. As RF PA's are designed to deliver high output power, the allowable voltage level across the gate-oxide is a very important parameter in RF PA design. If multiple breakdown events can be accurately taken into account for predicting RF PA circuit lifetime, design guidelines may be relaxed, allowing higher voltage levels and hence higher output power of the PA. Therefore a reliability simulator that incorporates multiple breakdown effects may be a very beneficial tool, especially for RF PA design.

Reliability simulators for CMOS circuits have been under development since the late 1980's [70, 71, 72, 73] which resulted into commercially available tools [74, 75]. Rather than looking at the parameter degradation of the individual components of the circuit, a reliability simulator allows the prediction of circuit lifetime on the basis of circuit parameter degradation. In a reliability simulator information on the degradation rate of a device under use conditions is combined with a model describing degraded MOSFET behavior. For hot carrier and BTI effects this can be done by making use of the fact that device degradation is a continuous effect. Oxide breakdown on the other hand is a discrete event of stochastic nature and therefore requires a different approach. Present-day reliability simulators typically only include effects of hot carrier degradation and NBTI. Modules for incorporating gate-oxide breakdown have been developed such as in [76], but here circuit failure was assumed to coincide with the first breakdown event in a circuit.

In this chapter a new simulation methodology will be presented that allows for incorporating multiple breakdown events into a reliability simulator in order to assess circuit lifetime. First in section 4.2 the model used for describing MOSFET performance after breakdown will be discussed, after this in section 4.3 it will be explained how the stochastic nature of oxide breakdown may be implemented into a reliability simulator. The resulting simulator has been evaluated on three different RF PA designs, these results will be discussed in section 4.4.

4.2 Degraded MOSFET model

The development of the simulator described in this chapter is focused on application in RF PA design. For this reason, and the fact the RF PA's typically consist of only nMOSFETs and passive components, the reliability simulator includes gate-oxide breakdown and hot carrier degradation. BTI degradation is a mechanism mainly obscuring pMOSFET performance (NBTI). It could be added in a way similar to hot carrier degradation, but implementing this is of less relevance for the purpose of validating the applicability of the simulation tool in RF PA design.

In the previous chapter it was found that the degradation rates of individual MOSFETs under RF stressing conditions may be approximated using quasi-static models for hot carrier degradation. For gate-oxide breakdown such an approximation may be used for obtaining a worst-case estimation for device degradation. t_{63} values may increase with increasing stress frequencies as shown in the previous chapter, but in a first-order approach this effect can be omitted in order simplify the approach of simulating multiple breakdown events. In this chapter use will be made of such quasi-static approximations.

4.2.1 Breakdown paths

Model

The statistics of oxide breakdown path formation can be well described using percolation theory [8]. The hardness of the breakdown path that is formed has been shown to be strongly dependent on the amount of power that is subsequently dissipated in the percolation path [77]: the larger the dissipated power, the larger the probability of having induced a hard breakdown event. This dissipated power is dependent on the stress voltage applied; this means that the larger the stress voltage is, the larger the probability of having formed a hard breakdown path is. As PA's typically encounter relatively large voltage levels, in this chapter it is assumed that all breakdown events are hard breakdown events. While possibly the occurrence of soft breakdown paths cannot be fully neglected, the effect of hard breakdown events on circuit performance is much more severe. Assuming all breakdown events to be hard breakdown events results in a worst-case estimation of circuit lifetime. In conventional lifetime models, circuit lifetime is assumed to coincide with the first breakdown event in a circuit; the new simulation tool can

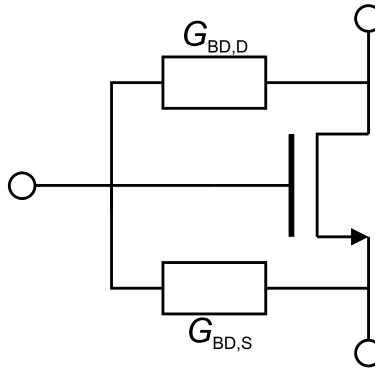

Figure 4.1: Model used for describing post-breakdown MOSFET behavior. Breakdown events are assumed to occur only at the gate-drain or gate-source overlap regions.

relax design guidelines considerably. Incorporating soft breakdown or progressive breakdown could further relax these guidelines, but in a first-order attempt to exploit the robustness of RF PA's against breakdown events this can be omitted.

The model used for describing MOSFET behavior after breakdown is based on the work done by Kaczer et al. [78]. In this work the breakdown path was modeled as a conductive path of $\sim$ 1 mS and the effect of this conductive path on MOSFET performance was related to the location of the path with respect to the channel. Breakdown paths located near the drain and source regions proved to exhibit the most disastrous effect on device performance. Breakdown paths in the middle of the channel are manifested as conductive paths with a conductance $\ll$ 1 mS.

The probability of a breakdown path being formed at a given location is strongly dependent on the voltage difference across the gate-oxide at that position. At present no accurate models exist for describing this failure probability alongside the channel for devices biased with a nonzero V_{DS}. The failure probability at the gate-drain respectively gate-source overlap regions on the other hand can be obtained from the values of V_{GD} respectively V_{GS}. Because the impact of breakdown events at these overlap regions on device performance is much higher than breakdown paths at positions along the channel, as a first-order approach only breakdown events at the gate-drain and gate-source overlap regions are taken into account. Although it is risky to make such a simplification, in the case of the PA's discussed in section 4.4 it appears that V_{GD} is the dominating voltage level in terms of breakdown probability. This is typical for PA's designed for delivering high output power signals. The development of the simulator as it is discussed in this chapter is focused on application in RF PA design; therefore this simplification is allowed.

The resulting model for describing breakdown behavior is shown in figure 4.1. In this figure two conducting paths are added to an unstressed MOSFET model.

The conductance of the paths is a multitude of the breakdown conductance of 1 mS:

$$G_{\mathrm{BD,D}} = n_{\mathrm{BD,D}} \cdot 1 \text{ mS}$$
$$G_{\mathrm{BD,S}} = n_{\mathrm{BD,S}} \cdot 1 \text{ mS} \quad (4.1)$$

$n_{\mathrm{BD,D}}$ and $n_{\mathrm{BD,S}}$ are the number of breakdown events that have occurred at the gate-drain respectively gate-source overlap region.

Degradation rate

Oxide breakdown is an instantaneous process of stochastic nature. Therefore it is not possible to identify a single moment in time at which a device will encounter gate-oxide breakdown with given stress conditions. It is, on the other hand, possible to describe the probability of encountering breakdown based on given stress conditions. In section 4.3 it will be discussed how this stochastic nature of gate-oxide breakdown can be implemented in a reliability simulator, making use of Poisson statistics. If only the first breakdown is considered, the cumulative Poisson distribution reduces to the cumulative Weibull distribution. This distribution is often used for characterizing oxide breakdown and is given by:

$$F_{\mathrm{BD}}(t) = 1 - e^{-\left(\frac{t}{t_{63}}\right)^{\beta}} \quad (4.2)$$

In this expression $F_{\mathrm{BD}}(t)$ represents the probability that at time t a breakdown event has occurred. t_{63} and β are the two parameters that can completely describe the failure probability as a function of time. These parameters apply for both the Weibull distribution of 4.2 as well as the Poisson distribution, which will be used in section 4.3.

From the percolation model it follows that β is dependent on the oxide thickness t_{ox}. A relation between t_{ox} and β can be found in [57]. t_{63} is a technology dependent parameter, that is also strongly dependent on stress conditions and the area of the gate-oxide. The area dependence of t_{63} follows from the percolation model and is found to be [8]:

$$t_{63} \sim A_{\mathrm{ov}}^{-1/\beta} \quad (4.3)$$

In this expression A_{ov} represents the area of the overlap region of interest. As was explained in chapter 3, in thin oxide devices t_{63} has a voltage dependency given by [56]:

$$t_{63} \sim |V_{\mathrm{GX}}|^{-m} \quad (4.4)$$

In this expression V_{GX} may refer to either V_{GD} or V_{GS}, depending on which overlap region is considered. As discussed above only the formation of breakdown paths in these two overlap regions is taken into account in the post-breakdown MOSFET model. Parameter m is a technology dependent parameter, typical values range between 30 and 40. Now combining expressions 4.3 and 4.4 the following expression for t_{63} can be derived.

$$t_{63} = \frac{d_0}{W^{1/\beta} \cdot |V_{\mathrm{GX}}|^{n}} \quad (4.5)$$

In this expression parameter d_0 is a technology dependent parameter, W is the device width. In equation 4.5 W is used rather than A_{ov}, as in equation 4.3. The reason for this is the fact that the exact length of the overlap region may not easily be determined, thereby making it difficult do choose an appropriate value for A_{ov}. The exact length of this overlap region does not need to be known, it is incorporated in parameter d_0.

Equation 4.5 is valid for DC voltage levels; since the design of the simulator is focused on application in RF PA's a translation should be made to time-varying stress conditions. For this purpose use is made of a quasi-static description of the oxide breakdown mechanism. In the previous chapter it was experimentally demonstrated that the use of such a quasi-static model is allowed for obtaining a worst-case prediction.

In order to determine t_{63} for a time-varying voltage signal, one must realize that it is the trap generation rate that determines t_{63}. In fact t_{63} describes the moment in time at which enough traps have been formed such that, on the basis of the percolation model, 63 % of all devices have encountered the first breakdown event. This trap generation rate has the voltage dependency as described in expression 4.5. If the trap generation rate under RF stress can be assumed to be a sum of quasi-static DC stress signals, the following expression can be derived:

$$\psi_{\mathrm{BD}}(t) = \frac{W^{1/\beta}}{d_0} \int_0^t |V_{\mathrm{GX}}|^m d\tau \tag{4.6}$$

In this expression the impact of hot-carriers on t_{63}, as observed in [79, 80] is not included. For the circuits evaluated in this chapter this is not required, because in these circuits no inversion channel exists at the peak stress conditions in terms of gate-drain voltage. Only at low drain voltage signals an inversion layer is present. The enhancement of gate-oxide breakdown due to hot carriers at these conditions is not expected to impact t_{63} as it follows from the high drain voltage levels.

In equation 4.6 parameter $\psi_{\mathrm{BD}}(t)$ is introduced, which is defined as:

$$\psi_{\mathrm{BD}}(t) = \frac{t}{t_{63}} \tag{4.7}$$

In section 4.3 it will be explained how this parameter is used in the multiple breakdown simulator described in this chapter.

4.2.2 Hot carrier degradation

Model

In reliability simulators the impact of hot carrier degradation is typically implemented by means of a change in the threshold voltage in combination with a model describing carrier mobility reduction in the channel. These models have been developed for use in DC and low frequency circuits. In recent years the applicability of these models for describing RF circuit performance degradation after hot carrier stress has been investigated [81, 82, 83, 84, 85, 86, 87, 68, 88]. It

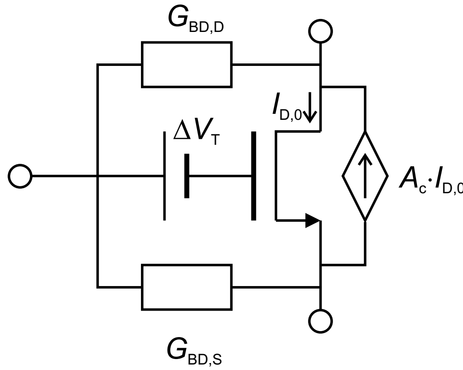

Figure 4.2: Degraded MOSFET model as it is used in the simulator for describing both oxide breakdown and hot carrier degradation. Conductances $G_{\mathrm{BD,D}}$ and $G_{\mathrm{BD,S}}$ model oxide breakdown at the gate-drain respectively gate-source overlap region. Voltage source ΔV_{T} models the threshold voltage shift and the CCCS with amplification factor A_{c} takes into account the effect of degraded carrier mobility.

was found that, in order to describe RF performance degradation also the degradation in gate-drain and gate-source capacitances C_{GD} and C_{GS} should be taken into account. Remarkably in [84, 86, 88] no change in these capacitances could be observed after hot carrier stress. Therefore, in order to reduce the complexity of the first-order model used in this chapter, it is assumed that hot carrier degradation in RF circuits may be modeled by means of a shift in V_{T} in combination with carrier mobility degradation. The hot carrier model is included to demonstrate the possibility of including an additional degradation mechanism with a continuous degradation rate, in combination with the instantaneous nature of oxide breakdown.

Often reliability simulators model the effects of hot carrier degradation by changing the compact parameters of the device. In the simulator discussed in this chapter a different approach is used that is more in line with the model describing oxide breakdown as in figure 4.1. Both V_{T} degradation and mobility degradation are modeled using external elements added to an unstressed MOSFET model as in figure 4.2. In this figure the two breakdown paths are included as explained in figure 4.1, this model is extended with a voltage source in series with the gate and a current controlled current source (CCCS) parallel to the channel. The voltage source is used to model a V_{T} shift and the CCCS takes into account carrier mobility degradation in the channel. In this chapter all hot carrier degradation is assumed to be originated from interface states generation. If the number of interface states generated under hot carrier stress is denoted as ΔN_{it}, the degradation in threshold

voltage, ΔV_{T} can be found to be [73]:

$$\Delta V_{\mathrm{T}} = \frac{1}{\eta_{\mathrm{VT}}} \cdot \frac{q \Delta N_{\mathrm{it}}}{A_{\mathrm{G}} C_{\mathrm{ox}}} \tag{4.8}$$

ΔV_{T} is the degradation of V_{T}, η_{VT} is a technology dependent parameter, q is the elementary charge and C_{ox} is the oxide capacitance. Furthermore A_{G} is the total gate area. Expression 4.8 can be rewritten as:

$$\Delta V_{\mathrm{T}} = C_{\mathrm{VT}} \cdot \Delta N_{\mathrm{it}}_W(t) \tag{4.9}$$

In this expression parameter ΔN_{it}_W was introduced, representing the total number of generated interface states per unit width. C_{VT} is a technology and geometry dependent parameter.

The degradation in channel mobility can be related to ΔN_{it} through [73]:

$$\frac{1}{\mu_{\mathrm{i}}} = \frac{1}{\mu_0} \cdot \left(1 + \frac{K_\mu \Delta N_{\mathrm{it}}}{A_{\mathrm{G}}}\right) \tag{4.10}$$

In this expression μ_0 is the channel mobility without the extra generated interface states and μ_{i} is the channel mobility where the effect of the extra generated interface states is included. K_μ is a technology dependent parameter. Similar to expression 4.9, expression 4.10 can be rewritten in terms of ΔN_{it}_W:

$$\frac{1}{\mu_{\mathrm{i}}} = \frac{1}{\mu_0} \cdot \left(1 + C_\mu \Delta N_{\mathrm{it}}_W\right) \tag{4.11}$$

Here C_μ is a technology and geometry dependent parameter. In the simulator described in this chapter, the compact parameters of the MOSFETs are not altered; degradation is modeled using the additional parasitic elements, as illustrated in figure 4.2. The CCCS in this model has a current amplification factor of A_{c}. It can be used to model mobility degradation as the total current flowing between source and drain with the CCCS included equals:

$$I_{\mathrm{D,T}} = I_{\mathrm{D,0}} \left(1 - A_{\mathrm{c}}\right) \tag{4.12}$$

In this expression $I_{\mathrm{D,T}}$ is the total drain current including the effect of degraded channel mobility and $I_{\mathrm{D,0}}$ is the drain current without the effect of the degraded channel mobility. $I_{\mathrm{D,0}}$ is the drain current as it results from the unstressed compact model of the MOSFET. Since the drain current is linearly proportional to the carrier mobility in the channel, and using expressions 4.11 and 4.12, A_{c} can be found to be:

$$A_{\mathrm{c}} = \frac{K_{\mathrm{A}} N_{\mathrm{it}}_W}{1 + K_{\mathrm{A}} N_{\mathrm{it}}_W} \tag{4.13}$$

This is the expression used for incorporating degraded carrier mobility in the channel after hot carrier stress using a current CCCS. If the amount of interface states generated by hot carrier stress is known, the model in figure 4.2, in combination with expressions 4.9 and 4.13, can be used for incorporating its effect on device performance.

Degradation rate

Expressions 4.9 and 4.13 make use of the number of generated interface states per unit width, N_{it}_W. In order to determine the appropriate values for ΔV_{T} and A_{c} at time t, the generation rate of interface states needs to be known. The reliability simulator discussed in this chapter makes use of the commonly used lucky electron model for this purpose [89]. (This model may not be fully accurate for MOSFETs with a channel length < 0.25 μm [34]). This lucky electron model is expressed as:

$$\frac{\tau_{\text{HC}} I_{\text{D}}}{W} \propto \left[\frac{I_{\text{sub}}}{I_{\text{D}}}\right]^{-\varphi_{\text{it}}/\varphi_{\text{i}}} \tag{4.14}$$

In this expression τ_{HC} is the hot carrier lifetime of a MOSFET. It can be defined as a given degradation of any device parameter, such as a 50 mV shift in V_{T}. W is the device width, I_{D} is the drain current and I_{sub} is the substrate current. φ_{it} represents the critical electron energy for generating an interface trap. In [89] it was obtained to be 3.7 eV. Furthermore φ_{i} is the minimum energy that an electron must have in order to create an impact ionization In [89] it was found to be 1.3 eV.

The basis of equation 4.14 is the the substrate current originates from avalanche multiplication of hot carriers. If the number of carriers with an energy sufficiently high for generating a given I_{sub} is known, the number of carriers with an energy sufficiently high for creating an interface state can also be found. Interface state generation is typically assumed to be the dominant degradation mechanism in hot carrier degradation. Especially in technologies with thin oxides, this assumption will hold; the effect of electron and hole trapping in the oxide can be assumed to be negligible.

Rather than finding an expression for device lifetime, a reliability simulator addresses circuit lifetime. For this to be available an expression is needed for device degradation as a function of time. In this chapter this is implemented using the widely used Takeda model [90]:

$$\Delta Par\,(t) = A \cdot t^{\text{n}} \tag{4.15}$$

In this expression $\Delta Par(t)$ is the degradation of a given device parameter at time t. n is a technology dependent parameter, it typically has a value around 0.4. Parameter A is dependent on both technology and the applied stress conditions. The link between the degradation as a function of time as in the Takeda model and the degradation rate as obtained from the lucky electron model for periodic time-varying stress conditions can be obtained using the AGE parameter [71]. This AGE parameter is defined as:

$$AGE(t) = \frac{1}{HW} \cdot \int_0^t I_{\text{D}}(\tau) \left[\frac{I_{\text{sub}}(\tau)}{I_{\text{D}}(\tau)}\right]^{-\varphi_{\text{it}}/\varphi_{\text{i}}} d\tau \tag{4.16}$$

In this expression H is a technology dependent parameter which is different for every device parameter monitored, W is the device width. For expression 4.16

to hold for RF circuits, it should be verified that such a quasi-static description holds for RF circuits. In the previous chapter it was experimentally shown that quasi-static models can indeed be used for modeling hot carrier degradation in RF circuits. For convenience a slightly different parameter, AGE_{r} is used in this chapter, defined as:

$$AGE_r(t) = \int_0^t I_{\mathrm{D}}(\tau) \left[\frac{I_{\mathrm{sub}}(\tau)}{I_{\mathrm{D}}(\tau)}\right]^{-\varphi_{\mathrm{it}}/\varphi_{\mathrm{i}}} d\tau \tag{4.17}$$

Now $\Delta N_{\mathrm{it}_}W(t)$ can be found to be:

$$\Delta N_{\mathrm{it}_}W\,(t) = \left(\frac{AGE_r(t)}{H_{\mathrm{N_{it}_W}}W}\right)^{\mathrm{n}} \tag{4.18}$$

Here $H_{\mathrm{N_{it}_W}}$ is a technology dependent parameter. Having a value for $AGE_{\mathrm{r}}(t)$ for given stress conditions, model parameters $\Delta V_{\mathrm{T}}(t)$ and $A_{\mathrm{c}}(t)$ can be found to be:

$$\Delta V_{\mathrm{T}}\,(t) = \left(\frac{AGE_r(t)}{H_{V_{\mathrm{T}}}W}\right)^{\mathrm{n}} \tag{4.19}$$

$$A_{\mathrm{c}}(t) = \frac{AGE_{\mathrm{r}}(t)^n H_\mu^{-n} W^{-n}}{1 + AGE_{\mathrm{r}}(t)^n H_\mu^{-n} W^{-n}} \tag{4.20}$$

In these expression $H_{V_{\mathrm{T}}}$ and H_μ are technology dependent parameters for describing V_{T} shifts respectively carrier mobility degradation using the AGE_{r} parameter. These expressions are used for describing hot carrier degradation as a function of time in the simulator discussed in this chapter.

4.3 Design of the simulator

Implementing the occurrence of multiple breakdown events in a reliability simulator requires a different simulation flow than for hot carrier and NBTI degradation. For given stress conditions, the probability of encountering breakdown may be very accurately described. However, as voltage signals in the circuit may change after a breakdown event has occurred, it is not allowed to describe the probability of inducing more than one breakdown event on the basis of simulated voltage signals of an unstressed circuit. Furthermore, it is not known beforehand how many breakdown events the circuit may tolerate before circuit failure takes place. As breakdown events can take place at various locations in the circuit, it is not possible to determine a fixed number of breakdown events at which circuit failure is encountered. This is further complicated by other degradation mechanisms such as hot carrier degradation. The exact number of breakdown events before circuit failure may vary between different circuits and between different samples of the same circuit.

In this section a new simulation approach will be presented where both oxide breakdown and hot carrier degradation effects are simultaneously taken into

account for circuits that can withstand multiple breakdown events. Use will be made of a Monte-Carlo approach, in which a large number of simulations can be performed in order to assess circuit-level failure probability.

4.3.1 Probability function for multiple breakdown events

As oxide breakdown cannot be described using time-continuous degradation rates as is possible for hot carrier degradation and NBTI a different approach is needed. Oxide breakdown is a discrete event of stochastic nature; a Monte-Carlo approach is suitable for analyzing such a kind of process. For this to be realized first the correct expression for the failure probability should be derived.

The probability of inducing breakdown events in the gate-oxide of a MOS structure is often described using Weibull statistics, as in expression 4.2; the dependency of t_{63} on stress conditions can be found using expressions 4.6 and 4.7. For expressing the probability of having more than 1 breakdown event in the gate-oxide, use can be made of Poisson statistics [91]. If this is done the cumulative probability of having induced at least k breakdown events is given by:

$$F_k(\nu) = 1 - \sum_{i=0}^{k-1} P_i(\nu) \tag{4.21}$$

In this expression $P_i(\nu)$ represents the probability of having induced exactly i breakdown events. It is given by:

$$P_i(\nu) = \frac{e^{-\nu}\nu^i}{i!} \tag{4.22}$$

Here parameter ν is related to time t through:

$$\nu = \psi_{\mathrm{BD}}^{\beta}(t) \tag{4.23}$$

where $\psi_{\mathrm{BD}}(t)$ is given by expression 4.7. One can easily derive that the probability function of 4.21 reduces to the Weibull distribution of 4.2 for the first breakdown event, i.e. $k = 1$. Expression 4.21 provides information the probability of inducing at least k breakdown events, if the stress conditions are constant. However, under actual circuit conditions, these stress conditions may change after a breakdown event has occurred. Therefore it is important to know the failure probability of inducing an additional breakdown event in a device. This is an important feature of the Monte-Carlo approach of the simulator discussed in this chapter. The simulator keeps track on the number of breakdown events that have already occurred for every possible breakdown location. Using Bayes' theorem the probability of inducing an additional breakdown event can be found for a device that has already encountered k breakdown events:

$$F_{\mathrm{k+1|k}}(\nu) = \frac{F_{\mathrm{k+1}}(\nu)}{F_{\mathrm{k}}(\nu)} \tag{4.24}$$

This expression is used in the simulator for finding the probability of inducing an additional breakdown event. In combination with a random generator and expressions 4.6, 4.7 and 4.23 this expression can be used to mimic oxide breakdown events as in an actual circuit.

4.3.2 Implementation

The simulator described in this chapter is implemented as an OCEAN script that can be run within the Cadence simulation environment. The simulator calls the SPECTRE simulator for obtaining circuit performance and time-domain stress signals. As the simulator is designed for the reliability evaluation of circuits with periodic signals, this time-domain information can be obtained from a periodic steady state (PSS) analysis. The use of a PSS analysis results in considerably shorter simulation times with respect to a transient analysis. Furthermore as with other reliability simulators, a translation should be made between the timescale of the periodicity of the signals, ns range, and the timescale at which circuit lifetime is expected to be, i.e. years. This can be done by acknowledging the fact that the amount of degradation at time t can be expressed as a function of the amount of degradation per period of the signal. The degradation per period can be obtained directly from the PSS analysis.

In order to assess circuit lifetime use is made of the simulation flow shown in figure 4.3. The simulation starts with a SPECTRE PSS simulation combined with a procedure that determines values of ν_{BD} for every possible breakdown location in the circuit, i.e. every gate-drain and gate-source overlap region. These ν_{BD} values are determined by first generating a random number; ν_{BD} is then determined to be that value of ν that results in $F_{\mathrm{k+1|k}}$ to be equal to the random number. For the unstressed circuit this results in F_1, the Weibull failure distribution. From the PSS analysis a value can be found for $\psi_{\mathrm{BD}}(T)$ using expression 4.6 for every possible breakdown location and T representing the periodic time of the stress signal. Using the values for $\psi_{\mathrm{BD}}(T)$, T and ν_{BD} a value for t_{BD} can be found for every possible breakdown location; t_{BD} is the time at which the next breakdown event will occur at the specified location.

After determining t_{BD} for every possible breakdown location, $t_{\mathrm{BD,min}}$, the time at which the first upcoming breakdown event in the entire circuit is determined. In order to investigate the significance of hot carrier degradation at $t_{\mathrm{BD,min}}$, the hot carrier parameters in the degraded MOSFET model of figure 4.2 are updated. To do this use is made of the time-domain information as obtained from the PSS analysis combined with expressions 4.19 and 4.20. With these updated hot carrier parameters a circuit simulation is performed. Circuit failure can be defined as the degradation in any circuit parameter. In section 4.4, discussing simulation results, the output power of a PA circuit is used for this purpose. If circuit failure occurs after the update of these parameters, hot carrier degradation is known to be the dominating degradation mechanism for this circuit. A separate routine is then called that finds the time for circuit failure, t_{fail}, making use of the continuous nature of hot carrier degradation.

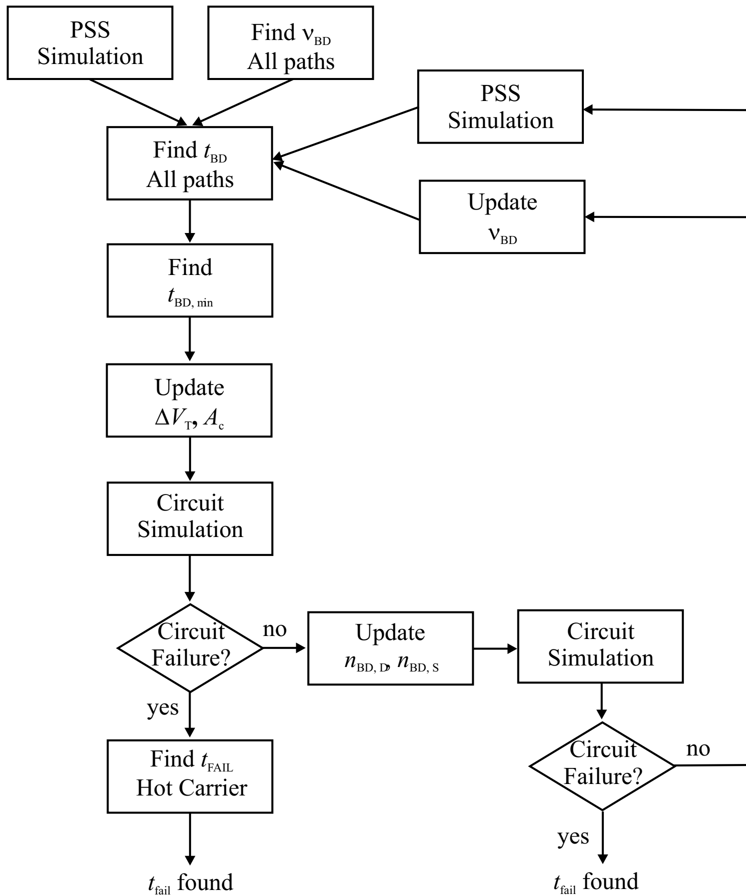

Figure 4.3: Simulation flow for the reliability simulator described in this chapter.

If the circuit remains functional at time $t_{\mathrm{BD,min}}$, after the update of the hot carrier parameters, a breakdown path is added at the location that had the lowest t_{BD} (the appropriate parameter $n_{\mathrm{BD,D}}$ or $n_{\mathrm{BD,S}}$ is increased with 1). After inserting this breakdown path again the circuit is tested for failure. If the circuit has failed the simulator stops and a value for t_{fail} is found. If the circuit remains functional a new PSS analysis with the updated hot carrier parameters and added breakdown path is performed. Also a new value for ν_{BD} is found for this breakdown location. Then new values for t_{BD} are determined for every breakdown location and the routine starts over again. This routine will stop when a value for t_{fail} has been found.

The simulation flow of figure 4.3 determines t_{fail} for one single circuit. As the simulator makes use of randomly generated numbers a single simulation result

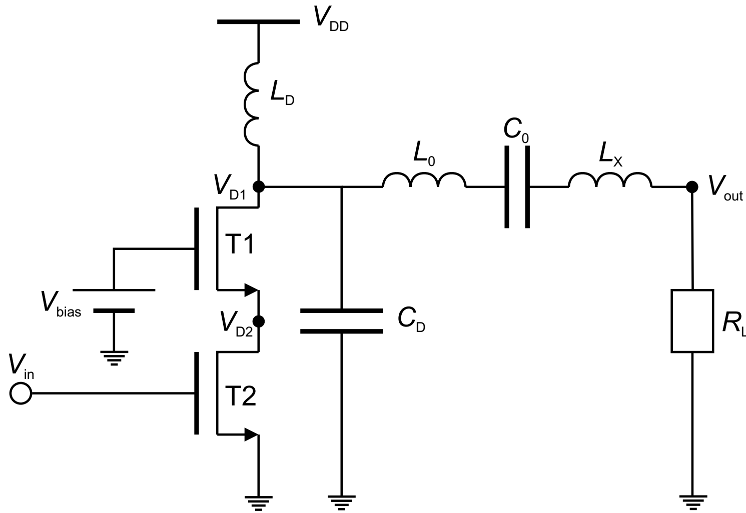

Figure 4.4: Schematic of the PA circuit design, used for evaluating the simulator.

is not suitable for describing circuit failure as a function of time. This can be achieved by performing this simulation cycle for a large number of times on the same circuit. In this way the circuit-level failure probability can be assessed as a function of time for circuits that can withstand multiple breakdown events and simultaneously suffer from hot carrier degradation.

4.4 Simulation results

4.4.1 Description of the evaluated circuits

The simulator of this chapter was evaluated using three different RF PA's, all using the circuit design of figure 4.4. The PA's consist of two nMOSFETs placed in a cascode structure. The different PA's were designed to operate in various class-E_{vv} mode as described in [92]. The operating frequency for which the circuits were designed was 900 MHz. Designs were made in a 90 nm CMOS process, the nominal supply voltage of this process is 1.2 V. V_{bias} was set to 1.2 V; V_{DD} was used as one of the design parameters of the three different PA's. V_{in} is generated by a pre-driving circuit that periodically switches between 0 and 1.2 V, thereby periodically switching the two nMOSFETs on and off. This is typical for class-E and class-E_{vv} operation. The passives are chosen such that a sinusoidal output voltage is delivered to the load R_L.

The three designs have the same specifications in terms of output power delivered to a 25 Ω load. The output power was chosen to be 25 mW. The circuits were designed with completely identical parameters of the two nMOSFETs, only the passive components L_D, C_D, L_0, C_0 and L_X were varied as well as V_{DD}. In

class-E_{vv} design a parameter α is used to vary between the different operation modes. When $\alpha = 0$, the circuit operates in class-E mode, at increasing values of α circuit operation deviates further away from class-E operation. Parameter α allows for tuning the voltage signals in a power amplifier in such a way that a trade-off can be made between initial circuit performance and circuit lifetime.

In figures 4.5 through 4.7 the performance in terms of voltage and current signals in the circuit is shown in an unstressed circuit for three different values of α: $\alpha = 0$, $\alpha = 1$, and $\alpha = 2$. The voltage signals in figure 4.5 reveal the comparable performance of the three different circuits in terms of output power. Finding the optimum operating mode of this PA is a trade-off between various circuit parameters. For example the Power Added Efficiency (PAE) for an unstressed device with $\alpha = 0$ is 93 %, for $\alpha = 1$ this is 88 % and for $\alpha = 2$ it is 67 %. PAE is defined as:

$$PAE = \frac{P_{\text{out,RF}} - P_{\text{in,RF}}}{P_{\text{DC}}} \cdot 100\% \tag{4.25}$$

Here $P_{\text{out,RF}}$ is the RF output power, delivered to R_{L}, $P_{\text{in,RF}}$ is the RF input power delivered by V_{in} and P_{DC} is the total DC power as it is provided to the circuit by V_{DD} and V_{bias}. PAE is a measure for the efficiency of transforming the available DC power into RF output power.

If one would only be interested in the PAE, the optimal design would be the class-E_{vv} with $\alpha = 0$. However, as can already be expected from the levels of stress induced by the voltage and current signals shown in figures 4.5 through 4.7, specification in terms of circuit lifetime, are significantly different between the three different designs, as will be shown below.

4.4.2 Stress conditions

Using the simulated voltage and current signals of figures 4.5 through 4.7 the stress monitoring parameters $\psi_{\text{BD}}(t)$ and $AGE_{\text{r}}(t)$, as defined in section 4.2, can be derived for every possible breakdown location respectively MOSFET in the circuit. In figure 4.8 the increase of parameter ψ_{BD}, $\Delta\psi_{\text{BD}}$ during one cycle of the periodic signal is plotted for every possible breakdown location and for all three values of α evaluated. Furthermore in figure 4.9 the increase of AGE_{r}, ΔAGE_{r} is plotted during one cycle of the periodic signal. This is done for both nMOSFETs of the PA and for all three values of α.

Figure 4.8 reveals that in terms of oxide breakdown, the voltage across the gate-drain overlap region for the T1 MOSFET results in the highest level of stress for all three values of α. Furthermore this figure also clearly reveals that in the $\alpha = 0$ mode the first oxide breakdown event is expected much earlier in time with respect to the other two modes for all possible breakdown locations. After 1 period of the stress signal ψ_{BD} has reached a value of $4.88 \cdot 10^{-14}$ for the $\alpha = 0$ mode in MOSFET T1. Using expression 4.7 this coincides with a t_{63} of $2.2 \cdot 10^{4}$ s for the first breakdown event to occur at this gate-drain overlap region. For the $\alpha = 1$ mode, ψ_{BD} of the gate-drain overlap region of the T1 MOSFET has a value

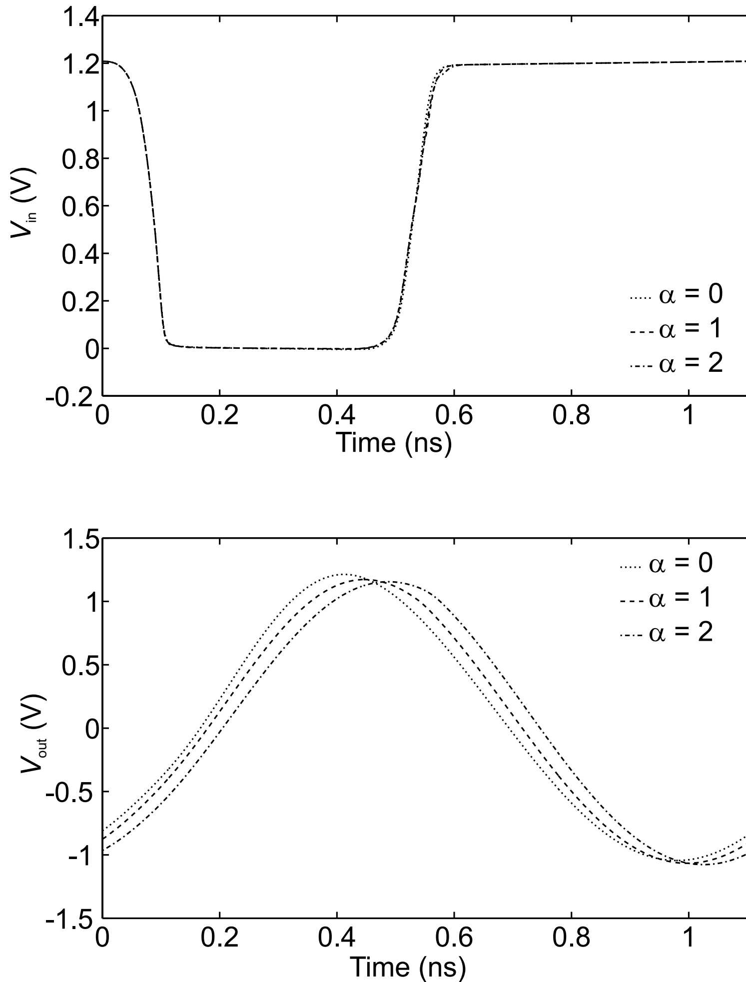

Figure 4.5: Simulated input and output voltage of the class-E_{vv} PA's in three different operation modes. V_{in} is generated by a pre-driver circuit and design parameters were chosen such that the output power is identical in all three operation modes for a 25 Ω load.

of $1.47 \cdot 10^{-22}$ after one period. This coincides with a value for t_{63} for the first breakdown event of $27.6 \cdot 10^{12}$ s, or $2.4 \cdot 10^{5}$ year. From this huge difference it is to be expected that the $\alpha = 1$ mode has a performance in terms of reliability exceeding that of the $\alpha = 0$ mode by far.

AGE_r is a measure for the hot carrier degradation rate of the individual MOSFETs of a circuit. Figure 4.9 reveals that AGE_r is the highest in the $\alpha = 2$ mode, for both MOSFETs in the circuit. Hence, by looking at both figures 4.8 and 4.9 it

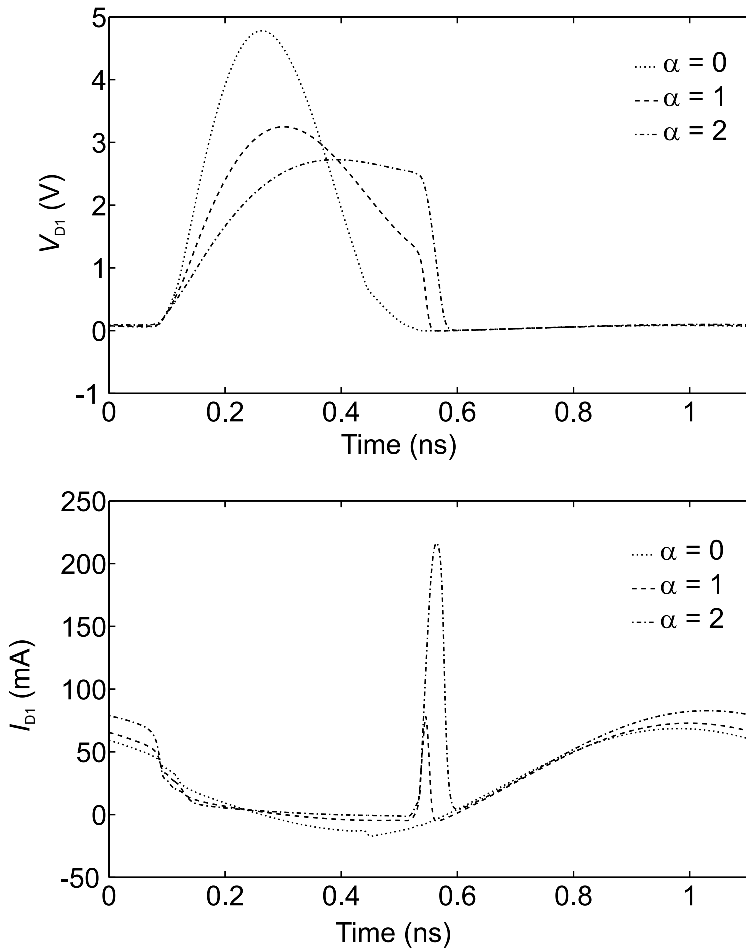

Figure 4.6: Simulated drain voltage and drain current signals for the upper nMOSFET in the circuit of figure 4.4, in three different operation modes. The $\alpha = 0$ mode coincides with class-E operation.

can be concluded that the $\alpha = 0$ mode has the worst properties in terms of oxide breakdown probability but in terms of hot carrier degradation the stress levels is less severe than with the other two modes. For the $\alpha = 2$ mode hot carrier degradation rate is the worst, while this mode is the least susceptible to gate-oxide breakdown. Both in terms of oxide breakdown and hot carrier degradation the reliability performance of the $\alpha = 1$ mode is in between the other two modes.

Now that the individual degradation rates for unstressed devices are known for the three different operation modes, the next step is to investigate the implications these degradation rates have on circuit performance. In order to signify the

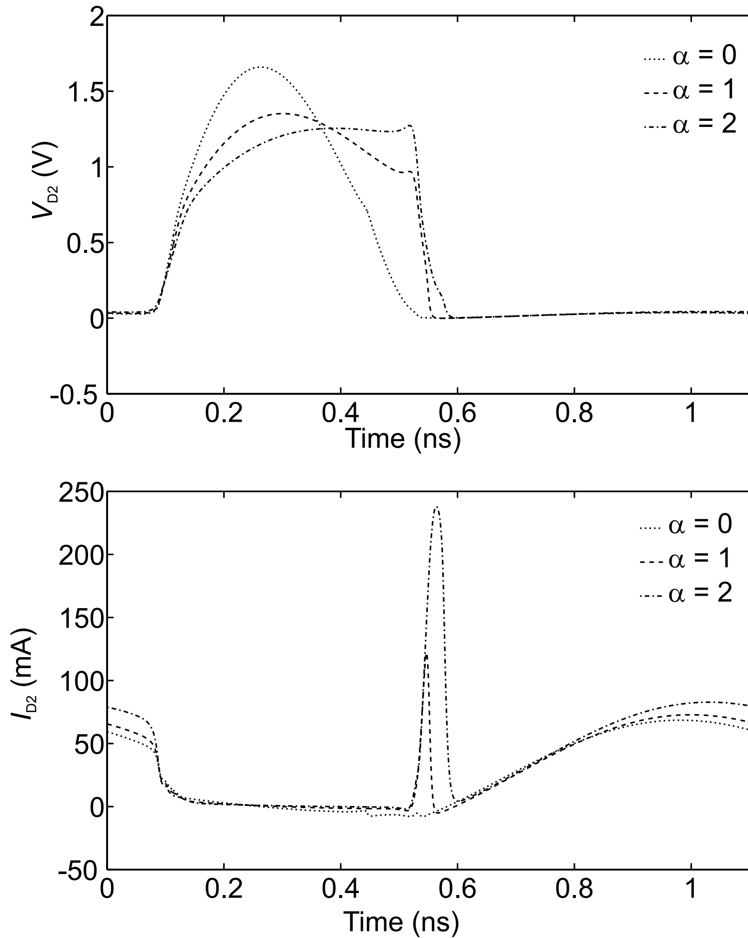

Figure 4.7: Simulated drain voltage and drain current signals for the lower nMOSFET in the circuit of figure 4.4, in the three different operation modes as used in this chapter.

importance of repeatedly determining the stress factors as in the simulation flow of figure 4.3 first the change of $\psi_{\rm BD}$ is evaluated after multiple breakdown events. As an example this is done for the gate-drain overlap region in the T1 MOSFET for the $\alpha = 0$ mode. For this overlap region, $\psi_{\rm BD}$ is directly related to $V_{\rm D1}(t)$. In figure 4.10 the degradation in $V_{\rm D1}(t)$ is shown after up to four breakdown events.

Figure 4.10 shows that the peak voltage of $V_{\rm D1}$ during one cycle of the periodic signal decreases. One consequence of this effect is that circuit performance may degrade. However, as long as circuit performance stays within specifications, the

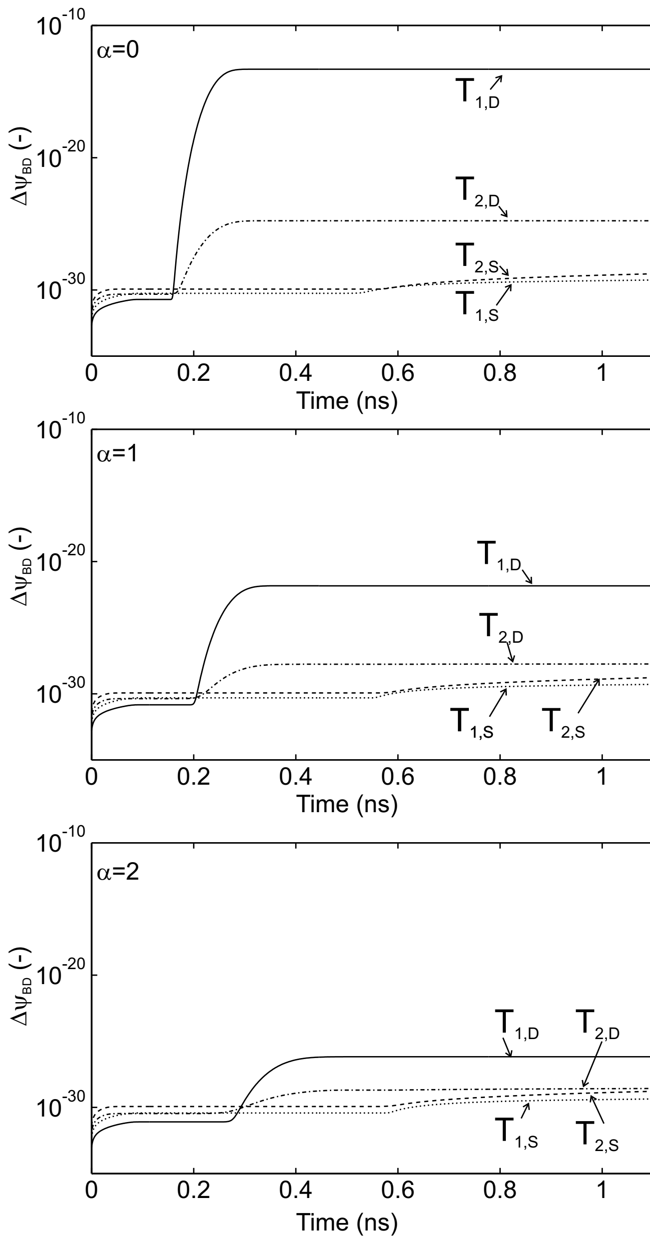

Figure 4.8: Increase of degradation parameter ψ_{BD} during one period of the periodic signal for all possible breakdown locations and all values of α.

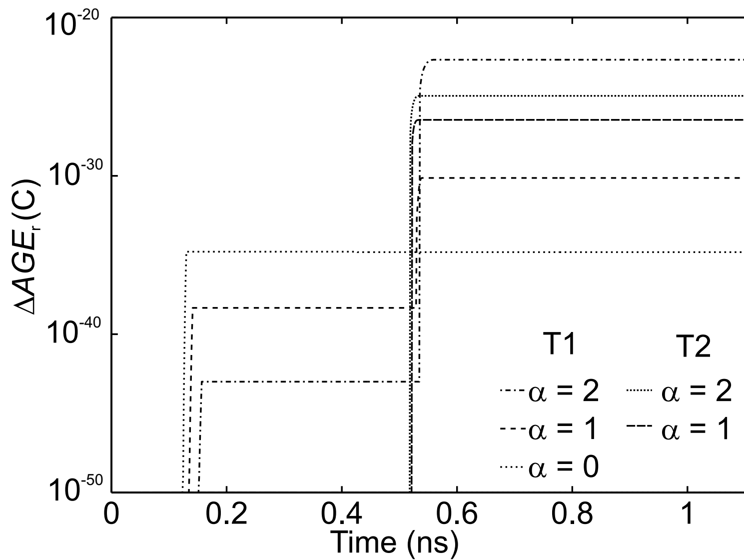

Figure 4.9: Increase of degradation parameter AGE_{r} during one period of the periodic signal for both MOSFETs and all values of α. The AGE_{r} parameter for MOSFET T2 in the $\alpha = 0$ mode appears to be so small, that it cannot be properly plotted on a the logarithmic scale of this figure. This indicates that hot carrier degradation for this MOSFET in this mode is negligible.

decrease in V_{D1} also has an important benefit for the reliability of the circuit: the stress factor ψ_{BD} per period decreases. This can be seen in figure 4.11. In this figure the increase of the ψ_{BD} parameter per period for the T1 MOSFET in the $\alpha = 0$ mode is plotted after degradation of the device has set in. $\Delta\psi_{\mathrm{BD}}$ is plotted for an unstressed circuit as well as after up to 4 breakdown events in the gate-drain overlap region of the T1 MOSFET. Clearly $\Delta\psi_{\mathrm{BD}}$ decreases with an increasing number of breakdown events. This is the result of a reduction in the voltage V_{D1}. This reduction in $\Delta\psi_{\mathrm{BD}}$ signifies the importance of determining the stress factors ψ_{BD} and AGE_{r} after every breakdown event in a circuit has occurred, as in the simulation flow of figure 4.3. If this is not done in the example of the $\alpha = 0$ mode, this would lead to a severe underestimation of circuit lifetime. In other circuits ψ_{BD} could possibly increase after a breakdown event has occurred; in this case this would lead to an overestimation of circuit lifetime.

4.4.3 Circuit lifetime prediction

With the stress factors $\psi_{\mathrm{BD}}(t)$ and $AGE_{\mathrm{r}}(t)$ known for all individual MOSFETs of the PA, it is now possible to determine circuit lifetime making use of the simulation flow of figure 4.3. To start off with, in figures 4.12 and 4.13 circuit

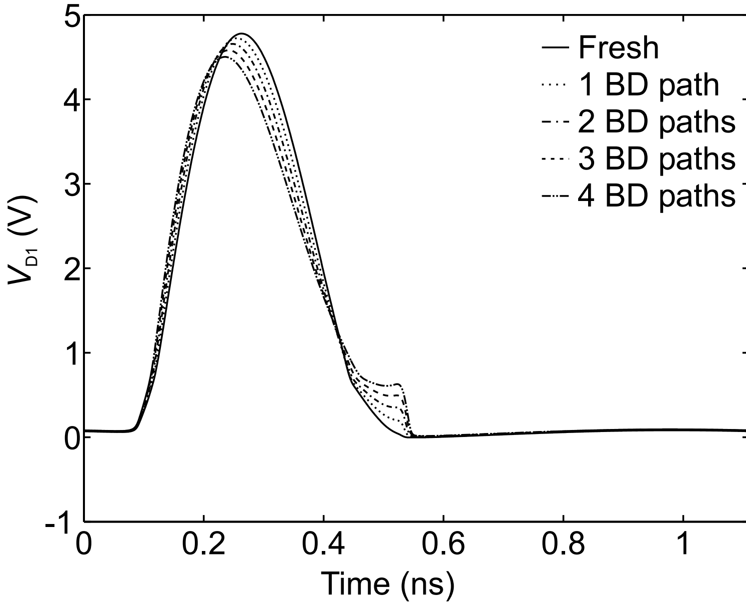

Figure 4.10: Degradation of $V_{\text{D1}}(t)$ after the occurrence of up to 4 breakdown events for the $\alpha = 0$ mode.

performance in terms of P_{out} and PAE is plotted as a function of stress time for the three different operation modes. This figure reveals a number of very interesting features. First of all one can observe that both P_{out} and PAE degrade rapidly in time for the $\alpha = 0$ mode, compared to $\alpha = 1$ and $\alpha = 2$ modes. A second feature that can be observed is the difference between the relative importance of oxide breakdown effects compared to hot carrier degradation. Oxide breakdown can be detected as a sudden drop in P_{out} or PAE, this can be seen for both the $\alpha = 0$ and $\alpha = 1$ mode. For both these operation modes three different simulations result in three different plots. In the $\alpha = 2$ mode the dominating degradation mechanism is hot carrier degradation for both P_{out} and PAE. Therefore circuit parameter degradation as function of time can be observed to be a continuously degrading effect for the $\alpha = 2$ mode. Different simulations result in exactly the same degradation of circuit performance in this mode.

For the $\alpha = 1$ mode, the effect of both oxide breakdown and hot carrier degradation can be observed; both discrete drops in performance as well as a continuous degradation between these discrete events can be seen. Finally the figures also show that the impact of an individual breakdown event on circuit performance varies between the different operation modes. For the $\alpha = 0$ mode P_{out} has moved below 22.5 mW after 5 breakdown events for all three simulations. In the $\alpha = 1$ mode, P_{out} does not drop below 22.5 mW after 8 breakdown events. This means that not only do the breakdown events occur earlier in time for the

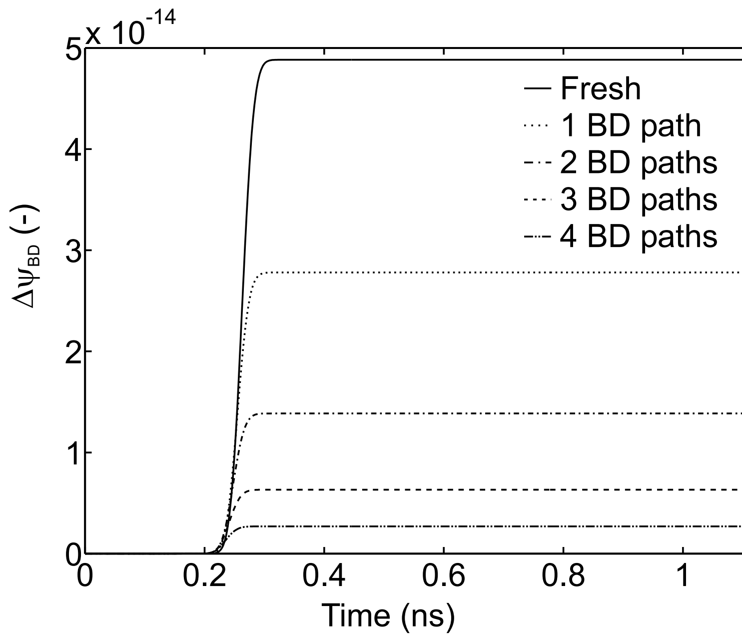

Figure 4.11: Degradation of the increase in ψ_{BD} parameter per period for MOSFET T1 in the $\alpha = 0$ mode for an unstressed circuit and for circuits with up to 4 breakdown events present in the gate-drain overlap.

$\alpha = 0$ mode, they also affect circuit performance more considerably.

As the oxide breakdown events are of stochastic nature it is desirable to have information on the probability of circuit failure as a function of time. This is obtained using the simulation flow of figure 4.3 for the three different operation modes. The result is shown in figure 4.14. The criterion for defining circuit lifetime used for this purpose was a 10 % decrease in output power, meaning that P_{out} has decreased below 22.5 mW. For the $\alpha =$ mode, hot carrier degradation is the dominating mechanism, therefore the time for circuit failure has the same value for every circuit. Again in figure 4.14 the high circuit lifetime for the $\alpha = 2$ and $\alpha = 1$ modes with respect to the $\alpha = 0$ mode can be observed. For the $\alpha = 0$ mode already after 0.09 years 63 % of all circuits has failed. Hence this circuit suffers from too high levels of stress for use in an actual commercial product. In the $\alpha = 2$ mode circuit lifetime has been reached after $1.8{\cdot}10^5$ years. The $\alpha = 1$ mode has the highest circuit lifetime: only after $4.3{\cdot}10^7$ years has 63% of all circuits failed. In contrast to the $\alpha = 0$ mode, this design is useful in an actual product based on its reliability specifications.

Besides the circuit lifetime, the simulator can also be used to obtain additional valuable information, such as the number of breakdown events before circuit failure and the location where most breakdown events take place. In the $\alpha = 2$

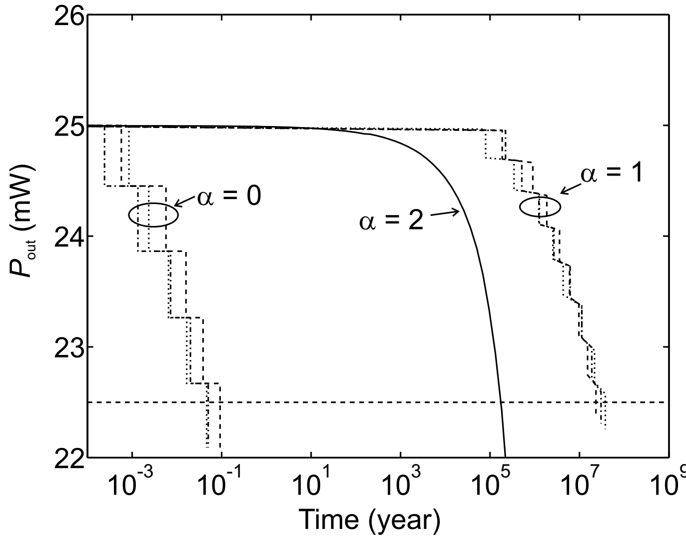

Figure 4.12: Output power plotted against stress time for the three different operation modes of the class-E_{vv} PA used in this chapter. For both the $\alpha = 0$ mode and the $\alpha = 1$ mode three different simulation results are shown, resulting in different performance in time. For the $\alpha = 2$ mode hot carrier degradation dominates, resulting in identical simulation results after different simulations. The dashed horizontal line shows the 22.5 mW criterion used for defining circuit failure in this chapter.

mode it appears that no breakdown event takes place before the circuit has already failed due to hot carrier degradation. For both the $\alpha = 0$ and the $\alpha = 1$ mode all breakdown events take place at the gate-drain overlap region of the T1 MOSFET. For the $\alpha = 0$ mode it appears that already 5 breakdown events imply circuit failure and in the $\alpha = 1$ mode this is 8 or 9.

The final property of the simulator that is evaluated is the fact that the simulator allows for a simultaneous simulation of breakdown effects and hot carrier degradation. The merits of this feature are made clear in figure 4.15. In this figure the circuit failure probability is plotted against time using simulation results that include both hot carrier and breakdown effects and after a simulation in which hot carrier degradation is not included. This is done for the $\alpha = 1$ mode. From this figure it is clear that if hot carrier degradation is not taken into account, a severe overestimation of circuit lifetime may be obtained. Furthermore if hot carrier degradation is not taken into account, the circuit can withstand 10 breakdown events, rather than 7 or 8, as is the case with hot carrier effects included.

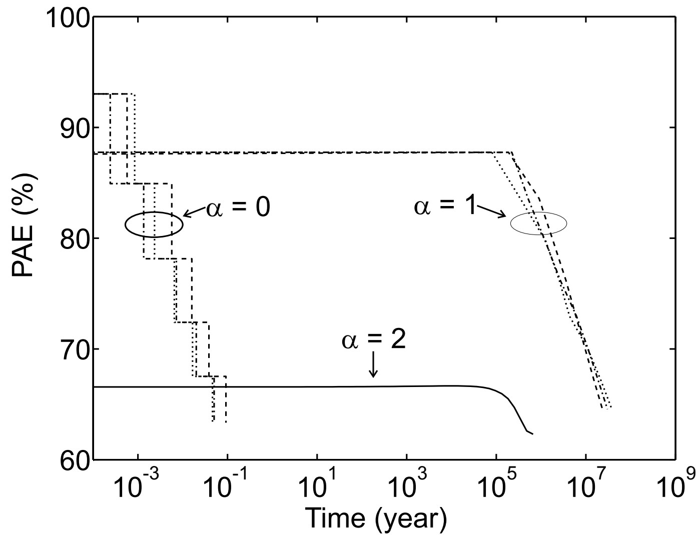

Figure 4.13: Power added efficiency plotted against stress time for the three different operation modes of the class-E_{vv} PA used in this chapter.

4.5 Discussion

The simulation results shown in this chapter show the merits of the newly developed simulator. Three operation modes of a class-E_{vv} PA have been compared on circuit lifetime performance. The results show that, although the $\alpha = 0$ mode has superior performance in terms of PAE for an unstressed circuit, already after 0.1 year the first breakdown event in the $\alpha = 0$ mode makes the performance in terms of both P_{out} and PAE for the $\alpha = 1$ mode superior.

If one would only consider the first breakdown event in a circuit to imply circuit failure, t_{63} would be $7.0 \cdot 10^{-4}$ year for the $\alpha = 0$ mode and $2.4 \cdot 10^{5}$ year for the $\alpha = 1$ mode. Making use of the simulator these values appear to be 0.09 year respectively $4.3{\cdot}10^{7}$ years. For both modes this signifies an increase in circuit lifetime of a factor of over 100. This means that the simulator can be used to relax design guidelines for RF PA's dramatically, thereby allowing PA's to be designed for higher output powers.

In order to determine the applicability of the simulator for other kinds of circuits besides RF PA's one must realize that the model used in the simulator is a first-order model. One of the important simplifications is the assumption that breakdown paths are only formed in the gate-source and gate-drain overlap regions. For RF PA's this assumption may indeed hold, as drain voltage levels are typically the dominating stress factor in these circuits. For other circuits a model that also allows for breakdown events in the oxide above the channel region may

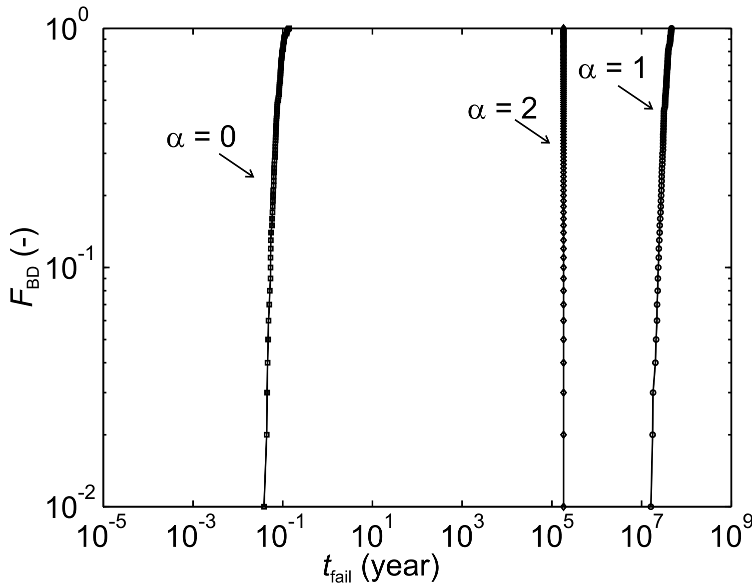

Figure 4.14: Failure probability plots of the class-E_{VV} PA's used in this chapter. Clearly the $\alpha = 1$ mode and $\alpha = 2$ mode have superior performance in circuit lifetime over the $\alpha = 0$ mode.

be required. However, as breakdown events at the overlap regions have the most deteriorating effect on device performance, this may also not be a real bottleneck. This needs to be investigated further. Also it should be verified whether the gate-source and gate-drain capacitances do not degrade under hot carrier stress if the simulator is applied for evaluating circuit lifetime in a given technology.

The simulator at present includes only oxide breakdown and hot carrier effects. BTI effects can be included in a manner similar to hot carrier degradation, provided that an accurate model is available for modeling device degradation under AC and RF BTI degradation. In the previous section it was shown that although the same models may be used for describing RF and AC NBTI degradation, these cannot be easily related to DC degradation models, due to recovery effects.

The simultaneous simulation of various degradation mechanisms is an important aspect of the simulator. It was shown that excluding hot carrier degradation from the lifetime prediction of the $\alpha = 1$ mode would result in an overly optimistic estimate. Also the number of breakdown events that the circuit could withstand without hot carrier degradation included would result in a number higher than with this effect included. This is one of the powerful elements of the simulator described in this chapter.

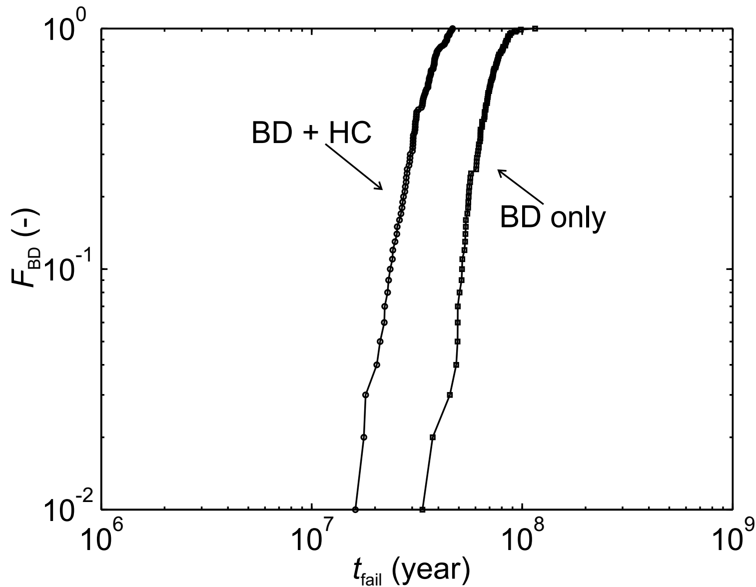

Figure 4.15: Failure probability plots for the $\alpha = 1$ class-E_{vv} PA with hot carrier degradation included (BD + HC) and left out (BD only). The figure clearly shows the importance of simulating hot carrier degradation simultaneously with oxide breakdown events.

4.6 Conclusions

In this chapter a new simulation methodology was presented for assessing circuit-level failure probability of RF PA's that can withstand multiple oxide breakdown events, up to present this has never been shown before. This simulator can be used by circuit designers in order to increase performance of RF circuits as the restriction on the maximum allowable voltage levels can be relaxed for individual circuits.

The simulator combines both multiple gate-oxide breakdown effects and hot carrier degradation. The model used for degraded MOSFET operation is based on well-accepted models for these effects. Adding additional degradation mechanisms such as NBTI could be included in a way similar to what is done for hot carrier degradation, thereby allowing the simulator to be used on more kinds of RF circuits.

The most important property of the new simulator is the fact that it updates circuit performance after both oxide breakdown events and hot carrier degradation. An update after breakdown has never been implemented before into a reliability simulator. In combination with the possibility of simulating hot carrier degradation the new simulator is very useful for an accurate prediction of RF PA

lifetime. The applicability of the simulator could be extended to other RF circuits by careful modeling of breakdown paths outside the overlap areas in combination with a good model for BTI degradation.

The simulator has been put under test by comparing a class-E_{vv} PA in three different operation modes. The results reveal that an increase in predicted circuit lifetime of a factor of over 100 can be achieved by making use of the simulator rather than looking at the first breakdown event only. This allows circuit designers to design their RF circuits much more at the limit of what can be allowed in terms of stress conditions. As a consequence, performance of RF circuits, especially RF PA's can be boosted significantly using the simulator described in this chapter.

Chapter 5

RF reliability characterization

5.1 Introduction

Two characterization techniques that are commonly used in the reliability evaluation of MOS devices are C-V and charge pumping (CP) measurements. With the decreasing dimensions of present day MOS devices, a considerable tunneling current can be observed to flow through the gate dielectric. For both C-V and CP characterization, the increase in tunneling current may severely affect the measurement accuracy [93, 94, 95, 96, 97, 98, 99]. The RF measurement techniques discussed in chapter 2 can be used to overcome the problems associated with this leakage current, as will be explained in this chapter.

In section 5.2 the RF C-V measurement technique will be discussed. First the basics of this technique will be summarized where it will be explained how small-signal RF measurements can solve the measurement difficulties caused by the high leakage current levels on ultra-thin dielectrics. This will be followed by a two-port analysis of the RF C-V measurement test structures. This two-port analysis will be used to develop a new gate capacitance extraction methodology. The accuracy of different extraction methodologies will be evaluated using device simulations as well as measurements.

In section 5.3 the RF CP technique will be discussed. This technique can be used to obtain CP curves on ultra-thin dielectrics. The RF CP technique makes use of the RF voltage generation procedure of section 2.4. Measurement results will be presented and a theoretical framework will be derived for explaining the observed frequency response of the RF CP measurements. This model will be used for getting a good understanding of the benefits and the limitations of the technique and different application areas will be discussed.

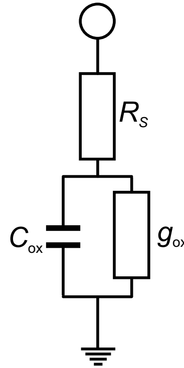

Figure 5.1: Three element circuit approximation of a leaky MOS capacitor.

5.2 Capacitance extraction from RF C-V measurements

C-V curves are used for extracting various device parameters, such as the (equivalent) oxide thickness, substrate doping, threshold voltage and flatband voltage. C-V curves are also very important in reliability experiments, for determining the interface state density and the oxide trap density of MOS devices. It is therefore a valuable characteristic of MOS devices and hence it is very important to extract these C-V curves accurately. The high leakage currents in ultra-thin oxide devices may complicate these measurements [93, 94, 95, 96, 97]. This leakage current may be overcome by making use of the RF C-V technique [97], as will be explained in section 5.2.1. In section 5.2.2 a theoretical two-port analysis of the RF C-V measurement test structures will be given. This analysis will be used for developing different gate capacitance extraction methodologies, discussed in section 5.2.3. In section 5.2.4 both simulation and measurement results are presented that compare the accuracy of different extraction methodologies.

5.2.1 Basics of the RF C-V technique

C-V curves are typically obtained by measuring the small-signal input impedance (z_{in}) of a DUT over a DC biasing voltage sweep. Using an appropriate model, the gate capacitance can subsequently be extracted from the measured z_{in}. For conventional C-V measurements a two-element model is generally considered sufficiently accurate. For ultra-thin oxide devices on the other hand the effect of the gate tunneling current should separately be taken into account. This can be done using a three-element circuit approximation as shown in figure 5.1 [93]. In this figure C_{ox} represents the oxide capacitance, which is the parameter of interest. The admittance g_{ox} models the effect of the gate tunneling current and R_{s} takes into account any series resistance in the test structure. In [97] it was explained

that accurate capacitance extraction is only straightforward if the quality factor Q is sufficiently high. This Q is defined as [97]:

$$Q = -\frac{\Im(z_{\mathrm{in}})}{\Re(z_{\mathrm{in}})} \tag{5.1}$$

In [97] the lower limit of Q was set to be 1 for an accurate extraction of the capacitance. For such a low value one must still be very cautious about the test structure parasitics and they should be carefully accounted for. Only for values of Q above 10, this becomes less of an issue. From the three-element model of figure 5.1 z_{in} can be found to be:

$$z_{\mathrm{in}} = R_{\mathrm{s}} + \frac{1}{g_{\mathrm{ox}} + j \cdot \omega C_{\mathrm{ox}}} \tag{5.2}$$

In this expression ω represents the angular frequency at which z_{in} is obtained. From expression 5.2 Q can be derived to be:

$$Q = \frac{\omega \cdot C_{\mathrm{ox}}}{g_{\mathrm{ox}} + R_{\mathrm{s}}\left(\omega^2 \cdot C_{\mathrm{ox}}^2 + g_{\mathrm{ox}}^2\right)} \tag{5.3}$$

This expression shows the dependence of Q on the measurement frequency and the magnitude of the model parameters of the three-element model of figure 5.1. For a 10 x 10 $\mu\mathrm{m}^2$ MOS capacitor structure typical model parameters of the three-element model are [100]: $C_{\mathrm{ox}} = 2$ pF, $g_{\mathrm{ox}} \approx 50 \cdot 10^{-6}\ \Omega^{-1}$ and $R_{\mathrm{s}} = 10\text{-}100\ \Omega$. In figure 5.2 Q is plotted against frequency for such a typical example with R_{s} set to 10 Ω, similar to what is done in [100]. In this example the lowest frequency for which $Q > 1$ is 4 MHz and only for frequencies higher than 42 MHz Q exceeds 10. As explained in chapter 2 measurements at such high frequencies require the use of RF measurement techniques. In [97, 100] it was shown that accurate C-V curves can be obtained on leaky dielectrics using two-port s-parameter measurements at radio frequencies using a VNA. These measurements were accompanied with SOLT calibration and OPEN-SHORT de-embedding.

5.2.2 Two-port analysis of the test structure

In contrast to conventional C-V measurements, the RF C-V technique as presented in [97, 100] makes use of two-port measurements rather than a one-port measurement. This means that more data is available, which can be used for a more accurate extraction of the gate capacitance, as previously explained in [101]. The three-element model of figure 5.1 is a one-port representation of an RF C-V test structure. It does not provide insight on the two-port behavior. In figure 5.3 a two-port circuit approximation is given of an RF C-V test structure biased in accumulation [97, 100]. In this illustration the overlap regions are separated from the intrinsic oxide region. $C_{\mathrm{ov,S}}$ and $g_{\mathrm{ov,S}}$ are the capacitance and leakage admittance of the gate to source overlap region, $C_{\mathrm{ov,D}}$ and $g_{\mathrm{ov,D}}$ are the capacitance and leakage admittance of the gate to drain overlap regions and C_{intr} and g_{intr}

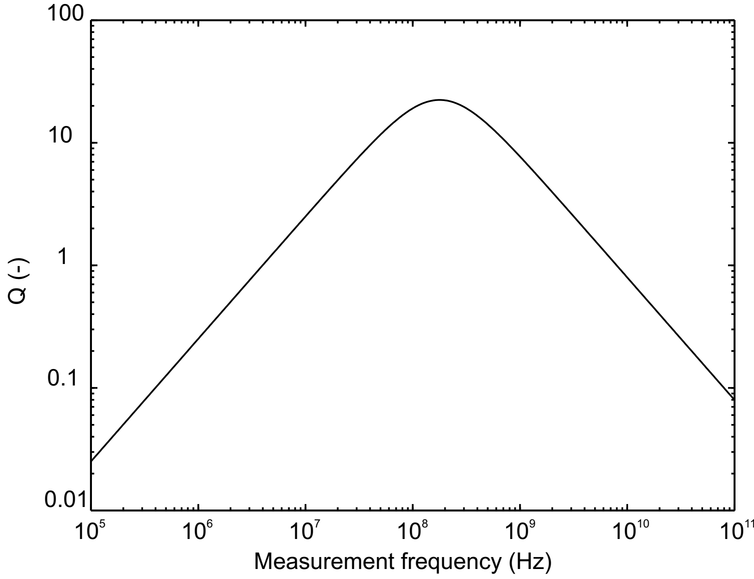

Figure 5.2: Quality factor plotted against frequency for the three-element model of figure 5.1 and using $R_s = 10\ \Omega$, $g_{ox} = 50{\cdot}10^{-6}\ \Omega^{-1}$ and $C_{ox} = 2$ pF.

are the capacitance and leakage admittance of the intrinsic oxide region. Furthermore R_G is the gate resistance, R_{well} the well resistance, R_S the series resistance in the source region and R_D is the series resistance in the drain region. Capacitances $C_{j,S}$ and $C_{j,D}$ are the junction capacitance between the well region and the source respectively the drain. This circuit approximation is more complex than the three-element model of figure 5.1, but it takes into account much better all important parasitic effects in an RF C-V measurement for devices with a small gate length and large gate width.

In order to simplify the derivation of the two-port parameters of the structure shown in figure 5.3, the circuit can be transformed into the circuit shown in figure 5.4. This circuit has the same two-port behavior as the circuit of figure 5.3; the effect of the two overlap regions is combined. R_{SD} takes into account the effect of both R_S and R_D. The effect of the two junction capacitances is modeled by the impedance z_j. In figure 5.4 the impedances z_{intr} and z_{ov} are introduced, defined as:

$$\begin{aligned} z_{intr} &= \frac{1}{g_{intr} + j \cdot \omega \cdot C_{intr}} \\ z_{ov} &= \frac{1}{g_{ov} + j \cdot \omega \cdot C_{ov}} \end{aligned} \tag{5.4}$$

For simplifying the derivation of the two-port parameters of the RF C-V test structure even further, the schematic of figure 5.4 can be transformed into the

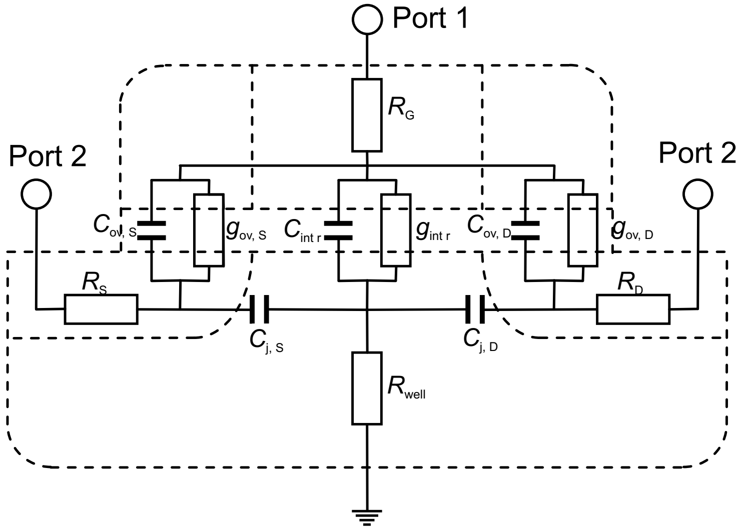

Figure 5.3: Two-port circuit approximation of an RF C-V test structure, biased in accumulation.

two port of figure 5.5 using a Π-T transformation (see e.g. [102]). In the T-network of figure 5.5 three new impedance are introduced: z_A, z_B, and z_C. They are related to the impedances in the Π-network of figure 5.4 through:

$$\begin{aligned} z_A &= \frac{z_{intr} \cdot z_{ov}}{z_{intr} + z_{ov} + z_j} \\ z_B &= \frac{z_{ov} \cdot z_j}{z_{intr} + z_{ov} + z_j} \\ z_C &= \frac{z_{intr} \cdot z_j}{z_{intr} + z_{ov} + z_j} \end{aligned} \quad (5.5)$$

Using the T-network of figure 5.5 the z-parameters of the two-port RF C-V measurement setup can be found to be:

$$\begin{aligned} z_{11} &= R_G + z_A + z_C + R_{well} \\ z_{12} &= z_C + R_{well} \\ z_{21} &= z_C + R_{well} \\ z_{22} &= R_{SD} + z_B + z_C + R_{well} \end{aligned} \quad (5.6)$$

These z-parameters can be transformed into other two-port parameters such as the y- or s-parameters. These two-port parameters can completely describe the two-port small signal behavior of the RF C-V test structure and can be used for extracting the gate capacitance from the measured two-port s-parameters.

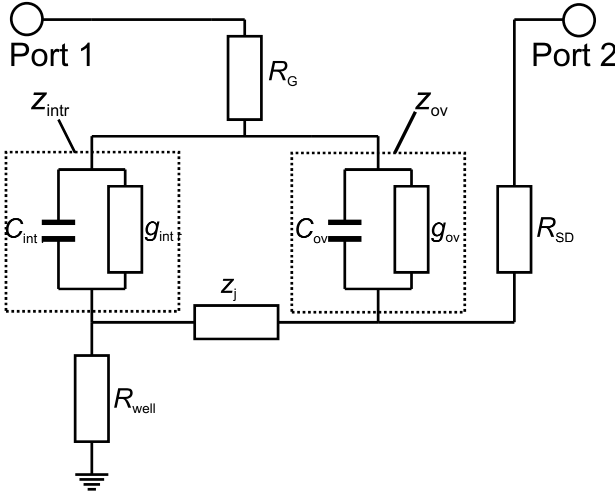

Figure 5.4: Simplified circuit approximation of figure 5.3 with identical two-port behavior, that can be used in a two-port analysis of the RF C-V test structure.

5.2.3 Capacitance extraction methodologies

In [97, 100] the gate capacitance was extracted using the y_{11} parameter of the test structure, similar to the methodology used in conventional C-V measurements. Using the two-port analysis discussed above and after some rearranging, it can be derived that:

$$\frac{1}{y_{11}} = R_{\mathrm{G}} + \frac{(Z_{\mathrm{ov}} + R_{\mathrm{SD}}) \cdot \left(z_{\mathrm{intr}} + \frac{R_{\mathrm{well}} \cdot z_{\mathrm{j}}}{R_{\mathrm{well+z_j}}}\right)}{z_{\mathrm{ov}} + R_{\mathrm{SD}} + z_{\mathrm{intr}} + \frac{R_{\mathrm{well}} \cdot z_{\mathrm{j}}}{R_{\mathrm{well+z_j}}}} \tag{5.7}$$

From this equation it is clear that the gate capacitance can not be extracted straightforwardly using the y_{11} parameter. It can however be derived that, if R_{G}, R_{well} and R_{SD} are negligible, the following expression holds:

$$C_{\mathrm{G}} \equiv C_{\mathrm{intr}} + C_{\mathrm{ov}} = \frac{\Im(y_{11})}{\omega} \tag{5.8}$$

This extraction methodology was used in[97, 100] for extraction the gate capacitance from RF C-V measurements. For guaranteeing that the test structure parasitics (i.e. R_{G}, R_{well} and R_{SD}) are indeed negligible, design guidelines are presented for minimizing these parasitics in [100]. However, with the high frequencies used in RF C-V measurements the impedance levels associated with the capacitances in the measurement setup may become very low. As a consequence it can become impracticable to design a test structure with parasitics sufficiently low for the extraction of equation 5.8 to be accurate. It can be very difficult to design a test structure in such a way that R_{well} is negligible. If a device is biased

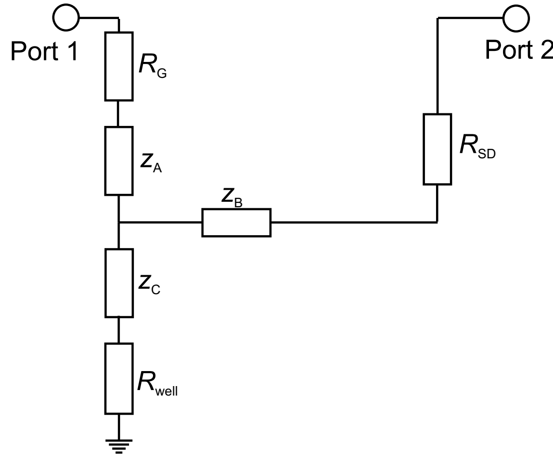

Figure 5.5: Equivalent circuit approximation of the RF C-V test structure after a Π-T transformation of the circuit of figure 5.4.

in inversion, the effect of a high R_{well} is not so pronounced; this can be understood by looking at figure 5.4, replacing z_{j} with a low-ohmic channel resistance and evaluating expression 5.7. If the device is biased in depletion or accumulation on the other hand, z_{j} consists of the two junction capacitances between the well and the drain and source regions as shown in figure 5.3. In this situation a high R_{well} will dramatically disturb the accuracy of the extracted gate capacitance.

Derivation of a new extraction methodology

In the previous section the two-port behavior of the RF C-V test structure was analyzed. This knowledge can be used to derive a more accurate extraction methodology on devices with a high R_{well} if the device is biased in accumulation or depletion. As a first step the following two assumptions can be made:

$$\begin{aligned} g_{\mathrm{intr}} &\ll \omega \cdot C_{\mathrm{intr}} \\ g_{\mathrm{ov}} &\ll \omega \cdot C_{\mathrm{ov}} \end{aligned} \tag{5.9}$$

These two assumptions state that at the measurement frequency used, the small signal impedance levels associated with the tunneling currents are negligible compared to the impedance levels associated with the intrinsic and overlap capacitances. These assumptions are valid as this resembles the key feature of RF C-V measurements: the measurement frequency in RF C-V measurements is chosen so high that the effect of the capacitances dominates over the tunneling current.

Using the assumptions in equation 5.9, z_{intr} and z_{ov} can be written as:

$$\begin{aligned} z_{\mathrm{intr}} &= \frac{1}{j \cdot \omega \cdot C_{\mathrm{intr}}} \\ z_{\mathrm{ov}} &= \frac{1}{j \cdot \omega \cdot C_{\mathrm{ov}}} \end{aligned} \tag{5.10}$$

Furthermore, in accordance with the two-port model of figure 5.3, z_{j} consists only of a junction capacitance if the device is biased in accumulation or depletion and hence may be written as:

$$z_{\mathrm{j}} = \frac{1}{j \cdot \omega \cdot C_{\mathrm{j}}} \tag{5.11}$$

In this expression C_{j} is the sum of the two junction capacitances of figure 5.3. Now, using the expressions given in 5.5, 5.6, 5.10 and 5.11 and after some rearranging it can be derived that:

$$C_{\mathrm{G}} \equiv C_{\mathrm{intr}} + C_{\mathrm{ov}} = \frac{1}{\omega} \cdot \frac{\Im(z_{22})}{\Im(z_{12}) \cdot \Im(z_{21}) - \Im(z_{11}) \cdot \Im(z_{22})} \tag{5.12}$$

This expression is a gate capacitance extraction method that is not disturbed by any of the resistances present in the two-port model of figure 5.3. Provided that the two-port model is an accurate description of the test structure and equations 5.10 and 5.11 hold, this extraction method will provide a much more accurate value of the gate capacitance than using expression 5.8

5.2.4 Comparison of the extraction methodologies

Simulation results

Based on the two-port models shown in figures 5.3, 5.4 and 5.5 it is to be expected that the extraction methodology of equation 5.12 is more accurate than the conventionally used expression 5.8 if the device is biased in depletion or accumulation. In order to investigate whether this is indeed true for typical RF C-V test structures the different extraction methodologies were investigated using device simulations. Use was made of the SILVACO ATLAS simulator for simulating the two-port behavior of RF C-V test structures at various frequencies and gate voltage levels. The simulated s-parameters were stored and subsequently the two different extraction equations were used to extract the gate capacitance, using MATLAB routines. A detailed description of this simulation flow can be found in [103].

Devices were defined that have a gate channel length of 1 μm and the length of the overlap regions was set to 0.2 μm. Although these may be relatively large values compared to actual device parameters in real RF C-V test structures, conceptually these devices may serve very well for comparing the different extraction methodologies. The thickness of the gate-oxide was chosen to be 3 nm. The devices were n-channel devices with an acceptor substrate doping of $1 \cdot 10^{16}$ cm^{-3}. The reason for choosing such a low value for the substrate doping is that in this

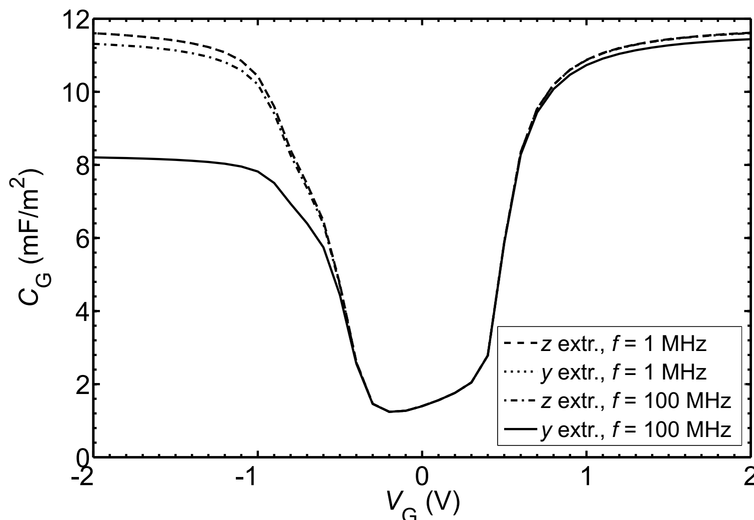

Figure 5.6: Extracted C-V curves from simulation results on an n-type device using both z-parameter and y-parameter extraction at two different frequencies. The two extracted lines at 1 MHz overlap over the entire curve. In the inversion regime also the z-parameter extraction at 100 MHz overlaps with these two curves. y-parameter extraction at 100 MHz is distinctly different.

way the resistance in the substrate is relatively large. This substrate resistance has the same effect as the well resistance in the models of figure 5.3, it is the resistance seen between the channel region and the ground connection. As it follows from equation 5.7 this high resistance may seriously affect the accuracy of the gate capacitance extraction method using the y_{11} parameter (equation 5.8) if the device is biased in depletion or accumulation. It is to be expected that the z-parameter extraction of expression 5.12 is not affected by this high resistance.

Two-port simulations were performed with the gate connected to port 1 and the source and drain tied together and connected to port 2. A connection to the substrate was made on top of the substrate, thereby mimicking two-port behavior of actual RF C-V test structures. This substrate connection was defined as the ground in the two-port s-parameter simulations. Simulations were performed without any model for the tunneling current included. While such simulations do not show the typical problems of C-V measurements at low frequencies on leaky MOS devices, they do serve very well for investigating the frequency response of the different extraction methodologies. Knowing that the simulations performed at low frequencies are not complicated by the tunneling current, the gate capacitance extracted at these low frequencies can safely be assumed to be correct.

In figure 5.6 the extracted C-V curves from simulation results using the two different extraction methodologies are compared at two different frequencies. The

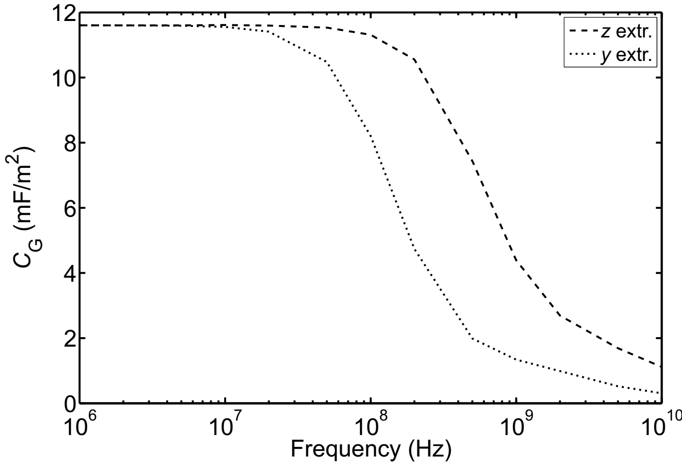

Figure 5.7: Extracted gate capacitance at -2 V obtained using two-port s-parameter simulations on the same device structure as in figure 5.6. The two extraction methodologies discussed in section 5.2.4 are compared. Tunneling currents were not included in the simulations.

C-V curves obtained at 1 MHz are completely identical to each other. This indicates that at this frequency both the y-parameter extraction of equation 5.8 and the z-parameter extraction of equation 5.12 provide accurate C-V curves, provided that the effect of the tunneling current is negligible.

The purpose of the RF C-V technique is, as discussed in section 5.2.1, to obtain C-V curves on devices with a high tunneling current, by making use of higher measurement frequencies. For this purpose, in figure 5.6 also C-V curves are shown, extracted from 2-port simulations at 100 MHz. At this frequency a clear difference can be seen between y-parameter extraction and z-parameter extraction. The z-parameter extraction results are very close to the results obtained at 1 MHz. This indicates that at this frequency z-parameter extraction is still allowed. The results obtained using y-parameter extraction on the other hand differ quite a lot from the other results, especially in the accumulation regime. As a result y-parameter extraction is not valid on this device at a frequency of 100 MHz.

To investigate the maximum frequency for which the two extraction methodologies is valid, the extracted capacitance is plotted against frequency at a gate biasing voltage of -2 V in figure 5.7. In this figure the superior accuracy of z-parameter extraction above y-parameter extraction can be recognized. At frequencies ranging from 1 MHz to 10 MHz the extracted gate capacitance remains constant for both extraction methodologies. Above 10 MHz, the extracted capacitance can be seen to decrease for y-parameter extraction while it appears to remain constant for z-parameter extraction. The frequency at which a 10 %

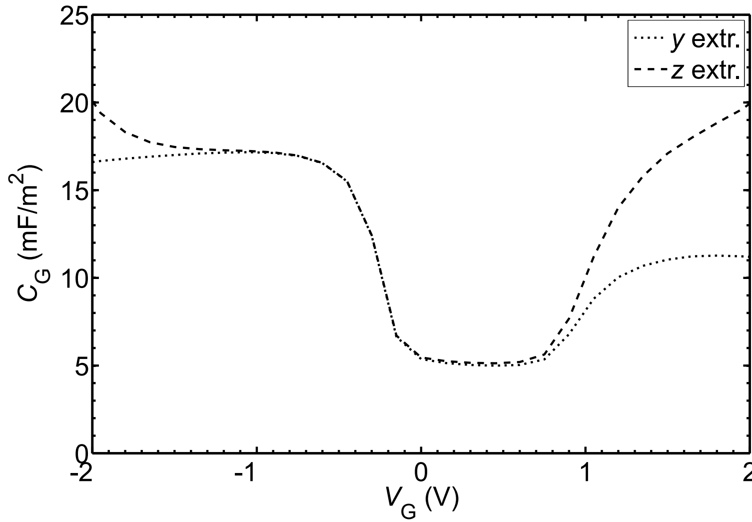

Figure 5.8: Extracted gate capacitance obtained after two-port s-parameter measurements on a p-type device at a measurement frequency of 187 MHz.

decrease in extracted capacitance can be observed is 51 MHz for y-parameter extraction and 209 MHz for z-parameter extraction. This is a large difference. This difference in accuracy is in accordance with the theory presented in section 5.2.3

Measurement results

The simulation results are a very useful tool for investigating the limitations of the two discussed extraction methodologies. Measurement results are required however, to fully appreciate the increase in accuracy achieved by making use of z-parameter extraction over y-parameter extraction. Two-port s-parameter measurements where obtained on a p-type device processed in a research process flow based on 0.18 μm CMOS. The device was laid out in a ground-signal-ground configuration following the design rules of [100]. The gate was connected to one of the signal ports, the drain and source were tied together and connected to the signal port and the well contact was connected to the ground plane. The device was designed with a channel length of 0.15 μm and a total gate width of 9360 μm. Measurements were performed using an HP 8510 C vector network analyzer, the measurements were accompanied with SOLT calibration and OPEN-SHORT de-embedding. In figure 5.8 the extracted gate capacitance is shown obtained at a frequency of 187 MHz using both y-parameter extraction and z-parameter extraction. The extracted gate capacitance is equal for both extraction methodologies for biasing voltages in the depletion region. In inversion a difference can be recognized between the two extraction methodologies. The y-parameter extraction results show the effect of gate depletion which is to be expected on

this device. As it follows from the theory presented in section 5.2.3 z-parameter extraction may be erroneous in the inversion regime, as condition 5.11 does not hold. On the other hand the results in the accumulation regime shows the typical accumulation behavior of ultra-thin dielectrics for z-parameter extraction. This can not be seen in y-parameter extraction. The results are very similar to the simulation results shown in figure 5.6. Therefore one can assume that accurate C-V extraction in the accumulation regime can only be done using the z-parameter extraction of expression 5.12.

5.2.5 Discussion

The two-port analysis presented in section 5.2.2 and the new z-parameter extraction methodology derived in section 5.2.3 provide an impressive increase in the measurement frequency at which accurate C-V curves can be obtained compared to the y-parameter extraction used in [97, 100]. At increasing levels of the gate tunneling current, increasing measurement frequencies are required for obtaining accurate C-V data. Series resistances present in the test-structure complicate accurate extraction of the gate capacitance at such high frequencies. This causes the use of y-parameter extraction to be not suitable for extracting the gate capacitance in the depletion and accumulation regime, as it follows from equation 5.7.

In section 5.2.4 simulation results are presented that show the increase in accuracy obtained with z-parameter extraction over y-parameter extraction. In these simulations tunneling effects are not included. Although this is not the most realistic simulation of actual RF C-V test structures with ultra-thin oxides, these simulations are very useful for comparing the different extraction methodologies. In the work described in this section, the maximum frequency at which gate capacitance is accurate is investigated. The problems associated with low frequency measurements are known to be caused by the gate tunneling current. Increasing the measurement frequency can solve this issue. If the tunneling current is not present, accurate C-V extraction can be done at low frequencies and these results can be used as a benchmark for evaluating the maximum frequencies for which the different extraction methodologies are valid. Figures 5.6 and 5.7 clearly show the increased frequency limit in z-parameter extraction compared to y-parameter extraction for devices biased in accumulation.

5.3 Charge pumping at radio frequencies

5.3.1 Introduction

Another characterization technique that is often used in the reliability evaluation of MOS devices is the CP technique [104]. It is widely used to quantify the interface state density at the Si-SiO_2 interface of MOS devices. With the decreasing thickness of the oxide layer in present day CMOS technologies a considerable gate tunneling current can be seen. This tunneling current can severely affect the correctness of the extracted interface state density from CP data [98, 99]. In

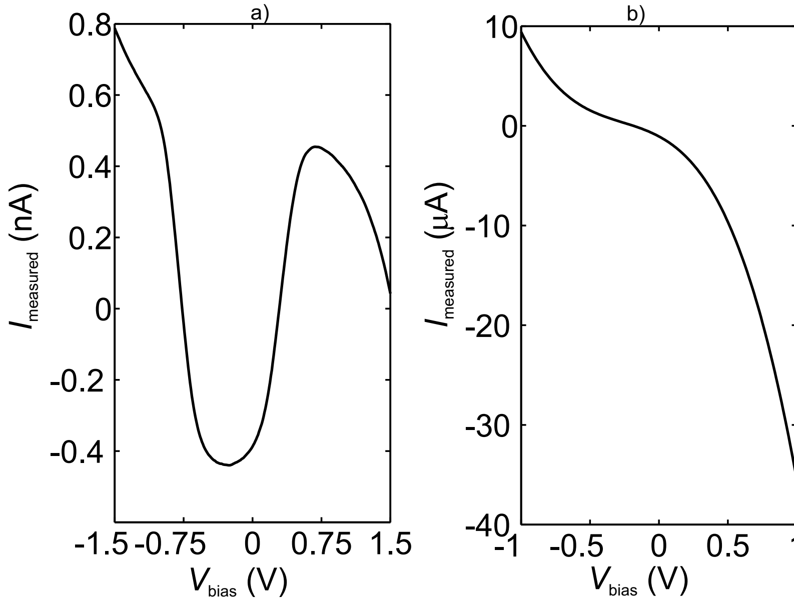

Figure 5.9: Charge pumping currents obtained on n-type devices with an oxide thickness of a) 3 nm and b) 1.4 nm. The data are obtained using a sinusoidal gate voltage with $V_{\text{pp}} = 2$ V and $f = 1$ MHz. The current is measured at the drain/source contact, resulting in negative values of the CP current. The CP current is completely overwhelmed by the leakage current on the 1.4 nm oxide.

figure 5.9 the problem is visualized by comparing CP data obtained on a device with 3 nm oxide thickness to data obtained on a 1.4 nm oxide device. On the 3 nm oxide the CP effect is still clearly visible, as a pronounced current that flows only when the device is modulated between accumulation and inversion. With a 1.4 nm gate-oxide, the CP current is overwhelmed by the leakage current.

In recent literature, several approaches have been presented to alleviate the gate leakage problem. In [98] it was shown that a large increase in accuracy can be obtained by correcting for the gate leakage component, as obtained from very low frequency CP data. Furthermore, the leakage current component can be minimized by carefully choosing the gate voltage window [105]. In [99] a small voltage swing approach was proposed that minimizes the leakage current component even further. A major drawback of this approach is, however, that only a very small portion of the bandgap is scanned, thereby probing only a very small subset of interface traps. This may lead to large inaccuracies in extracting the effective interface state density $\overline{D_{\text{it}}}$. Furthermore even if this issue can be overcome, this approach is still limited by the leakage current that is induced by the DC bias voltage. For a typical $\overline{D_{\text{it}}}$ of 10^{11} cm^{-2}eV^{-1}, a voltage window of 0.1 V and a measurement frequency of 1 MHz, the CP current density will be

approximately 1.6 mA·cm^{-2}. For an accurate measurement of the CP current, it must not be disturbed by the leakage current component. Even though the leakage current can be subtracted from the measured current, limitations in the resolution of the measurement equipment cause inaccuracies. Therefore the CP current must be sufficiently higher than the leakage current component. In this thesis a factor of 10 is used as a criterion. This implies an upper limit for the leakage current density of 0.16 mA · cm^{-2} for this example. For lower values of $\overline{D_{\mathrm{it}}}$ this upper limit is even further reduced.

If one wants to overcome larger leakage current densities, use can be made of the frequency dependence of the CP current as given by [104]:

$$I_{\mathrm{cp}} = fqA_{\mathrm{G}}\overline{D_{\mathrm{it}}}\Delta E \tag{5.13}$$

In this expression f is the frequency of the applied gate voltage signal, q is the elementary charge, A_{G} is the surface area of the device, $\overline{D_{\mathrm{it}}}$ is the interface state density (in cm^{-2}eV^{-1}) and ΔE is the energy window between which traps are located that contribute to the CP current. This energy window depends on the time available for the nonsteady-state emission of carriers during a CP cycle [104]. The frequency dependence of equation 5.13 can be applied in order to increase the CP current with respect to the leakage current moving far beyond the ~1 MHz signals used in conventional CP measurements. The problems due to distortion of these high frequency signals can e.g. be solved by designing a complete on chip pulse generator circuit as in the approach of [106] or by making use of the RF CP technique [107, 108]. In this section the use of the RF CP technique will be evaluated. This technique makes use of sinusoidal large voltage swing signals with frequencies into the GHz range, more than two orders higher than in the conventional CP approach.

5.3.2 Measurement setup and methodology

Setup

In a CP measurement a device is repeatedly switched between accumulation and inversion. Carriers, originating from the substrate are trapped during accumulation (holes in an n-type device, electrons in a p-type device) and released during inversion. These carriers recombine with carriers from the opposite polarity originating from the source/drain region. In this way a net amount of charge is transferred from the substrate towards the source/drain regions. By repeatedly performing such a CP cycle, a DC current can be observed at both the substrate and the source/drain connection. This DC current is the CP current I_{cp}.

The key idea of the RF CP technique is, that a higher excitation frequency will increase the CP current and thus its significance with respect to the tunneling current. Using frequencies above ~ 10 MHz the effects of impedance mismatch become noticeable thereby distorting the voltage signal between the source and the Device Under Test (DUT). Significant measurement or interpretation errors may thus arise. This may be solved by making use of the RF voltage generation

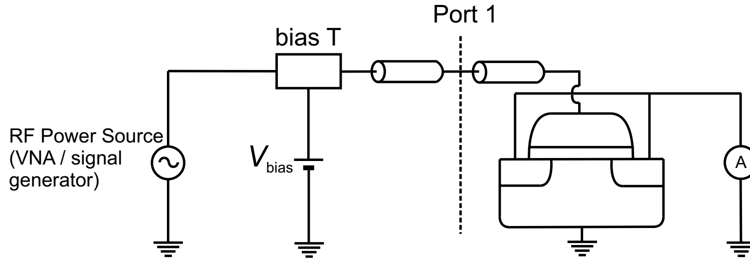

Figure 5.10: Schematic drawing of the RF CP measurement setup. An RF signal is superimposed on a DC voltage V_{bias} through the use of a bias T. Port 1 represents the location in the measurement setup where four of the seven necessary calibration measurements are performed. The CP current is measured at the drain/source connection.

procedure as discussed in section 2.4 in combination with the setup as illustrated in figure 5.10. The setup is different from the conventional CP measurement setup. A sinusoidal gate voltage signal is generated by making use of an RF power source. The RF power is superimposed on a DC voltage V_{bias} through the use of a bias T. A full CP curve (as in figure 5.9) is obtained by sweeping V_{bias}. For the results presented in this section a Rohde & Schwarz ZVB20 Vector Network Analyzer is used as an RF power source. An HP 4156A parameter analyzer generates V_{bias} and measures I_{cp}. For measurements below 10 MHz the same setup is used but the RF power source is replaced by an Agilent 33250A signal generator.

The test structures are designed in a two-port ground-signal-ground configuration similar to [100], optimized for accurate RF measurements. The structures consist of transistors with the source and drain connection shorted; the gate is connected to one of the signal pads, while the source/drain is connected to the other signal pad. The substrate is connected to the ground plane. This connection makes it impossible to measure the CP current at the substrate; therefore the CP current is measured at the drain/source connection. As a consequence the measured current has opposite polarity with respect to the various CP currents previously reported in literature (e.g. [104]).

Voltage generation

When radio frequencies are used to switch quickly between accumulation and inversion, a sinusoidal voltage signal is preferred. This waveform minimizes the effect of signal distortion, which may arise from an impedance mismatch between the measurement cables and the device under test. Distortion will change the precise waveform at the device, and this complicates the interpretation of CP currents. Using the signal integrity analysis discussed in section 2.4 it is possible to verify whether signal distortion is negligible for a sinusoidal voltage signal. This analysis was performed on the devices used for the RF CP measurements

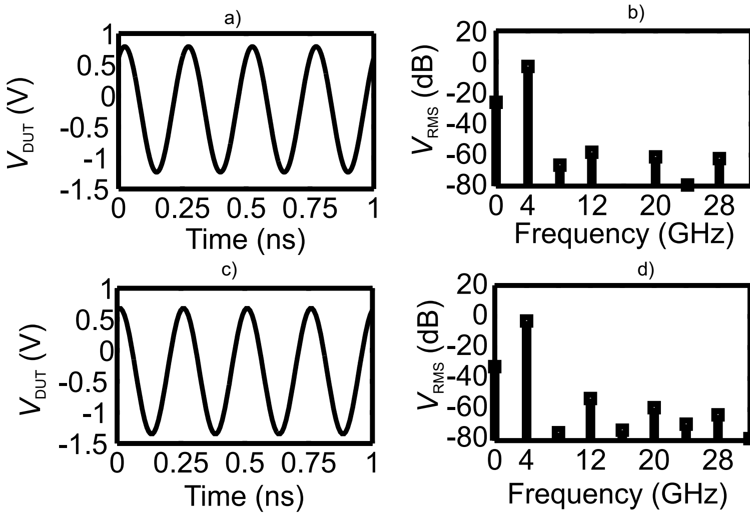

Figure 5.11: Time-domain signals as obtained using the discrete time solution to the transmission line equation as discussed in section 2.4. The harmonic content as it follows from a Fast Fourier Transform is also shown. In a), b) this is done for a 3nm oxide device and in c), d) for a 1.4 nm oxide device. The waveforms are determined with V_{bias} set for maximum CP condition at 4 GHz.

presented in this chapter and some typical results are shown in figure 5.11. In this figure the calculated waveforms are shown for both a 3 nm oxide n-type device as well as for a 1.4 nm oxide device. The calculations were performed at the highest frequency used in this section, i.e. 4 GHz. At this frequency any effects of a nonlinear input capacitance should be most visible. Furthermore the DC gate voltage was chosen such that the maximum CP condition was realized. At this condition also any distortion effects due to a nonlinear input capacitance should be expected to be the worst. Besides, it is also the voltage condition of interest for the results presented in this chapter. The figure clearly shows that the nonlinear behavior at the input of the devices is negligible. The voltage $V_{\mathrm{RMS}}(f)$ used in figure 5.11b) and d) is defined as the root mean square value of the harmonic component with frequency f of the gate voltage $V_{\mathrm{G}}(t)$.

The signal integrity analysis determines whether it is possible to generate a sinusoidal gate voltage signal with the desired frequency and amplitude. The next step is to set the base level and amplitude of the RF gate signal properly. V_{pp} is set by the power level of the RF source; the base level by the dc biasing. In [108] an easy to use method was proposed for obtaining the appropriate power level. This approach, however is only limited to very specific devices, which show a considerable tunneling current, but do not suffer from too high voltage levels. A more generally applicable approach is the RF voltage generation procedure that makes use of a VNA and the calibration procedure, as described in section 2.4.

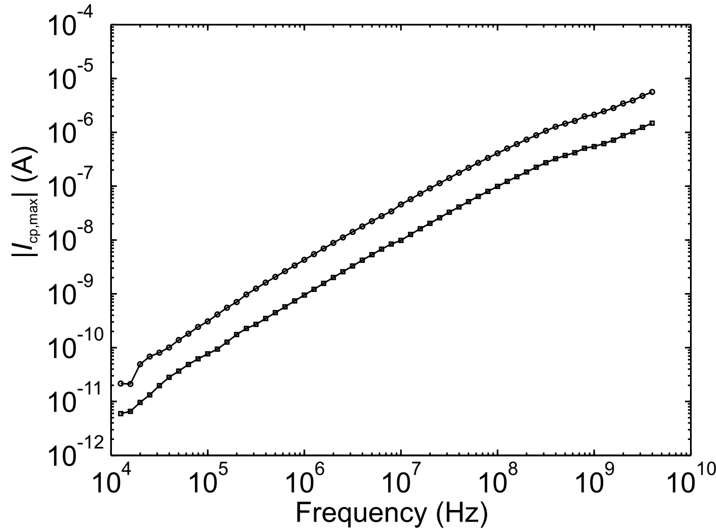

Figure 5.12: $I_{\mathrm{cp,max}}$ plotted against frequency, obtained on devices with $t_{\mathrm{ox}} = 3$ nm. The lower curve represents measurements made on an n-type device and the upper curve on a p-type device. The gate voltage signal used was sinusoidal with $V_{\mathrm{pp}} = 2$ V. The increase of the CP current with increasing frequency can clearly be seen.

5.3.3 Measurement results

Using the approach of section 2.4, gate voltage signals with well defined amplitude levels may be generated with frequencies of up to 4 GHz. This allows performing CP measurements at frequencies far beyond the frequencies used in conventional CP measurements. This is illustrated in figure 5.12 where $I_{\mathrm{cp,max}}$ is plotted against frequency on both an n-type as well as a p-type device with 3 nm oxide thickness. The maximum CP current, $I_{\mathrm{cp,max}}$, is defined as the largest (absolute value of the) CP current over an entire V_{bias} sweep. It is clearly visible that $I_{\mathrm{cp,max}}$ keeps increasing with frequency up to 4 GHz, over two orders of magnitude beyond conventional CP measurements. Hence, this frequency dependence, as predicted by equation 5.13, may be applied in order to perform CP measurements on dielectrics with a leakage current too high for conventional CP measurements.

This is shown in figure 5.13 where CP data are shown on the same 1.4 nm device as of figure 5.9 b), but at frequencies up to 4 GHz. Only at the highest frequencies, the CP effect is visible. The extension to GHz frequencies allows to determine $I_{\mathrm{cp,max}}$, and thus obtain information on the amount of fast interface states, on ultra-leaky dielectrics. Similar to figure 5.12 the frequency response of the measured CP current on both an n-type and p-type 1.4 nm device was also plotted. This is shown in figure 5.14. In this figure parameter $I_{\mathrm{measured,max}}$ was plotted rather than $I_{\mathrm{cp,max}}$ in order to prevent any confusion about the definition

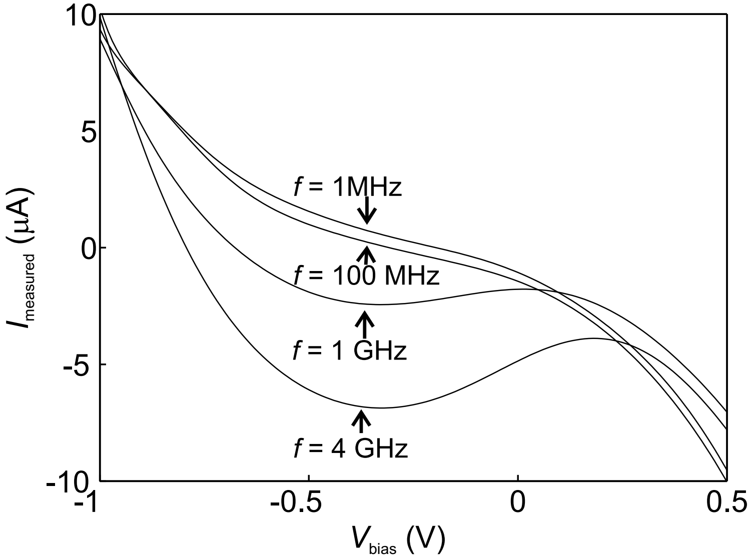

Figure 5.13: Charge pumping curves obtained at various frequencies on the same 1.4 nm oxide device as in figure 5.9b. The applied gate voltage level has a V_{pp} of 2 V. The CP effect is clearly visible for the 1 GHz and 4 GHz signals.

of $I_{\mathrm{cp,max}}$.

Similar to $I_{\mathrm{cp,max}}$, $I_{\mathrm{measured,max}}$ is also defined as the largest (absolute value of the) measured current over an entire V_{bias} sweep; for lower frequencies however, this measured current is dominated by the gate leakage current. The two horizontal lines in figure 5.14 represent the magnitude of the leakage component of the measured currents.

For extracting reasonably accurate CP data a CP component of at least 10 times higher than the leakage component of the measured current is needed, as explained earlier. This condition implies that for the 1.4 nm n-type device only measurements beyond 1.5 GHz may be used. This illustrates the ultra-high leakage current for this device. The 1.4 nm p-type device is less leaky, but still only frequencies beyond 80 MHz may be used for accurate extraction of the interface trap distribution. For extracting the number of traps that contribute to the CP current use can be made of the pumped charge per cycle, $Q_{\mathrm{cp,max}}$, defined as:

$$Q_{\mathrm{cp,max}} = \frac{|I_{\mathrm{cp,max}}|}{f} \tag{5.14}$$

$Q_{\mathrm{cp,max}}$ is a direct indicator of the amount of traps that contribute to the CP current and hence a measure of the Si-SiO_2 interface quality. In the conventional CP approach this $Q_{\mathrm{cp,max}}$ is assumed to be constant over frequency, when using trapezoidal gate voltage signals with constant rise and fall times for every fre-

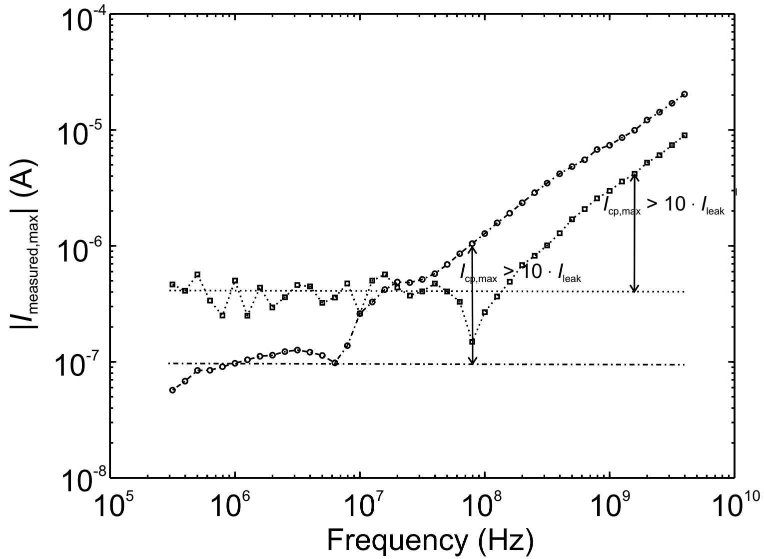

Figure 5.14: Maximum value of measured CP current plotted against frequency on both an n-type (lower curve) and a p-type(upper curve) 1.4 nm device. All measurements are performed with a sinusoidal voltage signal with $V_{\mathrm{pp}} = 2$ V. A plateau of minimum $I_{\mathrm{measured,max}}$ can be observed for low frequency data originating from the leakage current component present in these ultra-thin oxide devices. The two horizontal lines are drawn to indicate the absolute level of the leakage current.

quency. For the sinusoidal voltage signals used in this paper, the rise and fall times are equal, and proportional to the inverse of frequency.

In [109] it was shown that $Q_{\mathrm{cp,max}}$ increases linearly with $\log(f)$ when sinusoidal gate voltage signals are used. The increase in $Q_{\mathrm{cp,max}}$ with increasing frequency is attributed to the increase in the probed energy window ΔE. In figure 5.15 $Q_{\mathrm{cp,max}}$ is plotted against frequency up to 4 GHz, for the four different devices used in figures 5.12 and 5.14. The results presented are obtained from the measured current levels after a subtraction of low frequency measurement as in [98], in order to increase accuracy. On the 1.4 nm oxide devices gate leakage prohibits the quantification of $Q_{\mathrm{cp,max}}$ below 1.5 GHz and 80 MHz for the n-type and p-type devices respectively. On the 3 nm oxide devices $Q_{\mathrm{cp,max}}$ was extracted for frequencies ranging from 10 kHz to 4 GHz.

In order to obtain results below 10 MHz use was made of an Agilent 33250A signal generator for setting the sinusoidal gate voltage signals. The results at frequencies of 10 MHz and higher were obtained after subtracting a 1 nA DC offset of the measured current. The reason for performing this subtraction is that if this would not be done, a small transition point in the $Q_{\mathrm{cp,max}}$ vs. $\log(f)$ graphs could be noticed around 10 MHz, the frequency were the transition between the

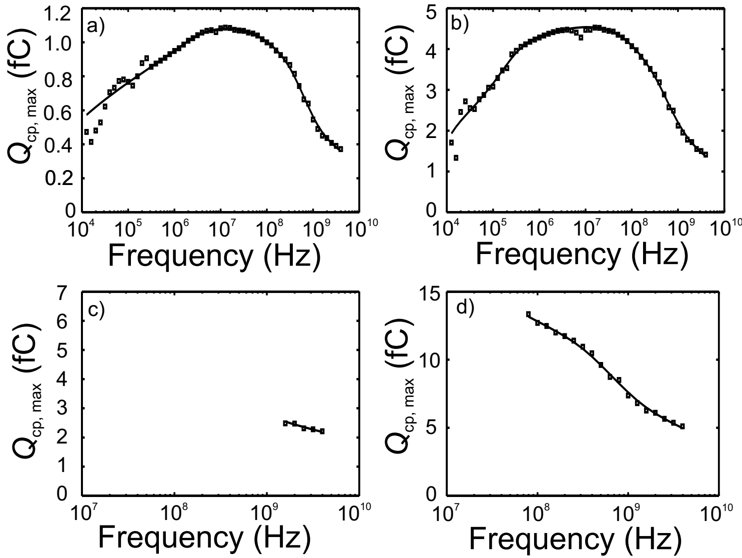

Figure 5.15: Pumped Charge per cycle plotted against frequency on a) a 3 nm n-type devices, b) a 3 nm p-type device, c) a 1.4 nm n-type device and d) a 1.4 nm p-type device. For the 1.4 nm devices $Q_{\mathrm{cp,max}}$ values are only obtained at frequencies where the CP effect is clearly visible (as indicated in figure 5.14). The data are obtained using a sinusoidal gate voltage with $V_{\mathrm{pp}} = 2$ V. The solid lines are meant as a guide to the eye.

two measurement setups is made. A possible explanation for this transition point could be the imperfect grounding of the VNA. This 1 nA DC offset is negligible for frequencies above 50 MHz, hence for the frequencies of interest in this section. For the results obtained on the 1.4 nm devices this small DC offset is negligible for all frequencies as can be seen in figure 5.14.

5.3.4 Trap Response

In figure 5.15 the expected linear slope for frequencies up to 10 MHz can be recognized. Above this frequency $Q_{\mathrm{cp,max}}$ starts to increase less than predicted by the theory of [109], and it even decreases above ~ 100 MHz. This indicates that an increasing number of traps is too slow to respond to the CP signal. In [108] it was shown that the decrease of $Q_{\mathrm{cp,max}}$ with increasing frequency as seen in figure 5.15 can not be explained by the classical CP theory, but by a distribution of traps in the oxide. The traps located far away from the interface are slow and those near the interface are fast traps (a commonly used assumption, see e.g. [110]). In this section it will be discussed how this trap distribution is exactly related to the observed frequency response of $Q_{\mathrm{cp,max}}$ for sinusoidal gate voltage signals. The derivation starts with the general expression for the CP current (see

e.g. [99]).

$$I_{\mathrm{cp}} = fqA_{\mathrm{G}} \int_{0}^{t_{\mathrm{ox}}} \int_{E_{\mathrm{low}}}^{E_{\mathrm{high}}} C_{\mathrm{it}}\left(E_{\mathrm{T}}, x_{\mathrm{T}}\right) \cdot \Delta f_{\mathrm{T}}\left(E_{\mathrm{T}}, x_{\mathrm{T}}, f\right) dE_{\mathrm{T}} dx_{\mathrm{T}} \tag{5.15}$$

In this expression a pure tunneling mechanism [110] is assumed to govern the capture and emission processes. Expression 5.13 is basically a simplification of expression 5.15. Where in expression 5.13 the effective interface state density $\overline{D_{\mathrm{it}}}$ (in $\mathrm{cm}^{-2}\mathrm{eV}^{-1}$)is used, in expression 5.15 use is made of the interface state concentration C_{it} (in $\mathrm{cm}^{-3}\mathrm{eV}^{-1}$). All traps are described with an energy level E_{T} and distance from the interface x_{T}. Parameter t_{ox} is the integration limit for x_{T}. In expression 5.15 it is rather arbitrarily chosen as the oxide thickness; later another integration limit will be derived, thereby separating the fast interface traps from the slow traps. Furthermore E_{low} and E_{high} represent the lowest and highest Fermi energy levels of the silicon surface during an entire CP cycle. Parameter Δf_{T} is the difference in trap occupancy level f_{T} between inversion and accumulation condition. In the classical CP theory this Δf_{T} is assumed to equal unity for all traps that are located between the energy levels $E_{\mathrm{em,e}}$ and $E_{\mathrm{em,h}}$, the limits of the energy levels where the emission process of carriers is negligible during a CP cycle. Parameter x_{T} represents the distance from a particular trap towards the Si-SiO_2 interface. The capture cross section of a trap is related to the distance of the trap from the interface through [110]:

$$\sigma(x_{\mathrm{T}}) = \sigma_0 e^{-x_{\mathrm{T}}/\lambda} \tag{5.16}$$

This expression is based on the first order trapping model of [110] where the capture cross section is assumed to be independent on energy level. In expression 5.16, λ represents the tunneling attenuation constant, which is approximately 0.07 nm [110]. The capture cross section of a trap has a direct influence on the speed at which charge carriers are captured by and emitted from a trap and is therefore a crucial parameter in understanding the trap response to RF CP measurements. In order to find a solution to the integral equation of 5.15, an expression is needed for Δf_{T}. From S-R-H statistics a differential equation can be found that describes the occupancy level of a trap as a function of time (see e.g. [111]):

$$\frac{df_{\mathrm{T}}}{dt} = [1 - f_{\mathrm{T}}(t)] \cdot [c_{\mathrm{n}}(t) + e_{\mathrm{p}}] - f_{\mathrm{T}}(t) \cdot [c_{\mathrm{p}}(t) + e_{\mathrm{n}}] \tag{5.17}$$

In this expression $c_{\mathrm{n}}(t)$ and $c_{\mathrm{p}}(t)$ represent the capture rates of electrons and holes respectively. Parameters e_{n} and e_{p} are the emission rates of electrons and holes. The capture rates $c_{\mathrm{n}}(t)$ and $c_{\mathrm{p}}(t)$ are directly related to the amount of charge carriers at the Si-SiO_2 interface and therefore dependent on gate voltage and hence time. Emission rates e_{n} and e_{p} are only dependent on the energy level of the trap with respect to the conduction band respectively valence band.

No general closed from solution can be found for the differential equation of 5.17; a similar approach will be adopted as in [112] in finding an expression for Δf_{T} from expression 5.17. In this approach use is made of three basic assumptions:

1. Traps outside ΔE do not contribute to the CP current and all traps within ΔE have negligible emission rates [104].
2. The capture processes are negligible outside inversion and accumulation, i.e. at voltage levels between V_{T} and V_{FB} [104].
3. At maximum CP conditions the integral of $c_{\mathrm{n}}(t)$ over time equals the integral of $c_{\mathrm{p}}(t)$ over time [112].

Using assumptions 1) and 2), the increase in capture rate during inversion may be written as:

$$\frac{df_{\mathrm{T}}}{dt} \approx [1 - f_{\mathrm{T}}(t)] \cdot c_{\mathrm{n}}(t) \tag{5.18}$$

Similarly, for the capture process during accumulation expression 5.17 reduces to:

$$\frac{df_{\mathrm{T}}}{dt} \approx -f_{\mathrm{T}}(t) \cdot c_{\mathrm{p}}(t) \tag{5.19}$$

Using expression 5.18 the maximum trap occupancy level, $f_{\mathrm{T,max}}$ can be found as a function of the minimum trap occupancy level, $f_{\mathrm{T,min}}$; $f_{\mathrm{T,min}}$ can be found as a function of $f_{\mathrm{T,max}}$ using expression 5.19:

$$f_{\mathrm{T,max}} = (1 - f_{\mathrm{T,min}}) \cdot e^{-\zeta_{\mathrm{n}}} \tag{5.20}$$

$$f_{\mathrm{T,min}} = f_{\mathrm{T,max}} \cdot e^{-\zeta_{\mathrm{p}}} \tag{5.21}$$

In these expression parameters ζ_{n} and ζ_{p} are defined as:

$$\zeta_{\mathrm{n}} = \int_{t_{\mathrm{inv,start}}}^{t_{\mathrm{inv,stop}}} c_{\mathrm{n}}(t)dt \tag{5.22}$$

$$\zeta_{\mathrm{p}} = \int_{t_{\mathrm{acc,start}}}^{t_{\mathrm{acc,stop}}} c_{\mathrm{p}}(t)dt \tag{5.23}$$

Parameters $t_{\mathrm{inv,start}}$ and $t_{\mathrm{inv,stop}}$ are the times at the onset and end of inversion condition respectively. Parameters $t_{\mathrm{acc,start}}$ and $t_{\mathrm{acc,stop}}$ signify the start and end times of accumulation.

Using expressions 5.20 and 5.21 an expression for Δf_{T} can be derived:

$$\Delta f_{\mathrm{T}} = f_{\mathrm{T,inv}} - f_{\mathrm{T,acc}} = \frac{\left(1 - e^{-\zeta_{\mathrm{n}}}\right)\left(1 - e^{-\zeta_{\mathrm{p}}}\right)}{\left(1 - e^{-\zeta_{\mathrm{n}} - \zeta_{\mathrm{p}}}\right)} \tag{5.24}$$

This expression cannot be solved analytically for all values of ζ_{n} and ζ_{p}. However it can be further simplified by making use of assumption 3 given above for maximum CP conditions. This means that at the maximum CP condition over an entire V_{bias} sweep one may assume that $\zeta_{\mathrm{np}} \approx \zeta_{\mathrm{n}} \approx \zeta_{\mathrm{p}}$. Using parameter ζ_{np}, Δf_{T} at maximum CP condition can be expressed as:

$$\Delta f_{\mathrm{T}} = \frac{\left(1 - e^{-\zeta_{\mathrm{np}}}\right)^2}{\left(1 - e^{-2\zeta_{\mathrm{np}}}\right)} \tag{5.25}$$

This function is approximated with a step function around $\zeta_{\mathrm{np}} = \ln(3)$, after [112], giving a very simple expression for Δf_{T} at maximum CP conditions. Now parameter $x_{\mathrm{T,max}}$ can be defined as the maximum value of x_{T} where traps are located that are fast enough to contribute to the CP current. It is the value that x_{T} has to have for Δf_T to equal ln(3). An expression for $x_{\mathrm{T,max}}$ can be derived by using the expression used for the capture rate of electrons $c_{\mathrm{n}}(t)$ (or similarly from the capture rate for holes, as assumption 3) is allowed). This capture rate is given by:

$$c_{\mathrm{n}}(t) = v_{\mathrm{th}}\sigma(x_{\mathrm{T}})n_{\mathrm{s}}(t) \tag{5.26}$$

In this expression v_{th} is the thermal velocity of charge carriers and $n_{\mathrm{s}}(t)$ is the electron concentration at the Si-SiO$_2$ interface. Using the definition of ζ_{np} one can find:

$$\zeta_{\mathrm{np}} = \zeta_{\mathrm{n}} = \sigma(x_{\mathrm{T}})v_{\mathrm{th}} \int_{t_{\mathrm{inv,start}}}^{t_{\mathrm{inv,stop}}} n_{\mathrm{s}}\left(V_{\mathrm{G}}(t)\right) dt \tag{5.27}$$

The solution to this integral can be found by expressing the sinusoidal gate voltage signal $V_{\mathrm{G}}(t)$ as:

$$V_{\mathrm{G}}(t) = V_{\mathrm{bias}} + \frac{V_{\mathrm{pp}}}{2}\sin\left(2\pi f t\right) \tag{5.28}$$

ζ_{np} may now be rewritten as:

$$\zeta_{\mathrm{np}} = \frac{\sigma(x_{\mathrm{T}})v_{\mathrm{th}}}{2\pi f} \cdot \xi \tag{5.29}$$

Parameter ξ can be found by integrating $n_{\mathrm{s}}(V_{\mathrm{G}}(t))$ over time with $V_{\mathrm{G}}(t)$ given by 5.28. The result of this integration equals:

$$\xi = 2\int_{V_{\mathrm{T}}}^{V_{\mathrm{bias}}+\frac{V_{\mathrm{pp}}}{2}} \frac{n_{\mathrm{s}}(V_{\mathrm{G}})}{\sqrt{1-\left(\frac{2(V_{\mathrm{G}}-V_{\mathrm{bias}})}{V_{\mathrm{pp}}}\right)^2}} dV_{\mathrm{G}} \tag{5.30}$$

Parameter ξ may be interpreted as the effective product of the surface electron concentration and the gate voltage during inversion. $x_{\mathrm{T,max}}$ can now be expressed as:

$$x_{\mathrm{T,max}}(f) = -\lambda \cdot \ln\left(\frac{2\pi f \ln(3)}{\sigma_0 \cdot v_{\mathrm{th}} \cdot \xi}\right) \tag{5.31}$$

Using this definition of $x_{\mathrm{T,max}}$ and introducing the energy window ΔE expression 5.15 for maximum CP conditions can be rewritten into:

$$I_{\mathrm{cp,max}} = fqA_{\mathrm{G}} \int_0^{x_{\mathrm{T,max}}} \overline{C_{\mathrm{it}}}\left(x_{\mathrm{T}}\right) \cdot \Delta E\left(x_{\mathrm{T}}, f\right) dx_{\mathrm{T}} \tag{5.32}$$

In this expression $\overline{C_{\mathrm{it}}}(x_{\mathrm{T}})$ is the mean trap concentration level over energy. Parameter ΔE is defined as the energy window between which traps are located for which the carrier emission term is negligible, it is given by [104]:

$$\Delta E = E_{\mathrm{em,e}} - E_{\mathrm{em,h}} \tag{5.33}$$

Energy level $E_{\mathrm{em,e}}$ represents the upper energy level at which the (non steady state) emission process of electrons to the conduction band is negligible. It can be expressed as [104]:

$$E_{\mathrm{em,e}} - E_{\mathrm{i}} = -kT \ln \left(v_{\mathrm{th}} n_i t_{\mathrm{em,e}} \sigma \left(x_{\mathrm{T}} \right) + e^{\frac{E_{\mathrm{i}} - E_{\mathrm{F,inv}}}{kT}} \right) \tag{5.34}$$

Similarly energy level $E_{\mathrm{em,h}}$ is the lowest energy level at which the (non steady sate) emission process of holes towards the valence band is negligible. It is given by: [104]

$$E_{\mathrm{em,h}} - E_{\mathrm{i}} = kT \ln \left(v_{\mathrm{th}} n_{\mathrm{i}} t_{\mathrm{em,h}} \sigma \left(x_{\mathrm{T}} \right) + e^{\frac{E_{\mathrm{F,acc}} - E_{\mathrm{i}}}{kT}} \right) \tag{5.35}$$

In expressions 5.34 and 5.35, $E_{\mathrm{F,inv}}$ and $E_{\mathrm{F,acc}}$ are the Fermi levels at inversion and accumulation condition respectively. The exponential term limits the highest possible ΔE to the difference in Fermi levels between inversion and accumulation condition. Parameters $t_{\mathrm{em,e}}$ and $t_{\mathrm{em,h}}$ are the times available for the nonsteady state emission of electrons and holes; these are the times that the device is between accumulation and inversion conditions. For sinusoidal gate voltages these depend on frequency and are given by: [109]

$$t_{\mathrm{em,e}} = t_{\mathrm{em,h}} = \frac{Z}{2\pi f} \tag{5.36}$$

In this expression parameter Z is given by [109]:

$$Z = \sin^{-1} \left(\frac{2 \left| V_{\mathrm{FB}} - V_{\mathrm{bias}} \right|}{V_{\mathrm{pp}}} \right) + \sin^{-1} \left(\frac{2 \left| V_{\mathrm{T}} - V_{\mathrm{bias}} \right|}{V_{\mathrm{pp}}} \right) \tag{5.37}$$

The CP current at maximum CP conditions can now be described using expression 5.32 with ΔE given by equations 5.33 through 5.37 and $x_{\mathrm{T,max}}$ given by 5.31. Using this theoretical framework for describing $I_{\mathrm{cp,max}}$ as a function of frequency the observed roll-off in $Q_{\mathrm{cp,max}}$ as a function of frequency can be explained both qualitatively as well as quantitatively. This roll-off is caused by the interface traps that are too slow to respond to the CP signal applied at the gate. In the next section the implications this trap response has on the applicability of the RF CP technique will be investigated.

5.3.5 Applications of RF CP measurements

As can be seen from figures 5.13 and 5.14, the RF CP technique allows the extraction of an $I_{\mathrm{cp,max}}$ on dielectrics with a leakage current too high for conventional CP measurements. The theoretical framework given in the previous section states that at increasing frequencies an increasing number of interface traps is too slow to respond to the applied gate voltage signal. In this section the implications this trap response has on the applicability of CP measurements at radio frequencies will be investigated.

Trap distribution extraction

In the previous section a model was derived that is able to accurately describe the trap response to the CP measurements as a function of frequency, thereby providing an accurate explanation in the observed roll-off in the obtained $Q_{\mathrm{cp,max}}$ at frequencies above ~ 100 MHz. The model states that the number of traps that is fast enough to respond decreases with increasing frequency. From figure 5.14 it is known that for the leaky 1.4 nm devices used in this section $I_{\mathrm{cp,max}}$ can only be determined at frequencies above 1.5 GHz and 80 MHz for the n-type respectively p-type device. Both these frequencies are too high for all traps to respond. In order to get a good estimate of the number of traps that is not fast enough to respond with respect to the total number of interface traps it is needed to find a good description of the trap distribution within the oxide. $x_{\mathrm{T,max}}$ as it follows from 5.31 can then be used to evaluate the ratio of the number of traps fast enough to the total number of traps for the given measurement frequency.

As one does not know beforehand the total number of interface states on the leaky 1.4 nm devices it is impossible to perform such an analysis on these devices. The 3 nm devices used in this chapter, however do have a leakage current sufficiently low for CP measurements to be performed at frequencies where all traps are fast enough to respond. Therefore such an accuracy evaluation is performed on one of these 3 nm devices. This result may subsequently be used to evaluate the accuracy that can be achieved in extracting $\overline{D_{\mathrm{it}}}$ on the leaky 1.4 nm devices. In this section use is made of the results obtained on the 3 nm n-type device for this purpose. A complete trap distribution can be extracted by making use of the frequency response of $Q_{\mathrm{cp,max}}$ and the expressions given in equations 5.31 to 5.37, provided that values of $Q_{\mathrm{cp,max}}$ are available at frequencies low enough for all traps to respond. The latter can be verified by looking at figure 5.15 a): The linear slope of the $Q_{\mathrm{,cp,max}}$ vs. $\log(f)$ plot for frequencies up to at least 1 MHz is a direct indicator for this.

A trap distribution is extracted using the following approach: If one knows the amount of traps fast enough to respond to a gate voltage signal with frequency $f-\Delta f$, one can calculate the expected value of $Q_{\mathrm{cp,max}}$ at frequency f for the same number of traps. By comparing this value with the measured value of $Q_{\mathrm{cp,max}}$ at frequency f, $Q_{\mathrm{cp,meas}}$, one can extract the number of traps fast enough to respond to a gate voltage signal with frequency $f-\Delta f$, but too slow for frequency f. This value can be used to find the effective trap concentration $\overline{C_{\mathrm{it}}}$ in the interval $[x_{\mathrm{T,max}}, x_{\mathrm{T,max+\Delta x_T}}]$, where $x_{\mathrm{T,max}}$ and $x_{\mathrm{T,max}} + \Delta x_{\mathrm{T}}$ are related to $f-\Delta f$ respectively f through expression 5.31. For a sufficiently small value of Δx_{T}, $\overline{C_{\mathrm{it}}}$ may be assumed constant over the interval $[x_{\mathrm{T,max}}, x_{\mathrm{T,max}}+\Delta x_{\mathrm{T}}]$. This concentration over the entire interval will be denoted as $\overline{C_{\mathrm{it}}}(x_{\mathrm{T,max}})$. Using expressions 5.14 and 5.32 it can now be derived that:

$$\overline{C_{\mathrm{it}}}(x_{\mathrm{T,max}}(f)) = \frac{Q_{\mathrm{cp,all}}(f) - Q_{\mathrm{cp,slow}}(f) - Q_{\mathrm{cp,meas}}(f)}{q \cdot A_{\mathrm{G}} \cdot \int_{x_{\mathrm{low}}}^{x_{\mathrm{high}}} \Delta E\left(\sigma(x_{\mathrm{T}}), f\right) dx_{\mathrm{T}}} \tag{5.38}$$

In this expression $Q_{\mathrm{cp,all}}$ is the expected value of $Q_{\mathrm{cp,max}}$ if all interface traps

would be fast enough to respond; $Q_{\mathrm{cp,slow}}$ represents the contribution of all traps that are too slow to respond a gate voltage signal with frequency $f-\Delta f$. Furthermore integration limits x_{low} and x_{high} represent $x_{\mathrm{T,max}}(f)$ and $x_{\mathrm{T,max}}(f)+\Delta x_{\mathrm{T}}$ respectively. $Q_{\mathrm{cp,all}}$ and $Q_{\mathrm{cp,slow}}$ can be found using:

$$\begin{aligned} Q_{\mathrm{cp,all}}(f) &= q \cdot A_{\mathrm{G}} \cdot \overline{D_{\mathrm{it}}} \cdot \Delta E(\sigma_{\mathrm{eff}}, f) \\ Q_{\mathrm{cp,slow}}(f) &= q \cdot A_{\mathrm{G}} \cdot \int_{x_{\mathrm{high}}}^{x_{\mathrm{T,slow}}} \overline{C_{\mathrm{it}}}(x_{\mathrm{T}}) \cdot \Delta E(\sigma(x_{\mathrm{T}}), f) dx_{\mathrm{T}} \end{aligned} \tag{5.39}$$

Parameter $x_{\mathrm{T,slow}}$ in this expression represents any value of $x_{\mathrm{T,max}}$ beyond the location of the slowest interface traps. Here the value of $x_{\mathrm{T,max}}$ for a frequency of 100 kHz is used, as it follows from expression 5.31. As can be seen from expressions 5.38 and 5.39 the effective interface state density $\overline{D_{\mathrm{it}}}$ is needed in order to extract $\overline{C_{\mathrm{it}}}(x_{\mathrm{T}})$. This parameter can be extracted very accurately on the 3 nm n-type device as used in this section, by making use of the approach discussed in [109]: $\overline{D_{\mathrm{it}}}$ is proportional to the linear slope of the $Q_{\mathrm{cp,max}}$ vs. $\log(f)$ for frequencies of up to 1 MHz and is found to be $3.0 \cdot 10^{10}$ cm^{-2}eV^{-1}. Parameter σ_{eff}, representing the effective capture cross section of the complete trap population also follows directly from these low frequency results and is found to be $1.2 \cdot 10^{-16}$ cm^2.

Using these values, expressions 5.38 and 5.39, and the measured frequency response of $Q_{\mathrm{cp,max}}$, it is now possible to extract a trap distribution for the 3 nm n-type device. The result of this is shown in figure 5.16. The inset of figure 5.16 shows the frequency response of $Q_{\mathrm{cp,max}}$ for the 3 nm n-type device after the 1 nA DC offset correction for the results obtained with the VNA, as discussed in section 5.3.3. The figure makes use of 2 x-axes: The bottom x-axis represents trap distance (in this figure the distance away from the position where $\sigma(x_{\mathrm{T}}) = 10^{-14}$ cm^2 is used). The top x-axis represents the frequency for which $x_{\mathrm{T,max}}$ is the associated value of x_{T} at the bottom x-axis as it follows from equation 5.31. From figure 5.16 it can clearly be seen that for frequencies up to 2 MHz $x_{\mathrm{T,max}}$ is sufficiently high for the complete trap population to follow the gate voltage signal. For higher frequencies an increasing number is too slow to respond to the gate voltage signal and to contribute to the CP current. This extraction procedure is a very useful side benefit of the RF CP technique: for relatively thick oxides a complete trap distribution can be extracted. This feature is not available with conventional CP measurements. For very leaky oxides this trap extraction procedure may not be used, as information on both the fast as well as the slow interface traps is needed. The slow interface traps can not be probed on very leaky oxides as shown in figure 5.14.

Application in stress experiments

In order to show a typical example for which the RF CP technique proves very useful on ultra-leaky devices the RF CP technique will also be applied within a stress experiment. Because of the high sensitivity to even a very small increase in

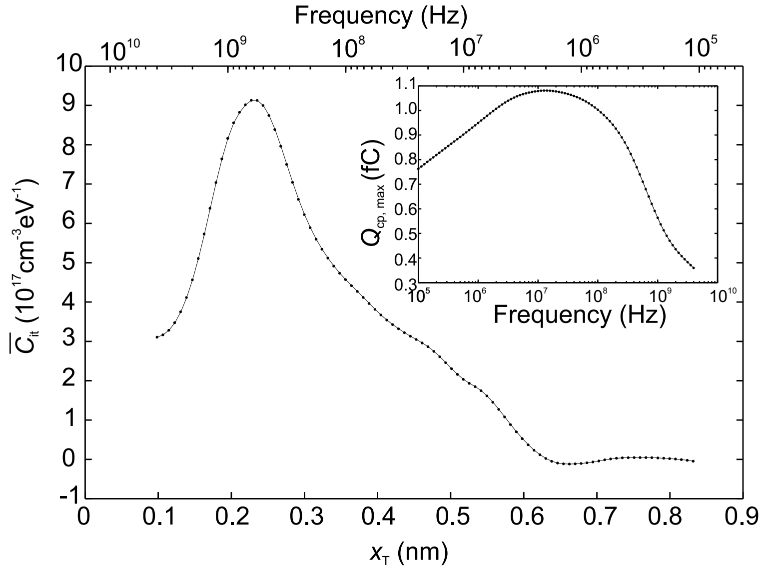

Figure 5.16: Extracted trap distribution obtained on an n-type device with $t_{ox} = 3$ nm. The inset shows the $Q_{cp,max}$ vs. f plot from which this distribution is extracted. The bottom x-axis is the distance from a trap towards the position where $\sigma(x_T) = 10^{-14}$ cm^2. The upper x-axis shows the frequency where the maximum probing depth equals the associated value on the bottom x-axis.

the interface traps density, CP measurements are often used in accelerated stress experiments for obtaining a device's lifetime. Use is made of a 3 nm n-type device that shows no considerably high leakage current. This allows us to compare results over a very wide frequency range. CP currents were measured using a V_{pp} of 2 V and frequencies ranging from 10 MHz to 4 GHz. Subsequently the device was stressed using a constant gate voltage stress of 3.25 V and the CP currents were measured after a stress of 10 s as well as 100 s. The result of this experiment is shown in figure 5.17. In this figure both the measured $I_{cp,max}$ as well as $Q_{cp,max}$ can be seen over the entire frequency range. Clearly a large increase in $I_{cp,max}$ as well as $Q_{cp,max}$ can be seen after stressing the device. For very leaky devices it may be needed to use frequencies into the GHz range before a CP current starts to emerge, such as the 1.5 GHz needed for the 1.4 nm n-type device used in this paper. Figure 5.17 shows that at these frequencies an increase in the interface state density can very well be detected by an increase in the CP current. This indicates that RF CP measurements may very well be applied for investigating the lifetime of very leaky devices.

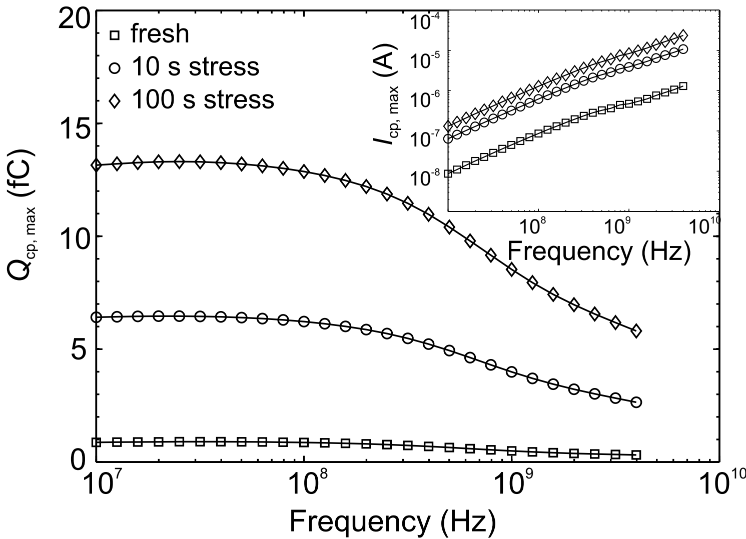

Figure 5.17: Measured $Q_{\mathrm{cp,max}}$ plotted against frequency before and after a constant gate voltage stress of 3.25 V. The inset shows the corresponding values of $I_{\mathrm{cp,max}}$. The figure clearly shows an increase in both $I_{\mathrm{cp,max}}$ and $Q_{\mathrm{cp,max}}$ over the entire frequency range after stress.

5.3.6 Discussion

The RF CP technique, as discussed in this section proves to be useful for obtaining CP curves on ultra-thin oxide devices. The voltage generation procedure of section 2.4 can be applied for generating RF voltage signals. These signals can be superimposed on a sweeping DC voltage source in order to obtain accurate CP curves.

The results presented in figures 5.15, 5.16 and 5.17 reveal that at frequencies exceeding 1 MHz an increasing number of interface states is too slow to respond to the applied gate voltage signal. As a consequence not all traps are probed at these high frequencies. Using the model discussed in section 5.3.4 the interface trap response to RF gate voltage signals can be described. The model is based on well-accepted models described in literature. A more thorough analysis would possibly lead to even more accurate description, but this would lead to impracticable expressions and for a first-order model the presented model is sufficiently accurate. It gives a physical insight on the frequency response of RF CP measurements.

On devices that show only moderate gate leakage, this trap response model, in combination with RF CP measurement results, can be used to extract the trap distribution in the oxide. This is an asset that is not available with conventional CP theory and measurements. The extraction procedure does not allow it to be used for ultra-thin oxide devices however.

In section 5.3.5 different applications of the RF CP technique are discussed. Besides the trap distribution extraction on moderately leaky devices, another important application is the use in stress experiments. It was shown that an increase in CP current can be observed over the entire frequency range from 10 MHz to 4 GHz after a constant gate voltage stress. This means that RF CP measurements over this entire frequency range can be used for determining a device lifetime under a gate voltage stress on ultra-thin oxide devices.

5.4 Conclusions

In this chapter the use of RF measurement techniques was presented as a solution to problems encountered with the characterization of ultra-thin oxide devices. C-V and CP measurements can provide accurate information on the number of oxide traps and the number of interface states in a MOS device and are therefore widely used in the reliability evaluation of MOS devices. With decreasing dimensions of the oxide thickness the gate tunneling current is known to become very large. This tunneling current complicates both C-V and CP measurements. This can be solved by increasing the measurement frequency with the use of RF measurement techniques.

Previously it had been shown that the RF C-V technique is a good solution to measurement inaccuracies due to the gate tunneling current in C-V measurements. In this chapter the gate extraction procedure in RF C-V measurements was evaluated. The two-port parameters of RF C-V test structures were derived and based on this two-port analysis a new extraction methodology was derived. It was shown the new extraction methodology is accurate in extracting the gate capacitance for devices biased in accumulation, over a broader frequency range compared to the conventionally used extraction procedure. This gain in frequency range means that even leakier devices can be characterized using RF C-V measurements than before. This was verified using both device simulations and measurements.

In this chapter it was furthermore shown that the RF CP technique can be used to overcome the leakage current issues in CP measurements. CP curves can be accurately obtained on devices that show a leakage current density far too high for use in conventional CP measurements. The RF CP technique makes use of the voltage generation procedure discussed in section 2.4. Measurements results are presented that show CP curves on ultra-thin oxide devices that exhibit a leakage current too high for use in conventional CP measurements. A model was derived that describes the interface trap response to the frequency of the applied sinusoidal gate voltage signals. It appears that at increasing frequencies an increasing number of interface traps is too slow to respond to the gate voltage signal. The model can be used to extract a trap distribution on devices that do not show a moderate gate leakage current. As an example for the area of application of the RF CP technique RF CP measurements were applied within a stress experiment. It was shown that an increase on the measured CP current can very well be used to evaluate the degradation of devices, using frequencies of

up to 4 GHz.

Both the RF C-V and the RF CP technique are very valuable techniques that can be applied in the reliability evaluation of MOS devices. The use of high frequencies in this chapter is appreciated in a very different way compared to e.g. the work discussed in chapter 3. In chapter 2 it was explained that the high frequencies posed problems on the measurements to be performed in chapter 3. In this chapter on the other hand the high measurement frequencies proved to be the solution.

Chapter 6

Conclusions and recommendations

6.1 General conclusions

In this thesis important advancements have been made in developing reliability engineering tools for RF CMOS. Reliability engineering covers a broad range of topics; in this thesis new developments in three of these topics have been presented, with an eye on application in RF CMOS. This includes a physical understanding of degradation mechanisms in RF CMOS, accurate lifetime prediction of RF CMOS circuits as well as the development of advanced characterization techniques for assessing device degradation in RF CMOS. The most important of these advancements are summarized below:

- Hot carrier degradation as well as NBTI degradation have been shown to be frequency independent mechanisms for frequencies up to 3.2 GHz. For the hot carrier effect this means that quasi-static models can be used for describing degradation in RF circuits. In NBTI degradation a good model for describing AC degradation is still needed. The experimental results of chapter 3 indicate that no difference should be expected between low frequency NBTI degradation and RF NBTI degradation

- Gate-oxide breakdown has been shown to exhibit a strong frequency dependence between the MHz range and the GHz range for unipolar, sinusoidal stress voltage signals. At increasing frequencies t_{bd} values increase, a lifetime enhancement of a factor of 14 has been observed between 5.6 MHz and 1.8 GHz, this has never been shown before. If properly utilized, this observed lifetime increase can allow circuit designers to design circuits with higher voltage levels for RF circuits, thereby boosting performance while maintaining reliability specifications.

- A new reliability simulator has been developed, that can further relax design guidelines for RF circuit designers, as explained in chapter 4. The new simulator is capable of predicting the lifetime of RF PA's that can withstand multiple breakdown events and simultaneously suffer from hot carrier degradation. Up to present no reliability simulator has ever provided the possibility of determining the lifetime of circuits that can withstand multiple breakdown events. A lifetime enhancement of a factor of over 100 has been observed by comparing simulation results with lifetime prediction of circuits on the basis of the first breakdown event, as conventionally done. This is a dramatic increase.

- A new extraction methodology has been developed for accurately extracting the gate capacitance from two-port RF C-V measurements. The new extraction methodology allows the use of much higher frequencies for an accurate extraction of the gate capacitance. This implies that the minimum allowable oxide thickness for which RF C-V provide accurate results can be reduced further. This development is useful for the reliability evaluation of both RF CMOS as well as digital CMOS.

- It has been shown that charge pumping curves can be obtained using measurement frequencies far beyond frequencies used in conventional charge pumping measurements. The interface state response to these measurements has also been accurately modeled. The benefit of using such high measurement frequencies is twofold. First it allows an extraction of the interface state density profile as a function of distance with respect to the interface. Second it can be used to evaluate the interface state increase after stress experiments on ultra-thin oxide devices. These thin-oxide devices exhibit a leakage current too high for conventional CP currents to be used. This may be applied for both RF CMOS as well as digital CMOS.

To summarize, a better understanding of MOSFET degradation under RF conditions is gained and tools are developed that allow for a better lifetime prediction of RF circuits. This in combination with the development of new RF characterization techniques for CMOS reliability, have initiated *reliability engineering in RF CMOS* as a new and important field of reliability engineering.

6.2 Recommendations for future work

While this thesis describes a lot of advancements made in the field of reliability engineering for RF CMOS, it also poses some new questions on how to proceed further. These questions will be discussed below.

- The RF stress experiments reveal that RF NBTI degradation may be accurately modeled using AC NBTI degradation models. However such models are not available at present. This means that, before RF NBTI can accurately be taken into account for lifetime prediction, AC NBTI degradation

should be modeled accurately. For this purpose models explicitly including time constants in recovery effects should be taken into account. Once AC NBTI models are available, the simulator of chapter 4 might be extended with NBTI degradation as well.

- A lifetime enhancement with respect to gate-oxide breakdown has been confirmed experimentally in this thesis. Making use of this observation in lifetime prediction is however not straightforward using only this experimental evidence. For accurate lifetime prediction to be available a good model describing this lifetime enhancement should be used. In this thesis possible explanations have been obtained from literature, relating this effect to hole trapping kinetics and the detrapping of holes during the off-state of the stress.

- The reliability simulator as discussed in this thesis makes use of a relatively simple model for degraded MOSFET operation. The development of the simulator was focused on the procedure to deal with multiple breakdown events rather than achieving high levels of accuracy. Now that the benefit of the new simulator approach has been verified by comparing different PA's, an important next step is to increase the accuracy of the used models. For instance the degradation of capacitances should be well investigated especially if RF circuits are investigated. Furthermore accuracy will be increased by applying more advanced models than the lucky electron model for describing hot carrier degradation in present-day CMOS technologies. Also the possibility of having breakdown paths outside the gate-drain or gate-source overlap regions should somehow be incorporated in the simulator for making the simulator more generically applicable besides in RF PA design.

- The two characterization techniques that have been discussed in chapter 5 can be used for determining the amount of degradation in MOSFETs. At present however these characterization methods have not been actually applied in the reliability evaluation of MOSFETs. These characterization methods might give a better insight on the RF degradation mechanisms discussed in chapter 3. For this purpose it might prove to be useful to combine RF C-V and RF CP measurements with RF stress experiments.

While even more directions of future work could be formulated, the above are the most important. If all these directions are followed, the topic of *Reliability Engineering in RF CMOS* will evolve into a mature field of reliability engineering.

Bibliography

[1] P.H. Woerlee, M.J. Knitel, R. van Langevelde, D.B.M. Klaassen, L.F. Tiemeijer, A.J. Scholten, and A.T.A. Zegers-van Duijnhoven, "RF-CMOS performance trends," *IEEE Trans. Electron Dev.*, vol. 48, no. 8, pp. 1776–1782, 2001.

[2] M. Ohring, *Reliability and failure of electronic materials and devices.* Academic Press, 1998.

[3] J.B. Bernstein, M. Gurfinkel, X. Li, J. Walters, Y. Shapira, and M. Talmor, "Electronic circuit reliability modeling," *Microelectronics Reliability*, vol. 46, no. 12, pp. 1957–1979, 2006.

[4] I.-C. Chen, S.E. Holland, and C. Hu, "Electrical breakdown in thin gate and tunneling oxides," *IEEE Trans. Electron Dev.*, vol. 32, no. 2, pp. 413–422, 1985.

[5] K.F. Schuegraf and C. Hu, "Hole injection SiO_2 breakdown model for very low voltage lifetime extrapolation," *IEEE Trans. Electron Dev.*, vol. 41, no. 5, pp. 761–767, 1994.

[6] D.J. DiMaria and J.W. Stasiak, "Trap creation in silicon dioxide produced by hot electrons," *J. Appl. Phys.*, vol. 65, no. 6, pp. 2342–2356, 1989.

[7] J.W. McPherson and H.C. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO_2 thin films," *J. Appl. Phys.*, vol. 84, no. 3, pp. 1513–1523, 1998.

[8] R. Degraeve, G. Groeseneken, R. Bellens, J.L. Ogier, M. Depas, P.J. Roussel, and H.E. Maes, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," *IEEE Trans. Electron Dev.*, vol. 45, no. 4, pp. 904–911, 1998.

[9] M. Depas, T. Nigam, and M.M. Heyns, "Soft breakdown of ultra-thin gate oxide layers," *IEEE Trans. Electron Dev.*, vol. 43, no. 9, pp. 1499–1504, 1996.

[10] F. Monsieur, E. Vincent, D. Roy, S. Bruyere, J.C. Vildeuil, G. Pananakakis, and G. Ghibaudo, "A thorough investigation of progressive breakdown in ultra-thin oxides. Physical understanding and application for industrial reliability assessment," in *Proceedings of the International Reliability Physics Symposium*, 2002, pp. 45–54.

[11] Y. Miura and Y. Matukura, "Investigation of silicon-silicon dioxide interface using MOS structure," *Japanese Journal of Applied Physics*, vol. 5, no. 2, p. 180, 1966.

[12] G. La Rosa, F. Guarin, S. Rauch, A. Acovic, J. Lukaitis, and E. Crabbe, "NBTI-channel hot carrier effects in pMOSFETs in advanced CMOS technologies," in *Proceedings of the International Reliability Physics Symposium*, 1997, pp. 282–286.

[13] R. Thewes, R. Brederlow, C. Schlünder, P. Wieczorek, A. Hesener, B. Ankele, P. Klein, S. Kessel, and W. Weber, "Device reliability in analog CMOS applications," in *Proceedings of the International Electron Devices Meeting*, 1999, pp. 81–84.

[14] N. Kimizuka, K. Yamaguchi, K. Imai, T. Iizuka, C.T. Liu, R.C. Keller, and T. Horiuchi, "NBTI enhancement by nitrogen incorporation into ultrathin gate oxide for 0.10-μm gate CMOS generation," in *Digest of Papers, symposium on VLSI technology*, 2000, pp. 92–93.

[15] M.A. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, "A comprehensive model for pMOS NBTI degradation: Recent progress," *Microelectronics Reliability*, vol. 47, no. 6, pp. 853–862, 2007.

[16] S. Chakravarthi, A.T. Krishnan, V. Reddy, and S. Krishnan, "Probing negative bias temperature instability using a continuum numerical framework: Physics to real world operation," *Microelectronics Reliability*, vol. 47, no. 6, pp. 863–872, 2007.

[17] D.K. Schroder, "Negative bias temperature instability: What do we understand?" *Microelectronics Reliability*, vol. 47, no. 6, pp. 841–852, 2007.

[18] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanism to modelling," *Microelectronics Reliability*, vol. 46, no. 6, pp. 1–23, 2006.

[19] *Low level measurements handbook*, 6th ed., Keithley Instruments Inc., 2004.

[20] D.M. Pozar, *Microwave Engineering.* John Wiley & Sons, 1998.

[21] K. Kurokawa, "Power waves and the scattering matrix," *IEEE Trans. Microwave Theory and Techniques*, vol. 13, no. 2, pp. 194–202, 1965.

[22] D. Rytting, “Calibration and error correction techniques for network analysis,” seminar held at OGI center for professional development, 30-09-2004. Slides are available at http://cpd.ogi.edu/IEEE-MTT-ED/DougRyttingSeminar.htm.

[23] M.C.A.M. Koolen, J.A.M. Geelen, and M.P.J.G. Versleijen, “An improved de-embedding technique for on-wafer high-frequency characterization,” in *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting*, 1991, pp. 188–191.

[24] J. Verspecht, “Large-signal network analysis,” *IEEE Microwave Magazine*, vol. 6, no. 4, pp. 82–92, 2005.

[25] M. Spirito and L.C.N. de Vreede, 2008, private communication.

[26] C. Hu, “Reliability phenomena under AC stress,” *Microelectronics Reliability*, vol. 38, no. 1, pp. 1–5, 1998.

[27] L. Larcher, D. Sanzogni, R. Brama, A. Mazzanti, and F. Svelto, “Oxide breakdown after RF stress: Experimental analysis and effects on power amplifier operation,” in *Proceedings of the International Reliability Physics Symposium*, 2006, pp. 283–288.

[28] C.D. Presti, F. Carrara, A. Scuderi, S. Lombardo, and G. Palmisano, “Degradation mechanisms in CMOS power amplifiers subject to radio-frequency stress and comparison to the DC case,” in *Proceedings of the International Reliability Physics Symposium*, 2007, pp. 86–92.

[29] E. Takeda, C.Y. Yang, and A.M. Miura-Hamada, *Hot-Carrier effects in MOS devices.* Academic Press, 1995.

[30] R. Woltjer, G.M. Paulzen, H.G. Pomp, H. Lifka, and P.H. Woerlee, “Three hot-carrier degradation mechanisms in deep-submicron pMOSFET’s,” *IEEE Trans. Electron Dev.*, vol. 42, no. 1, pp. 109–115, 1995.

[31] A. Acovic, G. La Rosa, and Y.-C. Sun, “A review of hot-carrier degradation mechanisms in MOSFETs,” *Microelectronics Reliability*, vol. 36, no. 7-8, pp. 845–869, 1996.

[32] G. Groeseneken, R. Degraeve, T. Nigam, G. Van den bosch, and H.E. Maes, “Hot carrier degradation and time-dependent dielectric breakdown in oxides,” *Microelectronic Engineering*, vol. 49, no. 1, pp. 27–40, 1999.

[33] G. La Rosa and S.E. Rauch III, “Channel hot carrier effects in n-MOSFET devices of advanced submicron CMOS technologies,” *Microelectronics Reliability*, vol. 47, no. 4-5, pp. 552–558, 2007.

[34] S.-G. Lee, J.-M. Hwang, and H.-D. Lee, “Experimental evidence for non-lucky electron model effect in 0.15-μm NMOSFETs,” *IEEE Trans. Electron Dev.*, vol. 49, no. 11, pp. 1876–1881, 2002.

[35] R. Bellens, P. Heremans, G. Groeseneken, H.E. Maes, and W. Weber, "The influence of the measurement setup on enhanced AC hot carrier degradation of MOSFET's," *IEEE Trans. Electron Dev.*, vol. 37, no. 1, pp. 310–313, 1990.

[36] E.S. Snyder, D.V. Campbell, S.E. Swanson, and D.G. Pierce, "Novel self-stressing test structures for realistic high-frequency reliability characterization," in *Proceedings of the International Reliability Physics Symposium*, 1993, pp. 57–65.

[37] V.-H. Chan, T.E. Kopley, P. Marcoux, and J.E Chung, "High-frequency AC hot-carrier degradation in CMOS circuits," in *Proceedings of the International Electron Devices Meeting*, 1994, pp. 299–302.

[38] C. Jiang and E. Johnson, "AC hot-carrier degradation in a voltage controlled oscillator," in *Proceedings of the International Reliability Physics Symposium*, 1993, pp. 53–56.

[39] S. Yoshida, T. Matsui, K. Okuyama, and K. Kubota, "AC hot-carrier degradation in the super-100MHz operation range," in *Digest of Papers, symposium on VLSI technology*, 1994, pp. 145–146.

[40] K.R. Mistry and B. Doyle, "AC versus DC hot-carrier degradation in n-channel MOSFET's," *IEEE Trans. Electron Dev.*, vol. 40, no. 1, pp. 96–104, 1993.

[41] P. Habas, "Hot-carrier aging of nMOST in analog circuits with large periodic drain signal," *Solid-State Electronics*, vol. 50, no. 5, pp. 727–732, 2006.

[42] M. Tsuno, M. Suga, M. Tanaka, K. Shibahara, M. Miura-Mattausch, and M. Hirose, "Physically-based threshold voltage determination for MOSFET's of all gate lengths," *IEEE Trans. Electron Dev.*, vol. 46, no. 7, pp. 1429–1434, 1999.

[43] T. Grasser and S. Selberherr, "Special Section: Modelling the negative bias temperature instability," *Microelectronics Reliability*, vol. 46, no. 6, pp. 839–911, 2007.

[44] B. Zhu, J.S. Suehle, Y. Chen, and J.B. Bernstein, "Negative bias temperature instability of deep-submicron p-MOSFETs under pulsed bias stress," in *Integrated Reliability Workshop Final Report*, 2002, pp. 125–129.

[45] W. Abadeer and W. Ellis, "Behavior of NBTI under AC dynamic circuit conditions," in *Proceedings of the International Reliability Physics Symposium*, 2003, pp. 17–22.

[46] S. Mahapatra, M.A. Alam, P. Bharath Kumar, T.R. Dalei, and D. Saha, "Mechanism of negative bias temperature instability in CMOS devices: Degradation, recovery and impact of nitrogen," in *Proceedings of the International Electron Devices Meeting*, 2004, pp. 105–108.

[47] A.T. Krishnan, C. Chancellor, S. Chakravarthi, P.E. Nicollian, V. Reddy, A. Varghese, R.B. Khamankar, and S. Krishnan, "Material dependence of hydrogen diffusion: Implications for NBTI degradation," in *Proceedings of the International Electron Devices Meeting*, 2005, pp. 688–691.

[48] S.S. Tan, T.P. Chen, and L. Chan, "Dynamic NBTI lifetime model for inverter-like waveform," *Microelectronics Reliability*, vol. 45, no. 7-8, pp. 1115–1118, 2005.

[49] T. Nigam and E.B. Harris, "Lifetime enhancement under high frequency NBTI measured on ring oscillators," in *Proceedings of the International Reliability Physics Symposium*, 2006, pp. 289–293.

[50] V. Huard and M. Denais, "Hole trapping effect on methodology for DC and AC negative bias temperature instability measurements in pMOS transistors," in *Proceedings of the International Reliability Physics Symposium*, 2004, pp. 40–45.

[51] G. Chen, M.F. Li, C.H. Ang, J.Z. Zheng, and D.L. Kwong, "Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling," *IEEE Electron Device Lett.*, vol. 23, no. 12, pp. 734–736, 2002.

[52] Y. Mitani, "Influence of nitrogen in ultra-thin SiON on negative bias temperature instability under AC stress," in *Proceedings of the International Electron Devices Meeting*, 2004, pp. 117–120.

[53] R. Fernández, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodríguez, M. Nafría, and G. Groeseneken, "AC NBTI studied in the 1 Hz-2 GHz range on dedicated on-chip CMOS circuits," in *Proceedings of the International Electron Devices Meeting*, 2006, pp. 1–4.

[54] D.L. Crook, "Method of determining reliability screens for time dependent dielectric breakdown," in *Proceedings of the International Reliability Physics Symposium*, 1979, pp. 1–7.

[55] E.S. Anolick and G.R. Nelson, "Low field time dependent dielectric integrity," in *Proceedings of the International Reliability Physics Symposium*, 1979, pp. 8–12.

[56] P.E. Nicollian, W.R. Hunter, and J.C. Hu, "Experimental evidence for voltage driven breakdown models in ultrathin gate oxides," in *Proceedings of the International Reliability Physics Symposium*, 2000, pp. 7–15.

[57] E.Y. Wu, A. Vayshenker, E. Nowak, J. Suñé, R-P. Vollertsen, W. Lai, and D. Harmon, "Experimental evidence of $t_{\rm bd}$ power-law for voltage dependence of oxide breakdown in ultrathin gate oxides," *IEEE Trans. Electron Dev.*, vol. 49, no. 12, pp. 2244–2253, 2002.

[58] M.-S. Liang, S. Haddad, W. Cox, and S. Cagnina, "Degradation of very thin gate oxide MOS devices under dynamic high field/current stress," in *Proceedings of the International Electron Devices Meeting*, 1986, pp. 394–398.

[59] Y. Fong, I.C. Chen, S. Holland, J. Lee, and C. Hu, "Dynamic stressing of thin oxides," in *Proceedings of the International Electron Devices Meeting*, 1986, pp. 664–667.

[60] E. Rosenbaum, Z. Liu, and C. Hu, "Silicon dioxide breakdown lifetime enhancement under bipolar bias conditions," *IEEE Trans. Electron Dev.*, vol. 40, no. 12, pp. 2287–2295, 1993.

[61] J.S. Suehle and P. Chaparala, "Low electric field breakdown of thin SiO_2 films under static and dynamic stress," *IEEE Trans. Electron Dev.*, vol. 44, no. 5, pp. 801–808, 1997.

[62] G. Ghidini, D. Brazelli, C. Clementi, and F. Pellizzer, "Charge trapping mechanism under dynamic stress and its effect on failure time," in *Proceedings of the International Reliability Physics Symposium*, 1999, pp. 88–92.

[63] B. Wang, J.S. Suehle, E.M. Vogel, and J.B. Bernstein, "Time-dependent dielectric breakdown of ultra-thin SiO_2 gate dielectrics under pulsed bias stress," *IEEE Electron Device Lett.*, vol. 22, no. 5, pp. 224–226, 2001.

[64] E.Y. Wu, J. Suñé, and W. Lai, "On the Weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination-part II: Experimental results and the effects of stress conditions," *IEEE Trans. Electron Dev.*, vol. 49, no. 12, pp. 2141–2150, 2002.

[65] B. Kaczer, R. Degraeve, G. Groeseneken, M. Rasras, S. Kubicek, E. Vandamme, and G. Badenes, "Impact of MOSFET oxide breakdown on digital circuit operation and reliability," in *Proceedings of the International Electron Devices Meeting*, 2000, pp. 553–556.

[66] L. Pantisano and K.P. Cheung, "The impact of postbreakdown gate leakage on MOSFET RF performances," *IEEE Electron Device Lett.*, vol. 22, no. 12, pp. 585–587, 2001.

[67] H. Yang, W. Smith, and J.S. Yuan, "Gate Oxide breakdown on low noise and power amplifier performance," in *Proceedings of the Radio Frequency Integrated Circuits Symposium*, 2003, pp. 663–666.

[68] C. Yu and J.S. Yuan, "MOS RF reliability subject to dynamic voltage stressmodeling and analysis," *IEEE Trans. Electron Dev.*, vol. 52, no. 8, pp. 1751–1758, 2005.

[69] J.S. Yuan and J. Ma, "Evaluation of RF-stress effect on class-E MOS power-amplifier efficiency," *IEEE Trans. Electron Dev.*, vol. 55, no. 1, pp. 430–434, 2008.

[70] D. S. Aur and and P. Yang, "Circuit hot electron effect simulation," in *Proceedings of the International Electron Devices Meeting*, 1987, pp. 498–501.

[71] C. Hu, "IC reliabilitv simulation," *IEEE J. Solid-State Circuits*, vol. 27, no. 3, pp. 241–246, 1992.

[72] W.-J. Hsu, B.J. Sheu, S.M. Gowda, and C-G. Hwang, "Advanced integrated-circuit reliability simulation including dynamic stress effects," *IEEE J. Solid-State Circuits*, vol. 27, no. 3, pp. 247–257, 1992.

[73] M.M. Lunenborg, "MOSFET hot-carrier degradation: Failure mechanisms and models for reliability circuit simulation," Ph.D. dissertation, University of Twente, 1996.

[74] "Reliability simulation in integrated circuit design," cadence white paper, available from http://www.cadence.com/whitepapers/5082_ReliabilitySim_FNL_WP.pdf.

[75] M. Karam, W. Fikry, H. Haddara, and H. Ragai, "Implementation of hot-carrier reliability simulation in Eldo," in *Proceedings of International Symposium on Circuits and Systems*, 2001, pp. 515–518.

[76] E. Rosenbaum, P.M. Lee, R. Moazzami, P.K. Ko, and C. Hu, "Circuit reliability simulator-oxide breakdown module," in *Proceedings of the International Electron Devices Meeting*, 1989, pp. 331–334.

[77] M.A. Alam, B.E. Weir, P.J. Silverman, Y. Ma, and D. Hwang, "The statistical distribution of percolation resistance as a probe into the mechanics of ultra-thin oxide breakdown," in *Proceedings of the International Electron Devices Meeting*, 2000, pp. 529–532.

[78] B. Kaczer, R. Degraeve, A. De Keersgieter, K. Van de Mieroop, V. Simons, and G. Groeseneken, "Consistent model for short-channel nMOSFET after hard gate oxide breakdown," *IEEE Trans. Electron Dev.*, vol. 49, no. 3, pp. 507–513, 2002.

[79] B. Kaczer, F. Crupi, R. Degraeve, P. Roussel, C. Ciofi, and G. Groeseneken, "Observation of hot-carrier-induced nFET gate-oxide breakdown in dynamically stressed CMOS circuits," in *Proceedings of the International Electron Devices Meeting*, 2002.

[80] F. Crupi, B. Kaczer, G. Groeseneken, and A. De Keersgieter, "New Insights Into the Relation Between Channel Hot Carrier Degradation and Oxide Breakdown in Short Channel nMOSFETs," *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 278–280, 2003.

[81] J.-T. Park, B.-J. Lee, D.-W. Kim, C.-G. Yu, and H.-K. Yu, "RF performance degradation in nMOS transistors due to hot carrier effects," *IEEE Trans. Electron Dev.*, vol. 47, no. 5, pp. 1068–1072, 2000.

[82] Q. Li, J. Zhang, W. Li, J.S. Yuan, Y. Chen, and A.S. Oates, "RF performance degradation in nMOS transistors due to hot carrier effects," *IEEE Trans. Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1546–1551, 2001.

[83] S. Naseh, M. Jamal Deen, and O. Marinov, "Effects of hot-carrier stress on the RF performance of 0.18 μm technology NMOSFETs and circuits," in *Proceedings of the International Reliability Physics Symposium*, 2002, pp. 98–104.

[84] L. Pantisano, D. Schreurs, B. Kaczer, W. Jeamsaksiri, R. Venegas, R. Degraeve, K.P. Cheung, and G. Groeseneken, "RF performance vulnerability to hot carrier stress and consequent breakdown in low power 90 nm RFCMOS," in *Proceedings of the International Electron Devices Meeting*, 2003, pp. 181–184.

[85] E. Xiao, J.S. Yuan, and H. Yang, "CMOS RF and DC reliability subject to hot carrier stress and oxide soft breakdown," vol. 4, no. 11, pp. 92–98, 2004.

[86] J.P. Walko and B. Abadeer, "RF s-parameter degradation under hot carrier stress," in *Proceedings of the International Reliability Physics Symposium*, 2004, pp. 422–425.

[87] W.-C. Lin, T.-C. Wu, Y.-H. Tsai, L.-J. Du, and Y.-C. King, "Reliability evaluation of class-E and class-A power amplifiers with nanoscaled CMOS technology," *IEEE Trans. Electron Dev.*, vol. 52, no. 7, pp. 1478–1483, 2005.

[88] S. Naseh, M. Jamal Deen, and C.-H. Chen, "Hot-carrier reliability of submicron NMOSFETs and integrated NMOS low noise amplifiers," *Microelectronics Reliability*, vol. 46, no. 2-4, pp. 201–212, 2006.

[89] C. Hu, S.C. Tam, F.-C. Hsu, P.-K. Ko, T.-Y. Chan, and K.W. Terril, "Hot-electron-induced MOSFET degradation-model, monitor, and improvement," *IEEE Trans. Electron Dev.*, vol. 32, no. 2, pp. 375–385, 1985.

[90] E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection," *IEEE Electron Device Lett.*, vol. 4, no. 4, pp. 111–113, 1983.

[91] J. Suñé, E.Y. Wu, and W.L. Lai, "Successive oxide breakdown statistics: Correlation effects, reliability methodologies, and their limits," *IEEE Trans. Electron Dev.*, vol. 51, no. 10, pp. 1584–1592, 2004.

[92] M. Acar, A.J. Annema, and B. Nauta, "Variable-voltage class-E power amplifiers," in *Proceedings of International Microwave Symposium*, 2007, pp. 1095–1098.

[93] K.L. Yang and C. Hu, "MOS capacitance measurements for high-leakage thin dielectrics," *IEEE Trans. Electron Dev.*, vol. 46, no. 7, pp. 1500–1501, 1999.

[94] W.K. Henson, K.Z. Ahmed, E.M. Vogel, J.R. Hauser, J.J. Wortman, R.D. Venables, M. Xu, and D. Venables, “Estimating oxide thickness of tunnel oxides down to 1.4 nm using conventional capacitance-voltage measurements on MOS capacitors,” *IEEE Electron Device Lett.*, vol. 20, no. 4, pp. 179–181, 1999.

[95] C.-H. Choi, J.-S. Goo, T.-Y. Oh, Z. Yu, R.W. Dutton, A. Bayoumi, M. Cao, P. Vande Voorde, D. Vook, and C.H. Diaz, “MOS C-V characterization of ultrathin gate oxide thickness (1.3 - 1.8 nm),” *IEEE Electron Device Lett.*, vol. 20, no. 6, pp. 292–294, 1999.

[96] D.W. Barlage, J.T. O’Keeffe, J.T. Kavalieros, M.M. Nguyen, and R.S. Chau, “Inversion MOS capacitance extraction for high-leakage dielectrics using a transmission line equivalent circuit,” *IEEE Electron Device Lett.*, vol. 21, no. 9, pp. 454–456, 2000.

[97] J. Schmitz, F.N Cubaynes, R.J. Havens, R. de Kort, A.J. Scholten, and L.F. Tiemeijer, “RF capacitance-voltage characterization of MOSFETs with high leakage dielectrics,” *IEEE Electron Device Lett.*, vol. 24, no. 1, pp. 37–39, 2003.

[98] P. Masson, J.-L. Autran, and J. Brini, “On the tunneling component of charge pumping current in ultrathin gate oxide MOSFET’s,” *IEEE Electron Device Lett.*, vol. 20, no. 2, pp. 92–94, 1999.

[99] D. Bauza, “Extraction of Si-SiO_2 interface trap densities in MOS structures with ultrathin oxides,” *IEEE Electron Device Lett.*, vol. 23, no. 11, pp. 658–660, 2002.

[100] J. Schmitz, F.N. Cubaynes, R.J. Havens, R. de Kort, A.J. Scholten, and L.F. Tiemeijer, “Test structure design considerations for RF-CV measurements on leaky dielectrics,” *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 2, pp. 150–154, 2004.

[101] G.T. Sasse, R. de Kort, and J. Schmitz, “Gate-capacitance extraction from RF C-V measurements,” in *Proceedings of the European Solid-state Device Research Conference*, 2004, pp. 113–116.

[102] N. Balabanian, *Electric circuits.* McGraw-Hill, 1994.

[103] E.W.M. van Luijk, “Comparison of RF C-V extraction methods using device simulations with focus on the influence of parasitics,” Master’s thesis, University of Twente, 2006.

[104] G. Groeseneken, H.E. Maes, N. Beltrán, and R.F. De Keersmaecker, “A reliable approach to charge-pumping measurements in MOS transistors,” *IEEE Trans. Electron Dev.*, vol. 31, no. 1, pp. 42–53, 1984.

[105] H.C. Lai, N.K. Zous, W.J. Tsai, T.C. Lu, T. Wang, Y.C. King, and S. Pan, "Reliable extraction of interface states from charge pumping method in ultra-thin gate oxide MOSFET's," in *Proceedings of the International Conference on Microelectronic Test Structures*, 2003, pp. 99–102.

[106] H.-H. Ji, Y.-G. Kim, I.-S. Han, K.-M. Kim, J.-S. Wang, H.-D. Lee, W.-J. Ho, S.-H. Park, H.-S. Lee, Y.-S. Kang, D.-B. Kim, C.-Y. Lee, I.-H. Cho, S-Y. Kim, S-B. Hwang, J-G. Lee, and J-W. Park, "On-chip charge pumping method for characterization of interface states of ultra thin gate oxide in nano CMOS technology," in *Proceedings of the International Electron Devices Meeting*, 2005, pp. 704–707.

[107] G.T. Sasse, H. de Vries, and J. Schmitz, "Charge pumping at radio frequencies," in *Proceedings of the International Conference on Microelectronic Test Structures*, 2005, pp. 229–233.

[108] G.T. Sasse and J. Schmitz, "Charge pumping at radio frequencies: methodology, trap Response and application," in *Proceedings of the International Reliability Physics Symposium*, 2006, pp. 627–628.

[109] J.L. Autran and C. Chabrerie, "Use of the charge pumping technique with a sinusoidal gate waveform," *Solid-State Electronics*, vol. 39, no. 9, pp. 1394–1395, 1996.

[110] F.P. Heiman and G. Warfield, "The effects of oxide traps on the MOS capacitance," *IEEE Trans. Electron Dev.*, vol. 12, no. 4, pp. 167–178, 1965.

[111] D. Bauza and G. Ghibaudo, "Analytical study of the contribution of fast and slow oxide traps to the charge pumping current in MOS structures," *Solid-State Electronics*, vol. 39, no. 4, pp. 563–570, 1996.

[112] Y. Manéglia, F. Rahmoune, and D. Bauza, "On the Si-SiO_2 interface trap time constant distribution in metal-oxide semiconductor transistors," *J. Appl. Phys.*, vol. 97, no. 1, pp. 014 502–1–014 502–8, 2004.

Appendix A

Calibration equations

A.1 Error terms

The 7-step calibration procedure makes use of the SFG of figure 2.6, containing 7 error terms. In this appendix the equations needed for relating $V_{\mathrm{DUT,pp}}$ to a_{m} and b_{m} are derived. First we start of with finding values for the 7 error terms, using the 7 calibration measurements. The first measurement consists of a SHORT measurement at Port 1. The SHORT standard is assumed ideal, which can be reflected as:

$$b_1 = -a_1 \tag{A.1}$$

$\Gamma_{\mathrm{S},1}$ is the reflection coefficient seen at the VNA, when Port 1 is terminated with a SHORT. From the SFG it can be derived that:

$$\Gamma_{\mathrm{S},1} = E_{\mathrm{DF}} - \frac{E_{\mathrm{1m}} \cdot E_{\mathrm{m1}}}{1 + E_{\mathrm{SF}}} \tag{A.2}$$

Similarly when an OPEN is connected to Port 1 this can be reflected as:

$$b_1 = a_1 \tag{A.3}$$

The reflection coefficient seen at the VNA when an OPEN is connected to Port 1, $\Gamma_{\mathrm{O},1}$ can be therefore derived to be:

$$\Gamma_{\mathrm{O},1} = E_{\mathrm{DF}} + \frac{E_{\mathrm{1m}} \cdot E_{\mathrm{m1}}}{1 - E_{\mathrm{SF}}} \tag{A.4}$$

If we connect a LOAD standard to Port 1, the system is perfectly matched and therefore no reflection will occur. This can be expressed as:

$$b_1 = 0 \tag{A.5}$$

This leads to the following expression for the measured reflection coefficient with a LOAD connected at Port 1, $\Gamma_{\mathrm{L},1}$:

$$\Gamma_{\mathrm{L},1} = E_{\mathrm{DF}} \tag{A.6}$$

Using expressions A.2, A.4 and A.6 we can find expressions for the error terms E_{DF}, E_{SF} and $E_{\mathrm{1m}} \cdot E_{\mathrm{m1}}$, using measurable parameters $\Gamma_{\mathrm{O,1}}$, $\Gamma_{\mathrm{S,1}}$ and $\Gamma_{\mathrm{L,1}}$:

$$E_{\mathrm{DF}} = \Gamma_{\mathrm{L,1}} \tag{A.7}$$

$$E_{\mathrm{SF}} = \frac{\Gamma_{\mathrm{O,1}} - \Gamma_{\mathrm{S,1}} - 2 \cdot \Gamma_{\mathrm{L,1}}}{\Gamma_{\mathrm{O,1}} - \Gamma_{\mathrm{S,1}}} \tag{A.8}$$

$$E_{\mathrm{1m}} \cdot E_{\mathrm{m1}} = 2 \cdot \frac{(\Gamma_{\mathrm{O,1}} - \Gamma_{\mathrm{L,1}}) \cdot (\Gamma_{\mathrm{L,1}} - \Gamma_{\mathrm{S,1}})}{\Gamma_{\mathrm{O,1}} - \Gamma_{\mathrm{S,1}}} \tag{A.9}$$

In order to find the absolute value of the voltage waves, a power calibration procedure must be performed. Using this power calibration a value can be found for $|E_{\mathrm{1m}}|$. This absolute power calibration can be performed using a power meter connected to Port 1. In fact this power level calibration is the only reason for making use of the 7-step calibration procedure for on wafer measurements; the power meter cannot be connected on wafer without the use of (non-ideal) probe needles. For off wafer measurements the common three-step de-embedding procedure can simply be extended with one power calibration step. If we connect the power meter to Port 1, the measured power wave a_{pow} can be found to be:

$$a_{\mathrm{pow}} = a_{\mathrm{m}} \cdot \frac{E_{\mathrm{1m}}}{1 - \Gamma_1 \cdot E_{\mathrm{SF}}} \tag{A.10}$$

So error term $|E_{\mathrm{1m}}|$ can be found by measuring a_{pow}, a_{m} and b_{m}:

$$|E_{\mathrm{1m}}| = \frac{|a_{\mathrm{pow}}|}{|a_{\mathrm{m}}|} \cdot |1 - \Gamma_1 \cdot E_{\mathrm{SF}}| \tag{A.11}$$

Parameter E_{SF} can be found from expression A.8. Γ_1 is the reflection coefficient seen at Port 1. An expression for this parameter will be derived later (A.24).

The final part of the 7-step calibration procedure is intended for finding error terms M_{C1}, M_{CD} and L_{C}. This can be performed using a SOL procedure with all calibration standards connected on wafer. If we connect the SHORT standard of the calibration substrate, the apparent reflection coefficient at Port 1, with a SHORT connected in place of the DUT, $\Gamma_{\mathrm{1,S,DUT}}$ can be found using:

$$\Gamma_{\mathrm{1,S,DUT}} = \frac{\Gamma_{\mathrm{S,DUT}} - E_{\mathrm{DF}}}{E_{\mathrm{1m}} \cdot E_{\mathrm{m1}} + E_{\mathrm{SF}} \cdot (\Gamma_{\mathrm{S,DUT}} - E_{\mathrm{DF}})} = M_{\mathrm{C1}} - \frac{L_{\mathrm{C}}^2}{1 + M_{\mathrm{CD}}} \tag{A.12}$$

Similarly the apparent reflection coefficient at Port 1 with an OPEN connected in place of the DUT, $\Gamma_{\mathrm{1,O,DUT}}$ is equal to:

$$\Gamma_{\mathrm{1,O,DUT}} = \frac{\Gamma_{\mathrm{O,DUT}} - E_{\mathrm{DF}}}{E_{\mathrm{1m}} \cdot E_{\mathrm{m1}} + E_{\mathrm{SF}} \cdot (\Gamma_{\mathrm{O,DUT}} - E_{\mathrm{DF}})} = M_{\mathrm{C1}} + \frac{L_{\mathrm{C}}^2}{1 - M_{\mathrm{CD}}} \tag{A.13}$$

Finally the apparent reflection coefficient at Port with a LOAD connected at the DUT, $\Gamma_{\mathrm{1,L,DUT}}$ is equal to:

$$\Gamma_{\mathrm{1,L,DUT}} = \frac{\Gamma_{\mathrm{L,DUT}} - E_{\mathrm{DF}}}{E_{\mathrm{1m}} \cdot E_{\mathrm{m1}} + E_{\mathrm{SF}} \cdot (\Gamma_{\mathrm{L,DUT}} - E_{\mathrm{DF}})} = M_{\mathrm{C1}} \tag{A.14}$$

Rewriting expressions A.13, A.12 and A.14 gives expressions for error terms M_{C1}, M_{CD} and L_{C}:

$$M_{\mathrm{C1}} = \Gamma_{1,\mathrm{L,DUT}} \tag{A.15}$$

$$M_{\mathrm{CD}} = \frac{\Gamma_{1,\mathrm{O,DUT}} + \Gamma_{1,\mathrm{S,DUT}} - 2\Gamma_{1,\mathrm{L,DUT}}}{\Gamma_{1,\mathrm{O,DUT}} - \Gamma_{1,\mathrm{S,DUT}}} \tag{A.16}$$

$$L_{\mathrm{C}}^2 = 2\frac{(\Gamma_{1,\mathrm{O,DUT}} - \Gamma_{1,\mathrm{L,DUT}}) \cdot (\Gamma_{1,\mathrm{L,DUT}} - \Gamma_{1,\mathrm{S,DUT}})}{\Gamma_{1,\mathrm{O,DUT}} - \Gamma_{1,\mathrm{S,DUT}}} \tag{A.17}$$

In expressions A.15, A.16 and A.17 quantities $\Gamma_{1,\mathrm{O,DUT}}$, $\Gamma_{1,S,DUT}$ and $\Gamma_{1,\mathrm{L,DUT}}$ are introduced. These are the reflection coefficients seen at Port 1 when an OPEN, SHORT respectively LOAD is connected at the DUT. The reflection coefficient seen at Port 1 can be related to the reflection coefficient actually seen by the VNA (A.24).

A.2 Calculating $V_{\mathrm{DUT,pp}}$

We are interested in the peak-to-peak value of the sinusoidal voltage at DUT level, $V_{\mathrm{DUT,pp}}$. Its value is related to the complex voltage wave V_{DUT} through:

$$V_{\mathrm{DUT,pp}} = 2\sqrt{2} \cdot |V_{\mathrm{DUT}}| \tag{A.18}$$

V_{DUT} can be found from basic transmission line theory:

$$V_{\mathrm{DUT}} = a_{\mathrm{DUT}} \cdot (1 + \Gamma_{\mathrm{DUT}}) \cdot \sqrt{Z_0} \tag{A.19}$$

So if we know both Γ_{DUT} and a_{DUT} we can find $V_{\mathrm{DUT,pp}}$. From the signal flow graph of 2.6 voltage wave a_{DUT} can be expressed in terms of a_1 and different error terms.

$$a_{\mathrm{DUT}} = a_1 \cdot \frac{L_{\mathrm{C}}}{1 - M_{\mathrm{CD}} \cdot \Gamma_{\mathrm{DUT}}} \tag{A.20}$$

Similarly voltage wave a_1 can be expressed in terms of a_{m}, a term that is known.

$$a_1 = a_{\mathrm{m}} \cdot \frac{E_{\mathrm{1m}}}{1 - E_{\mathrm{SF}} \cdot \Gamma_1} \tag{A.21}$$

Combining expressions A.19, A.20 and A.21 we can find an expression for $|V_{\mathrm{DUT}}|$ in terms of a_{m} and the different error terms and reflection coefficients.

$$|V_{\mathrm{DUT}}| = \sqrt{Z_0} \cdot |a_{\mathrm{m}}| \cdot |E_{\mathrm{1m}}| \cdot \left| \frac{L_{\mathrm{C}} \cdot (1 + \Gamma_{\mathrm{DUT}})}{(1 - E_{\mathrm{SF}} \cdot \Gamma_1) \cdot (1 - M_{\mathrm{CD}} \cdot \Gamma_{\mathrm{DUT}})} \right| \tag{A.22}$$

In expression A.22 both terms Γ_{DUT} and Γ_{DUT} can be recognized. They can also be derived from the measured quantities a_{m} and b_{m} and known error terms. The VNA can give both complex voltage waves a_{m} and b_{m}, from these the apparent measured reflection coefficient can be found:

$$\Gamma_{\mathrm{m}} = \frac{b_{\mathrm{m}}}{a_{\mathrm{m}}} = E_{\mathrm{DF}} + \frac{E_{\mathrm{1m}} \cdot E_{\mathrm{m1}} \cdot \Gamma_1}{1 - E_{\mathrm{SF}} \cdot \Gamma_1} \tag{A.23}$$

Rewriting this expression gives an expression for Γ_1:

$$\Gamma_1 = \frac{\Gamma_\mathrm{m} - E_\mathrm{DF}}{E_\mathrm{1m} \cdot E_\mathrm{m1} + E_\mathrm{SF} \cdot (\Gamma_\mathrm{m} - E_\mathrm{DF})} \tag{A.24}$$

Similarly an expression for Γ_DUT can be found:

$$\Gamma_\mathrm{DUT} = \frac{\Gamma_1 - M_\mathrm{C1}}{L_\mathrm{C}^2 + M_\mathrm{CD} \cdot (\Gamma_1 - M_\mathrm{C1})} \tag{A.25}$$

Now with expressions A.18, A.22, A.24 and A.25 it is possible to determine $V_\mathrm{DUT,pp}$ from the two measured voltage waves a_m and b_m and the 7 known error terms of figure 2.6.

Summary

In this thesis new developments are presented for reliability engineering in RF CMOS. Given the increase in use of CMOS technology in applications for mobile communication, also the reliability of CMOS for such applications becomes increasingly important. When applied in these applications, CMOS is typically referred to as RF CMOS, where RF stands for radio frequencies.

Performing reliability experiments requires the availability of well-defined voltage signals. This is the key challenge in performing reliability experiments at radio frequencies. In chapter 2 a methodology is described in which a sinusoidal voltage signal with a well-defined amplitude and a frequency of several GHz can be generated for on-wafer experiments. This method is used in various experiments discussed in this thesis.

In chapter 3 it was investigated whether the degradation of MOS transistors is dependent on the frequency of the voltage signal with which it is stressed. Three well-known degradation mechanisms have been investigated: hot-carrier degradation, NBTI and gate-oxide breakdown.

For both hot-carrier stress and NBTI these experiments revealed no frequency dependence for sinusoidal signals with frequencies ranging between several MHz and several GHz. This means that, for predicting degradation and lifetime of MOS transistors in RF circuits the same models apply as for circuits with lower frequency signals. Based on literature this implies that for hot-carrier degradation quasi-static models may be applied for RF circuits.

Gate-oxide breakdown has proven to be dependent on the frequency of the applied stress signal. At increasing frequency the lifetime of the device also increases.

In chapter 4 a new simulator is presented that allows for a better prediction of the lifetime of RF circuits, than what can be achieved with existing reliability simulators. The simulator has been designed with a special attention to RF power amplifiers. In these circuits even multiple gate-oxide breakdown events do not imply circuit failure. The new simulator makes use of a Monte-Carlo approach for utilizing this knowledge. Besides gate-oxide breakdown the simulator also incorporates hot-carrier degradation.

The simulator was evaluated by performing simulations on three different PA's. Simulation results show that the expected lifetime of these circuits increases considerably if this expected lifetime is obtained using the new simulator rather than

using conventional models. As a consequence higher voltage levels can be allowed in these circuits than what can be expected using conventional models.

In chapter 5 two measurement techniques are described that can be used for the reliability evaluation of MOS transistors in modern CMOS technologies. Due to the thin gate-oxide that is used in these technologies, a considerable tunneling current can be observed to flow through the gate-oxide. As a result of this tunneling current two widely used characterization techniques are no longer applicable: C-V and CP measurements. Performing these characterization techniques by making use of RF measurement techniques the effect of this tunneling current can be overcome.

The concept of RF C-V measurements is not new; in this thesis a methodology is presented that allows for a more accurate extraction of the gate-capacitance from data obtained using RF C-V measurements than before.

RF CP is a new technique; using this technique it is possible measure the degradation of a MOS transistor in terms of the number of interface states on transistors with a gate-oxide too thin for conventional charge pumping measurements. Besides this, the RF CP technique also allows for an extraction of the concentration of interface states as a function of the distance away from the substrate.

All these new developments contribute to developing reliability engineering for RF CMOS as a new and important field of engineering.

Samenvatting

In dit proefschrift zijn enkele nieuwe ontwikkelingen beschreven voor betrouwbaarheids-engineering van RF CMOS. Gezien het toenemende gebruik van CMOS technologie voor draadloze communicatietechniek is ook de betrouwbaarheid van CMOS voor dergelijke toepassingen een alsmaar belangrijker wordend onderwerp. Indien CMOS in deze toepassingen wordt gebruikt spreekt men doorgaans over RF CMOS, waar RF staat voor radiofrequenties.

Voor het uitvoeren van experimenten voor betrouwbaarheidsmetingen is een goed gedefiniëerde spanning nodig. Dit vormt de grootste uitdaging om betrouwbaarheidsexperimenten uit te voeren bij radiofrequenties. In hoofdstuk 2 is een methode beschreven waarmee een sinusvormige spanning met een goed gedefinierde amplitude en een frequentie van enkele GHz gegenereerd kan worden voor experimenten die uitgevoerd worden op een wafer. Deze methode is gebruikt bij verschillende experimenten die in dit proefschrift beschreven zijn.

In hoofdstuk 3 is onderzocht of de degradatie van MOS transistoren afhankelijk is van de frequentie van het spanningssignaal waarmee de transistor belast wordt. Drie bekende mechanismen die degradatie van MOSFETs kunnen veroorzaken zijn onderzocht: degradatie ten gevolge van hoog-energetische ladingsdragers, NBTI en gate-oxide doorslag.

Voor zowel degradatie ten gevolge van hoog-energetische ladingsdragers als NBTI is uit deze experimenten gebleken dat er geen frequentieafhankelijkheid bestaat voor sinusvormige signalen met een frequentie varierend van enkele MHz tot enkele GHz. Voor het voorspellen van de degradatie en de levensduur van MOS transistoren in een RF circuit mogen daarom de modellen worden gebruikt die ook worden toegepast voor circuits opererend op lagere frequenties. Op basis van kennis uit de literatuur betekent dit in het geval van degradatie door heteladingsdragers dat deze voorspelling gedaan kan worden met behulp van quasistatische modellen.

Het is gebleken dat doorslag van het gate-oxide van MOS transistoren wel afhankelijk is van de frequentie van de spanning waarmee de transistor belast wordt. Bij toenemende frequentie van het belastingssignaal neemt de levensduur van transistoren toe.

In hoofdstuk 4 is een nieuwe simulator beschreven waarmee de levensduur van RF circuits beter voorspeld kan worden dan op basis van bestaande simulatoren mogelijk is. De simulator is specifiek ontworpen voor het evalueren van RF

vermogensversterkers. Van dergelijke circuits gaat de functionaliteit niet altijd verloren bij de vorming van één of meerdere doorslagpaden in het gate-oxide van de transistoren. De nieuwe simulator maakt gebruik van een Monte-Carlo aanpak om deze kennis optimaal te benutten. Naast doorslagpaden wordt ook degradatie ten gevolge van hoog-energetische ladingsdragers gemodelleerd in deze simulator.

De simulator is geëevalueerd door simulaties uit te voeren op drie verschillende vermogensversterkers. De simulaties hebben laten zien dat de verwachte levensduur van deze circuits drastisch toeneemt indien deze verwachte levensduur op basis van deze nieuwe simulator wordt gedaan, in vergelijking met conventionele modellen. Het gevolg hiervan is dat hogere spanningsniveaus in het circuit kunnen worden toegestaan dan op basis van conventionele modellen verwacht kan worden.

In hoofdstuk 5 zijn twee meettechnieken beschreven die gebruikt kunnen worden voor de betrouwbaarheidsevaluatie van MOS transistoren in moderne CMOS technologieën. Tengevolge van het dunne gate-oxide dat bij deze technologieën gebruikt wordt, kan er een grote tunnelstroom lopen door dit gate-oxide. Deze tunnelstroom zorgt ervoor dat twee veelgebruikte karakterisatietechnieken niet langer gebruikt kunnen worden: C-V en charge pumping metingen. Door deze karakterisatietechnieken uit te voeren gebruik makend van RF meettechnieken kunnen de problemen ten gevolge van deze tunnelstroom teniet gedaan worden.

Het principe van RF C-V metingen is al langer bekend; in dit proefschrift is een nieuwe methode beschreven waarme de gate capaciteit nauwkeuriger dan voorheen kan worden geëxtraheerd uit data verkregen middels RF C-V metingen.

RF charge pumping is een nieuwe techniek; met behulp van deze techniek is het mogelijk om degradatie van een MOS transistor met betrekking tot het aantal oppervlaktetoestanden te meten op transistoren met een gate-oxide dat te dun is voor conventionele charge pumping metingen. Ook is het met de RF charge-pumping techniek mogelijk om de concentratie van deze oppervlaktetoestanden te bepalen als functie van de afstand tot het substraat.

Al deze nieuwe ontwikkelingen dragen bij aan het ontwikkelen van betrouwbaarheidsengineering voor RF CMOS als een nieuwe en belangrijke tak van engineering.

List of Publications

G.T. Sasse, M. Acar, F.G. Kuper, J. Schmitz, "RF CMOS reliability simulations", *European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF)*, Maastricht, the Netherlands, 2008, accepted for publication.

G.T. Sasse, F.G. Kuper, J. Schmitz, "MOSFET Degradation under RF stress", *IEEE Transactions on Electron devices*, submitted for publication.

G.T. Sasse, J. Schmitz, "Application and evaluation of the RF charge-pumping technique", *IEEE Transactions on Electron devices, vol.55, no.3, March 2008, pp. 881-889.*

G.T. Sasse, R.J de Vries, J. Schmitz, "Methodology for performing RF reliability experiments on a generic test structure", *International Conference on Microelectronic Test Structures (ICMTS)*, Tokyo, Japan, 2007.

G.T. Sasse, J. Schmitz, "Interface trap response to RF charge pumping measurements", *Semiconductor Advances for future Electronics (SAFE)*, Veldhoven, The Netherlands, 2006.

G.T. Sasse, J. Schmitz, "Charge Pumping at radio frequencies: methodology, trap response and application", *International Reliability Physics Symposium (IRPS)*, San Jose, Ca, USA, 2006 (BEST POSTER AWARD).

R.G. Bankras, M.P.J. Tiggelman, M.A. Negara, **G.T. Sasse**, J. Schmitz, "C-V test structures for metal gate CMOS", *International Conference on Microelectronic Test Structures (ICMTS)*, Austin, TX, USA, 2006.

G.T. Sasse, J. Schmitz, "The RF charge pump technique for measuring the interface state density on leaky dielectrics", *Semiconductor Advances for Future Electronics (SAFE)*, Veldhoven, The Netherlands, 2005.

G.T. Sasse, H. de Vries, J. Schmitz, "Charge pumping at radio frequencies", *International Conference on Microelectronic Test Structures (ICMTS)*, Leuven, Belgium, 2005.

G.T. Sasse, R. de Kort, J. Schmitz, "Gate-capacitance extraction from RF C-V measurements", *Semiconductor Advances for Future Electronics (SAFE)*, Veldhoven, The Netherlands, 2004.

L.F. Tiemeijer, R.J. Havens, R. de Kort, A.J. Scholten, R. van Langevelde, D.B.M. Klaassen, **G.T. Sasse**, Y. Bouttement, C. Petot, S. Bardy, D. Gloria, P. Scheer, S. Boret, B. van Haaren, C. Clement, J-F. Larchanche, I-S. Lim, A. Zlotnicka, A. Duvallet, "Record RF performance of standard 90 nm CMOS technology", *International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2004.

G.T. Sasse, R. de Kort, J. Schmitz, "Gate-capacitance extraction from RF C-V measurements", *European Solid-State Device Research Conference(ESSDERC)*, Leuven, Belgium, 2004.

Dankwoord

Het meest gelezen onderdeel van ieder proefschrift is ongetwijfeld het dankwoord. Dit wetende ligt er een enorme druk op mijn schouders om dit met veel zorg te schrijven: ik wil natuurlijk niemand over het hoofd zien. In de afgelopen jaren ben ik door vele mensen in meer of mindere mate geholpen, ik ben iedereen hier dankbaar voor. Ik wil hier enkele personen uitlichten zonder wiens bijdrage dit proefschrift zeker niet tot stand was gekomen. Daarnaast zijn er nog vele mensen die een bijdrage hebben geleverd, maar wie ik niet bij naam noem; als jouw naam er niet tussen staat, weet dan dat ik ook jou dankbaar ben.

Allereerst wil ik mijn promotor Jurriaan Schmitz bedanken. Jurriaan, ik heb de afgelopen jaren met veel plezier met jou samen gewerkt, je hebt mij regelmatig op goede ideeën weten te brengen en door jouw enthousiasme heb je me vaak weten te motiveren, ook op de momenten dat het even niet meezat.

Verder wil ik mijn tweede promotor Fred Kuper bedanken. Fred, de verschillende besprekingen die we in de loop van de jaren hebben gehad heb ik als erg nuttig ervaren. Ook ben ik blij dat je ondanks je drukke werkzaamheden bij NXP toch als promotor bij mij wil optreden.

Ook wil ik graag de leden van de promotiecomissie bedanken voor hun nuttige commentaar op mijn proefschrift. In het bijzonder wil ik Paul van der Wel bedanken voor het aanleveren van enkele van de samples die ik heb gebruikt voor dit proefschrift. Ook het bezoek aan de IRPS in San Jose en Phoenix samen met jou (en de bijbehorende tripjes) waren leuke ervaringen.

Het onderzoek dat ik heb uitgevoerd was in het kader van het STW project Reliable RF. Ik wil hiervoor graag de STW en de gebruikerscommissie bedanken. Ik wel met name Rudolf Velghe noemen voor het aanleveren van verschillende samples die ik heb gebruikt voor metingen. Het project was een samenwerkingsproject met de IC-design groep. Ik wil graag Bram Nauta en Anne-Johan Annema bedanken voor deze samenwerking.

In dit kader wil ik zeker ook Mustafa Acar bedanken. Mustafa, bedankt voor de hulp die je me hebt gegeven bij het implementeren van jouw ontwerpen in mijn simulator, maar niet minder belangrijk voor de gezelligheid buiten het werk om. Ook wil ik je bedanken voor het feit dat jij mijn paranimf wil zijn.

In het eerste jaar dat ik promovendus was heb ik enkele maanden stage gelopen bij de groep device modelling and characterization binnen Philips Research in Eindhoven om ervaring op te doen met het uitvoeren van RF metingen. Ik heb

dit als een zeer prettige en nuttige periode ervaren; ik wil dan ook iedereen die destijds bij deze groep werkte bij deze bedanken voor de prettige werksfeer. In het bijzonder wil ik graag Reinout Woltjer, Luuk Tiemeijer en Ramon Havens bedanken voor hun hulp.

Als het gaat om meten, RF metingen in het bijzonder, wil ik zeker mijn dank betuigen aan Henk de Vries. Henk, zonder jouw hulp zou ik nooit toegekomen zijn aan de metingen die ik in dit proefschrift beschrijf. Als een door mij bedachte meting weer eens niet werkte wist jij vaak vrij snel het probleem te verhelpen.

Verder ben ik dank verschuldigd voor de technische ondersteuning die ik heb gekregen van Gerard Wienk en Sander Smits. Ik wil Cor Bakker en Frederik Reenders bedanken voor het onderhoud en beheer van PC's, netwerken en alle daaraan gerelateerde zaken. Ik wil ook graag Annemiek Janssen bedanken voor de ondersteuning die ik van haar heb gekregen in de afgelopen jaren. Annemiek, je hebt me altijd geholpen als ik daarom vroeg, maar vaak juist ook zonder daarom te hoeven vragen. Als Annemiek er niet was kon ik terecht bij Gerdien Lammers, Gerdien ook jij bedankt.

Er zijn in de afgelopen jaren een aantal studenten geweest wiens werk een zinvolle bijdrage heeft opgeleverd aan dit proefschrift: Erik van Luijk, Rein de Vries en Akis Rogotis. Ik wil jullie bij deze bedanken.

Ik heb ruim vier jaar doorgebracht in kamer 3138 van de Hogekamp. Ik wil al mijn kamergenoten door de jaren heen bedanken voor de leuke werksfeer: Hieu, Gratiela, Ihor, Joost en Giulia. We hebben vele nuttige discussies gehad, maar minstens zo belangrijk waren de geanimeerde gesprekken over bijvoorbeeld de nieuwe pata's van Nina. Dit zorgde voor de nodige afleiding. Joost: ik wil jou speciaal bedanken, omdat je ook mijn paranimf wil zijn.

Hiernaast zijn er natuurlijk de overige mensen van vloer 3, in het bijzonder de SC groep, die ik wil bedanken voor enkele leuke jaren. Ik denk dat ik de dagelijkse lunchroep van Ray en de komkommer van Arjen nog ga missen.

Verder wil ik vooral niet vergeten mijn familie en vrienden te bedanken voor alle steun die in de afgelopen jaren (en de jaren daarvoor) heb gekregen, bedankt!

Jeroen wil ik bedanken voor het ontwerpen van het omslag van dit proefschrift. Verder, al zijn er meer mensen die me hiermee hebben geholpen wil ik toch Harald speciaal bedanken voor zijn uitleg over hoe een proefschrift te schrijven in LaTeX.

Tot slot wil ik natuurlijk Esther bedanken. Ik ben in de vrij unieke postitie dat ik mijn vriendin kan bedanken voor de hulp die ik heb gekregen voor het oplossen van problemen met Matlab en Labview. Maar nog veel belangrijker is de steun en liefde die je me al jaren geeft.