

Application and Evaluation of the RF Charge-Pumping Technique

Guido T. Sasse, *Student Member, IEEE*, and Jurriaan Schmitz, *Senior Member, IEEE*

Abstract—In this paper, we will discuss the extendibility of the charge-pumping (CP) technique toward frequencies up to 4 GHz. Such high frequencies are attractive when a significant gate leakage current flows, obscuring the CP current at lower pumping frequencies. It is shown that using RF gate excitation, accurate CP curves can be obtained on MOS devices with a leakage current density exceeding $1 \text{ A}\cdot\text{cm}^{-2}$. A theoretical analysis of the trap response to RF gate voltage signals is presented, giving a clear insight on the benefits and limitations of the technique.

Index Terms—Characterization, charge pumping (CP), CMOS, dielectrics, RF, trap response, tunneling.

I. INTRODUCTION

THE CHARGE-PUMPING (CP) technique [1] is widely used to quantify the interface state density at the Si–SiO₂ interface of MOS devices. With the decreasing thickness of the oxide layer, in present day CMOS technologies, a considerable gate leakage current can be seen. This leakage current can severely affect the correctness of the extracted interface state density from CP data [2], [3]. In Fig. 1, 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 [2], 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 [4]. In [3], 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}_{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}_{it} of $10^{11} \text{ cm}^{-2}\cdot\text{eV}^{-1}$, a voltage window of 0.1 V, and a measurement frequency of 1 MHz, the

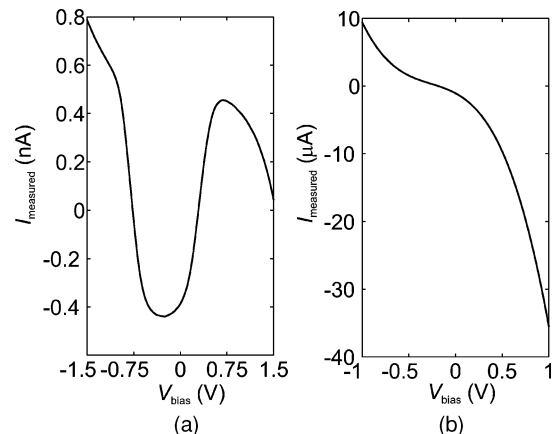


Fig. 1. CP currents obtained on *n*-type devices with different oxide thickness. (a) 3 nm. (b) 1.4 nm. The data are obtained using a sinusoidal gate voltage with $V_{pp} = 2 \text{ V}$ and $f = 1 \text{ MHz}$. The current is measured at the drain/source contact, resulting in negative values of the CP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

CP current density will be approximately $1.6 \text{ mA}\cdot\text{cm}^{-2}$. 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 paper, we use a factor of 10 as a criterion. This implies an upper limit for the leakage current density of $0.16 \text{ mA}\cdot\text{cm}^{-2}$ for the given example. For lower values of \overline{D}_{it} , this upper limit is even further reduced.

If one wants to overcome larger leakage current densities, the frequency dependence of the CP current can be used as given by [1]:

$$I_{cp} = fqA_G\overline{D}_{it}\Delta E. \quad (1)$$

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}_{it} is the interface state density (in per square centimeter per electron volt), 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 an CP cycle [1]. The frequency dependence of (1) can be applied in order to increase the CP current w.r.t. the leakage current moving far beyond the $\sim 1 \text{ MHz}$ 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 [5], or by making use of the RF CP technique [6], [7]. In this paper, we will elaborate

Manuscript received June 1, 2007; revised November 1, 2007. This work was supported by the Dutch Technology Foundation STW. The review of this paper was arranged by Editor C.-Y. Lu.

The authors are with the Group of Semiconductor Components, MESA+ Institute for Nanotechnology, University of Twente, 7500 AE, Enschede, The Netherlands (e-mail: g.t.sasse@utwente.nl; j.schmitz@utwente.nl).

Digital Object Identifier 10.1109/TED.2007.915088

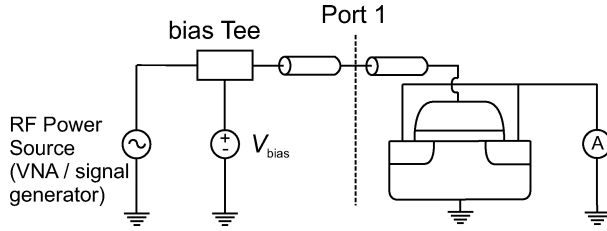


Fig. 2. Schematic drawing of the RF CP measurement setup. An RF signal is superimposed on a dc voltage V_{bias} through the use of a bias tee. Port 1 represents the location in the measurement setup where four of the seven necessary calibration measurements are performed. The CP current is measured at the drain/source connection.

on the application of the RF CP technique. This technique makes use of sinusoidal large voltage swing signals with frequencies into the gigahertz range, more than two orders higher than in the conventional CP approach.

II. MEASUREMENT SETUP AND METHODOLOGY

A. Setup

In an 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 toward the source/drain regions. By repeatedly performing such an 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 w.r.t. 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. Therefore, we use a measurement setup different from conventional CP measurements, as illustrated in Fig. 2. We generate a sinusoidal gate voltage signal 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 tee. A full CP curve (as in Fig. 1) is obtained by sweeping V_{bias} . For the results presented in this paper, a Rohde & Schwarz ZVB20 vector network analyzer (VNA) 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 [8], 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, we measure the CP current 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, such as, e.g., in [1].

B. Signal Integrity

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 that may arise from an impedance mismatch between the measurement cables and the DUT. Distortion will change the precise waveform at the device, and this complicates the interpretation of CP currents. In this section, we will summarize our distortion analysis presented in [6] and [9], showing that distortion effects do not play a significant role in RF CP measurements under normal, practical circumstances.

From basic transmission line theory (see, e.g., [10]), it is known that a sinusoidal input voltage on a linear impedance will only lead to a shift in phase and amplitude, not to higher order harmonics. Yet, the MOS gate capacitance is voltage-dependent, and the gate exhibits a voltage-dependent tunneling current, so this component is not linear. To assess the significance of distortion, we measured the small-signal input impedance of the DUT as a function of applied voltage using a linear VNA in one-port setup. The measurement was carried out at a wide range of bias voltages (V_{bias}) and for all relevant frequencies. The small-signal input impedance $z_{11}(V_{bias})$ is obtained from the small-signal reflection coefficient s_{11} [corrected with a SHORT–OPEN–LOAD (SOL) calibration], using

$$z_{11}(V_{bias}) = Z_0 \frac{1 + s_{11}(V_{bias})}{1 - s_{11}(V_{bias})}. \quad (2)$$

In this equation, Z_0 is the characteristic impedance of the measurement cables. Note that z_{11} is obtained with the SOL calibration, but without deembedding; this is done deliberately as we want to investigate the linearity at the input of our test structures, i.e., at the tip of our probe needles. z_{11} includes the bond pad capacitance and line inductance.

With z_{11} known, we can calculate the gate voltage signal resulting from a high-power RF signal. With the reasonable assumption that the device is quasi-static at the frequency range of interest, we can find the time-dependent voltage signal by solving the transmission line equation (3) [9]

$$\frac{V_{DUT}^+(t) - V_{DUT}^-(t)}{Z_0} = V_{DUT}(t) \left[G_{DUT}(t) + \frac{dC_{DUT}(t)}{dt} \right] + \frac{dV_{DUT}(t)}{dt} C_{DUT}(t). \quad (3)$$

This expression is a time-domain representation of the linear transmission line equation that can be derived using basic transmission line theory [10]. It holds for sinusoidal voltage signals and test structures where capacitive effects in the input impedance dominate over inductive effects (i.e., a negative imaginary part of the input impedance); it is, therefore, applicable for our measurement setup. In this expression, $V_{DUT}^+(t)$ and $V_{DUT}^-(t)$ represent the incoming and reflected voltage waves

[10] at the DUT level, respectively, if the DUT would be connected as in Fig. 2 and the RF power source would deliver an RF signal with an appropriate frequency and power; $V_{\text{DUT}}(t)$ is the actual voltage signal at the DUT level that would result in this setup using this specific test structure. G_{DUT} and C_{DUT} are the device input conductance and capacitance derived from the input impedance z_{11} using

$$G_{\text{DUT}}(t) = \Re \left(\frac{1}{z_{11}(V_{\text{DUT}}(t))} \right) \quad (4)$$

$$C_{\text{DUT}}(t) = \frac{1}{2\pi f} \Im \left(\frac{1}{z_{11}(V_{\text{DUT}}(t))} \right). \quad (5)$$

Equation (3) can be solved by realizing that [10]

$$V_{\text{DUT}}(t) = V_{\text{DUT}}^+(t) + V_{\text{DUT}}^-(t). \quad (6)$$

The voltage signal at the gate can be related to V_{DUT} by modeling the test structure parasitics using a series line impedance and a parallel bond pad capacitance (similar to what is done in OPEN-SHORT deembedding). The parallel bond pad capacitance does not influence the voltage division between V_{DUT} and V_G , and therefore, the following relation holds

$$V_G = V_{\text{DUT}} \frac{Z_G}{Z_{\text{line}} + Z_G} \approx V_{\text{DUT}}. \quad (7)$$

In this expression, V_G and V_{DUT} are the complex phasor notations of $V_G(t)$ and $V_{\text{DUT}}(t)$, respectively. Z_G is the gate impedance consisting of the gate capacitance in parallel with a gate conductance due to gate tunneling. Z_{line} represents the parasitic impedance existing with the test devices, which is generally modeled by series resistance and line inductance and may be dominated by the inductance at very high frequencies. The approximation made in (7) is valid as long as $Z_{\text{line}} \ll Z_G$. For properly designed test structures (see, e.g., [8]), this assumption is valid for the frequencies used in this paper.

Using the measured values of $z_{11}(V_{\text{bias}})$, we solve (3) with (6) in discrete-time domain using an iterative solution technique embedded in a MATLAB routine. The frequency and amplitude of $V_{\text{DUT}}^+(t)$ are chosen to generate the desired voltage signal at the DUT level $V_{\text{DUT}}(t)$. Fig. 3 shows examples of this signal integrity analysis obtained on a 3-nm oxide device and a 1.4-nm oxide device. The figure clearly shows that the nonlinear behavior at the input of the devices is negligible. The voltage $V_{\text{RMS}}(f)$ used in Fig. 3(b) and (d) is defined as the root mean square value of the harmonic component with frequency f of the gate voltage $V_G(t)$.

C. Voltage Generation

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. The peak-peak amplitude V_{pp} is set by the power level of the RF source and the base level by the dc biasing. In order to find the appropriate power level, we proposed an easy to use method in [7], which makes use of the frequency-independent tunneling current. This approach, however, is only limited to very specific devices that

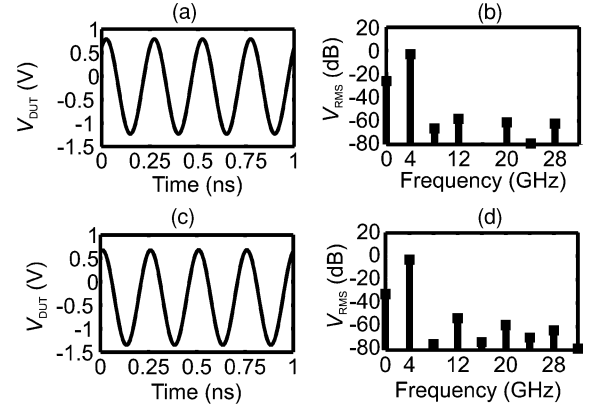


Fig. 3. Time-domain signals as obtained using the discrete-time solution to (3) and their harmonic content as it follows from a fast Fourier transform: (a) and (b) a 3-nm oxide device, (c) and (d) a 1.4-nm oxide device. The waveforms are determined with V_{bias} set for maximum CP condition at 4 GHz.

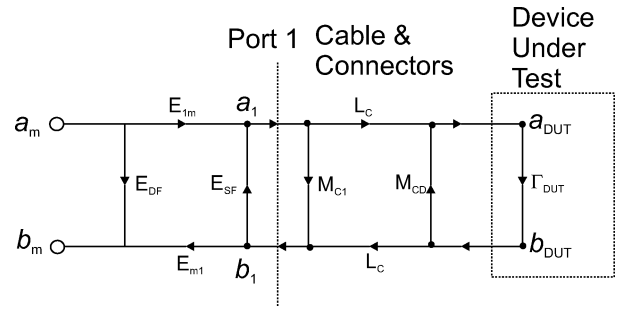


Fig. 4. SFG representing all the RF signal paths of the measurement setup of Fig. 2, as it is used for determining the realized voltage at the DUT.

show a considerable tunneling current but do not suffer from too high voltage levels. Here, we make use of the more generally applicable voltage generation procedure discussed in [9]. In this approach, we use an VNA in continuous wave (CW) mode as an RF power source with the appropriate frequency and the RF power set for generating the desired voltage signal at the DUT level. The VNA is capable of measuring both the incoming and reflected complex power waves (as defined in [11]). This means that if the complete measurement setup is accurately characterized and nonlinear behavior is negligible, $V_G(t)$ can be determined directly from the measured complex power waves.

The appropriate power level for setting the desired V_{pp} , the peak-to-peak voltage of $V_{\text{DUT}}(t)$ [and hence, of $V_G(t)$ through (7)], can be found using a Labview routine that gradually increases the CW power of the VNA and determines the corresponding value of V_{pp} until the desired V_{pp} is reached. In this approach, the measurement setup is modeled using the signal flow graph (SFG) shown in Fig. 4 [9]. In this SFG, a_m and b_m represent the complex power wave vectors [11] as they are measured by the VNA. Quantities a_{DUT} and b_{DUT} are the corresponding complex power wave vectors as they are present at the DUT level. Γ_{DUT} is the complex reflection coefficient of the DUT. V_{pp} is related to a_{DUT} and Γ_{DUT} through (7) [9]

$$V_{\text{pp}} = \sqrt{8Z_0} |a_{\text{DUT}}| |1 + \Gamma_{\text{DUT}}|. \quad (8)$$

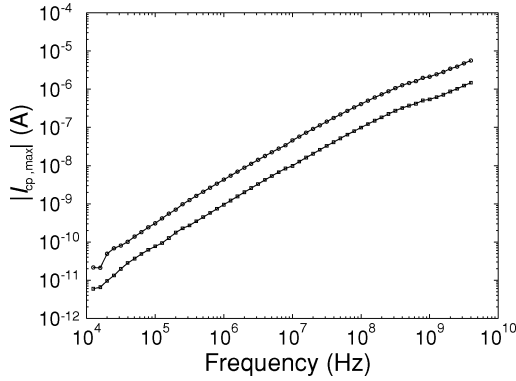


Fig. 5. $I_{cp,max}$ plotted against frequency, obtained on devices with $t_{ox} = 3$ -nm. The gate voltage signal used was sinusoidal with $V_{pp} = 2$ V. The bottom line represents measurements made on an n -type device and the upper line on a p -type device. The increase of the CP current with increasing frequency can clearly be seen.

The values of $|a_{DUT}|$ and $|1 + \Gamma_{DUT}|$ can be found directly from the measured complex values of a_m and b_m and the error terms of Fig. 4 as is explained in more detail in [9]. The error terms are determined beforehand by performing seven individual calibration measurements for all desired power and frequency levels used. The first three calibration measurements consist of an OPEN, SHORT, and LOAD calibration at port 1 using coaxial calibration standards. The next step is an absolute power measurement at port 1 using an external power meter. The last three calibration measurements consist of an OPEN, SHORT, and LOAD measurement at the tip of the probe needle, using a calibration substrate. In this calibration procedure, port 1 is an arbitrarily chosen location between the VNA and the coaxial connection of the probe needle. It is needed to perform the absolute power measurement as the power meter can only be connected using a coaxial connection and it is not possible to do this on wafer.

Based on these calibration measurements, values can be found for E_{DF} , E_{SD} , $E_{1m}E_{m1}$, $|E_{1m}|$, M_{C1} , M_{CD} , and L_C for all power levels and frequencies used. We will not discuss the exact expressions needed to extract these error terms, as they can be derived straightforwardly using the SFG of Fig. 4. For more details, we refer to [9].

III. MEASUREMENT RESULTS

Using the aforementioned approach, gate voltage signals with well-defined amplitude levels may be generated with frequencies of up to 4 GHz. This allows us to perform CP measurements at frequencies far beyond the frequencies used in conventional CP measurements. This is illustrated in Fig. 5 where $I_{cp,max}$ 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 $I_{cp,max}$ is defined as the largest (absolute value of the) CP current over an entire V_{bias} sweep. It is clearly visible that $I_{cp,max}$ keeps increasing with frequency up to 4 GHz, over two orders of magnitude beyond conventional CP measurements. Hence, this frequency dependence, as predicted by (1), may be applied in order to perform CP measurements on dielectrics with

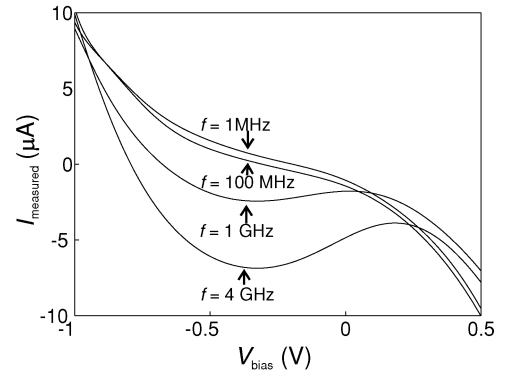


Fig. 6. CP curves obtained at various frequencies on the same 1.4-nm oxide device as in Fig. 1(b). The applied gate voltage level has a V_{pp} of 2 V. The CP effect is clearly visible for the 1 and 4 GHz signals.

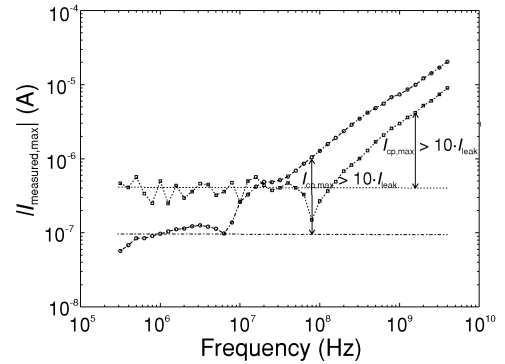


Fig. 7. Maximum value of measured CP current plotted against frequency on both an n -type and a p -type 1.4-nm device. The squares represent measurements made on an n -type device and the circles on a p -type device. All measurements are performed with a sinusoidal voltage signal with $V_{pp} = 2$ V. A plateau of minimum $I_{measured,max}$ can be observed for low frequency data on the 1.4-nm devices. This minimum $I_{measured,max}$ is, in fact, not an CP current but the leakage current component present in these ultrathin oxide devices. The two horizontal lines are drawn to indicate the absolute level of the leakage current.

a leakage current too high for conventional CP measurements. This is shown in Fig. 6 where CP data are shown on the same 1.4-nm device as in Fig. 1(b), but at frequencies up to 4 GHz. Only at the highest frequencies, the CP effect is visible. The extension to gigahertz frequencies allows to determine $I_{cp,max}$, and thus, obtain information on the amount of fast interface states on ultraleaky dielectrics. Similar to Fig. 5, we also plotted the frequency response of the measured CP current on both an n -type and p -type 1.4-nm device. This is shown in Fig. 7. In this figure, we plotted parameter $I_{measured,max}$ rather than $I_{cp,max}$ in order to prevent any confusion about the definition of $I_{cp,max}$. Similar to $I_{cp,max}$, $I_{measured,max}$ 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 Fig. 7 represent the magnitude of the leakage component of the measured currents. For extracting reasonably accurate CP data, we need at least an CP component ten times higher than the leakage component of the measured current, as explained in Section I. This condition implies that for the 1.4-nm n -type device, only measurements beyond 1.5 GHz may be used. This illustrates the ultrahigh leakage current for this device. The

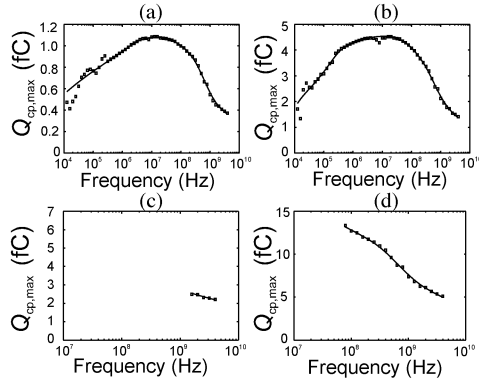


Fig. 8. Pumped charge per cycle plotted against frequency. (a) 3-nm n -type device. (b) 3-nm p -type device. (c) 1.4-nm n -type device. (d) 1.4-nm p -type device. For the 1.4-nm devices, $Q_{cp,max}$ values are only obtained at frequencies where the CP effect is clearly visible (as indicated in Fig. 7). The data are obtained using a sinusoidal gate voltage with $V_{pp} = 2$ V. The data at frequencies below 10 MHz are obtained using an Agilent 33250A signal generator, the other data points are obtained using a Rohde & Schwarz ZVB20 VNA. The data obtained at 10 MHz and higher are obtained after a 1 nA dc offset subtraction from the measured current. The solid lines are meant as a guide to the eye.

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, we make use of the pumped charge per cycle $Q_{cp,max}$, defined as

$$Q_{cp,max} = \frac{|I_{cp,max}|}{f}. \quad (9)$$

$Q_{cp,max}$ is a direct indicator of the amount of traps that contribute to the CP current, and hence, a measure of the Si-SiO₂ interface quality. In the conventional CP approach, this $Q_{cp,max}$ is assumed to be constant over frequency, when using trapezoidal gate voltage signals with constant rise and fall times for every frequency. For the sinusoidal voltage signals used in this paper, the rise and fall times are equal and proportional to the inverse of frequency. In [12], it was shown that $Q_{cp,max}$ increases linearly with $\log(f)$ when sinusoidal gate voltage signals are used. The increase in $Q_{cp,max}$ with increasing frequency is attributed to the increase in the probed energy window ΔE . In Fig. 8, we plotted $Q_{cp,max}$ against frequency up to 4 GHz for the four different devices used in Figs. 5 and Fig. 7. The results presented are obtained from the measured current levels after a subtraction of low frequency measurement as in [2], in order to increase accuracy. On the 1.4-nm oxide devices, gate leakage prohibits the quantification of $Q_{cp,max}$ below 1.5 GHz and 80 MHz for the n -type and p -type devices, respectively. On the 3-nm oxide devices, we extracted $Q_{cp,max}$ for frequencies ranging from 10 kHz to 4 GHz. In order to obtain results below 10 MHz, we made use 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 this is that if this would not be done, a small transition point in the $Q_{cp,max}$ versus $\log(f)$ graphs could be noticed around 10 MHz, the frequency were the transition between the two measurement

setups is made. A possible explanation for this transition point could be found in issues related to the grounding of the VNA. This 1 nA dc offset is negligible for frequencies above 50 MHz, hence, for the frequencies of interest in this paper. For the results obtained on the 1.4-nm devices, this small dc offset is negligible for all frequencies as can be seen in Fig. 7.

IV. TRAP RESPONSE

In Fig. 8, we can recognize the expected linear slope for frequencies up to 10 MHz. Above this frequency, $Q_{cp,max}$ starts to increase less than predicted by the theory of [12], 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 [7], we have shown that the decrease of $Q_{cp,max}$ with increasing frequency as seen in Fig. 8 cannot 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., [13]). In this section, we will discuss how this trap distribution is exactly related to the observed frequency response of $Q_{cp,max}$ for sinusoidal gate voltage signals. Our derivation starts with the general expression for the CP current (see, e.g., [3])

$$\frac{I_{cp}}{fqAG} = \int_0^{t_{ox}} \int_{E_{low}}^{E_{high}} N_{it}(E_T, x_T) \Delta f_T(E_T, x_T, f) dE_T dx_T. \quad (10)$$

In this expression, a pure tunneling mechanism [13] is assumed to govern the capture and emission processes. Expression (1) is basically a simplification of expression (10). Where in expression (1), the effective interface state density \overline{D}_{it} is used, we use the interface state concentration N_{it} in expression (10). All traps are described with an energy level E_T and distance from the interface x_T . \overline{D}_{it} is the integral of N_{it} over distance and energy. Parameter t_{ox} is basically the integration limit for x_T . In expression (10), it is rather arbitrarily chosen as the oxide thickness; later, we will derive another integration limit, 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 be equal to unity for all traps that are located between the energy levels $E_{em,e}$ and $E_{em,h}$, the limits of the energy levels where the emission process of carriers is negligible during an CP cycle. Parameter x_T represents the distance from a particular trap toward the Si-SiO₂ interface. The capture cross section of a trap is related to the distance of the trap from the interface through [13]

$$\sigma(x_T) = \sigma_0 e^{-x_T/\lambda}. \quad (11)$$

This expression is based on the first-order trapping model of [13] where the capture cross section is assumed to be independent of energy level. In expression (11), λ represents the tunneling attenuation constant, which is approximately 0.07 nm [13]. 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 (10), we need an expression 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., [14])

$$\frac{df_T}{dt} = [1 - f_T(t)] [c_n(t) + e_p] - f_T(t) [c_p(t) + e_n]. \quad (12)$$

In this expression, $c_n(t)$ and $c_p(t)$ 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 $c_n(t)$ and $c_p(t)$ are directly related to the amount of charge carriers at the Si–SiO₂ interface, and therefore, dependent on the 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 and valence band, respectively.

No general closed-form solution can be found for the differential equation of (12); we will adopt a similar approach as in [15] in finding an expression for Δf_T from expression (12). In this approach, we make use of three basic assumptions.

- 1) Traps outside ΔE do not contribute to the CP current and all traps within ΔE have negligible emission rates [1].
- 2) The capture processes are negligible outside inversion and accumulation, i.e., at voltage levels between V_T and V_{FB} [1].
- 3) At maximum CP conditions, the integral of $c_n(t)$ over time equals the integral of $c_p(t)$ over time [15].

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

$$\frac{df_T}{dt} \approx [1 - f_T(t)] c_n(t). \quad (13)$$

Similarly, for the capture process during accumulation, expression (12) reduces to

$$\frac{df_T}{dt} \approx -f_T(t) c_p(t). \quad (14)$$

Using expression (13), the maximum trap occupancy level $f_{T,\max}$ can be found as a function of the minimum trap occupancy level $f_{T,\min}$; $f_{T,\min}$ can be found as a function of $f_{T,\max}$ using expression (14)

$$f_{T,\max} = (1 - f_{T,\min}) e^{-cnt} \quad (15)$$

$$f_{T,\min} = f_{T,\max} e^{-cpt}. \quad (16)$$

In these expressions, parameters cnt and cpt are defined as

$$cnt = \int_{t_{inv,start}}^{t_{inv,stop}} c_n(t) dt \quad (17)$$

$$cpt = \int_{t_{acc,start}}^{t_{acc,stop}} c_p(t) dt. \quad (18)$$

Parameters $t_{inv,start}$ and $t_{inv,stop}$ are the times at the onset and end of inversion conditions, respectively. Parameters $t_{acc,start}$ and $t_{acc,stop}$ signify the start and end times of the accumulation. Using expressions (15) and (16), we can derive an expression

for Δf_T

$$\Delta f_T = f_{T,inv} - f_{T,acc} = \frac{(1 - e^{-cnt})(1 - e^{-cpt})}{(1 - e^{-cnt-cpt})}. \quad (19)$$

This expression cannot be solved analytically for all values of cnt and cpt . However, we can further simplify the expression by making use of assumption 3) given earlier for maximum CP conditions. This means that at the maximum CP condition over an entire V_{bias} sweep, we may assume that $ct \approx cnt \approx cpt$. Using parameter ct , we can express Δf_T at maximum CP condition as

$$\Delta f_T = \frac{(1 - e^{-ct})^2}{(1 - e^{-2ct})}. \quad (20)$$

We will approximate this function with a step function around $ct = \ln(3)$, after [15], giving a very simple expression for Δf_T at maximum CP conditions. We can now define parameter $x_{T,\max}$ 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 $\ln(3)$. We can derive an expression for $x_{T,\max}$ by using the expression used for the capture rate of electrons $c_n(t)$ (or similarly from the capture rate for holes, since we make use of assumption 3). This capture rate is given by

$$c_n(t) = v_{th} \sigma(x_T) n_s(t). \quad (21)$$

In this expression, v_{th} is the thermal velocity of charge carriers and $n_s(t)$ is the electron concentration at the Si–SiO₂ interface. Using the definition of ct , we can find

$$ct = cnt = \sigma(x_T) v_{th} \int_{t_{inv,start}}^{t_{inv,stop}} n_s(V_G(t)) dt. \quad (22)$$

The solution to this integral can be found by expressing the sinusoidal gate voltage signal $V_G(t)$ as

$$V_G(t) = V_{bias} + \frac{V_{pp}}{2} \sin(2\pi ft). \quad (23)$$

We may now write ct as

$$ct = \frac{\sigma(x_T) v_{th}}{2\pi f} n_s V_{eff}. \quad (24)$$

Parameter $n_s V_{eff}$ can be found by integrating $n_s(V_G(t))$ over time with $V_G(t)$ given by (23)

$$n_s V_{eff} = 2 \int_{V_T}^{V_{bias} + \frac{V_{pp}}{2}} \frac{n_s(V_G)}{\sqrt{1 - [2(V_G - V_{bias})/V_{pp}]^2}} dV_G. \quad (25)$$

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

$$x_{T,\max}(f) = -\lambda \ln \left(\frac{2\pi f \ln(3)}{\sigma_0 v_{th} n_s V_{eff}} \right). \quad (26)$$

Using this definition of $x_{T,\max}$ and introducing the energy window ΔE , we can rewrite expression (10) for maximum CP conditions into

$$I_{\text{cp,max}} = fqA_G \int_0^{x_{T,\max}} \overline{N_{\text{it}}}(x_T) \Delta E(x_T, f) dx_T. \quad (27)$$

In this expression, $\overline{N_{\text{it}}}(x_T)$ 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 [1]

$$\Delta E = E_{\text{em},e} - E_{\text{em},h}. \quad (28)$$

Energy level $E_{\text{em},e}$ represents the upper energy level at which the (nonsteady state) emission process of electrons to the conduction band is negligible. It can be expressed as [1]

$$E_{\text{em},e} - E_i = -kT \ln \left(v_{\text{th}} n_i t_{\text{em},e} \sigma(x_T) + e^{\frac{E_i - E_{F,\text{inv}}}{kT}} \right). \quad (29)$$

Similarly, energy level $E_{\text{em},h}$ is the lowest energy level at which the (nonsteady state) emission process of holes toward the valence band is negligible. It is given by [1]

$$E_{\text{em},h} - E_i = kT \ln \left(v_{\text{th}} n_i t_{\text{em},h} \sigma(x_T) + e^{\frac{E_{F,\text{acc}} - E_i}{kT}} \right). \quad (30)$$

In expressions (29) and (30), $E_{F,\text{inv}}$ and $E_{F,\text{acc}}$ are the Fermi levels at inversion and accumulation conditions, respectively. The exponential term limits the highest possible ΔE to the difference in Fermi levels between the inversion and accumulation conditions. Parameters $t_{\text{em},e}$ and $t_{\text{em},h}$ are the times available for the nonsteady-state emission of electrons and holes; these are the times that the device is between the accumulation and inversion conditions. For sinusoidal gate voltages, these depend on frequency and are given by [12]

$$t_{\text{em},e} = t_{\text{em},h} = \frac{Z}{2\pi f}. \quad (31)$$

In this expression, parameter Z is given by [12]

$$Z = \sin^{-1} \left(\frac{2|V_{\text{FB}} - V_{\text{bias}}|}{V_{\text{pp}}} \right) + \sin^{-1} \left(\frac{2|V_T - V_{\text{bias}}|}{V_{\text{pp}}} \right). \quad (32)$$

The CP current at maximum CP conditions can now be described using expression (27) with ΔE given by (28) through (32) and $x_{T,\max}$ given by (26). Using this theoretical framework for describing $I_{\text{cp,max}}$ as a function of frequency, we can both qualitatively and quantitatively explain the observed roll in $Q_{\text{cp,max}}$ as a function of frequency. This rolloff is caused by the interface traps that are too slow to respond to the CP signal applied at the gate. In the next section, we will investigate the implications this trap response has on the applicability of the RF CP technique.

V. APPLICATION

As we can see from Figs. 6 and Fig. 7, the RF CP technique allows us to extract an $I_{\text{cp,max}}$ 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, we will investigate the implications this trap response has on the applicability of CP measurements at radio frequencies.

A. Trap Distribution Extraction

In the previous section, we have derived a model 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 rolloff in the obtained $Q_{\text{cp,max}}$ at frequencies above ~ 100 MHz. The model states that the number of traps that is fast enough to respond decreases with increasing frequency. From Fig. 7, we know that, for the leaky 1.4 nm devices used in this paper, we can only determine $I_{\text{cp,max}}$ at frequencies above 1.5 GHz and 80 MHz for the n -type and p -type devices, respectively. 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 w.r.t. the total number of interface traps, we need to find a good description of the trap distribution within the oxide. $x_{T,\max}$, as it follows from (26), 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 we do 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 paper, however, do have a leakage current sufficiently low for CP measurements to be performed at frequencies where all traps are fast enough to respond. We can, therefore, perform such an accuracy evaluation on one of these 3 nm devices. This result may subsequently be used to evaluate the accuracy that can be achieved in extracting $\overline{D_{\text{it}}}$ on the leaky 1.4 nm devices. In this paper, we make use of the results obtained on the 3-nm n -type device for this purpose. We can extract a complete trap distribution by making use of the frequency response of $Q_{\text{cp,max}}$ and the expressions given in (26)–(32), provided that values of $Q_{\text{cp,max}}$ are available at frequencies low enough for all traps to respond. The latter can be verified by looking at Fig. 8(a). The linear slope of the $Q_{\text{cp,max}}$ versus $\log(f)$ plot for frequencies up to at least 1 MHz is a direct indicator for this.

We extract a trap distribution using the following approach. If we know the amount of traps fast enough to respond to a gate voltage signal with frequency $f - \Delta f$, we can calculate the expected value of $Q_{\text{cp,max}}$ at frequency f for the same number of traps. By comparing this value with the measured value of $Q_{\text{cp,max}}$ at frequency f , $Q_{\text{cp,meas}}$, we 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 $\overline{N_{\text{it}}}$ in the interval $[x_{T,\max}, x_{T,\max} + \Delta x_T]$, where $x_{T,\max}$ and $x_{T,\max} + \Delta x_T$ are related to $f - \Delta f$ and f , respectively, through expression (26). For a sufficiently small value of Δx_T , $\overline{N_{\text{it}}}$ may be assumed constant over the interval $[x_{T,\max}, x_{T,\max} + \Delta x_T]$. We will denote this concentration over the entire interval as $\overline{N_{\text{it}}}(x_{T,\max})$.

Using expressions (9) and (27), we can now derive that:

$$\overline{N}_{it}(x_{T,max}(f)) = \frac{Q_{cp,all}(f) - Q_{cp,slow}(f) - Q_{cp,meas}(f)}{qA_G \int_{x_{low}}^{x_{high}} \Delta E(\sigma(x_T), f) dx_T} \quad (33)$$

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

$$\frac{Q_{cp,all}(f)}{qA_G} = \overline{D}_{it} \Delta E(\sigma_{eff}, f)$$

$$\frac{Q_{cp,slow}(f)}{qA_G} = \int_{x_{high}}^{x_{T,slow}} \overline{N}_{it}(x_T) \Delta E(\sigma(x_T), f) dx_T. \quad (34)$$

Parameter $x_{T,slow}$, in this expression, represents any value of $x_{T,max}$ beyond the location of the slowest interface traps. We use the value of $x_{T,max}$ for a frequency of 100 kHz, as it follows from expression (26). As can be seen from expressions (33) and (34), we need the effective interface state density \overline{D}_{it} in order to extract $\overline{N}_{it}(x_T)$. We can extract this parameter very accurately on the 3 nm n -type device as used in this paper, by making use of the approach discussed in [12]: \overline{D}_{it} is proportional to the linear slope of the $Q_{cp,max}$ versus $\log(f)$ for frequencies of up to 1 MHz and is found to be $3.0 \times 10^{10} \text{ cm}^{-2} \cdot \text{eV}^{-1}$. 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 $1.2 \times 10^{-16} \text{ cm}^2$. Using these values, expressions (33) and (34) and the measured frequency response of $Q_{cp,max}$, we may now extract a trap distribution for the 3 nm n -type device. The result of this is shown in Fig. 9. The inset of Fig. 9 shows the frequency response of $Q_{cp,max}$ 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 III. The figure makes use of two x -axes: The bottom x -axis represents trap distance [in this figure, we use the distance away from the position where $\sigma(x_T) = 10^{-14} \text{ cm}^2$]. The top x -axis represents the frequency for which $x_{T,max}$ is the associated value of x_T at the bottom x -axis as it follows from (26). From Fig. 9, we clearly see that for frequencies up to 2 MHz, $x_{T,max}$ 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 we need information on both the fast as well as the slow interface traps. The slow interface traps cannot be probed on very leaky oxides as we have shown in Fig. 7.

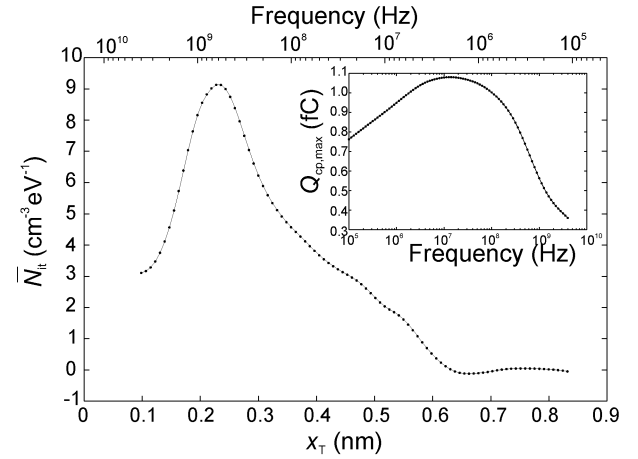


Fig. 9. Extracted trap distribution obtained on an n -type device with $t_{ox} = 3 \text{ nm}$. The inset shows the $Q_{cp,max}$ versus f plot from which this distribution is extracted. The bottom x -axis is the distance from a trap toward the position where $\sigma(x_T) = 10^{-14} \text{ cm}^2$. The upper x -axis shows the frequency where the maximum probing depth equals the associated value on the bottom x -axis. The figure clearly shows that all traps are able to respond to a signal with $f = 2 \text{ MHz}$. At increasing frequencies, an increasing number of traps is located too deep to respond.

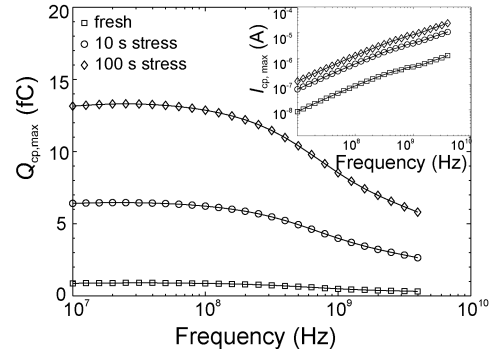


Fig. 10. Measured $Q_{cp,max}$ plotted against frequency before and after a constant gate voltage stress of 3.25 V. The inset shows the corresponding values of $I_{cp,max}$. The figure clearly shows an increase in both $I_{cp,max}$ and $Q_{cp,max}$ over the entire frequency range after stress.

B. Application in Stress Experiments

In order to show a typical example for which the RF CP technique proves very useful on ultraleaky devices, we will also apply the RF CP technique within a stress experiment. Because of the high sensitivity to even a very small increase in the interface traps density, CP measurements are often used in accelerated stress experiments for obtaining a device's lifetime. We make use of a 3 nm n -type device that shows no considerable high leakage current. This allows us to compare results over a very wide frequency range. We measure CP currents using a V_{pp} of 2 V and frequencies ranging from 10 MHz to 4 GHz. Subsequently, we stressed the device using a constant gate voltage stress of 3.25 V and we measured CP currents after a stress of 10 s as well as 100 s. The result of this experiment is shown in Fig. 10. In this figure, we see both the measured $I_{cp,max}$ as well as $Q_{cp,max}$ over the entire frequency range. We clearly see a large increase in $I_{cp,max}$ as well as $Q_{cp,max}$ after stressing the device. For very leaky devices, we may need frequencies into

the gigahertz range before an CP current starts to emerge, such as the 1.5 GHz needed for the 1.4 nm *n*-type device used in this paper. Fig. 10 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.

VI. CONCLUSION

In this paper, we have shown that using the RF CP technique, CP curves can be accurately obtained on devices that show a leakage current density far too high for use in conventional CP measurements. We have given a complete explanation on how to perform such RF CP measurements accurately. An VNA is used to generate high-frequency voltage signals, and the amplitude of the signals can be accurately set by careful modeling of errors in the measurement setup. We have shown RF CP measurement results on various devices, some of which show a considerable leakage current density and some only moderate. Conventional CP measurements cannot be performed on the devices with leakage current densities exceeding $\sim 1 \text{ mA}\cdot\text{cm}^{-2}$. The RF CP measurements, on the other hand, allow us to extract the number of fast interface traps on devices with leakage current densities exceeding $\sim 1 \text{ A}\cdot\text{cm}^{-2}$. This is a significant improvement of the RF CP technique over conventional CP measurements.

We have derived an accurate model for describing the trap response to the gate voltage signals with frequencies into the gigahertz range. From this model, it follows that at increasing frequencies, an increasing number of traps is too slow to respond to the gate voltage signal. RF CP measurements can provide the frequency response of the CP signal at frequencies where this effect is significant. Using RF CP data in combination with the trap response model, a complete trap distribution can be derived, provided that the leakage current does not exceed $\sim 1 \text{ mA}\cdot\text{cm}^{-2}$, the leakage current limit for conventional CP measurements.

As an example for the area of application of the RF CP technique, we have applied RF CP measurements 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 a device, using frequencies of up to 4 GHz. This means that the RF CP technique can be a very powerful tool in the reliability evaluation of leaky devices.

ACKNOWLEDGMENT

The authors would like to thank NXP Research for the use of their samples.

REFERENCES

- [1] 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 Devices*, vol. 31, no. 1, pp. 42–53, Jan. 1984.
- [2] P. Masson, J.-L. Autran, and J. Brini, "On the tunneling component of charge pumping current in ultrathin gate oxide MOSFETs," *IEEE Electron Device Lett.*, vol. 20, no. 2, pp. 92–94, Feb. 1999.
- [3] D. Bauza, "Extraction of Si–SiO₂ interface trap densities in MOS structures with ultrathin oxides," *IEEE Electron Device Lett.*, vol. 23, pp. 658–660, Nov. 2002.

- [4] H. C. Lai, N. K. Zous, W. J. Tsai, T. C. Lu, T. Wang, Y. C. King, and S. Pam, "Reliable extraction of interface states from charge pumping method in ultra-thin gate oxide MOSFET's," in *Proc. ICMTS*, 2003, pp. 99–102.
- [5] H.-H. Ji, Y.-G. Kim, I.-S. Han, H.-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 *IEDM Tech. Dig.*, 2005, pp. 704–707.
- [6] G. T. Sasse, H. de Vries, and J. Schmitz, "Charge pumping at radio frequencies," in *Proc. ICMTS*, 2005, pp. 229–233.
- [7] G. T. Sasse and J. Schmitz, "Charge pumping at radio frequencies: Methodology, trap response and application," in *Proc. IRPS*, 2006, pp. 627–628.
- [8] J. Schmitz, F. N. Cubaynes, R. De Kort, R. J. Havens, 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, May 2004.
- [9] G. T. Sasse, R. J. de Vries, and J. Schmitz, "Methodology for performing RF reliability experiments on a generic test structure," in *Proc. ICMTS*, 2007, pp. 177–182.
- [10] D. M. Pozar, *Microwave Engineering*. New York: Wiley, 1998.
- [11] K. Kurokawa, "Power waves and the scattering matrix," *IEEE Trans. Microw. Theory Tech.*, vol. 13, no. 2, pp. 194–202, Mar. 1965.
- [12] J. L. Autran and C. Chabrierie, "Use of the charge pumping technique with a sinusoidal gate waveform," *Solid-State Electron.*, vol. 39, pp. 1394–1395, 1996.
- [13] F. P. Heiman and G. Warfield, "The effects of oxide traps on the MOS capacitance," *IEEE Trans. Electron Devices*, vol. 12, no. 4, pp. 167–178, Apr. 1965.
- [14] 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 Electron.*, vol. 39, pp. 563–570, 1996.
- [15] Y. Manèglia, F. Rahmoune, and D. Bauza, "On the Si–SiO₂ interface trap time constant distribution in metal–oxide–semiconductor transistors," *J. Appl. Phys.*, vol. 97, pp. 014502-1–014502-8, 2005.



Guido T. Sasse (S'01) received the M.Sc. degree in electrical engineering in 2003 from the University of Twente, Enschede, The Netherlands, where he is currently working toward the Ph.D. degree at the Semiconductor Components Group, MESA+ Institute for Nanotechnology.

His current research interests include CMOS device and circuit reliability and the development of advanced CMOS characterization techniques, including RF measurements.



Jurriaan Schmitz (M'02–SM'06) received the M.Sc. (*cum laude*) and Ph.D. degrees in experimental physics from the University of Amsterdam, Amsterdam, The Netherlands, in 1990 and 1994, respectively.

He joined Philips Research as a Senior Scientist, where he was engaged in the study of CMOS transistor scaling, characterization, and reliability. Since 2002, he has been a Full Professor at the University of Twente, Enschede, The Netherlands. He is the author or coauthor of over 120 journal and conference papers and holds 16 U.S. patents.

Prof. Schmitz has been a Technical Program Committee (TPC) member of the International Electron Devices Meeting (IEDM), the International Reliability Physics Symposium (IRPS), the European Solid State Device Research Conference (ESSDERC), and the International Conference on Microelectronic Test Structures (ICMTS), and is a board member of the Dutch Physical Society (NNV).