Novel Electrical Measurement Techniques for Silicon Devices

by

Arash Elhami Khorasani

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved January 2015 by the
Graduate Supervisory Committee:

Terry Alford, Chair
Michael Goryll
David Theodore

ARIZONA STATE UNIVERSITY

May 2015

ABSTRACT

Semiconductor manufacturing economics necessitate the development of innovative device measurement techniques for quick assessment of products. Several novel electrical measurement techniques will be proposed for screening silicon device parameters. The studied parameters range from oxide reliability, and carrier lifetime in MOS capacitors to the power MOSFET reverse recovery.

It will be shown that positive charge trapping is a dominant process when thick oxides are stressed through the ramped voltage test (RVT). Exploiting the physics behind positive charge generation/trapping at high electric fields, a fast I-V measurement technique is proposed that can be used to effectively distinguish the ultra-thick oxides' intrinsic quality at low electric fields.

Next, two novel techniques will be presented for studying the carrier lifetime in MOS Capacitor devices. It will be shown that the deep-level transient spectroscopy (DLTS) can be applied to MOS test structures as a swift mean for screening the generation lifetime. Recombination lifetime will also be addressed by introducing the optically-excited MOS technique as promising tool.

The last part of this work is devoted to the reverse recovery behavior of the body diode of power MOSFETs. The correct interpretation of the LDMOS reverse recovery is challenging and requires special attention. A simple approach will be presented to extract meaningful lifetime values from the reverse recovery of LDMOS body-diodes exploiting their gate voltage and the magnitude of the reverse bias.

For my beloved family

Majid Elhami Khorasani, Monir Zarei, and Nazanin Elhami Khorasani

and

In memory of

Dieter K. Schroder

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my PhD advisor, Professor Terry Alford. Without his utterly professional guidance and support, I would have not been able to arrive at this stage. I must also take a moment to honor the memory of my late advisor, Dr. Dieter K. Schroder, to whom I am greatly indebted my knowledge. I would like to also thank Dr. Michael Goryll and Dr. David Theodore for being my committee members, and taking the time and interest for evaluating my work. I am specifically thankful to Mark Griswold from ON Semiconductor Co., Phoenix, AZ for being an inspiring manager and mentor throughout my PhD experience.

I would also like to extend my deepest gratitude to my family. Without their encouragement and support, I would not have the chance to follow my graduate studies at Arizona State University.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Positive Charge Trapping and Time Dependent Dielectric Breakdown Behavior in Thick Inter-Layer Dielectrics

1.1.1 Electrical Conduction in Oxides

Electrical conduction mechanisms in dielectrics, based on their limiting factor, can be divided to two general categories [1.1]: electrode-limited and bulk-limited. In electrode-limited mechanisms – such as tunneling – the conduction is limited by the amount of electrons that get injected into the oxide from the cathode. In the bulk-limited, on the other hand, the factor that limits the conduction is the transport of electrons through the dielectric. Electrode-limited mechanisms assume that electrons can move freely in the oxide, while this is not the case for bulk-limited. The bulk-limited mechanisms start to play a role when the dielectric contains considerable amount of shallow level defects.

For the case of thick oxides, the major electrode-limited conduction is through Fowler-Nordheim tunneling [1.1]. It is a specific type of tunneling in which the electrons tunnel through a triangular barrier as opposed to the full oxide thickness in the conventional tunneling. The FN current is expressed through relation (1.1):

$$J_{FN} = AE_{ox}{}^{2}\exp(-\frac{B}{E_{ox}}) \qquad (1.1)$$

where A and B are Fowler-Nordheim constants that after replacing for their constants with numerical values can be calculated as [1.2]:

$$A = 1.54 \times 10^{-6} \times \frac{\left(\frac{m}{m_{ox}}\right)}{\Phi_B} \ [\mathrm{A/V^2}] \tag{1.2}$$

$$B = 6.83 \times 10^{7} \sqrt{m_{ox}{\Phi_B}^3/m} \ [\mathrm{V/cm}] \tag{1.3}$$

with Φ_B being an effective barrier height in eV that takes into account the barrier height lowering and electron quantization effects [1.2], E_{ox} the average electric field in the oxide (V/t_{ox}), m_{ox} the effective mass of electrons in the oxide and m the rest mass of electrons. The ratio m_{ox}/m is often denoted as m^* in the literature that for SiO_2 is approximately equal to 0.5. Equation (1.1) can also be written in the form of (1.4):

$$\ln\left(\frac{J_{FN}}{{E_{ox}}^2}\right) = ln(A) - \frac{B}{E_{ox}} \tag{1.4}$$

Based on (1.4) the *I-V* data plotted as $\ln(J_{FN}/E_{ox})$ vs. $1/E_{ox}$ should be linear if the conduction is purely due to Fowler-Nordheim tunneling. Such plot is often referred to as the *Folwer-Nordheim* plot whose slope yields B and the intercept ln(A). The minimum electric field for the FN current to become detectable is ~5-6 MV/cm and hence care should be taken that at lower electric fields any leakage current is most likely due to other conduction mechanisms, *e.g.* the Poole-Frenkel [1.3]. It has been also reported [1.4] that in the presence of deep-level defects in the oxide a pre-Fowler-Nordheim conduction may be observable at lower fields, as showed in Figure 1.1.

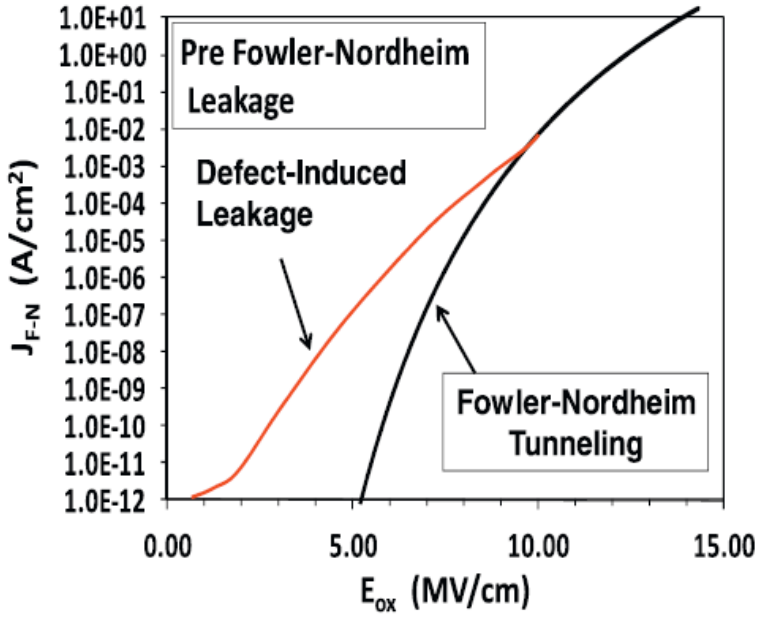

Figure 1.1. The Fowler-Nordheim and pre Fowler-Nordheim conduction through oxide. Reported by McPherson 2012 [1.4].

It must be noted here that the electron/hole trapping can cause the current to deviate from ideal Fowler-Nordheim described by (1.4). Hence a 10% difference from ideal FN tunneling is considered to be an acceptable variation [1.5].

1.1.2 Impact ionization and positive charge generation/creation effect in thick oxides

Both positive and negative charges – which we will refer to them as holes/electrons interchangeably in this context – can be generated and get trapped in an oxide during the tunneling [1.5]. The presence of any charge in any form in the oxide will perturb the local electric field based on Poisson's equation, following:

$$\frac{dE}{dx} = \frac{-Q}{k_{ox}\varepsilon_0} \tag{1.5}$$

Such a change in the oxide's electric field would result the anode and cathode electric fields (E_{an} and E_C) to deviate from what is measured as average electric field in the oxide $E_{ox} = V/t_{ox}$. Since the Fowler-Nordheim tunneling is an electrode-limited process and a function of E_C, deviation of E_C from E_{ox} would result in deviation of leakage current from theoretical FN current [1.6] described by Eq. (1.4).

Based on Eq. (1.5) trapping of positive charges in the oxide would result in an increase in the local E_C, while negative charge would cause a decrease in E_C. Whether if the trapped charge is electron or hole depends primarily on how the stress is applied to the oxide. In case of a constant voltage stress (CVS), electrons would be the major trapped charges [1.4]. For ramped voltage stress (RVT) tests, on the other hand, holes would get trapped more [1.4]. Since our study is focused on ramped voltage tests, we will put more emphasize on hole generation/trapping physics.

There are two major theories on how holes are being generated and consequently get trapped in the oxide at high electric fields. The most widely accepted model by DiStefano and Shatzkes for thick oxides proposes that the creation of positive charge happens via the impact ionization of the tunneling electrons [1.7]. According to this model those tunneling electrons that have higher energy than a threshold energy $E_{TH} \sim 9$ eV (that is the band gap for SiO2) will cause impact ionization giving rise to an excess electron hole pair. The generated electron would be swapped out of the oxide under the applied electric field, while the hole - that has considerably lower mobility than the electron - will get trapped giving

rise to a fixed positive charge density in the oxide Q^+. This model also suggests that the trapped holes may later recombine with subsequent tunneling electrons causing a steady-state density for Q^+ in the oxide. This model is often referred to as the impact ionization-recombination (IR) model [1.6]. A rather older theory in this regard is the impaction ionization of the electrons opposed with the drift of holes (ID model) by O'Dwyer [1.8, 1.9]. These two models comprise a category that explains the hole generation as being an impact process. There are other theories that utilize a non-impact phenomenon, happening at Si/SiO2 interface, as the source for hole generation. However, these models will not be used in this work since the devices that are studies are Metal-Oxide-Metal and do not have any semiconductor/oxide interface. DiStefano and Shatzkes utilize a microscopic or discrete approach to calculate the kinetics of how holes are being generated in the oxide [1.10]. Later on, and due to some deficiency of their model, Solomon [1.11] utilized a macroscopic approach for the same problem through solving the continuum equations. We found that although both approaches can be well used for our purposes, the macroscopic approach provide a more tangible interpretation to our data.

The trapped electron/holes can be thought of as a charge sheet, to the first order approximation, with a characteristic centroid $\overline{x}$ (measure from the cathode) and a density Q^+ with units of C/cm^2. The centroid and the density can be used for characterizing the trapped charge and identifying the changes it imposes on the cathode electric field and so the leakage current. In practice, however, the trapping will likely have spatial distribution that will further complicate the interpretations. Solving (1.5) the new electric field under positive charge trapping can be defined as [1.12]:

$$E_C = E_{ox} + \Delta E_c \tag{1.6}$$

$$\Delta E_c = \frac{Q^+}{k_{ox}\varepsilon_0}\left(1 - \frac{\overline{x_+}}{t_{ox}}\right) \tag{1.7}$$

$$\Delta E_a = \frac{Q^+}{k_{ox}\varepsilon_0}\left(\frac{\overline{x}}{t_{ox}}\right) = \Delta V_{FB}/t_{ox} \tag{1.8}$$

where $E_{ox} = V_{ox}/t_{ox}$ is the average electric field in the oxide with the thickness of t_{ox} under the applied bias of V_{ox}, k_{ox} the oxide dielectric constant, ε_0 the permittivity of vacuum, and ΔE_c and ΔE_a are the electric field changes at the cathode and anode due to positive charge trapping. The change of electric field in the oxide due to trapping is depicted in Figure 1.2 [1.13].

The charge properties are usually investigated through *C-V* and *I-V* measurements. While *C-V* measurements are needed to be done on MOS-C test structures, the *I-V* measurements can be done on simple capacitor test structures. The change in the anode electric field is mostly studied through *C-V* measurements and by looking at the change in flat-band voltage shift ΔV_{FB}. The change in cathode electric field, on the other hand, is often calculated through *I-V* measurements and by looking at the deviation of the leakage current from ideal FN tunneling conduction. It should be noted that in the case of high impact ionization, the generated excess electrons would further complicate the deviation from FN Tunneling making a straightforward measurement of cathode electric field difficult. Another issue during charge study is the "*turn around*" effect while using Al electrodes where interface traps are generated at Si/SiO_2 interface [1.14]. Since simple capacitors are used in this study, the turn-around effect would not be a concern. Moreover,

it has been shown that the created interface states barely influence the electrical conduction through the oxide [1.11].

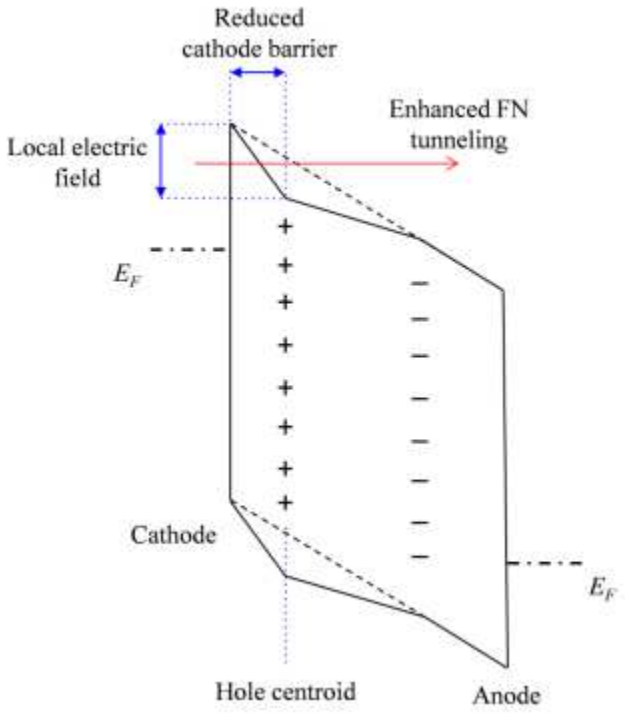

Figure 1.2. Hole/electron trapping and locally increased/decreased cathode electric field.

Although the positive charge trapping is predicted to be a problem in thick oxide (t_{ox} >14 nm), it has been reported as a problem in oxides as thin as 10 nm [1.15]. Positive charge trapping, if underestimated, can lead to difficulties both in interpretation of oxide quality assessments and reliability projections. Most of the reliability models do not take into account the excess positive charge trapping at high electric fields and hence their projection into operating condition might become severely impaired when hole generation is present

at highly accelerated testing conditions [1.16]. Although it is believed that there is a relationship between the rate of hole generation and the time/charge to breakdown [1.4], the problem arises when the origin of these holes is not identified correctly. If the origin of trapped holes is the high field impact ionization, then at operating condition one would never come across breakdown through generation of such holes. In other terms, a change of mechanism in breakdown would happen with acceleration in these cases which would invalidate reliability results. Therefore it is necessary to check for signs of impact ionization before choosing the appropriate reliability model. This includes checking for conduction mechanisms and the deviation of *I-V* data from theoretical values.

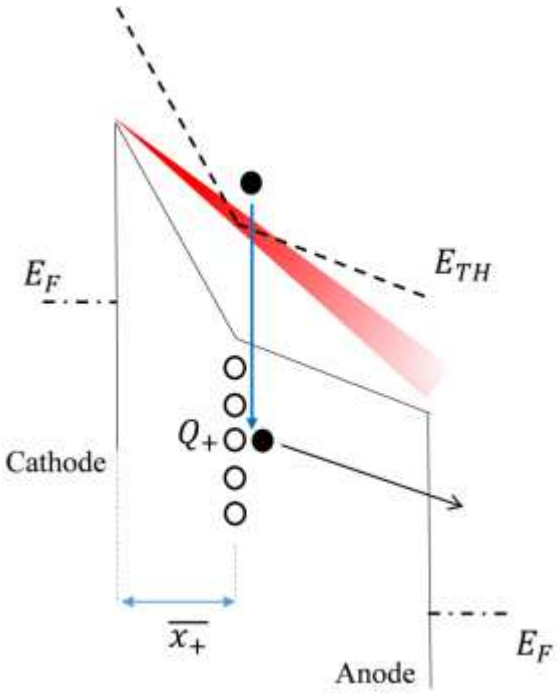

Figure 1.3. Proposed model by DiStefano and Shetzkes [1.17]. The red ray represents the tunneling electrons whose density reduces with distance due to scattering. Those

electrons above impact ionization threshold energy E_{TH} would cause impact ionization and contribute to generation of the positive trapped charge

The electric field above which the impact ionization will become detectable is called E_{TH}. Knowledge about the value of E_{TH} is important from reliability point of view, especially in TDDB constant voltage stress (CVS) tests. If in TDDB-CVS an accelerated electric field higher than E_{TH} is chosen the projection capabilities would become severely impaired. The first sign of being higher than E_{TH} is the increase of current with time due to positive charge trapping. Hence it is necessary to obtain enough information on E_{TH} values before referring to TDDB-CVS results and in order to choose a safe region for stressing.

1.1.3 Time Dependent Dielectric Breakdown of Oxides

Time dependent dielectric breakdown of silica (SiO2) based dielectrics is perhaps one of the more important failures that is of concern in semiconductor industry. This is mainly due to the large utilization of SiO_2 in integrated circuits. A generic definition for TDDB is the failure with time of an oxide that is stressed under an electric field which is lower than its breakdown limit.

There have been various models presented to describe the TDDB behavior. However, a unanimous agreement on a specific model has yet to be achieved [1.18]. Therefore, the choice of an appropriate model is more based on the experience and rather a personal choice for the engineer. The model that will be used is Thermochemical E model

[1.4] with the main reason that we are interested in the TDDB behavior at low electric fields. Moreover, the *E*-model is more pessimistic than the others and therefore safer to use for lifetime predictions.

Most models can generally be categorized to those that are current based and recognize the leakage current as the culprit that lead to the oxide break-down, and those that identify the electric field as the responsible factor behind the breakdown breaking the oxide bonds. Although there is not a unanimous vote on a single model, there are certain facts accepted regarding the TDDB behavior [1.19]:

- TDDB is strongly field dependent
- TDDB can be strongly temperature dependent
- The activation energy for breakdown can be temperature dependent
- Filed acceleration increases with dielectric constant
- Irreversible broken bonds/defects are created during the TDDB
- Generation of defects lead to creation of conductive percolation path
- Statistics of TDDB can be described with Weibull distribution
- Polar dielectrics such as SiO2 show TDDB behavior, while this behavior is very uncommon amongst nonpolar dielectrics.

1.1.4 Thermochemical *E* Model

Amorphous SiO2 (as well as the crystalline SiO_2) is made up of SiO_4 tetrahedrons and is comprised of a continuous random network with short range order (SRO) [1.20]. Within each tetrahedron, one silicon atom is bonded to 4 oxygen atoms with a fixed bonding angel

of 109.5^{o}. These tetrahedrons are connected to each other through their corner oxygen atoms. The intertetrahedron Si-O-Si bond has a dynamic angel of 120-180^{o}. This ever varying angel is one of the factors that gives rise to the amorphous structure of the SiO2. Maximum bond strength exists when this bonding angel is 150^{o}.If an oxygen is absent in the SiO2 network, an oxygen vacancy (or *E'* center) would be created. As a result of an oxygen vacancy the ionic Si-O-Si bond has to be replaced by a rather weak covalent Si-Si bond. The Si-O bond is a polar bond with a bond energy of 5.4 eV, while Si-Si bond is purely covalent and has a bonding energy of 1.8 eV. In other terms, the presence of oxygen vacancy gives rise to weak bonds throughout the oxide. Such weak bonds are believed to act as the nuclei for the oxide breakdown process since they can be broken with much lower energies. The oxygen vacancies can be detected using Electron-Spin Resonance (ESR) spectroscopy [1.21].

Based on Electron-Spin Resonance studies [1.22] the number of broken bonds (dangling bonds) in the oxide increases significantly during the TDDB behavior. This is also in agreement with the percolation theory as the dangling bonds produce a conductive path for the electrons through the oxide. These dangling bonds are believed to be generated through the breakage of weak bonds in the oxide.

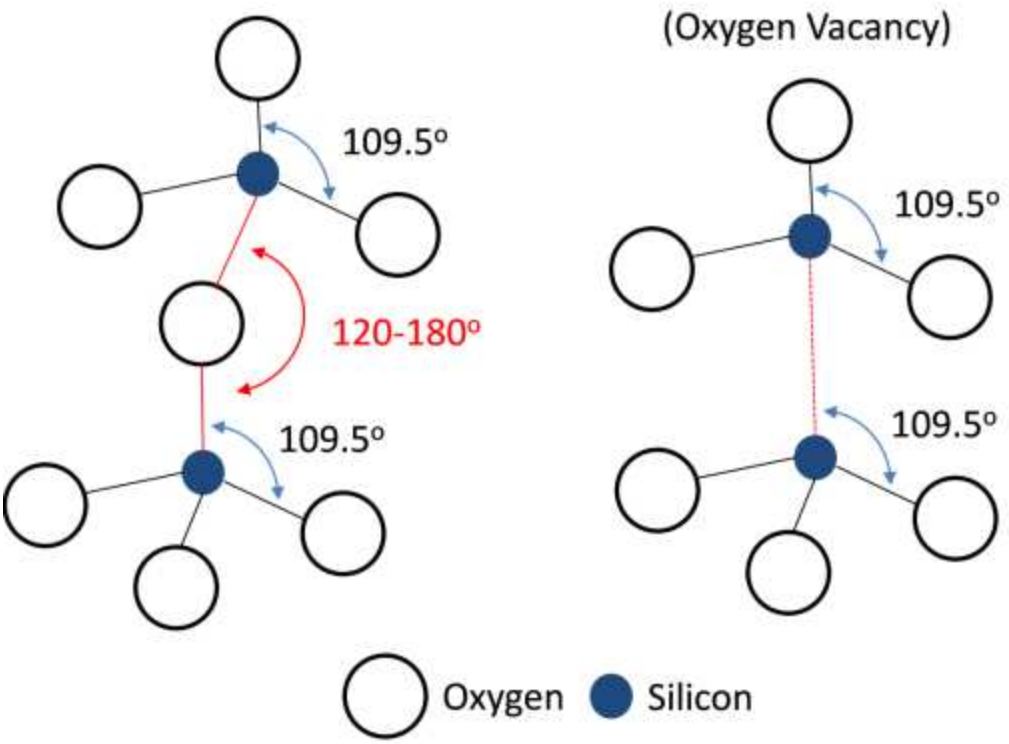

Figure 1.4. Left: two SiO_4 tetrahedrons connected via a Si-O-Si bond. Right: An oxygen vacancy causing a weak Si-Si bond to replace the Si-O-Si bond.

Under an applied electric field (E), due to the polar nature of oxide bonding, a local electric field (E_{loc}), also called the Mossotti field [1.23], will develop throughout the material:

$$E_{loc} = (\frac{2 + K_{ox}}{3})E \tag{1.9}$$

with k_{ox} being the oxide's dielectric constant. The reason for such phenomenon, is the induced polarization created by the electric field and the dipole moment that the molecules feel in addition to the externally applied electric field.

Considering the activation energy that is necessary to break a weak bond (ΔH), based on the Boltzmann statistics and thermodynamics consideration, the rate at which the bond breakage would happen (k) then can be defined as:

$$k = \vartheta_0 exp(\frac{-\Delta H}{k_B T}) \tag{1.10}$$

where ϑ_0 is a vibrational frequency by which the bond is being hit by energetic phonons (or perhaps electrons), and k_B is the Boltzman constant. ΔH is also referred to as the enthalpy of bond breakage. As previously mentioned, in the presence of an electric field (E_{ox}) a much larger local electric field would prevail in the oxide as the created dipole moment tend to distort the bonds. This would lower the enthalpy of bond breakage from typical bond energies to a value given by:

$$\Delta H = \Delta H_0 - p_{eff}(m,n)E_{ox}; \tag{1.11}$$

$$p_{eff}(m,n) = z^* q r_0 [\rho(m,n)]^{-1}(\frac{2+K}{3}) \tag{1.12}$$

with z^* being the number of electron charges on the Si^{4+} cation, r_0 the equilibrium Si-O bond length, and ρ being a function of the bond exponents. The values of ρ function are given in reference [1.24].

It is already mentioned that the dangling bonds (*i.e.,* unbounded-bonds bonds) are responsible for the catastrophic breakdown observed during the TDDB. This can be

quantified stating that there is a critical dangling bond/broken bond density N (1.14) necessary for the percolation path to form and for the breakdown to be observed. Knowing the rate of bond breakage k (1.10), an initial weak bond density $N(t=0)$, and the critical broken bond density, we can write a relationship for the time it takes for the breakdown to happen (1.16). This can be easily done employing a first order rate equation as defined in eqn. (1.13) [1.22]:

$$\frac{dN}{dt} = -k.N(t); \tag{1.13}$$

$$f_c = \frac{N(t=TF)}{N(t=0)}; \tag{1.14}$$

$$TF = \frac{ln(1/f_c)}{k}; \tag{1.15}$$

$$TF = A_0 exp\left(-\frac{\Delta H_0 - \gamma E}{kT}\right). \tag{1.16}$$

Equation (1.16) is the main relationship for the Thermochemical E model that correlates the time to failure (TF) to the electric field (E), activation energy (ΔH_0), and acceleration factor (γ). This equation is defined for the case when there is only one type of weak bonds. It can also be modified to incorporate the existence of various weak bonding modes. Such treatment is covered in details in Ref. 1.22.

1.2 Carrier Lifetime in Silicon

All semiconductor materials and devices contain defects. Some of these are metallic impurities others are structural defects such as dislocations, stacking faults, *etc*. Fortunately, the defect density in virgin silicon has been decreased substantially in recent years [1.25]. Although the impurity density sometimes increases during device fabrication, gettering reduces the defect density to sufficiently low levels. Nevertheless, these defects have, at times, a significant influence on device operation. For example, interface traps at the SiO_2/Si interface and oxide traps in the oxide of MOS devices lead to noise, random telegraph signals, leakage current, stress-induced leakage current, non-volatile memory retention degradation, dynamic random access memory storage time degradation and other deleterious effects [1.2]. Impurities in the silicon lead to reduced recombination/generation lifetimes and increased leakage or dark current. In general, device performance after fabrication is generally very good to excellent, but operation-induced stress during device operation can lead to degradation [1.2].

Defects come in point, line, area, and volume configurations. Examples of these are: point defects: metallic impurity atoms, oxygen atoms, dopants; line defects: edge dislocations; area defects: stacking faults; volume defects: oxide and metal precipitates. The bulk defect density in Si has decreased over the years and Si today contains on the order of 10^9-10^{10} cm^{-3} metallic impurities after crystal growth. The SiO_2/Si interface defect density, on the other hand, has hardly changed with interface trap densities of ~10^{10} cm^{-2} being about the same as 20-30 years ago [1.26]. As a result, device parameters such as junction leakage current, are usually dominated by interface, rather than bulk, effects.

1.2.1 Carriers Generation and Recombination Lifetimes

Carrier lifetime is one of few parameters that can give information about the low defect densities in semiconductors [1.27]. No other technique can detect defect densities as low as 10^9-10^{11} cm^{-3} in a simple, contactless room-temperature measurement. Although impeded by experimental constraints, in theory there is no lower limit to the defect density that can be identified using lifetime measurements. It is for these reasons that the IC community, largely concerned with unipolar MOS devices, uses lifetime measurements as a "process cleanliness monitor."

Different measurement methods can give widely differing lifetimes for the same material or device [1.28]. In many cases, the reasons for these discrepancies are fundamental and are not due to a deficiency of the measurement. The difficulty with defining a lifetime is that we are describing a property of a carrier within the semiconductor rather than the property of the semiconductor itself. Although we usually quote a single numerical value, we are measuring some weighted average for the behavior of carriers under the influence of surfaces, interfaces, energy barriers, and the density of carriers besides the properties of the semiconductor material and its temperature. Lifetime measurements yield effective values influenced by the bulk and surfaces/interfaces.

Lifetimes fall into two primary categories: recombination lifetimes (τ_r) and generation lifetimes (τ_g) [1.29]. The concept of recombination lifetime holds when there is an excess density of carriers in the semiconductor. The excess carrier in this case will decay through recombination so that their density decreases to the equilibrium value. Generation lifetime, on the other hand, applies when there is a paucity of carriers. This is the case in the space-charge region (scr) of a reverse-biased device where the device tries to attain equilibrium.

During recombination an electron-hole pair ceases to exist on average after a time τ_r, illustrated in Figure 1.5(a). The generation lifetime, by analogy, is the time that it takes on average to generate an electron-hole pair (ehp), illustrated in Figure 1.5(b). Thus, generation lifetime is a misnomer, since the creation of an ehp is measured and generation time would be more appropriate. Nevertheless, the term "generation lifetime" is commonly accepted.

When these recombination and generation events occur in the bulk, they are characterized by τ_r and τ_g. When they occur at the surface, they are characterized by the surface recombination velocity s_r and the surface generation velocity s_g, also illustrated in Figure 1.5. Both bulk and surface recombination or generation occurs simultaneously and their separation is sometimes quite difficult. The measured lifetimes are always effective lifetimes consisting of bulk and surface components [1.30].

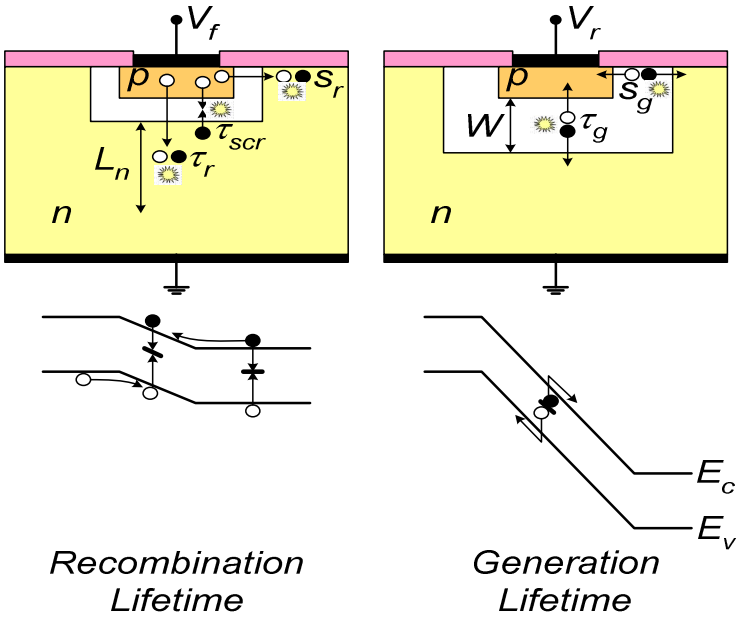

Figure 1.5. Cross section and energy band diagrams for forward- and reverse-biased junctions illustrating recombination and generation. W is the scr width, L_n is the minority carrier diffusion length. V_f and V_r represent the forward and reverse biases.

1.2.2 Recombination Lifetime/Surface Recombination Velocity

Three main recombination mechanisms determine the recombination lifetime: Shockley-Read-Hall (SRH) or multiphonon recombination characterized by τ_{SRH}, radiative recombination characterized by τ_{Rad} and Auger recombination characterized by τ_{Auger}. The three recombination mechanisms are illustrated in Fig 1.6. The actual recombination lifetime (τ_r) is a combination of the mentioned mechanisms and is determined according to the relationship:

$$\tau_r = \frac{1}{\tau_{SRH}^{-1} + \tau_{rad}^{-1} + \tau_{Auger}^{-1}} \tag{1.17}$$

During SRH recombination, electron-hole pairs recombine through deep-level impurities or traps, characterized by trap density N_T, trap energy level E_T, and capture cross-sections σ_n and σ_p for electrons and holes, respectively. The energy liberated during the recombination event is dissipated by lattice vibrations or phonons, as illustrated in Figure 1.6(a). The SRH lifetime is given by:

$$\tau_{SRH} = \frac{\tau_p(n_o + n_1 + \Delta n) + \tau_n(p_o + p_1 + \Delta p)}{p_o + n_o + \Delta n} \tag{1.18}$$

with n_1, p_1, τ_n, and τ_p being defined as:

$$n_1 = n_i \exp\left(\frac{E_T - E_i}{kT}\right);\ p_1 = n_i \exp\left(-\frac{E_T - E_i}{kT}\right);\ \tau_p = \frac{1}{\sigma_p v_{th} N_T};\ \tau_n = \frac{1}{\sigma_n v_{th} N_T} \tag{1.19}$$

and n_0 and p_0 the equilibrium carrier concentrations. It worth noting that n_1 and p_1 represent the magnitude of carrier density if the Fermi level would have been placed at E_T instead of E_F.

The radiative lifetime is defined as:

$$\tau_{rad} = \frac{1}{B(p_o + n_o + \Delta n)} \tag{1.20}$$

with B being the radiative recombination coefficient. It is important to notice that the radiative lifetime is inversely proportional to the carrier density because in band-to-band recombination both electrons and holes must be present simultaneously.

During Auger recombination, showed in Figure 1.6(c), the recombination energy is absorbed by a third carrier. The Auger lifetime is inversely proportional to the carrier density squared. The Auger lifetime is given by:

$$\tau_{Auger} = \frac{1}{C_p(p_o^2 + 2p_o\Delta n + \Delta n^2) + C_n(n_o^2 + 2n_o\Delta n + \Delta n^2)} \approx \frac{1}{C_p(p_o^2 + 2p_o\Delta n + \Delta n^2)} \qquad (1.21)$$

where C_p is the Auger recombination coefficient for a holes and C_n for electrons.

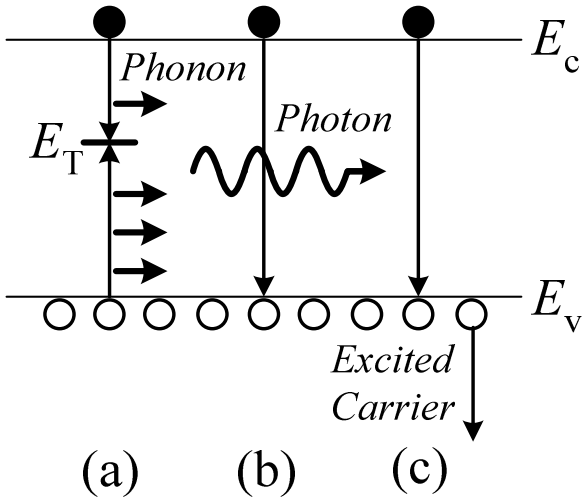

Figure 1.6. Three main recombination mechanisms: (a) SRH, (b) Radiative, and (c) Auger recombination.

All of the three mentioned lifetime mechanisms mentioned above can be devided into two major categories based on the magnitude of the minority carriers: Low-level injection when the excess minority carrier density (Δn) is low compared to the equilibrium majority carrier density (p_o); and high-level injection when minority carrier density is high compared to the equilibrium majority carrier density.

The injection level is important during lifetime measurements. The appropriate expressions for low-level (ll) and for high-level (hl) injection become

$$\tau_{SRH}(ll) \approx \frac{n_1}{p_o}\tau_p + \left(1+\frac{p_1}{p_o}\right)\tau_n \approx \tau_n;\ \tau_{SRH}(hl) \approx \tau_p + \tau_n \tag{1.22}$$

$$\tau_{rad}(ll) = \frac{1}{Bp_o};\ \tau_{rad}(hl) = \frac{1}{B\Delta n};$$

$$\tau_{Auger}(ll) = \frac{1}{C_p p_o^2};\ \tau_{Auger}(hl) = \frac{1}{(C_p + C_n)\Delta n^2} \tag{1.23}$$

The Si recombination lifetimes are plotted in Figure 1.7. At high carrier densities, the lifetime is controlled by Auger recombination and at low densities by SRH recombination. Auger recombination has the characteristic $1/n^2$ dependence. The high carrier densities may be due to high doping densities or high excess carrier densities. Whereas SRH recombination is controlled by the cleanliness of the material, Auger recombination is an intrinsic property of the semiconductor. Radiative recombination plays almost no role in Si because τ_{rad} is so high.

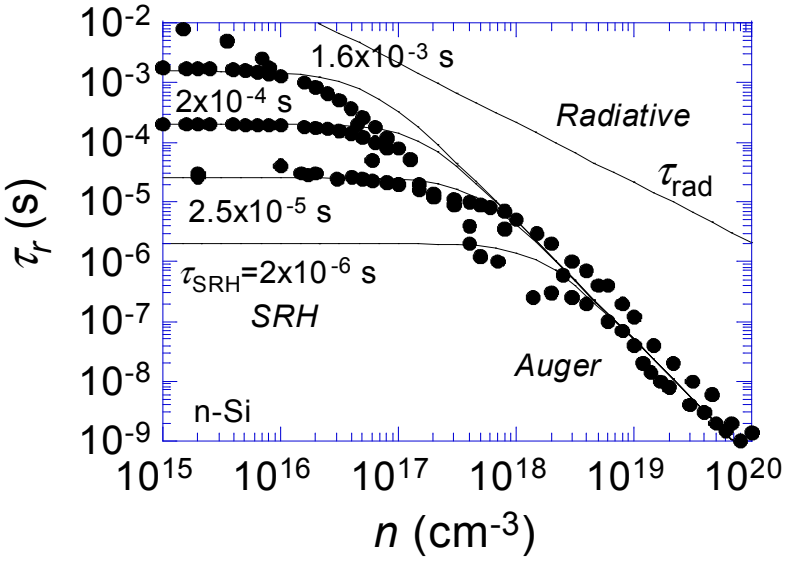

Figure 1.7. Recombination lifetimes due to three different recombination mechanisms in n-type silicon. SRH dominates at low doping density, while auger takes over at higher

doping concentrations. Radiative recombination lifetime is usually higher than the other two for silicon and plays no role [1.2-2]. *Courtesy of D. K. Schroder*.

The surface recombination velocity s_r is:

$$s_r = \frac{s_n s_p (p_{os} + n_{os} + \Delta n_s)}{s_n (n_{os} + n_{1s} + \Delta n_s) + s_p (p_{os} + p_{1s} + \Delta p_s)} \tag{1-24}$$

Just like the bulk lifetime, the surface recombination velocity for low-level and high-level injection becomes:

$$s_r(ll) = \frac{s_n s_p}{s_n (n_{1s} / p_{os}) + s_p (1 + p_{1s} / p_{os})} \approx s_n ; \; s_r(hl) = \frac{s_n s_p}{s_n + s_p} \tag{1.25}$$

where

$$s_n = \sigma_{ns} v_{th} N_{it} ; s_n = \sigma_{ns} v_{th} N_{it} \tag{1.26}$$

and σ_{ns} and σ_{ps} are the capture cross sections of interface traps with density N_{it}.

It is interesting to mention that the surface recombination velocity (s_r), in contract to what may come to mind at first, is not a property that is defined at the surface. In fact, s_r is the speed of electrons who are drifting toward the surface at $x = W$ (and not at $x = 0$). Although the surface recombination happens at $x = 0$, the behavior of minority carriers are defined with $s = s_r$ at $x = W$ [1.25]. The actual surface recombination velocity is s_{eff} that is barely equal to the measured s_r. However, it is a routine practice to use s_r as an estimate for s_{eff}.

1.2.3 Generation Lifetime/Surface Recombination Velocity

Each of the recombination processes of Fig. 1.6 has a generation counterpart. The inverse of multi-phonon recombination is thermal ehp generation. The inverse of radiative and Auger recombination are optical and impact ionization generation. Optical generation is negligible for a device in the dark and with negligible blackbody radiation from its surroundings. Impact ionization is usually considered to be negligible for devices biased sufficiently below their breakdown voltage. However, impact ionization at low ionization rates can occur at low voltages, and care must be taken to eliminate this generation mechanism during τ_g measurements.

$$\begin{aligned}\tau_g &= \tau_p \exp\left(\frac{E_T - E_i}{kT}\right) + \tau_n \exp\left(-\frac{E_T - E_i}{kT}\right) \\ &= 2\sqrt{\tau_n \tau_p}\cosh\left[\left(\frac{E_T - E_i}{kT}\right) + 0.5\ln\left(\frac{\tau_p}{\tau_n}\right)\right]\end{aligned} \tag{1.27}$$

τ_g depends inversely on the impurity density and on the capture cross-section for electrons and holes, just as recombination does. It also depends exponentially on the energy level E_T. The generation lifetime can be quite high if E_T does not coincide with E_i. Typically $\tau_g \approx$ (50-100) τ_r. The *surface generation velocity* is given by

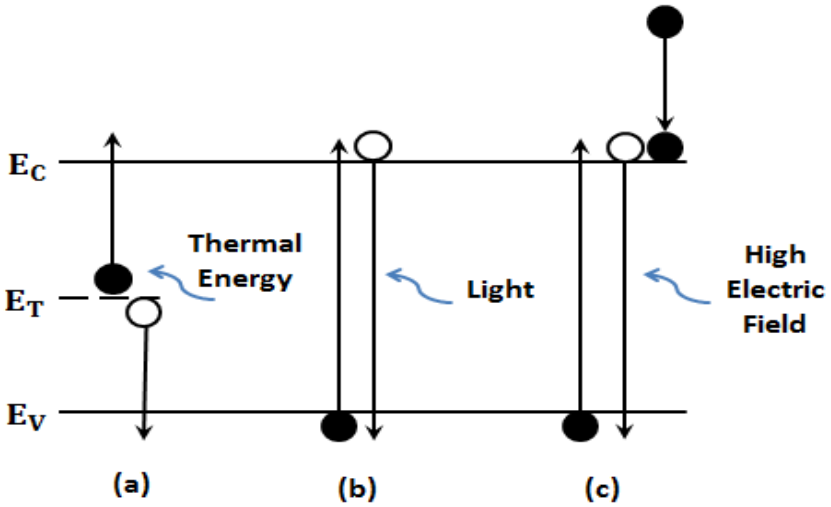

Figure 1.8. (a) thermal generation, (b) optical generation and (c) carrier multiplication.

$$s_g = \frac{s_n s_p}{s_n \exp((E_{it} - E_i)/kT) + s_p \exp(-(E_{it} - E_i)/kT)} \qquad (1.28)$$

based on (1.28) if $E_{it} \neq E_i$ then $s_r > s_g$.

1.3 Summary

The economics of semiconductor industry necessitate the invention and use of measurement techniques that can be implemented in a fast way, and yield critical information on how a device operate, or how a material's properties become affected by known or unknown changes in the process flow. This report will cover several innovative electrical measurement techniques that have been developed in this regard.

Chapter 2 will be focused on the oxide reliability. Employing the physics behind positive charge trapping that is observed while ramping thick oxides, a method will be presented to predict the TDDB behavior of thick inter-layer-dielectrics from fast ramped-

voltage-tests. Given that the TDDB diagnostics tests take weeks to complete, a method that can be performed in shorter time frame would be valuable for screening purposes.

Chapter 3 will be focused on lifetime study using standard Metal-Oxide-Semiconductor Capacitors (MOS-C). A fast technique will be proposed by which generation lifetime can be screened using Deep-Level Transient Spectroscopy (DLTS) on MOS-C devices. It will be tried to come up with a new technique named "*Optically Excited MOS-C*" as a mean to screening the recombination lifetime (τ_r) on MOS-C test structures.

Chapter 4 will be devoted to the reverse recovery study of the body diode of the Laterally Diffused MOSFET (LD-MOS) using Transmission-Line-Pulse technique. An investigation for better understanding of these device's reverse recovery will be presented. Since reverse recovery of LDMOS devices can be important when used in DC-DC converters, such study would greatly help design teams to come-up with efficient power supplies by avoiding switching losses. Moreover, it will be tried to correlate the lifetime with the reverse recovery behavior.

1.4 References

[1.1] H. Bentarzi, *Transport in metal-oxide-semiconductor structures: Mobile ions effect on the oxide properties*, Springer, 2011.

[1.2] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., Wiley-Interscience, New York, 2006.

[1.3] S. M. Sze, *Physics of semiconductor devices*, 3rd ed. Wiley-Interscience, New York, 2007.

[1.4] J. W. McPherson, "Time dependent dielectric breakdown physics – Models revisited," Microelectronics Reliability, vol 52, pp. 1753-1760, 2012.

[1.5] E. Rosenbaum, J. C. King, and C. Hu, "Accelerated testing of SiO_2 reliability," *IEEE Transactions on Reliability*, vol. 43, pp. 70-80, 1996.

[1.6] M. Shatzkes, and M. Av-Ron, "Impact ionization and positive charge in thin SiO_2 films," *Journal of Applied Physics*, vol. 47, pp. 3192- 3202, 1976.

[1.7] M. Shatzkes, M. Av-Ron, and R. M. Anderson, "On the nature of conduction and switching is SiO_2*," J. Appl. Phys.*, vol. 45, pp. 2065-2078, 1974.

[1.8] J. J. O'Dwyer, "Theory of high field conduction in a dielectric," *J. Appl. Phys.*, vol. 40, pp. 3887-3890,1969.

[1.9] Kashat, and N. Klein, "Current runaway in insulators affected by impact ionization and drift," *J. Appl. Phys.*, vol. 58, pp. 5217-5226, 1977.

[1.10] T. H. DiStefano, and M. Shatzkes, "Dielectric instability and breakdown in wide bandgap insulators," *J. Vac. Sci. Technol.*, vol. 12, pp. 37-46, 1975.

[1.11] N. Klein, and P. Solomon," Current runaway in isulators affected by impact ionization and recombination," *J. Appl. Phys.,* vol. 47, pp. 4364-4372, 1976.

[1.12] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "High-field and current-induced positive charge in thermal SiO_2 layers," *J. Appl. Phys.*, vol. 57, pp. 2830-2839, 1985.

[1.13] Aal, "Fast wafer level reliability assessment of ultra-thick oxides under impact ionization," *IEEE Trans. Dev. Mat. Rel.*, vol. 7, pp. 117-120, 2007.

[1.14] D. R. Young, E. A. Irene, D. J. Dimaria, R. F. Dekeersmaecker, and H. Z. Masoud, "Electron trapping in SiO_2 at 295 and 77 °K," *J. Appl. Phys.*, vol. 50, pp. 6366-6372, 1979.

[1.15] R. Bottini, "Oxide reliability: a new methodology for reliability evaluation at parametric testing," *2006 IEEE International Integrated Reliability Workshop Final Report*, pp. 142-145, 2006.

[1.16] U. Schwalke, M. Polzl, T. Sekinger, and M. Kerber, "Ultra-thick gate oxides: charge generation and its impact on reliability," *Microelectronics Reliability*, vol. 41, pp. 1007-1010, 2001.

[1.17] T. H. DiStefano, and M. Shatzkes, "Impact ionization model for dielectric instability and breakdown," *Appl. Phys. Lett.*, vol. 25, pp. 685-687, 1974.

[1.18] C. Hu, and Q. Lu, "A unified gate oxide reliability model," *IEEE 37th International Reliability Symposium*, pp. 47-51, 1999.

[1.19] J. W. McPherson, "Time dependent dielectric breakdown physics – Models revisited," *Microelectronics Reliability*, vol 52, pp. 1753-1760, 2012.

[1.20] Y. Chiang, D. P. Birnie, W. D. Kingery, *Physical Ceramics*, John Wiley & Sons, Inc., 1996.

[1.21] P. Lenahan, J. F. Conley, and B. D. Wallace, "A model of hole trapping in SiO_2 films on silicon," *J. Appl. Phys.*, vol 81, pp. 6822-6824, 1997.

[1.22] R. Degraeve, J. L. Ogier, R. Bellens, P. J. Roussel, G. Groeseneken, and H. E. Maes, "A new model for the field dependence of intrinsic and extrinsic time-dependent dielectric breakdown," *IEEE Trans. Electron Dev.*, vol. 45, pp. 472-481.

[1.23] C. Kittel, *Introduction to solid-state physics*, 4th ed., Wiley, New York, 1971.

[1.24] 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, pp. 1513-1523, 1998.

[1.25] K. V. Ravi, "*Imperfections and impurities in semiconductor silicon*," John Wiley & Sons, 1981.

[1.26] D. Fleetwood, "Defects in microelectronic materials and devices," CRC Press, 2009.

[1.27] D. K. Schroder, "Carrier lifetime in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[1.28] D. K. Schroder, B. D. Choi, S. G. Kang, W. Ohashi, K. Kitahara, G. Opposits, T. Pavelka, and J. Benton, "Silicon epitaxial layer recombination and generation lifetime characterization," *IEEE Trans. Electron Dev.,* vol. 50, pp. 906-912, Apr. 2003.

[1.29] D.K. Schroder, "The concept of generation and recombination lifetimes in semiconductors," *IEEE Trans. Electron Dev.,* vol. ED-29, p. 1336, 1982.

[1.30] D. K. Schroder, "effective lifetime in high quality silicon devices," *Solid-St. Electron.*, vol. 27, pp. 247-251, Mar. 1984.

CHAPTER 2

POSITIVE CHARGE TRAPPING AND ITS APPLICATION IN DIFFERENTIATING THICK INTER-LAYER- DIELECTRICS

2.1 Introduction

Semiconductor manufacturing economics necessitate the development of novel and innovative techniques that can replace the traditional time consuming reliability methods (*i.e.* the constant voltage stress time-dependent dielectric breakdown (CVS-TDDB) tests). We show that positive charge trapping is a dominant process when thick oxides are stressed through the ramped voltage test (RVT). Exploiting the physics behind positive charge generation/trapping at high electric fields, a fast *I-V* measurement technique is proposed that can be used to effectively distinguish the ultra-thick oxides' intrinsic quality at low electric fields. It will be demonstrated, based on experimental data, that our proposed technique can be a suitable replacement for the CVS-TDDB as a quality screening tool.

2.2 Experimental Setup

Metal comb fingers with oxide in between were used as test structures in this study. These metal comb fingers have spacing (*i.e.*, oxide thicknesses) of 0.35, 0.45, 0.55, 0.65, 0.75 and 0.85 μm. We will use the terms "oxide" and "inter-layer-dielectric" (ILD) interchangeably throughout this report. The oxide is a fluorinated SiO_2 deposited via PE-CVD with an approximate dielectric constant of $k_{ox} = 3.315$. The area of each structure was determined from capacitance measurements performed with an Agilent 4284A precision LCR meter and utilizing the $C = k_{ox}\varepsilon_0 A/t_{ox}$ relation. The density-factors of the test structures are identical for all samples (*i.e.*, the consequent fingers are of equal width). The test

structures are fabricated on 8'' wafers for wafer level testing. Figure 2.2 shows the typical metal comb finger test structures we used in this study.

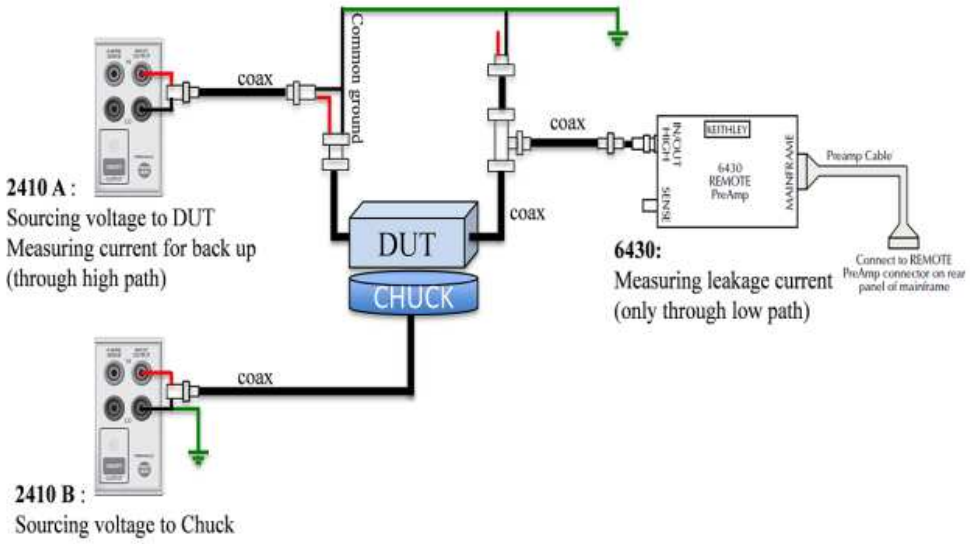

Figure 0.1. The measurement setup used in this study.

We were provided with three different wafer splits with different oxide fabrication process flow –referred to as oxide-A, -B and -C. The exact result of the change in fabrication specifications was neither investigated nor validated before our study. It is assumed that the process change might have led to some unintentional changes in the oxide quality.

The measurement setup was designed to guarantee the detection of a break down, when happens, through ILD instead of secondary paths such as the pad to substrate. Since the oxide break-down happens through the weakest link, design flaws may – especially at

wafer level – lead to availability of secondary paths for which are not of interest (*e.g.* a break down through pad to substrate). This turned out to be the case for the wafers we were provided with as the design had made it possible for the break down to happen in the pad to substrate path due to that spacing being comparable with the finger's spacing. Our tests showed that this path would break down at a bias level of around 320 V corresponding to a spacing of about ~ 0.5 μm. Such problem would have rendered all test structures with t_{ox} greater than 0.5 μm useless as the detected break down was not through the ILD and hence not valuable for reliability studies. In order to avoid this technical issue a constant bias was applied on the substrate during all the measurements to avoid break down in pad-to-substrate path. For this to happen, two Keithley 2410 source-meters were used for biasing the substrate as well as the test structures. A Keithley 6430 Sub-Femtoamp meter was used to monitor the leakage current through ILDs with high precision. The Sub-Femtoamp meter was placed on the low-path side of the structures to insure the reading being from the ILDs. The current was also read intermittently from the 2410 source-meter to ensure a 100% breakdown detection capability regardless of its path. The later was necessary to be able to stop the test after the breakdown to protect the instruments against high potential that would have released as oxide became shorted. The measurement configuration is showed in Fig. 2.1. All of the measurement instruments were controlled through GPIB board using a Matlab platform.

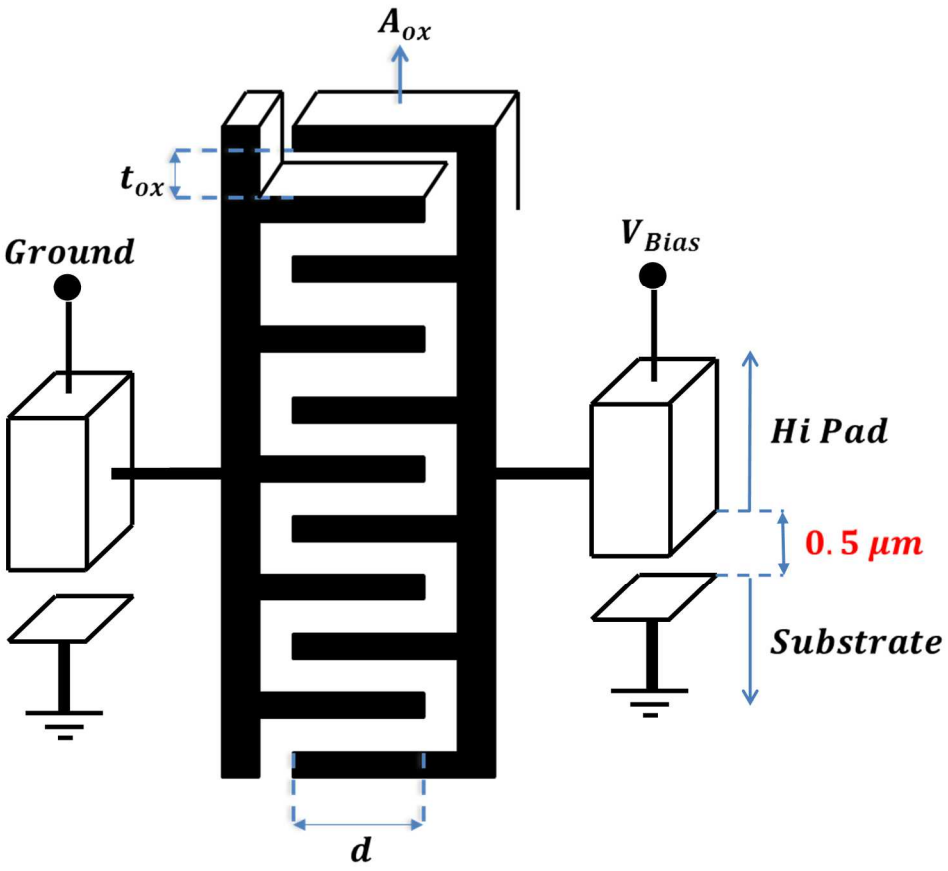

Figure 2.2. Metal comb finger test structures used for wafer level testing in this study. *d* is the finger length, A_{ox} the capacitor area, and t_{ox} is the oxide thickness.

2.2 *I-V* Behavior of Ramped Voltage Tests (RVT)

The first step for correctly assessing an oxide's quality is to study its corresponding leakage *I-V* characteristics and to identify the active electrical conduction mechanisms. Fig. 2.3 shows a typical measured *I-V* plot for the structures we studied. Although, at first glance, one may ascribe this *I-V* behavior to an electrode-limited (*i.e.*, FN tunneling)

conduction at high fields and a bulk-limited (*i.e.* Poole-Frenkel) conduction at intermediate fields, as will be explained, this is not true. The major problem with such identification is the fact that the leakage current at intermediate fields was observed not to have any strong temperature dependence, as showed in Fig. 2.7. The lack of temperature dependence rules out the Poole-Frenkel conduction as a highly temperature dependent mechanism, as (2.1) suggests [2.1]:

$$J_{PF} = CE_{ox}\left[\frac{-q(\varphi_B - \sqrt{qE_{ox}/\pi\varepsilon_0})}{k_{ox}T}\right] \tag{2.1}$$

with φ_B being the barrier height, $\varepsilon_0$the permittivity of vacuum, k_{ox} oxide's dielectric constant, and C the Poole-Frenkel constant.

Moreover, as will be explained shortly, it was observed that the leakage current is a strong factor of the ramp rate used in the RVT. Such behavior cannot be easily justified with traditional conduction mechanisms and beckons that a more complex phenomenon is happening in the oxide.

Surprisingly, it turned out that the rise in current observable at high electric fields (Fig. 2.3) is due to excessive positive charge trapping in the oxide followed by an enhanced FN tunneling. The leakage at intermediate electric fields, on the other hand, was identified to be the ideal FN tunneling utilizing the Fowler-Nordheim plots. Figure 2.3 shows the perfect fit of theoretical FN equation to the measured data at intermediate fields and the deviation of current happening at higher electric fields. Such deviation is in good agreement with

what one would expect when positive charges pile up in the oxide at high electric fields, as explained in chapter 1.

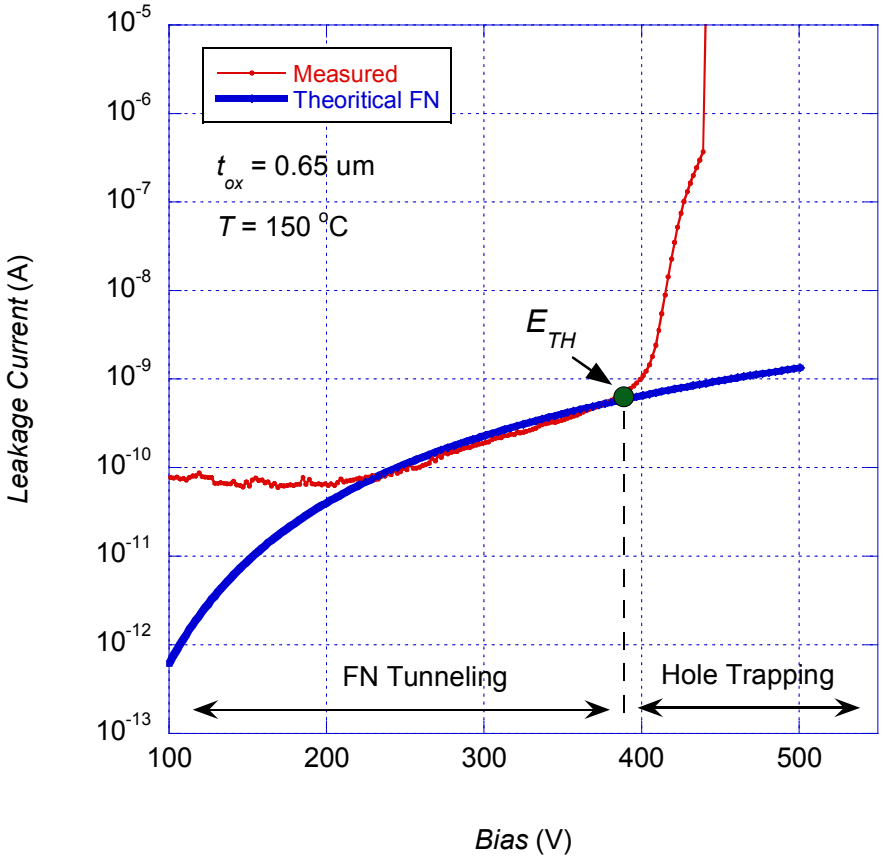

Figure 0.3. Measured *I-V* plot of oxide and the theoretical Fowler-Nordheim tunneling current. There is a good agreement between theory and measured data at moderate electric fields.

The proposed conduction mechanism, can well explain various effects that are observed in the measured *I-V* plots. The FN tunneling conduction has poor (if any) temperature dependence, which is in harmony with our experimental data, as showed in

Fig. 2.7. It is also interesting to notice that the electric field at which the leakage starts to deviate from theoretical FN tunneling (E_{TH}) depends strongly on the oxide thickness, as showed in Fig. 2.5. Such dependence arises from the fact that thicker oxides are more sensitive to positive charge trapping. The reason behind this is that, as reported extensively by others [2.2, 2.3, 2.4, 2.5], as the oxide thickness increases the trapped positive charge centroid gets closer to the cathode giving rise to a more pronounced change in the cathode electric field (E_C) based on (1.6) and (1.7).

Figure 2.4 shows another interesting sign of positive charge trapping: the break down voltage V_{BD} becomes a function of ramp rate under hole trapping conditions. Knowing that the trapped charge density is a function of time, such behavior is justified by taking into account that under a lower ramp rate more time is provided for positive charges to be generated and trapped, in contrast to higher ramp rates. It was observed that using very high ramp rates would lead to arbitrary V_{BD} which we think is due to localized break down.

Having excessive positive charge trapping in the oxide due to impact ionization at high fields will cause the local field at cathode E_C to become considerably larger than the average electric field in the oxide E_{OX}. Since the property that is measured in the RVT is the E_{OX}, the measured electric field at break down becomes a function of oxide thickness. This is due to the fact that the break down is dominated by E_C rather than E_{OX}. Such thickness dependence is again because the oxide thickness increases the effect of trapped positive charge density on E_C becomes more pronounced as equations (1.6) and (1.7) show. We observed such behavior as demonstrated in Fig. 2.6 for oxides with varying thicknesses of 0.35 through 0.85 μm.

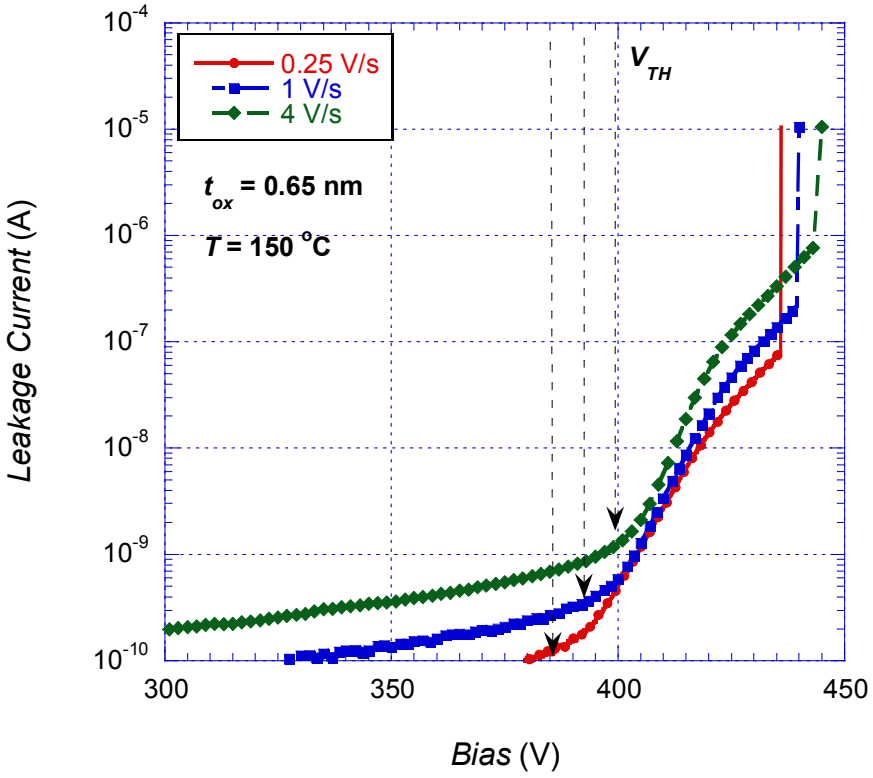

Figure 0.4. Ramp-rate dependence of *I-V* plots in the presence of positive charge trapping for oxide thickness of 0.65 μm. The higher the ramp-rate the higher V_{TH}, V_{BD}, and I_{BD}.

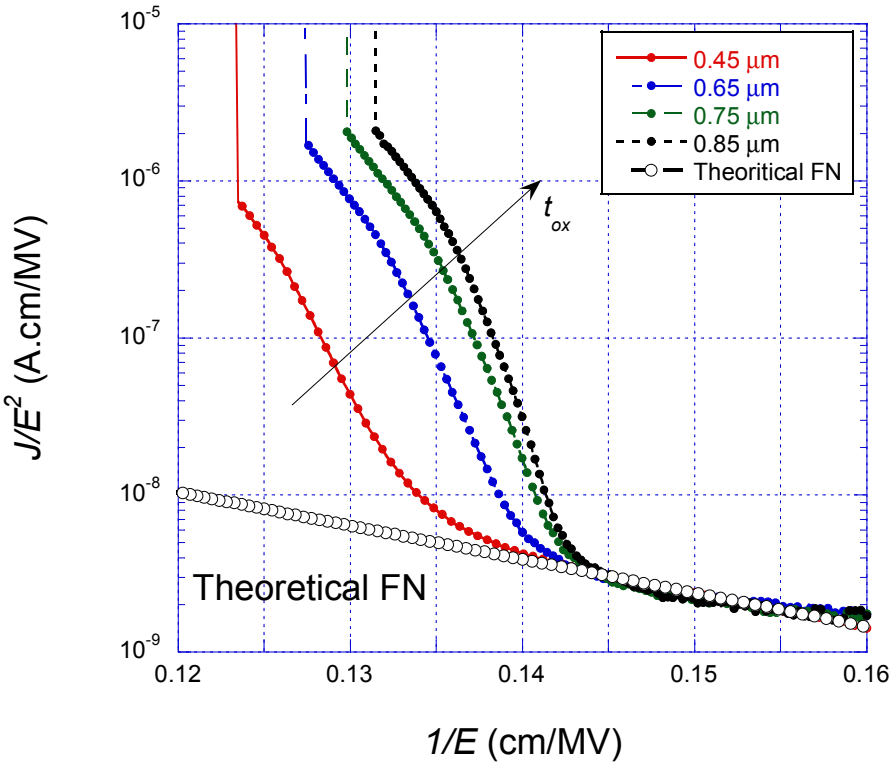

Figure 0.5. Thickness dependence of the threshold electric field, at which the electrical conduction in the oxide deviates from theoretical FN tunneling. The thicker the oxide the lower the threshold electric field.

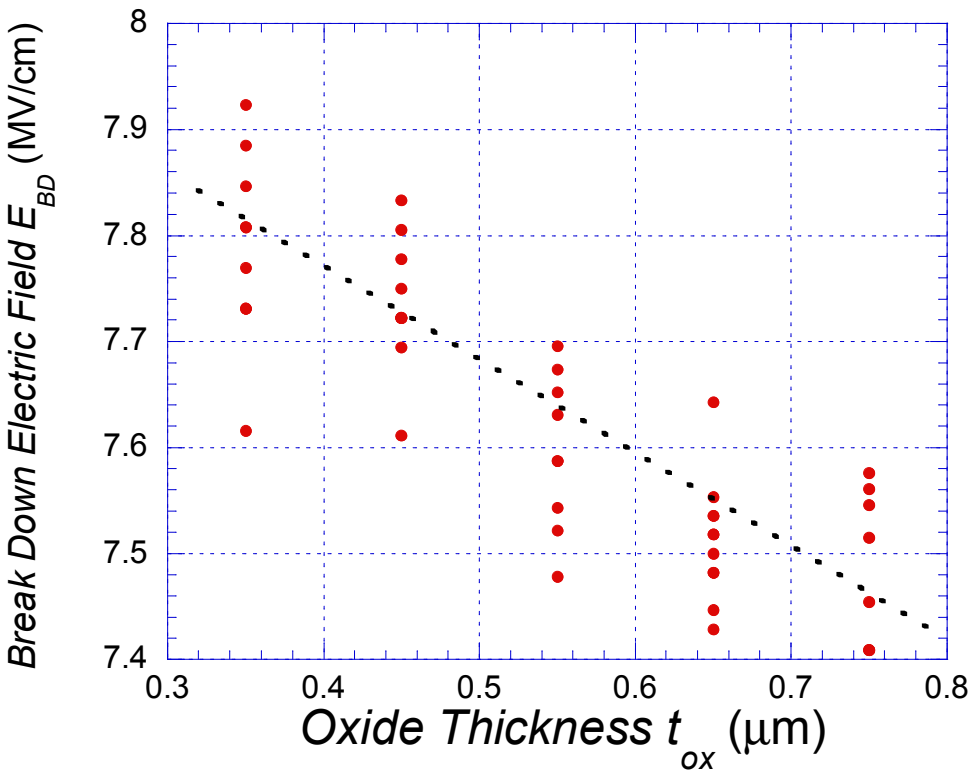

Figure 0.6. Thickness dependence of average electric field at break down. The data is obtained from 80 structures tested through a RVT with ramp rate of 2 V/s at T =150 °C.

Figure 2.7 shows the *I-V* plots for an oxide with a thickness of 0.65 μm measured at two different temperatures of 150 and 200 ºC. As mentioned before, electron trapping is expected to be present at intermediate electric field levels giving rise to a decreased cathode electric field. The electron trapping has been shown to reduce with increasing temperature due to the enhanced remission for shallow levels and reduced efficiency for deep levels [2.6]. This justifies the rise in the FN tunneling current with temperature. It is worth noting that there is no considerable difference in E_{TH} and E_{DB} as the temperature changes. However, the differential resistance in the positive charge-trapping region of the *I-V* (the

slope) is lower for higher temperatures. The *I-V* slope, on which we will focus in the next section, is a direct measure of the impact ionization rate α which itself is related to the mean free path of electrons in the oxide λ [2.4]:

$$\alpha = \alpha_0 \exp(\frac{-K}{E}) \tag{2.2}$$

$$K = \frac{\Delta E}{\lambda} \tag{2.3}$$

where α is the impact ionization rate, α_0 impact ionization constant, ΔE energy of electrons above E_{TH}, and λ is the mean free path for electrons in the oxide (MFP). The MFP, although a intricate concept, has some widely accepted characteristics like being temperature dependent. The λ of electrons is believed to decrease with temperature giving rise to an increase in resistivity with temperature [2.7]. Such temperature dependence of λ justifies the change in *I-V* slope with temperature, showed in Fig. 2.7.

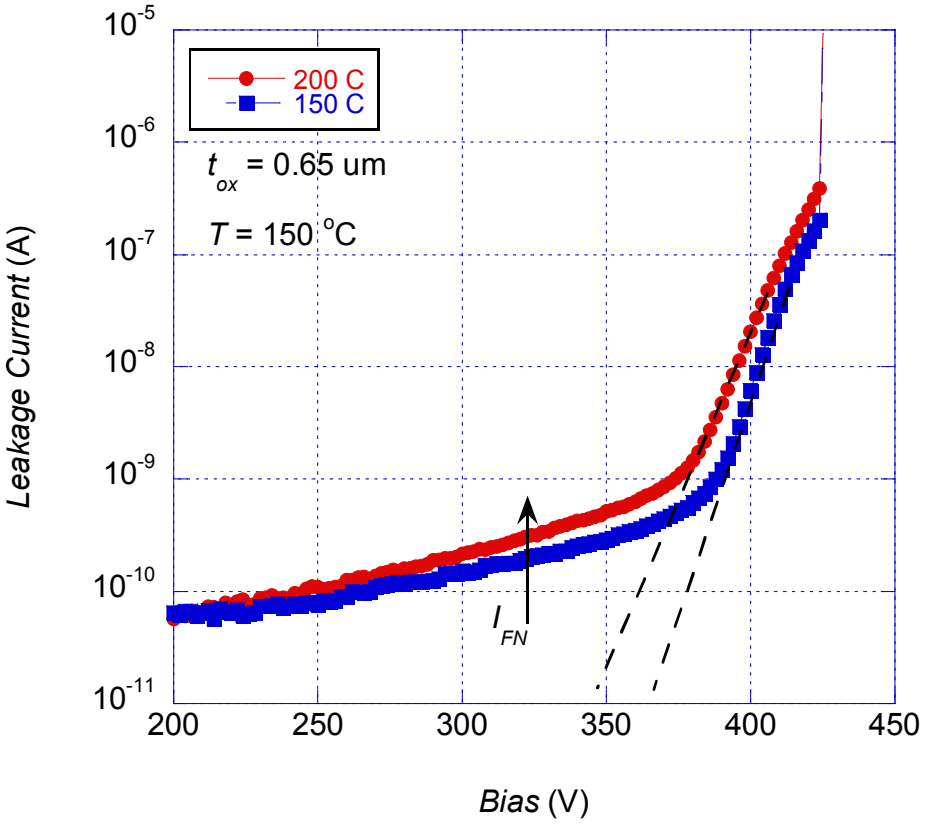

Figure 2.7. Effect of temperature on charge trapping for T = 150 and 200 °C. Oxide thickness is 0.65 μm.

2.3 A Fast I-V Screening Measurement For TDDB Assessment of Ultra-Thick Inter-Layer Dielectrics

The breakdown voltage V_{BD} that is detected through RVT in the measured I-V plots does not necessarily always correlate with the time dependent dielectric breakdown TDDB, not at least directly. This break down voltage is often attributed to fast joule heating processes due to the high current density that passes through the oxide at high electric fields [2.8]. In other terms, in this case the break down is controlled by the density of the current

passing through the oxide. At operating condition, however, the time-dependent break-down follows a more complicated process and is believed to be under the control of the electric field and better understood at molecular levels [2.9]. Therefore the RVT breakdown voltage distribution cannot be a good indicator of oxide's intrinsic quality at low electric fields where TDDB is of great concern. In order to correlate the *I-V* behavior of an oxide to its intrinsic quality (*i.e.* median time to failure MTTF at low electric fields) it is necessary to use the *I-V* as a tool for understanding the dielectric's physics and use them to predict the TDDB properties.

Time dependent dielectric breakdown (TDDB) is one of the most important failure mechanisms in SiO_2-based dielectrics. Although still open to debate, the most widely accepted model for explanation of TDDB behavior at low electric fields is the thermochemical *E* model [2.8] that correlates the time to break down t_{BD} with the effect of the applied electric field on the dielectric's molecular bonds, as briefly explained in chapter 1. Extraction of any TDDB model's parameters requires, at least, two unique distributions of TDDB constant voltage stress (CVS) test results. The inherent problem with the CVS tests is however the long time they take to complete as one tries to avoid changes in failure mechanism by running them close to the operating condition. These times can easily reach some thousands of hours for thick oxides. It is for this reason that the IC community, largely concerned with SiO_2, uses ramped voltage tests (RVT) - that are done in minute's time frame - for process screening purposes. However, since the RVT breakdown is merely a Joule heating failure it cannot provide an adequate insight into the TDDB failure at low electric fields, as mentioned before. When dealing with thick oxides, on the other hand, the excess positive charge generation/trapping that happens at high electric fields can be well

characterized through RVT *I-V* plots. Since the positive charge generation is an impact process [2.4] that is related to the mean free path (MFP, λ) for electrons [2.10], it can be employed to predict the oxide's time to breakdown (t_{BD}) based on thermochemical *E* model.

Figure 2.8 shows the *I-V* plots for two oxides that have different process flow specifications but are similar in other parameters (*e.g.* the RVT ramp rate, the t_{ox}, k_{ox}, and *T*). It can be easily observed that the RVT V_{BD} is essentially the same for these two oxides. However, one finds the threshold voltage V_{TH} at which the current starts to deviate from theoretical Fowler-Nordheim and the slope of *I-V* in the enhanced Fowler-Nordheim region to be completely different for these two oxides.

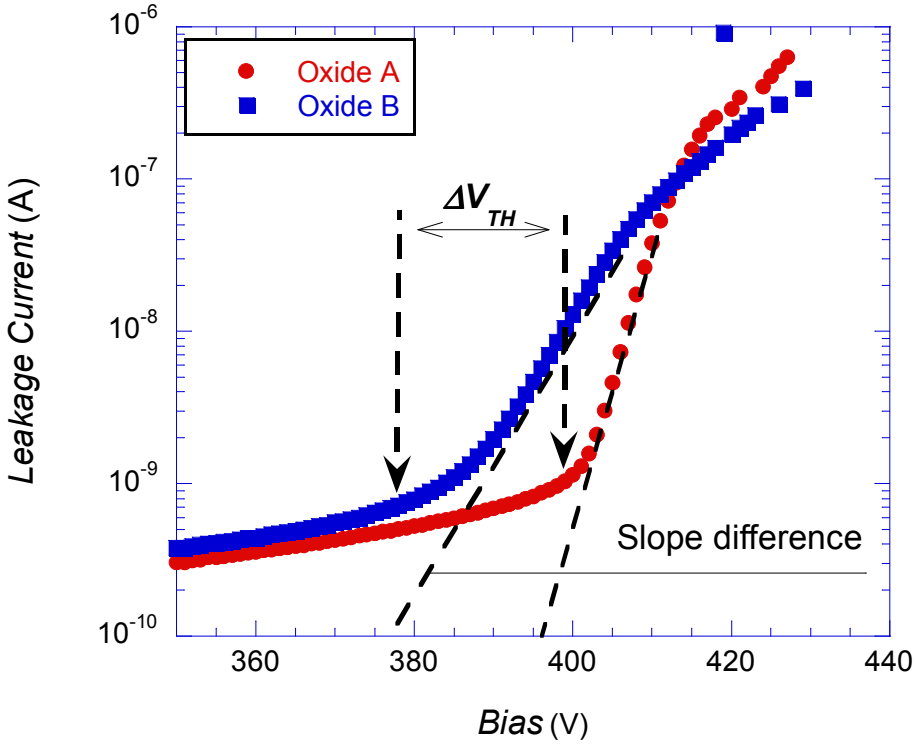

Figure 0.8. *I-V* plots for the leakage current of two different oxides with different processing specifications. The variation in V_{TH} for impact ionization as well as the slope of the *I-V* in the

enhanced FN region is obvious. Both wafers are showing same break down behavior. Plots are for mean data of 16 measurements on each structure.

Before starting to further scrutinize these differences, one key concern has to be resolved: May such difference be due to variation of oxide thickness from wafer to wafer? In order to make sure that the difference in V_{TH} and the slope is not a consequence of oxide thickness variation, capacitance measurements were performed on both wafers. Figure 2.9 shows the capacitance vs. structures of different t_{ox}. The capacitance difference is less that 0.1% which corresponds to a less than 0.1% difference in A_{ox}/t_{ox}, where A_{ox} is the capacitor's area following:

$$C = \frac{k_{ox}\varepsilon_0 A_{ox}}{t_{ox}} \tag{2.4}$$

Figure 2.10 shows two *I-V* plots for same oxides with different spacing. It can be easily observed that when the *I-V* curves belong to the same oxide the V_{TH} and slope variations are not observed any more. It is also interesting to notice that, as depicted in Fig. 2.10, the variation of V_{TH} is well projected into V_{BD} for the two samples. Such comparison reassures us that the two above mentioned *I-V* differences should be due to a difference in the oxide bulk material rather than design differences.

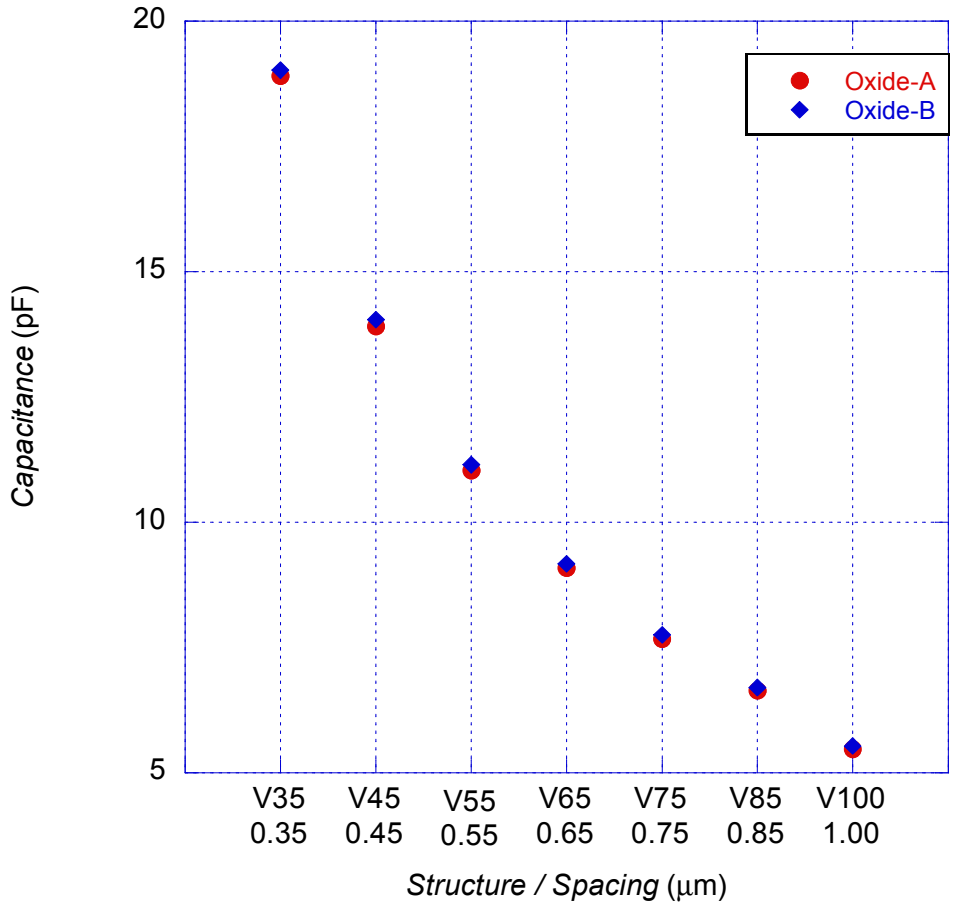

Figure 0.9. Capacitance values for different structures with different spacing for two Oxide-A and Oxide-B. The capacitance difference is less than 0.1% indicating no thickness difference.

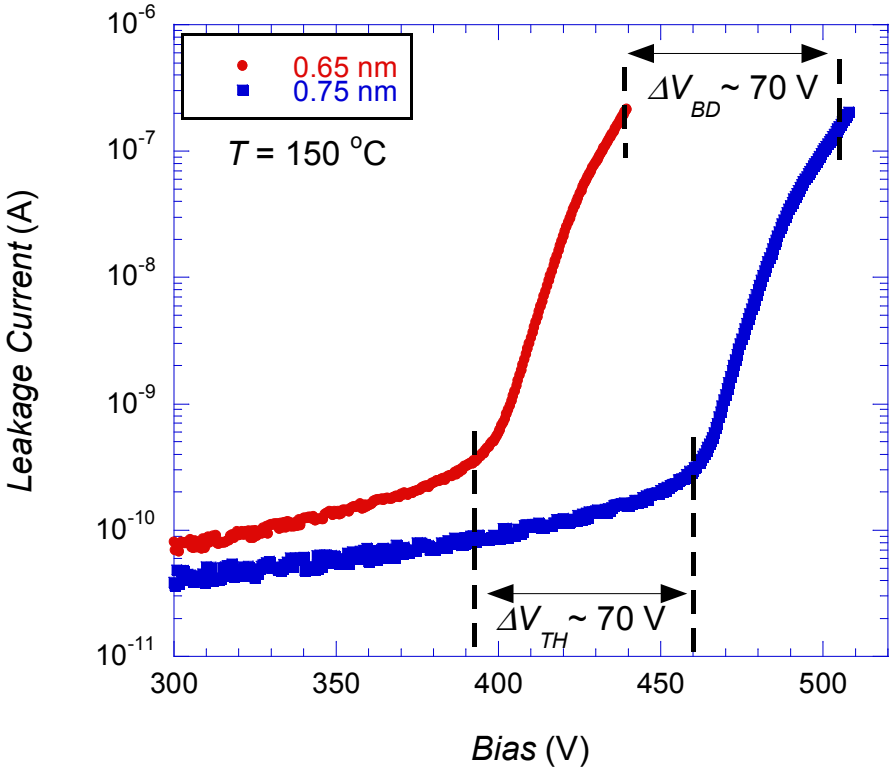

Figure 0.10. *I-V* plots of capacitors with different oxide thicknesses from the same wafer. The V_{TH} variation due to thickness is well projected into differences of V_{DB}. T=150 C, t_{ox}=0.65um, and E_{ox}=6.5 MV/cm

Now the fundamental point is the interpretation of the observed differences in the *I-V* plots for different oxides. In other terms, there is a need to describe the *I-V* slope variations in the enhanced FN region is related to the oxide's properties.

For this purpose we will employ the macroscopic approach to study the hole generation phenomenon in the oxide [2.10], [2.4], as it provides a more mathematically flexible insight into the problem.

It has been shown that, at high electric fields, the impact ionization due to the tunneling electrons can give rise to the formation of trapped positive charges in thick oxides [2.11]

(t_{ox} >14 nm). Based on the Poison's equation, the accumulation of such positive charge would cause the local electric field at the cathode E_C to deviate from the average electric field in the oxide $E_{ox}=V/t_{ox}$ following [2.4]:

$$E_C = E_{ox} + \Delta E_c \tag{2.5}$$

$$\Delta E_c = \frac{Q^+}{k_{ox}\varepsilon_0}\left(1 - \frac{\overline{x}}{t_{ox}}\right) \tag{2.6}$$

where Q^+ is the charge density, $\overline{x}$ the charge centroid measured from the cathode, k_{ox} the dielectric constant of the oxide, and ε_0 is the vacuum permittivity. As it is showed in Fig. 2.3, the local increase in E_C would result in the oxide's leakage current density to deviate from the theoretical Fowler-Nordheim (FN) tunneling current defined by:

$$J_{FN} = AE_{ox}{}^2 \exp\left(-\frac{B}{E_{ox}}\right) \tag{2.7}$$

with A and B being the FN constants, and E_{ox} the average electric field. This deviation is partly due to the locally reduced triangular tunneling barrier as the cathode electric field increases due to the positive charge accumulation [2.3].

Since the positive charge generation by impact ionization has a recombination counterpart (IR model), the rate of positive charge accumulation in the oxide can be written in terms of [2.10]:

$$\frac{\partial Q^+}{\partial t} = \frac{J}{q}(\alpha - \sigma Q^+) \tag{2.8}$$

where J is the tunneling current density, q charge of an electron, α impact ionization rate, and σ is the recombination capture cross section. Using the lucky-electron concept, Solomon [2.10] proposed α to be proportional to $\exp(-1/\lambda)$.

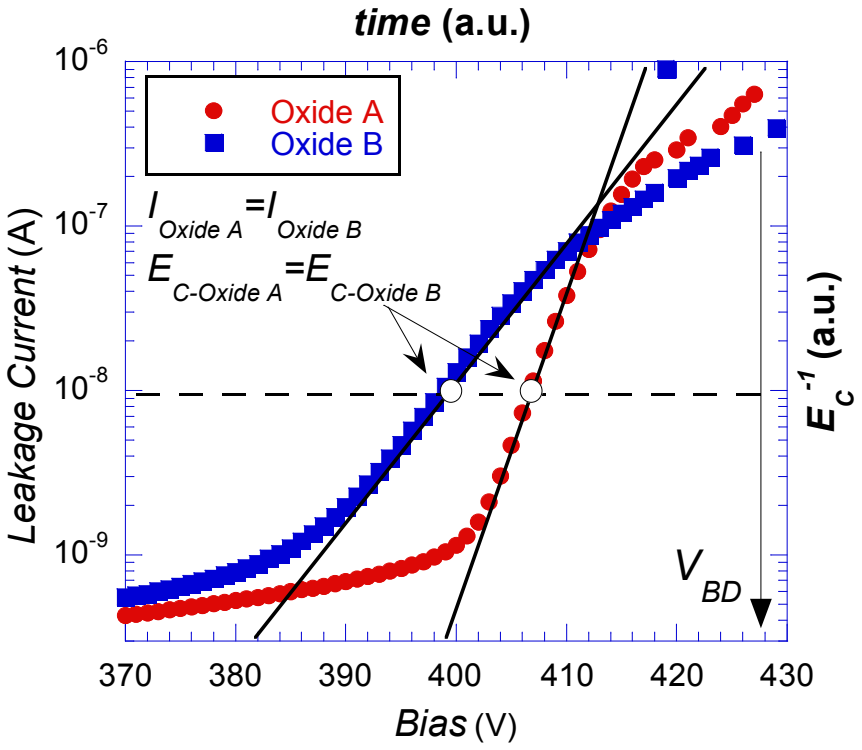

Figure 0.11. Lines of equi-current, equi-E_C crossing the I-V for two different oxides with different slopes. T=150 ºC, t_{ox}=0.65um, and E_{ox}=6.5 MV/cm.

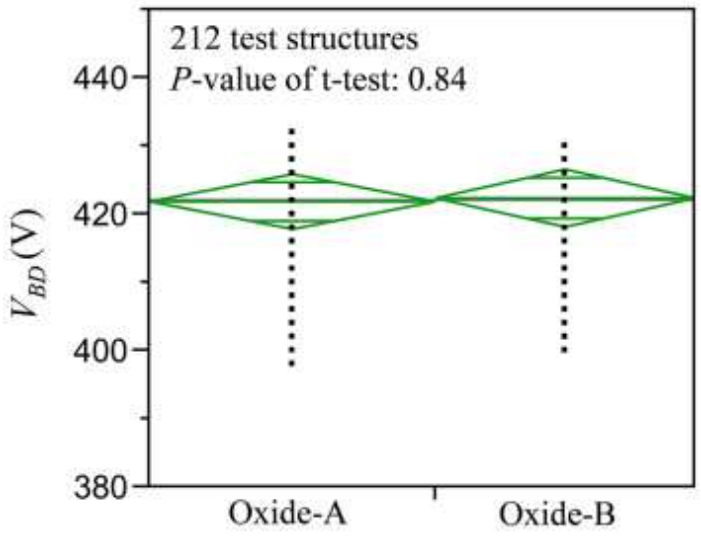

Figure 0.12. The corresponding V_{BD} distribution for 212 test structures of these two oxides in Fig. 2.11.

As mentioned before, the deviation from theoretical FN tunneling is primarily due to the deviation of the cathode electric field E_C from the average electric field in the oxide E_{ox} $=V/t_{ox}$. Assuming that the barrier's properties remain unchanged as the cathode electric field increases locally, the current deviation should demonstrate FN tunneling characteristics as a function of E_C instead of E_{ox}. Hence equation (2.7) can be rewritten for the enhanced FN tunneling as:

$$J = AE_C{}^2 \exp(-B/E_C) \tag{2.9}$$

where A and B are the FN constants as in (2.7). Based on (2.9), if two oxides of same t_{ox} have the same leakage current densities, the cathode electric field E_C must be the same for both of them.

Therefore, any equi-current line (*e.g.,* the dashed horizontal line in Fig. 2.11) is crossing the *I-V* curves at points where E_C is the same for both oxides. Consequently, one can think of the ordinate of the *I-V* plot (in Fig. 2.11) as being proportional to the E_C. This concept can be better visualized mathematically in the following. Given that, to the first order approximation, the E_C^2 term in (2.9) can be ignored when compared to its exponential term [2.2], the current density J can be replaced by $A\times\exp(-B/E_C)$. This substitution makes the J axis - in logarithmic scale - equal to $-B/E_C$. The *I-V*'s abscissa, on the other hand, can be viewed as a representative of time, as the ramp rate has been held constant during the RVT. Therefore, the slope of *I-V* can be calculated as $\partial(-B/E_C)/\partial t$. At this point, care should be taken that the parameter in which we are interested is ΔE_C as it is the one that is related to the trapped charge density based on (2.6). For this reason we would replace E_C with $E_{ox} + \Delta E_C$. With such substitution the slope of *I-V* can now be defined as:

$$\frac{\partial(-B/(E_{ox}+\Delta E_C))}{\partial t} = \frac{-B}{E_C^2}\frac{\partial(\Delta E_C + E_{ox})}{\partial t} = \frac{-B}{E_C^2}(R + \frac{\partial \Delta E_C}{\partial t}) \qquad (2.10)$$

where $\partial E_{ox}/\partial t$ is equal to the ramp rate R. The above relation shows that when one compares the slope of the two *I-V* curves along an equi-current line, since R and E_C are constants, any difference in the slope would be solely due to a difference in $\partial \Delta E_C/\partial t$. Assuming that the charge centroid is only a factor of thickness and remains essentially unchanged in the region of study [2.3], the $1 - \overline{x}/t_{ox}$ term in (2.6) can be considered as a constant that would yield:

$$\partial(\Delta E_C)/\partial t = \partial(Q^+)/\partial t \tag{2.11}$$

Following the IR model, since the charge density is the integration of the rate of positive charge generation, equation (2.11) can be written as a function of impact ionization coefficient α using (2.12):

$$Q^+ = \int_{t=t_1}^{t_1+dt} \frac{J}{q}\alpha.dt \tag{2.12}$$

$$\partial(\Delta E_C)/\partial t = \alpha(J/q) \tag{2.13}$$

As we are making a comparison along an equi-current line, equations (2.10) and (2.13) show that any difference in the slope of the two *I-V* plots can be interpreted as a difference in the impact ionization coefficients and consequently in the electrons' MFP in the corresponding oxides.

Up to this point, we were able to successfully correlate the *I-V* slope with a key parameter of the oxide bulk: the Mean Free Path of electrons which itself is an indication of how electrons get scattered as they drift in the conduction band of the oxide. Based on the thermochemical *E* model – which is the most widely accepted model for explaining the dielectric breakdown at low electric fields – the TDDB behavior is related to the bonding states of the dielectric molecules. The building blocks of SiO_2 are SiO_4 tetrahedrons. The *E* model suggests that the weak bonds in the oxide are the nucleus for TDDB. When the Si-Si bonds (often referred to as *E'* centers) break an irreversible dangling bond would be created. The accumulation of these dangling bonds will create a percolation path that can act as a perfect conduction path for electrons shorting the oxide.

Based on the thermochemical *E*-model, a critical number of dangling bonds (broken bonds) is necessary for the percolation path to form and for the oxide to break down [2.8]. This critical number is often referred to in terms of a critical fraction (*F*):

$$F = N_{(BD)}/N_0 \tag{2.14}$$

where N_0 is the initial density of weak bonds at $t = 0$ and N_{BD} is the density of weak bonds at the breakdown point $t = t_{BD}$. The *E*-model suggests that t_{BD} can be calculated through:

$$t_{BD} = \frac{F}{v_0}\exp\left(\frac{\Delta H_0}{k_B T} - \gamma E_{ox}\right) \tag{2.15}$$

where γ is the voltage acceleration factor, ΔH_0 the enthalpy of bond breakage in the absence of any external electric field (for Si-Si ~ 1.15 eV), v_0 a characteristic vibration frequency ~ 10^{13} s^{-1}, and k_B is the Boltzmann's constant. Based on equation (2.15) the ratio *F* can have a considerable effect on the time to breakdown t_{BD}: The oxide with a higher initial weak bond (*E'* centers) density will have a shorter t_{BD}. The density of weak bonds, on the other hand, can be characterized through carrier scattering frequency in the oxide ($1/\tau$) following [2.12]:

$$\frac{1}{\tau} \sim \frac{N_T q^4}{2\pi\sqrt{2m^*(E_C - E_T)}hk_{ox}\varepsilon_0} \tag{2.16}$$

where m^* is the mass of electron in oxide, E_C conduction band of oxide, E_T the energy level associated with the weak bond, N_T weak bonds density, and *h* is the plank constant. Based on (12), as the density of weak bonds (N_T) in SiO_2 increases, one should expect the carrier (electron) scattering to increase respectively. The above equation in conjunction with

(2.15) and (2.13) demonstrate how the *I-V* measurements can be correlated to the TDDB t_{BD} based on the information they yield about MFP.

2.4 Statistical Comparison of TDDB Results

In order to confirm the validity of the proposed relationship between the *I-V* slope under positive charge trapping and the TDDB lifetime, an experiment was performed with three oxides that are of different fabrication quality but similar otherwise. Figure 2.13 shows the *I-V* plots for these oxides obtained using the ramp rate of 2 V/s, and the temperature of 150 °C.

The oxides' *I-V* plots demonstrate considerable slope difference in the region where positive charge trapping is dominant. Based on our hypothesis, such a difference is an indication that these three oxides have different MFPs: $\lambda_A > \lambda_B > \lambda_C$. Therefore, oxide-A having the minimum electron scattering is predicted to have the longest t_{BD} as opposed to oxide-C being predicted to have the shortest t_{BD}. A TDDB-CVS test on these three oxides confirms our prediction about their lifetimes. The test is performed at an electric field of 6.5 MV/cm and the temperature of 225 °C using the same oxides. Figure 2.14 shows the Weibull distribution of the measured TDDB data analyzed with JMP software [2.13]. The Weibulls' scale parameters (α) and their 95% non-parametric confidence intervals are different for three oxides. The mean time to failure (MTTF) for oxide-A, -B and -C are calculated to be 7×10^4, 3.6×10^4 and 1.6×10^4 s, respectively. This is in agreement with the t_{BD} comparison predicted by our model solely from the *I-V* slopes.

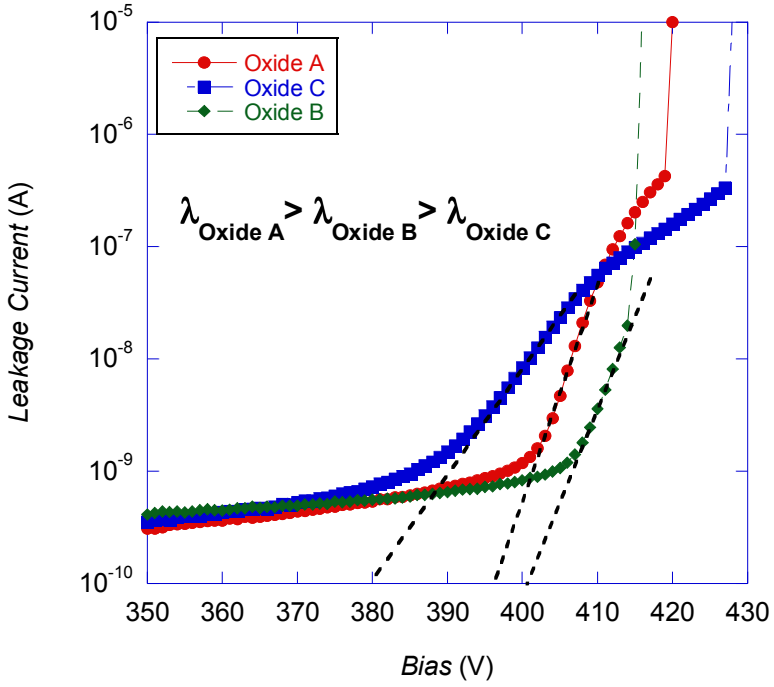

Figure 0.13. *I-V* plots for three different oxides. The *I-V* slopes are different in the positive charge generation regions. Based on the proposed theory, such difference can be related to the difference in MFP of electrons in the oxide. T=150 °C, t_{ox}=0.65 μm, and E_{ox}=6.5 MV/cm.

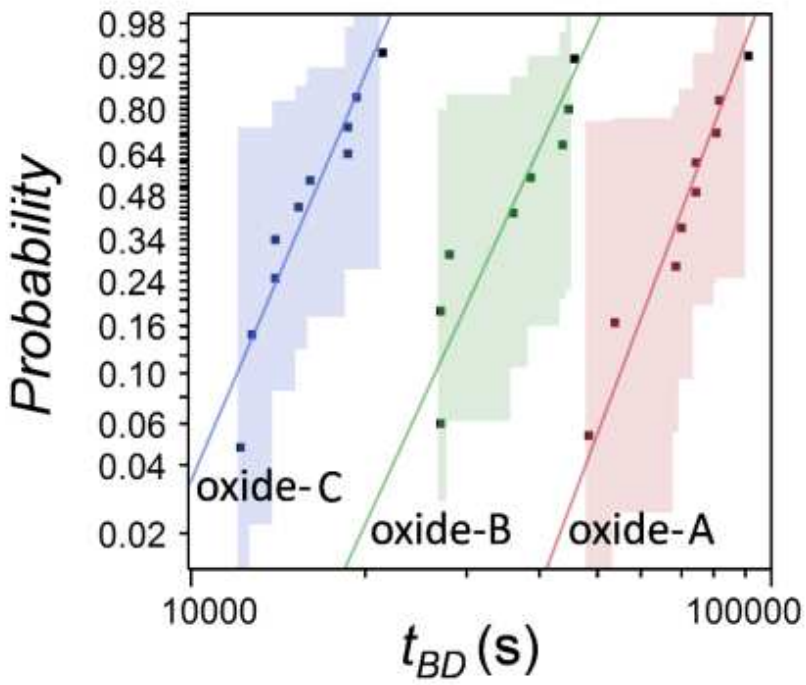

Figure 2.14. Weibull distribution and non-parametric confidence intervals of TDDB t_{BD} data for the same oxides. Oxide-A has the longest lifetime as opposed to oxide-C with shortest. T=225 °C, t_{ox}=0.65 µm, and E_{ox}=6.5 MV/cm

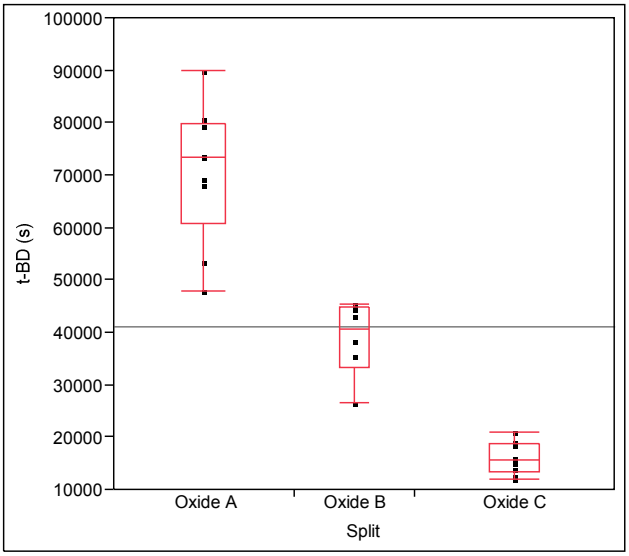

Figure 0.15. Lifetime distribution of three different oxides. The distribution matches well the prediction based on our model. T=225 C, t_{ox}=0.65um, and E_{ox}=6.5 MV/cm.

Although TDDB lifetime data are usually presented with a Weibull distribution and plots [2.14], one is generally safe to use ANOVA or *t*-test to compare them statistically. The reason behind such safety is the fact that the assumption of normality is not a critical prerequisite for validity of ANOVA and *t*-test [2.15]. The ANOVA comparison results of the measured data are showed in Table 2.1.

Table 2.1. Analysis of Variance for TDDB t_{BD} of three oxides

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**	**Prob > F**
Split	2	1.4125e+10	7.0625e+9	90.4263	<.0001*
Error	22	1718247682	78102167		
C. Total	24	1.5843e+10			

Level	**Number**	**Mean**	**Std Error**	**Lower 95%**	**Upper 95%**
Oxide-A	9	70540.9	2945.8	64432	76650
Oxide-B	6	38848.3	3607.9	31366	46331
Oxide-C	10	16012.8	2794.7	10217	21809

Using the thermochemical *E* model, the Weibull distribution function is defined as:

$$F = 1 - \exp(-(\frac{t_{BD}}{A_0 \exp\left(\frac{\Delta H_0}{kT} - \gamma E\right)})^{\beta}) = 1 - \exp(-\left(\frac{t_{BD}}{\alpha}\right)^{\beta}) \qquad (2.17)$$

where β is referred to as the Weibull shape parameter and α is the scale parameter defined through thermochemical model. β is a parameter that can be used to compare various data and to make sure they are following the same distribution and are comparable to each other. An unreasonably low β can be an indication of having too much extrinsic failure. It is interesting to mention that as the oxide thickness increases the β parameter is also expected to increase [2.16].

Table 2.2 shows the estimates and 95% confidence intervals for three oxides' lifetime distribution. Care should be taken that although the estimate of β is not exactly the same for three oxides but the 95% confidence intervals overlap reasonably (*i.e.* with a risk of only 5% the Weibull shape parameters are the same for these distribution). Obviously, as predicted by our model, there is a huge difference between α values of t_{BD} distribution for the three oxides. It is interesting, however, to check the origin of this difference in α values since our model predicts that the difference should be from a change in F ratio. Based on (2.17) a difference in α parameter has two distinct interpretation: Such difference may be due to change in either ΔH_0 or γ that signals a change in bonding nature of the oxide (*i.e.* the mixing effect), or, it may be due to a change in F ratio that signals the presence of different E' centers density in the two oxides.

Table 2.2. Weibull parameter estimates for three oxides and their 95% confidence intervals.

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
Oxide-A Weibull α	75604.66	3846.87	68428.70	83533.15
Oxide-A Weibull β	6.89	1.83	4.52	14.42
Oxide-B Weibull α	41447.10	2185.21	37378.03	45959.15
Oxide-B Weibull β	8.08	2.84	4.78	26.07
Oxide-C Weibull α	17249.78	952.39	15480.57	19221.19
Oxide-C Weibull β	6.07	1.49	4.09	11.71

Table 2.3 contains the detailed *E*-model parameters for oxide-B and -C. These parameters are calculated from the t_{BD} distributions of CVS-TDDB measurements that are

performed at two temperatures of 225 °C and 200 °C and at two electric fields of 6.5 MV/cm and 6.8 MV/cm (*i.e.,* at three distinct acceleration conditions for each oxide). As our technique predicted, the noticeable difference in t_{BD} of these two oxides is mainly due to the huge variation of F/υ_0 parameter for them. In other terms, the difference in the initial weak-bond density, that leads to having different t_{BD}, has been well characterized using the *I-V* slopes in the enhanced FN tunneling region.

Table 2.3. Thermochemical E model parameters for oxides –B and -C

Parameter	Oxide-C	Oxide-B
ΔH_0 (eV)	2.43	1.88
γ (cm/MV)	5.29	5.13
p_{eff} (eÅ)	22	22
f/υ_0 (s^{-1})	2.6×10^{-6}	0.49

2.4. Conclusion

Utilizing the positive charge generation/trapping kinetics, we proposed a novel fast screening *I-V* measurement technique that can distinguish between the intrinsic quality of ultra-thick oxides and their corresponding TDDB t_{BD}. It was demonstrated that there is a good agreement between the quality assessment based on our proposed technique and the

actual t_{BD} values that are calculated by running CVS-TDDB for three oxides of different qualities.

2.5 References

[2.1] H. Bentarzi, *Transport in metal-oxide-semiconductor structures: Mobile ions effect on the oxide properties*, Springer, 2011.

[2.2] E. Rosenbaum, J. C. King, and C. Hu, "Accelerated testing of SiO_2 reliability," *IEEE Transactions on Reliability*, vol. 43, pp. 70-80, 1996.

[2.3] M. Shatzkes, and M. Av-Ron, "Impact ionization and positive charge in thin SiO_2 films," *Journal of Applied Physics*, vol. 47, pp. 3192- 3202, 1976.

[2.4] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "High-field and current-induced positive charge in thermal SiO_2 layers," J. Appl. Phys., vol. 57, pp. 2830-2839, 1985.

[2.5] A. Aal, "Fast wafer level reliability assessment of ultra-thick oxides under impact ionization," *IEEE Trans. Dev. Mat. Rel.*, vol. 7, pp. 117-120, 2007.

[2.6] T. H. Ning, "Thermal reemission of trapped electrons in SiO_2." *J. Appl. Phys.*, vol. 49, pp. 5997-6003, 1978.

[2.7] O. Gunnarsson, and J. E. Han, "The mean free path for electron conduction in metallic fullerenes," Nature, vol. 405, pp. 1027-1030, 2000.

[2.8] J. W. McPherson, 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, pp. 1513-1523, 1998.

[2.9] J. W. McPherson, "Time dependent dielectric breakdown physics – Models revisited," Microelectronics Reliability, vol. 52, pp. 1753-1760, 2012.

[2.10] N. Klein, and P. Solomon, "Current runaway in isulators affected by impact ionization and recombination," *J. Appl. Phys.,* vol. 47, pp. 4364-4372, 1976.

[2.11] U. Schwalke, M. Polzl, T. Sekinger, and M. Kerber, "Ultra-thick gate oxides: charge generation and its impact on reliability," *Microelectronics Reliability*, vol. 41, pp. 1007-1010, 2001.

[2.12] K. Hess*, Advanced theory of semiconductor devices*. Wiley-IEEE Press, 2000.

[2.13] JMP, Version10. SAS Institute Inc., Cary, NC, 1989-2012.

[2.14] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., Wiley-Interscience, New York, 2006.

[2.15] D. C. Montgomery, *Design and analysis of experiments*, 8th ed., John Wiley & Sons, Inc., 2013.

[2.16] D. M. Fleetwood, S. T. Pantelides, and R. D. Schrimpf, *Defects in microelectronic materials and devices*, CRC press, 2008.

CHAPTER 3

A FAST TECHNIQUE TO SCREEN CARRIER GENERATION LIFETIME USING DLTS ON MOS CAPACITORS

3.1 Introduction

Generation lifetime (τ_g) is not only a key device parameter that has a dominating role in how a variety of semiconductor devices (*e.g.*, CCD, DRAM, p-n junctions, *etc.*) operate [3.1]; but, it is also a materials' parameter that is extensively used as a process cleanliness monitoring tool [3.2]. It has been shown that measuring the generation lifetime is an eminently suitable technique for characterization of thin epitaxial layers [3.3]. The most widely used technique for measuring the generation lifetime is the pulsed MOS-capacitor (MOS-C) in which the τ_g is extracted from the capacitance transition (*C-t*) of a pulsed MOS-C as it moves from the non-equilibrium deep-depletion state into the equilibrium inversion state [3.1]. This technique has gained popularity since the deep-depleted MOS-C test structures are, in fact, building blocks for a variety of semiconductor devices [3.4] and so the extracted lifetime parameters are of direct design importance. Although various modifications over time have increased its accuracy, the pulsed MOS-C method is still a time consuming technique especially when applied to ultraclean wafers [3.5]; where, the transition times can easily reach some thousands of seconds. This necessitates the need for other innovative techniques that can be used to screen generation lifetime on MOS-C test structures in much shorter time frames.

When a p-n junction or MOS capacitor is biased in the dark, the electron-hole pairs (ehps) are generated thermally with a rate defined as $G=n_i/\tau_g$. Therefore, and to first order approximation, the recovery time of a pulsed MOS can be estimated as [3.6]:

$$t_f = 10\frac{N_A}{n_i}\tau_g \tag{3.1}$$

Substituting for N_A~10^{15} cm^{-3}, n_i=10^{10} cm^{-3} (for Silicon at room temperature) and τ_g~10^{-5} s yields t_f = 250 s. The problem become worse when the goal is to map the whole wafer area or to measure tens of devices to produce a statistically sound estimate of the lifetime on a test-vehicle. The time it takes to do these measurement then can rapidly increase beyond the satisfaction of industry. Figure 3.1 shows a typical measured *C-t* plot of a clean wafer where the lifetime is estimated to be in milliseconds range. Notice the ~ 6500 s that it takes for this *C-t* to saturate at room temperature.

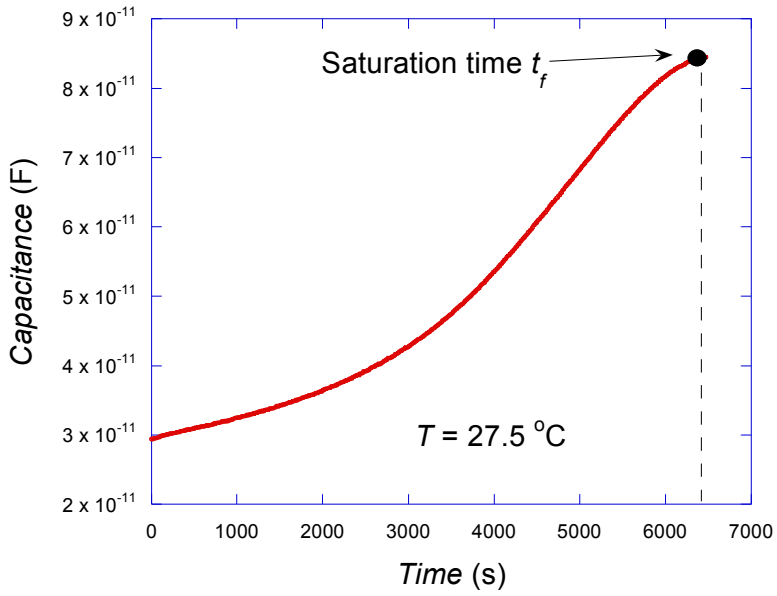

Figure 3.1. A typical *C-t* plot measured through pulsed MOS-C technique on a clean wafer where the lifetime is estimated to be in the milliseconds range. Notice the long time it takes for the capacitance to saturate and for the measurement to complete.

Deep-level transient spectroscopy (DLTS) is an alluring technique for the study of carrier emission properties from deep level traps [3.7]. Its fast application, automation, and simple fabrication of its test structures have turned DLTS into an unparalleled electrical measurement technique among semiconductor defect characterization methods. The DLTS technique, however, is specifically engineered for the study of exponential transients [3.7], [3.8] (*e.g.*, the capacitance transient of a reversed biased Schottky diode or a p-n junction). Unlike the carrier emission, the capacitance transient due to the minority carrier generation (i.e., the *C–t* of a pulsed MOS-C) does not have an explicit exponential form. Although it has been showed that *C–t* transients of pulsed MOS capacitors due to minority carrier

generation can give rise to a high-temperature peak in thier DLTS spectrum [3.9], [3.10], the corresponding peak has not yet been characterized for its potential application in screening the τ_g. Instead, the focus had been on the characterization of other peaks, which corresponds to the emission of majority carriers from traps and consequently studying individual deep-level impurities [3.11]. We show that the rate window concept used in DLTS can be effectively employed to determine the generation lifetime variations in MOS-C test structures. A mathematical description will be presented to correlate the DLTS peak corresponding to the minority carrier generation to the generation lifetime. It will be demonstrated that the high sensitivity of lock-in DLTS enables us to study very short MOS-C transients that is otherwise impossible. This would enable us to greatly simplify the classic equations (that are used to describe the pulsed MOS capacitance transient) to analyze the measurements. As a result of this one can screen the lifetime by directly monitoring the saturation times in an advantageous way where the results are least affected by other nuisance parameters, as will be shown by experimental data.

3.2 Theory

3.2.1 Pulsed MOS Capacitor (MOS-C) Technique

Pulsing the MOS-C gate bias from accumulation (or depletion) to the inversion will drive the device into a non-equilibrium state called the deep-depletion. If the gate voltage is held constant, the device will then start to gradually get back into the inversion as a result of the thermal generation of minority carriers and as the inversion layer forms beneath the gate. The capacitance of MOS-C, when measured with respect to time, will hence show a

transition as the device moves from deep-deletion (C_i) into inversion (C_f), as depicted in Fig. 3.1.

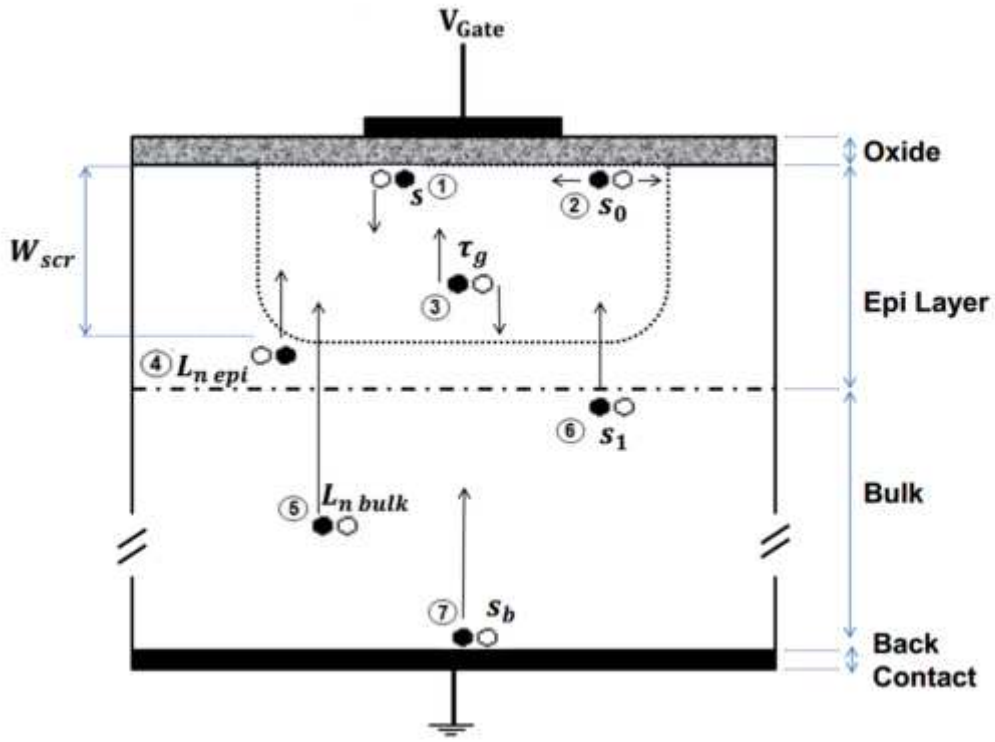

Figure 3.2. Different mechanisms and their contribution to the thermal generation of minority carrier and the formation of an inversion channel beneath the gate under deep-depletion state.

The thermal generation of minority carriers can be divided to seven different components. These seven thermal generation mechanisms work together to form an inversion layer beneath the gate and to bring the device back to the equilibrium state (*i.e.,* inversion), as shown in Fig. 3.2 [3.12].

Surface generation is characterized by surface generation velocity which is made up of two different components (*i.e.*, mechanisms 1 and 2). Mechanism 1 is the surface generation velocity in the lateral portion of the space charged region (scr) beyond the inversion layer which remains always depleted of minority carriers and is characterized by parameter s_0. Mechanism 2, on the other hand, is the surface generation velocity in the lateral portion of the scr under the gate that is initially depleted but becomes inverted as the inversion layer forms and is characterized by parameter s. The latter surface generation velocity s then varies with time from its maximum value s_0 at $t = 0$ to zero at the end of the capacitance transient. There is also carrier generation in the quasi neutral region (qnr) of the epitaxial layer as well as the substrate. These generations are characterized by their respective diffusion length $L_{n\text{-}eff}$ (mechanisms number 4 and 5) due to their contribution in supplying minority carriers to the inversion layer. The epi/bulk interface and the ohmic back contact can also be considered for their surface generation velocities and are respectively characterized by s_1 and s_b (mechanisms numbers 6 and 7). And finally, τ_g (mechanism number 3) that corresponds to the generation lifetime in the depletion region width (scr) where electron hole pairs are generated via deep level impurity energy levels. We are interested in τ_g since it is directly related to deep level impurity density in the epitaxial layer as shown in the following equation [3.3]:

$$\tau_g \approx \tau_r \times \exp\left(\frac{|E_T - E_i|}{kT}\right) ; \ \tau_r = \frac{1}{\sigma_n \vartheta_{th} N_T} \tag{3.2}$$

where E_T is the deep level impurity energy level, E_i the intrinsic energy level, σ_n the minority carrier capture cross section, v_{th} the thermal velocity, and N_T is the deep-level impurity concentration.

As the inversion layer forms and the device moves back to the equilibrium, its capacitance will increase as a result of the depletion region width becoming narrower. Taking into account the seven mentioned mechanisms, the following equations have been proposed in describing the capacitance transient with respect to time (*C*-*t*) [3.1]:

$$(1-\gamma)ln\frac{(C_F/C_i-1+\gamma)}{(C_F/C-1+\gamma)}+\frac{C_F}{C_i}-\frac{C_F}{C}=\frac{C_F n_i t}{C_{ox}N_A\tau_{g-eff}}\,, \tag{3.3}$$

$$\gamma = C_F s'\tau_{g\,eff}/K_s\varepsilon_0\,, \tag{3.4}$$

$$s' = s + n_i D_n/N_A L_{n\,eff}, \tag{3.5}$$

$$\tau_{g\,eff} = \frac{\tau_g}{1+2\tau_g s_0/r} \tag{3.6}$$

where C_i is the initial capacitance in deep depletion, C_F the final equilibrium capacitance in inversion, C_{ox} the oxide capacitance, *r* the gate radius, *t* the oxide thickness, N_A the epi layer doping concentration, ε_0 the vacuum permittivity, and K_s is the silicon's dielectric constant. Figure 3.3 demonstrates how well this mathematical model fits the measured data.

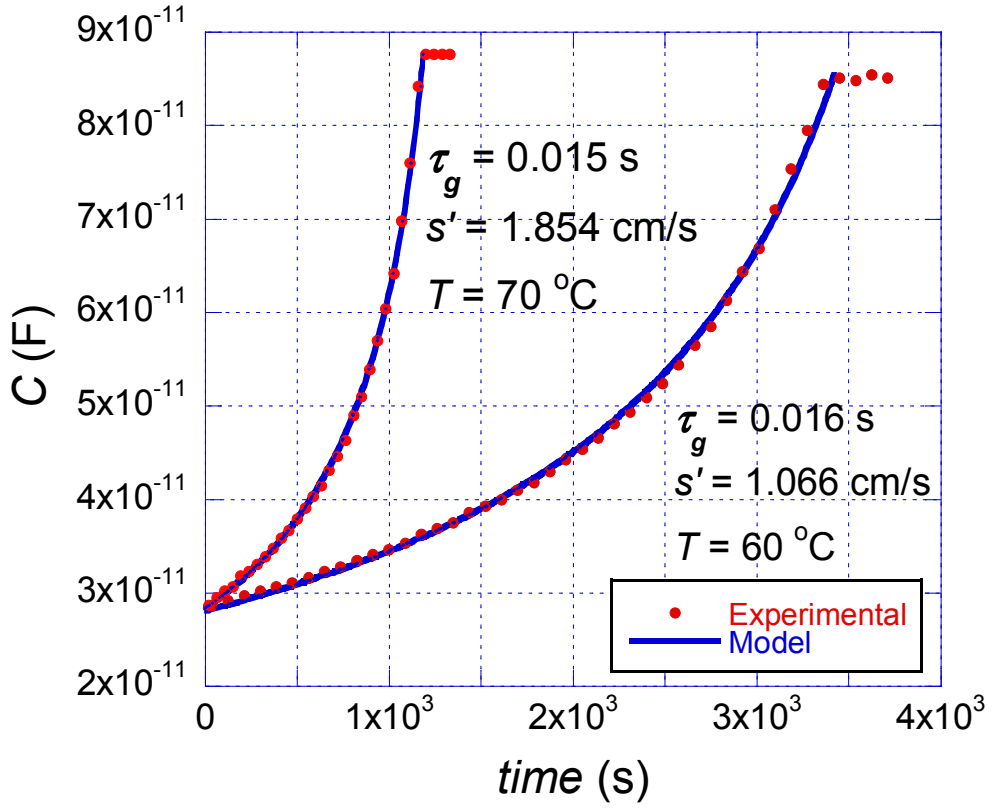

Figure 3.3. Equation (3.3) has been fitted into the measured *C-t* plots for T =60 ºC and 70 ºC where the device is pulsed from -5 V to +5 V. C_{ox} =1.35×10^{-9} F, N_A =9.8×10^{14} cm^{-3}, t_{ox} =20 nm.

One of the main challenges in measuring the bulk generation lifetime (τ_g) has always been to properly and efficiently separate it from the surface (s) and bulk generation (L_{n-eff}) components. I recently proposed a new method as a mean for such accurate separation [3.12] where the lifetime is extracted from multiple *C–t* measurements performed at different temperatures. Another way to address this issue, but in considerably shorter time frames, is to exploit the C_f/C_i ratio. This ratio for classic pulsed MOS-C is approximately close to 3 [3.13]. However, if the accumulation pulse is applied in a way such that the C_i converges to C_f then (3.3) can be simplified to (3.7). This simplification can be done since

the logarithmic $\ln(1 + (C_f/C_i - 1)/\Upsilon)$ term in (3.3) becomes negligible and, to the first approximation, can be ignored compared to the non-logarithmic $C_f/C_i - 1$ term. From mathematical point of view, this results into the elimination of the Υ parameter from any future analysis. Equation (3.7) is of great importance as it implies that the saturation time (t_f) is now a direct measure of τ_g and can be used to screen the lifetime. Such correlation can only be obtained when the log term is neglected from (3.3), otherwise considerable errors may be introduced due to the surface term, as will be shown later on. In contrast, if the t_f parameter is measured from *C–t* plots obtained from classic pulsed MOS-C technique (like the ones demonstrated in Fig. 3.3), then it will not have any direct correlation with the lifetime – based on (3.3) – and should not be used as a measure of τ_g. In other words, the simplification of (3.3) to (3.7) transforms t_f from a mere meaningless parameter to an entity that can be used to screen lifetime. Such simplification can only come true in the light of our new technique.

$$t_f = \frac{N_A C_{ox}}{n_i C_f} \tau_{geff} \left(\frac{C_f}{C_i} - 1\right) \tag{3.7}$$

The physical interpretation of what happens is that when the accumulation pulse is applied, the majority carriers will follow it instantaneously. Their response time is determined by their velocity ($\sim 10^7$cm/s) and the width of the depletion region (a few micrometers) and is typically in the order of 10−10s [3.3]. The minority carriers in the inversion layer, on the other hand, need some time before they can recombine. This time frame can be as long as few tens of microseconds in clean silicon wafers. Consequently, if

the width of the accumulation pulse is chosen short enough (a few microseconds), at the beginning of the subsequent depletion bias there would be still some minority carriers in the inversion layers. This will lead to a an initial device capacitance C_i that is fairly close to C_f. Figure 3.4 shows schematically the pulse and the corresponding C–t transients for classic and DLTS techniques.

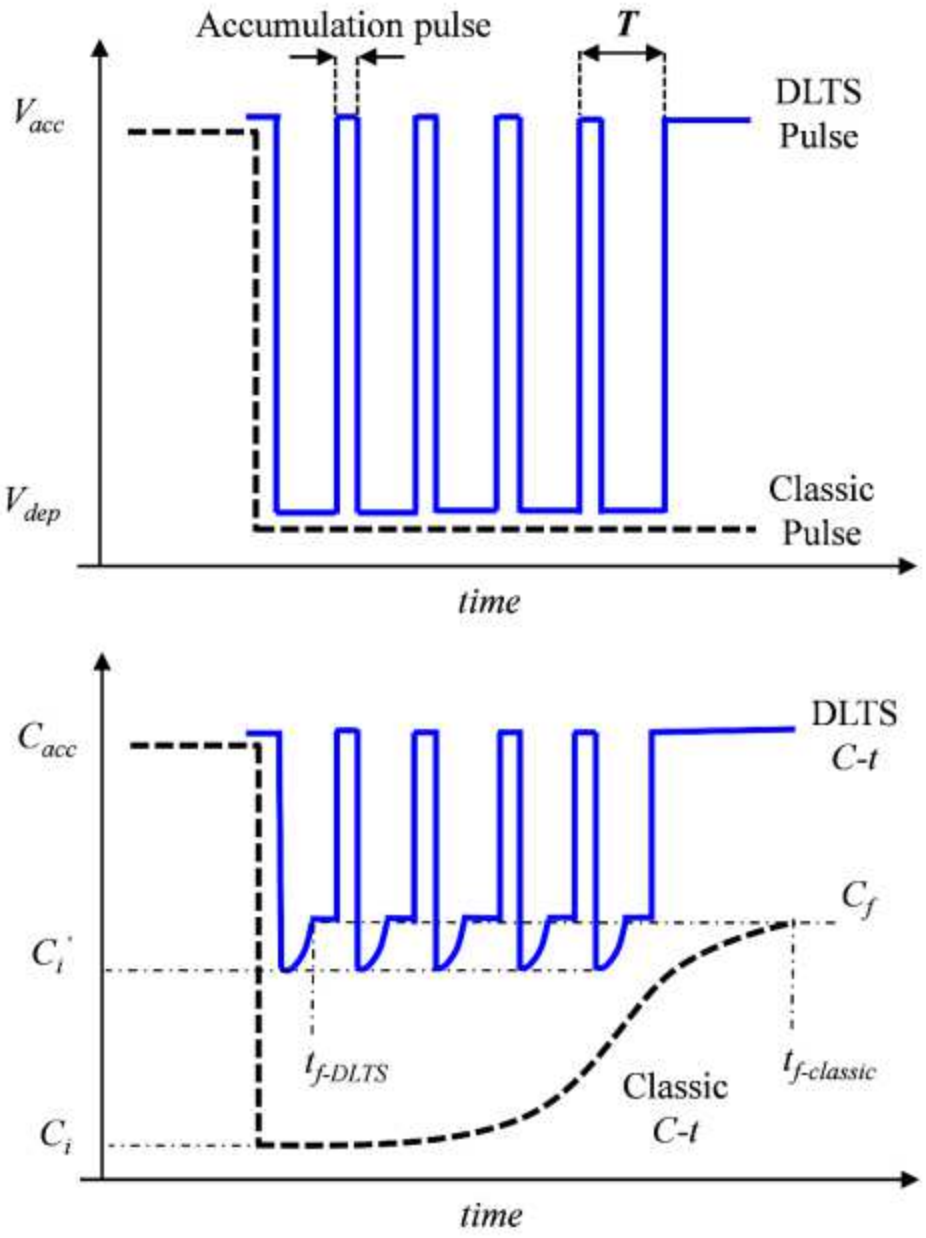

Figure 3.4. Schematic comparison between the applied pulses and consequent *C–t* transitions for the classic and the new DLTS techniques. In both cases, the device is initially in equilibrium inversion before the accumulation pulse is applied. As a result of using short accumulation pulses, the C_i values (identified in the figure by C_i') will

converge to the C_f. V_{acc} and V_{dep} represent the accumulation and depletion biases, respectively

However, although such approach may seem feasible in theory, in practice as the C_i gets close to C_f the C–t transient becomes undetectably short and noisy. Even with the resolution of modern capacitance meters recording such transitions would be a difficult task to do. Here, the signal processing power of a DLTS system can be employed to study these kinds of transients and to extract the required parameters in an accurate way

3.2.2 Lock-in DLTS

The DLTS can be thought as an automated mean to study *C–t* transients employing the gating electronics (*e.g.*, the lock-in amplifier and box-car integrators) [3.14]. Although the very first *C–t* wave-forms for which the DLTS technique was engineered belonged to the exponential emission transients, it is possible to apply the method to any other capacitance transients. The challenge, however, would be how to analyze the resultant spectrum.

The DLTS is basically a signal-processing technique [3.6]. The gating electronics receive the *C–t* signal as an input (S_{in}), then multiply it by a weighting function *w(t)*, and finally average the outcome using a linear filter to form the output signal (S_{out}) as:

$$S_{out} = \frac{1}{T}\int_0^T S_{in}(t).w(t).dt \tag{3.8}$$

The form of the output signal in this process is greatly influenced by the weighting function used and the filtering method applied. Lock-in amplifiers are the most commercially available gating electronics that are used in DLTS systems. They provide acceptable resolutions and good signal-to-noise ratio [3.15]. The lock-in amplifier, in

broadband mode, uses a continuous square weighting function as defined in (3.9). The period of this weighting function, that is the same as the period of the applied pulse, is defined by the LIA's frequency. A DLTS peak would be observed when this frequency bears a specific relationship with the input *C*-*t* transient. Although such relationship is extensively studied for exponential transients, as we will mention, it does not have a closed-form mathematical expression for the case of a MOS-C.

$$w(t) = \begin{cases} +1 & ; \ 0 < t < T/2 \\ -1 & ; \ T/2 < t < T \end{cases} \tag{3.9}$$

The DLTS measurements are performed in various modes depending on how the scans are performed and how the perturbation is applied. We will expand our technique based on the frequency scan mode [3.16] where the sample's temperature is kept constant, while the lock-in frequency is swept over a finite range.

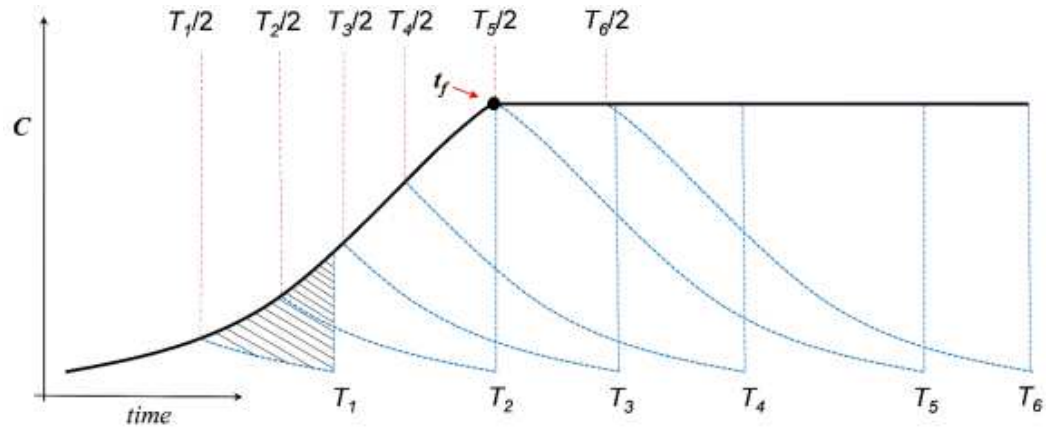

Figure 3.5. Schematic presentation of the input signal of the lock-in integrator and the integration process that is done on it after being multiplied by the weighting function in the multiplier for the case of constant temperature and varying frequency

3.2.3 Application of DLTS to Pulsed MOS-C C-t Transient

Figure 3.5 demonstrates the typical *C-t* transient response of a pulsed MOS-C. Multiplying this transient with the weighting function as defined by (3.9) and integrating the result over an increasing period is also illustrated schematically in this figure. The absolute value of integration result over the first period (T_1) is identified as a hatched area in Fig. 3.5. As is evident, the integrations increase in value as the period increases. Care should be taken, however, that the lock-in output is this integration result divided by the individual period at which the integration has been taken. A numerically calculated DLTS output of a *C–t* transient, using MATLAB, is showed in Fig. 3.6 for various temperatures.

As it is demonstrated in Fig. 3.6, the application of the lock-in DLTS concept on the pulsed MOS-C *C–t*, described by (3.3), gives rise to a temperature sensitive DLTS peak.

The way such a peak is generated can be better understood graphically. Let us consider the *C–t* transient in Fig. 3.5 once again. The integration in (3.8) – whose result is identified in Fig. 3.5 by areas enclosed by dashed lines – increases in magnitude as T increases toward T_5. On the other hand, when T is further increased beyond T_5, although the integration value remains constant at a maximum, the total output (S_{out}) decreases since the integration result is being divided by an ever-increasing T value. Such behavior produces a spectrum with a peak like the one shown in Fig. 3.6. Given that S_{out} is the same for T_2 and T_5 (if the corresponding integration areas are considered as triangles), the location of this peak must be between these two periods – *i.e.*, $T_2 < T_{max} < T_5$. The peak's location can be numerically calculated for the ideal case to be:

$$t_f = 0.68 \times T_{max} = 0.68/\vartheta_{max} \tag{3.10}$$

where T_{max} and ϑ_{max} are the lock-in period and frequency at which the peak is observed. Equation (3.10) provides us with the saturation time (t_f) for a *C–t* transition whose C_f and C_i are fairly close to each other. Such convergence of C_i toward C_f is achieved by employing a sequence of accumulation depletion pulses with high frequencies. Employing (3.7), the extracted t_f can be used to directly screen the generation lifetime.

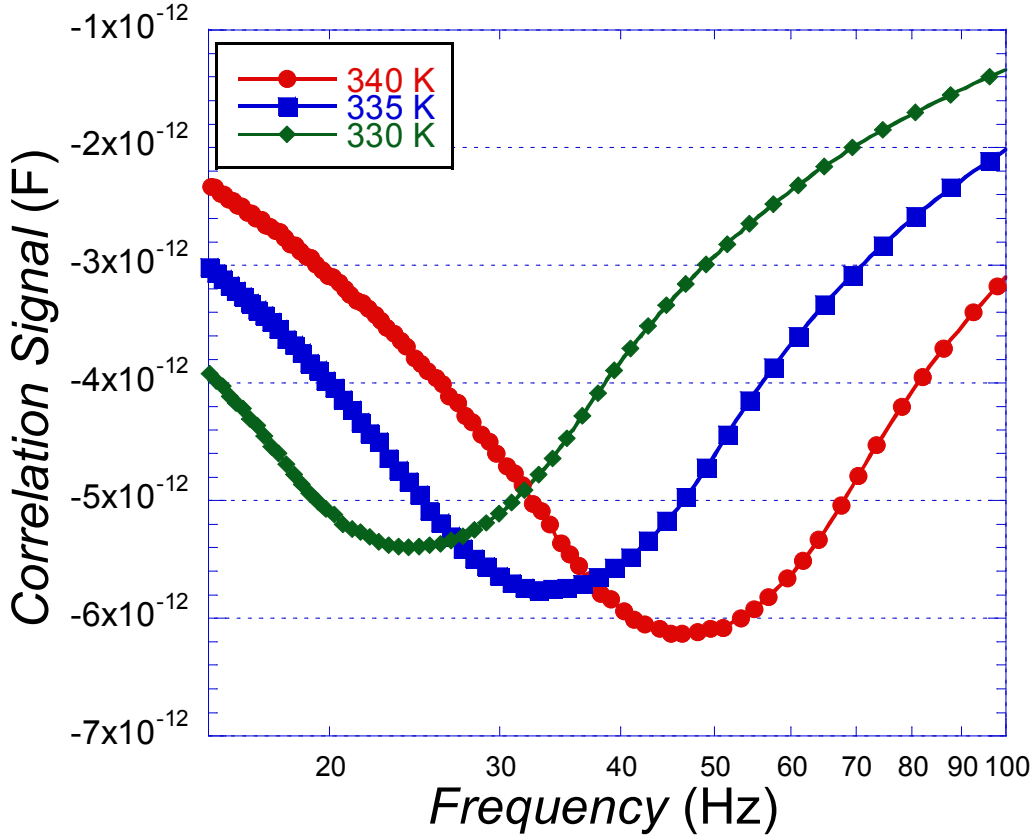

Figure 3.6. Calculated frequency scan DLS correlation signal (S_{out}) using equations (1) and (6) for thee different temperatures of 330, 335, and 340 K respectively. The signal is simulated for τ_g=2 ms, s=100 cm/s, Ci=80 pF, C_f=90 pF, C_{ox} =1.2×10^{-9} F, and N_A=9.8×10^{14} cm^{-3}.

Although for all practical purposes, the t_f parameter is adequate to screen the lifetime, the τ_g value itself can also be calculated using the DLTS spectrum. We already know all parameters in (3.6) except for C_f/Ci whose value needs to be determined before τ_g can be extracted. This can be done exploiting the magnitude of the DLTS peak. Based on (3.8),

and to the first order approximating, one can express the DLTS peak's magnitude |DLS| as:

$$|DLS| = (3/8) \times \Delta C \tag{3.11}$$

where ΔC is the ($C_f - C_i$). Since C f is the inversion capacitance and independent of how the pulse is applied, it can be easily obtained from a routine *C–V* measurement. Now with C_f and ΔC in hand C_i and therefore (C_f / C_i) can be calculated.

Care should be taken, though, that the numerical value of τ_g extracted this way, in general, is shorter than the one extracted from the classic pulsed MOS-C version. This is partly due to the fact the *C–t* transition that we study does not account for the field-enhanced generation that can cause the generation lifetime to decrease dramatically [3.1], [3.17], [3.18]. In addition, our analysis is for the ideal case where the instrument effects such as gate-off time is ignored. Nevertheless, the obtained t_f can be used for detailed comparison of silicon wafers, as we will demonstrate in the results section.

3.3 Experimental

MOS-C devices were fabricated on 200 mm wafers of 4 μm lightly p-doped epitaxial layer of 30 ohm-cm resistivity over 725 μm heavily p-doped substrate of 10^{-3} Ohm-cm resistivity. Two different sets of wafers with reported Tungsten ($E_T = E_V$ +0.41 eV) contamination levels of less than 10^{10} and 3×10^{10} cm^{-3} were used for comparison in this study. The Aluminum gates were evaporated and deposited through a shadow mask with circular holes of 1 mm in diameter on top of a 20 nm of thermal oxide layer, as showed in

Fig 3.2. The deposition was done in an E-Beam Lesker-75 PVD evaporator. The back contact was deposited directly on the p-doped HF cleaned substrate with the same routine. The thickness of gates and back contact are 200 nm, deposited at a rate of 0.2 nm/s. Forming gas annealing, to reduce interface traps, was done at 450 °C in an ambient of 95% nitrogen and 5% hydrogen for 30 minutes.

Figure 3.7. MOS capacitors fabricated on the surface of the wafer with circular Aluminum gates deposited with PVD and through a shadow mask.

The doping concentration profile of samples was determined from differential capacitance method [3.6] using the deep-depletion C-V and $1/C^2$-V curves, and was deemed to be constant throughout the epitaxial layer and equal to 9.8×10^{14} cm^{-3}, as depicted in Fig. 3.8. This measurement was in good agreement with the resistivity data provided by the vendor.

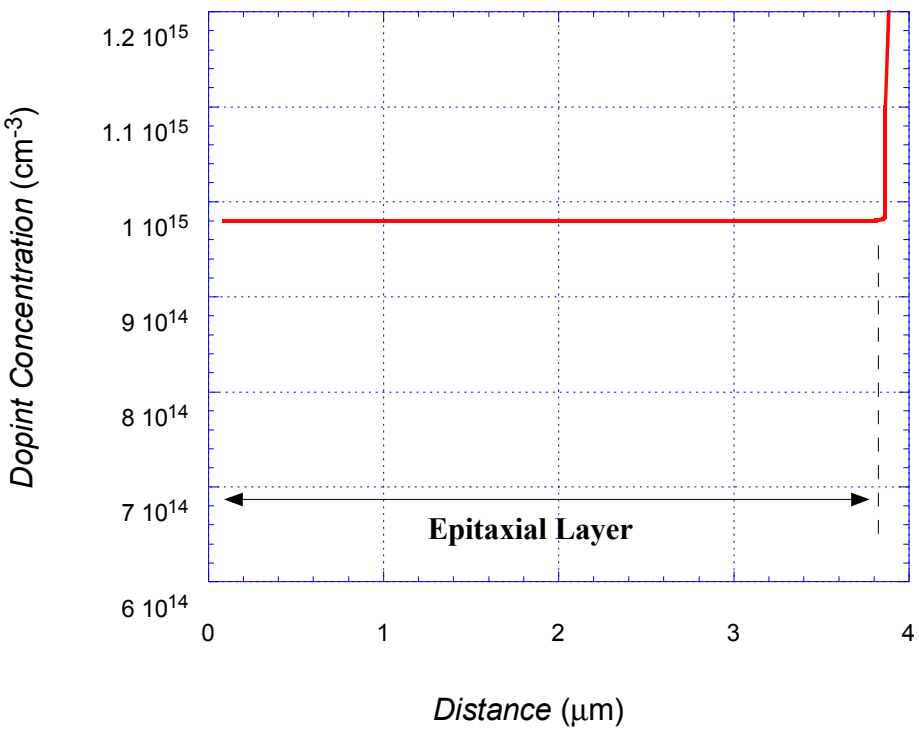

Figure 3.8. Doping concentration measured through differential capacitance method.

The capacitance measurements were made in part using an Agilent 4294-A precision impedance analyzer and in part a 4284 LCR meter. The DLTS measurements were done using a Semilab DLS-83D unit. The 4284 LCR meter was controlled on a LabVIEW platform. A Keithley model 82 *C-V* system was used to do quasi-static *C-V* measurements.

Quasi-static interface measurements [3.6] were performed on devices fabricated on more- and less- contaminated wafers to check against variation of interface states density. As it is shown in Fig. 3.9 both wafers share the same density of interface states.

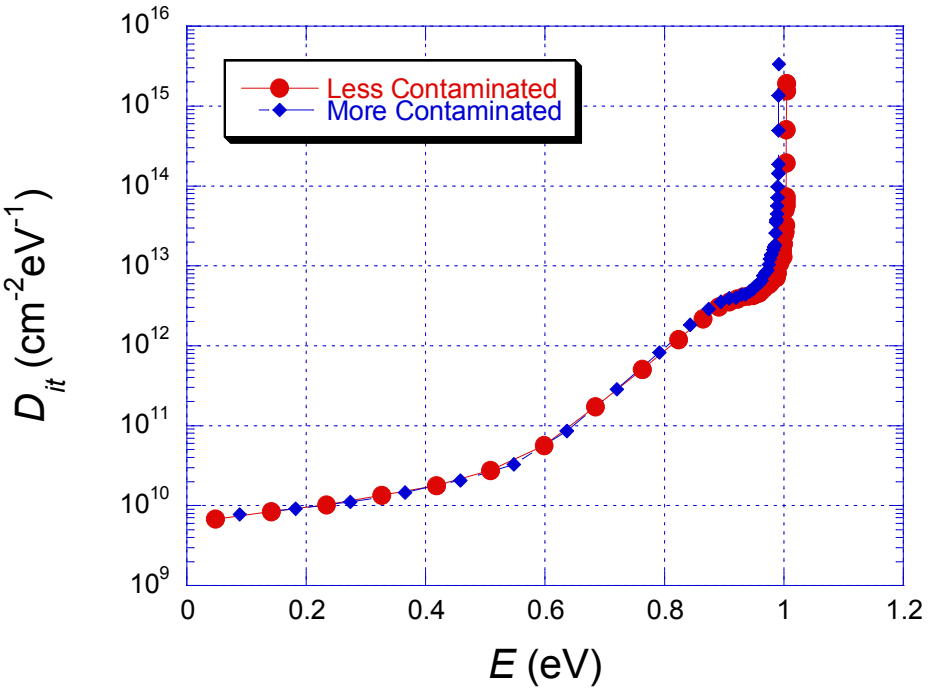

Figure 3.9. Interface state density measured with quasi-static method for two more- and less- contaminated wafers. The data show the interface states to be of similar distribution and density for both samples.

3.4 Results and Discussion

3.4.1. Classic Measurements

Classic pulsed MOS-C measurements were performed on the two more- and less-contaminated device families in order to see if the expected difference in their measured lifetime could be observed following (3.2). Devices were pulsed from -5 (accumulation) to +5 (inversion) and the capacitance transition was measured with an ac probing frequency of 1 MHz and a dc probing level of 25 mV. Subsequently, the lifetime was extracted from

the *C-t* data using the simplified relation in (3.12) that is a derivative of (3.3) assuming a γ of 3 [3.13]. Such an assumption is sound for clean silicon wafers. Lifetimes obtained with the classic pulsed MOS-C measurement are shown in Figs 3.10 and 3.11. The lifetime distribution is not revealing the variation one would expect for the less- and more-contaminated wafers.

$$\tau_{g\,eff} = \frac{n_i}{N_A} \times \frac{1}{1 + C_{ox}/C_i} t_f \tag{3.12}$$

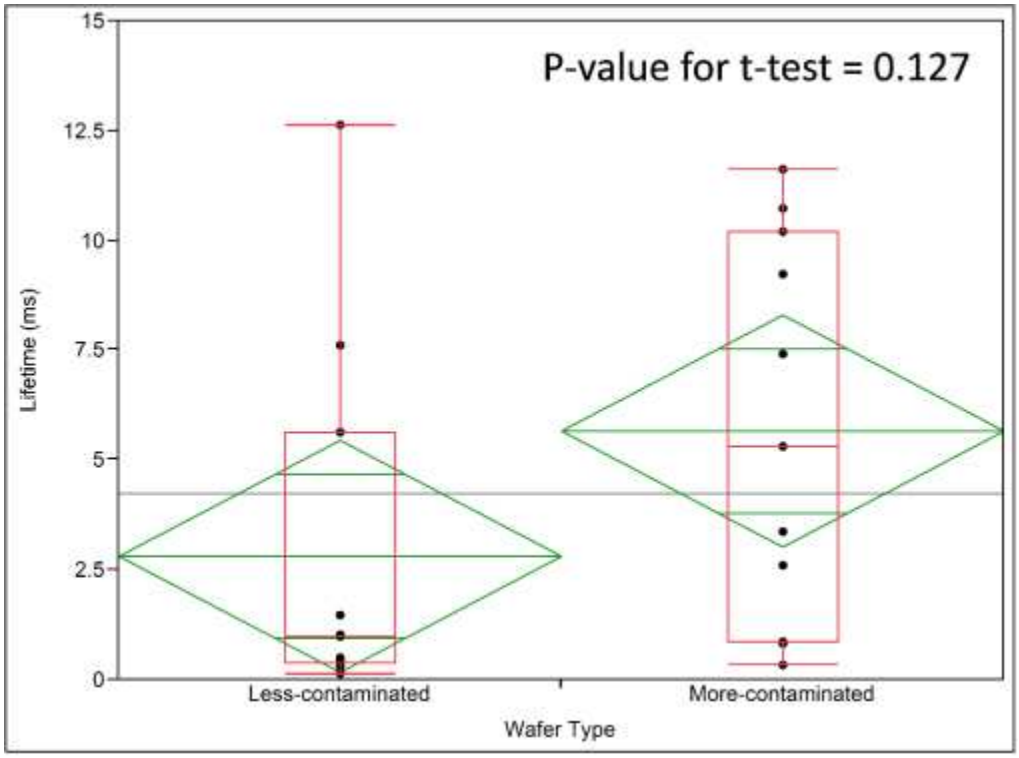

Figure 3.10. Lifetime extracted using (3.12) from measured *C-t* plot for more- and less-contaminated samples. The lifetime parameter does not show the expected variation.

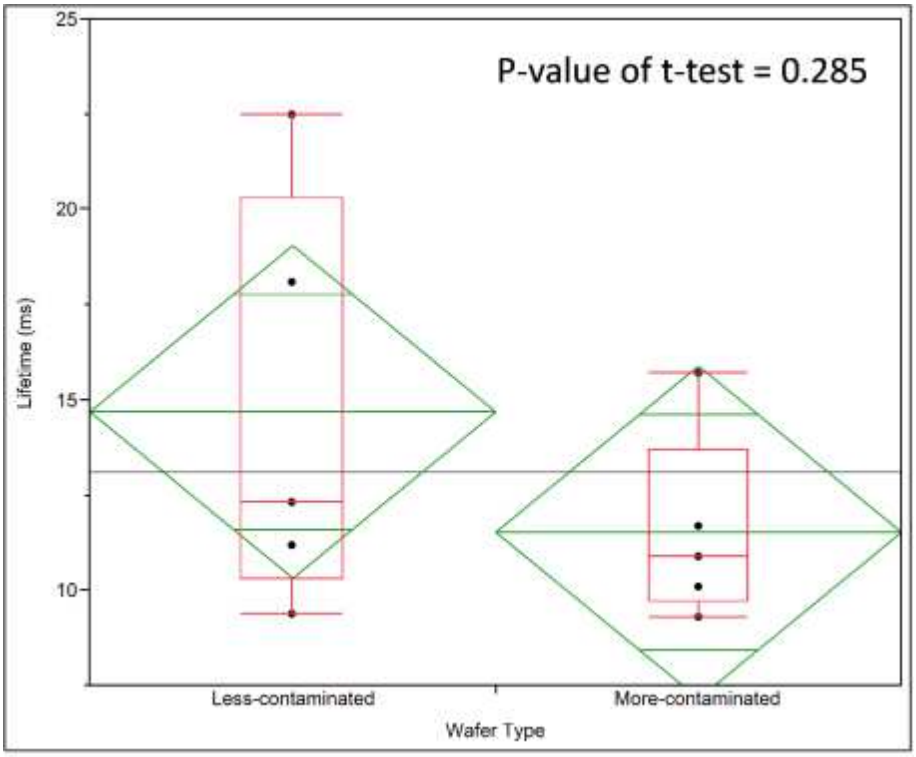

Figure 3.11. Lifetime parameter extracted by fitting equation (3.3) to the *C-t* plot using the method of least-squares. The parameter still does not show the difference as expected.

Even when equation (3.3) is directly fitted to the experimental data through the method of least-squares, the extracted lifetime, as depicted in Fig. 3.11, does not show the expected variations. It was not until the more advanced "*Modified pulsed MOS-C technique*" was used that the lifetime difference could be actually observed [3.12]. In this method the lifetime is extracted from multiple *C-t* measurements performed at different temperatures [3.19]. Care should be taken, that although accurate, the modified pulsed MOS-C not only fails to reduce the measurement time but it increases the time it takes to extract the lifetime by at least a factor of 2. In order to be able to extract the lifetime this way, the *C-t*

measurements have to be done for at least two different temperatures. The results obtained using this technique is demonstrated in Fig. 3.12.

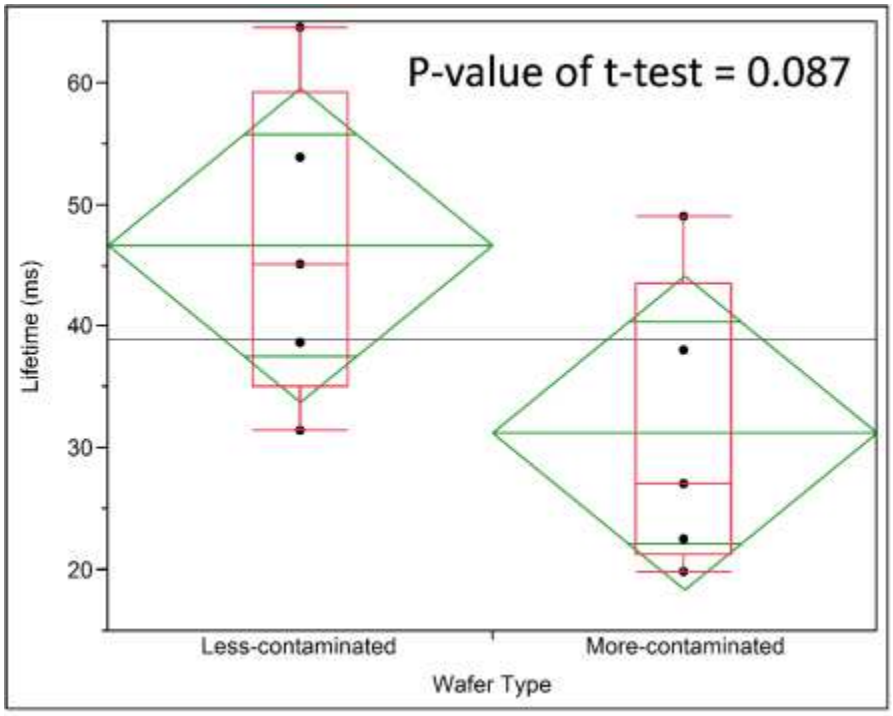

Figure 3.12. Lifetime parameter using the more advanced "modified pulsed MOS-C" technique. The parameter well reveals the variation between more- and less-contaminated wafers as expected, but in longer time frames.

3.4.2. DLTS Measurements

The frequency-scan DLTS measurements were performed on MOS-C devices employing filling/accumulation pulses of 20 μs in width. The lock-in frequency was swept from 1 to 2500 Hz, with the temperature being kept constant at finite values. Figure 3.13 shows a typical measured DLTS spectrum for the samples.

The first step before analyzing the results, when DLTS measurements are done on MOS-C devices, should be to determine the source of observed peaks in the DLTS spectrum. Care has to be taken that a MOS-C DTLS peak may correspond to the emission of majority carriers from bulk

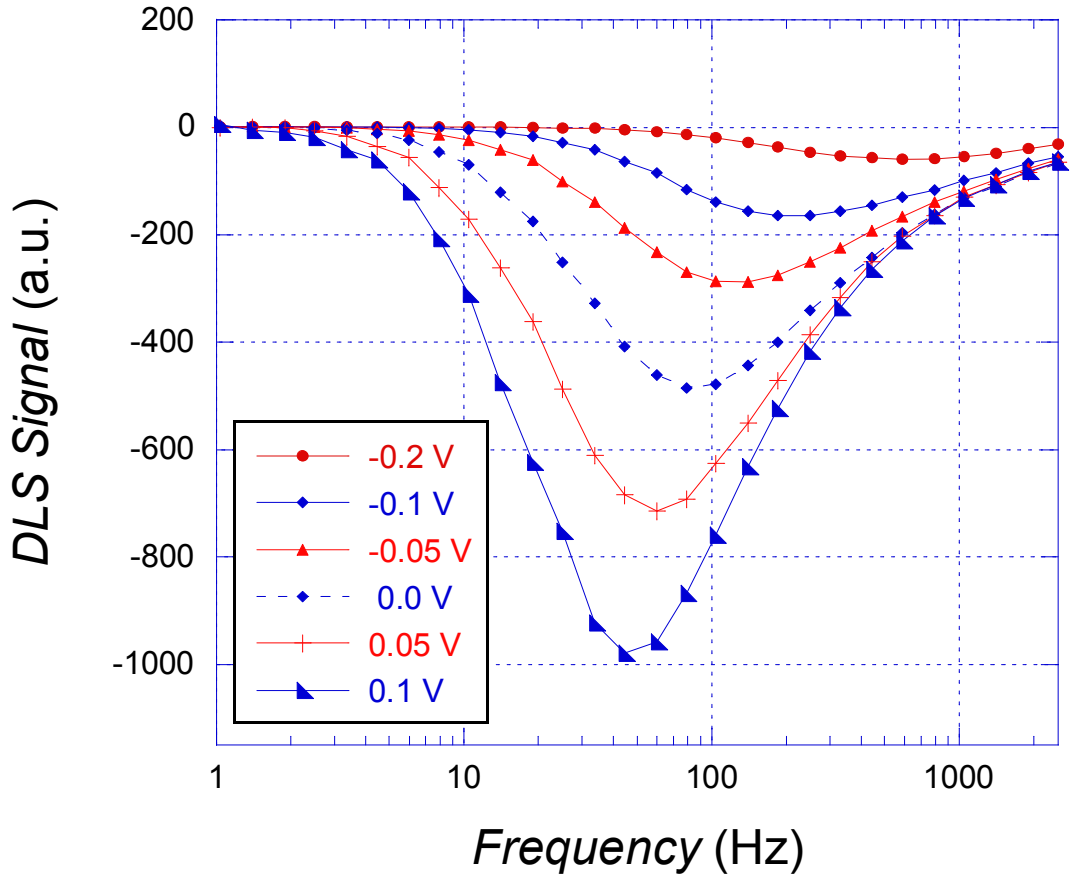

Figure 3.13. The observed DLTS peaks increases in magnitude as the depletion bias is pushed more into inversion. This indicates that the observed peak is due to minority carrier generation. The accumulation pulse is set to -2 V.

or interface traps, or it may be due to the generation of minority carriers (*i.e.*, the pulsed MOS-C *C-t* transient). Fortunately, each of these individual peaks behaves in a unique way

when the applied accumulation pulse or the depletion bias is manipulated. It has been shown [3.11] that the peak due to the generation of minority carriers – in which we are interested – is very sensitive to the magnitude of the depletion bias. The peak due to the minority carrier generation increases in size as the depletion bias is pushed deep into the inversion; while, the other two peak types are insensitive in this regard. This is the observed behavior in Fig. 3.13. The increase in peak's magnitude as the depletion bias is moved toward more positive values (*i.e.*, inversion) clearly indicates that it is due to the generation of minority carriers. Figure 3.14 shows the behavior of this peak with respect to temperature. The trend is in accordance with the peak that was numerically calculated in Fig. 3.6. The decrease in the peak's period with temperature (*i.e.*, the decrease of t_f with increasing temperature) is mainly associated with the strong temperature dependence of n_i in (3.7).

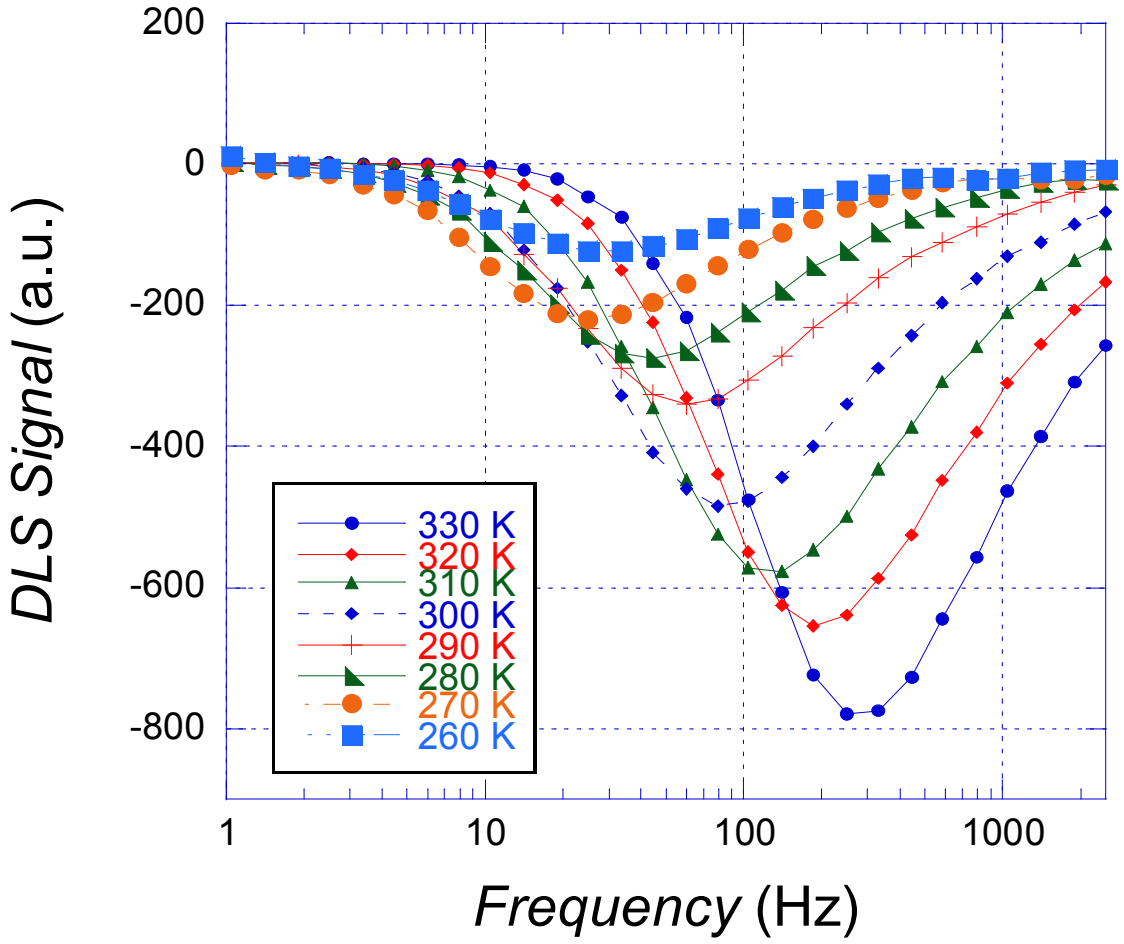

Figure 3.14. Effect of temperature on the DLTS peak. The effect is the same as what was predicted by numerical calculation and the simulated peak demonstrated in Fig. 3.12.

Quasi-static interface measurements that were performed on devices from more- and less- contaminated wafers (as demonstrated in Fig. 3.9) showed that both wafers share the same density of interface states. This is important because, as stated in (3.6), the measured generation lifetime is an effective value that incorporates both the surface (s_0) and the bulk (τ_g) generation components. Knowing that both wafers have the same interface states

density it can be concluded that any variation of deep-level impurity concentrations – that is related to the generation lifetime τ_g through equation (3.13) – has to be directly reflected in the measured lifetime values. In other terms, one has to expect to measure considerably different lifetimes for the two more- and less- contaminated wafers based on:

$$\tau_g = \frac{\exp(\Delta E_T / kT)}{\sigma_n \vartheta_{th} N_T} \tag{3.13}$$

where ΔE_T is the energy difference between deep impurity's energy level E_T and the intrinsic energy level E_i, σ_n is the minority carrier capture cross section, ϑ_{th} is the carriers thermal velocity, and N_T is the deep-level impurity concentration. In fact, it was already shown in Fig. 3.12 that the bulk lifetime is different when modified pulsed MOS-C capacitor is used on the samples.

Figure 3.15 shows DLTS spectra and the minority carrier generation peaks for the two more- and less-contaminated wafers. The measured peaks occur at frequencies of ~25 and 90 Hz for less- and more- contaminated wafers, respectively. Based on equations (3.10) and (3.11) such a difference in the minority carrier generation peak can be interpreted as different saturation times (t_f) for the two wafers. The corresponding t_f parameter is calculated to be 0.027 and 0.007 ms in this case which, as expected, well reflects the difference between lifetimes in the tested structures. These saturation times corresponds to lifetimes in microsecond range.

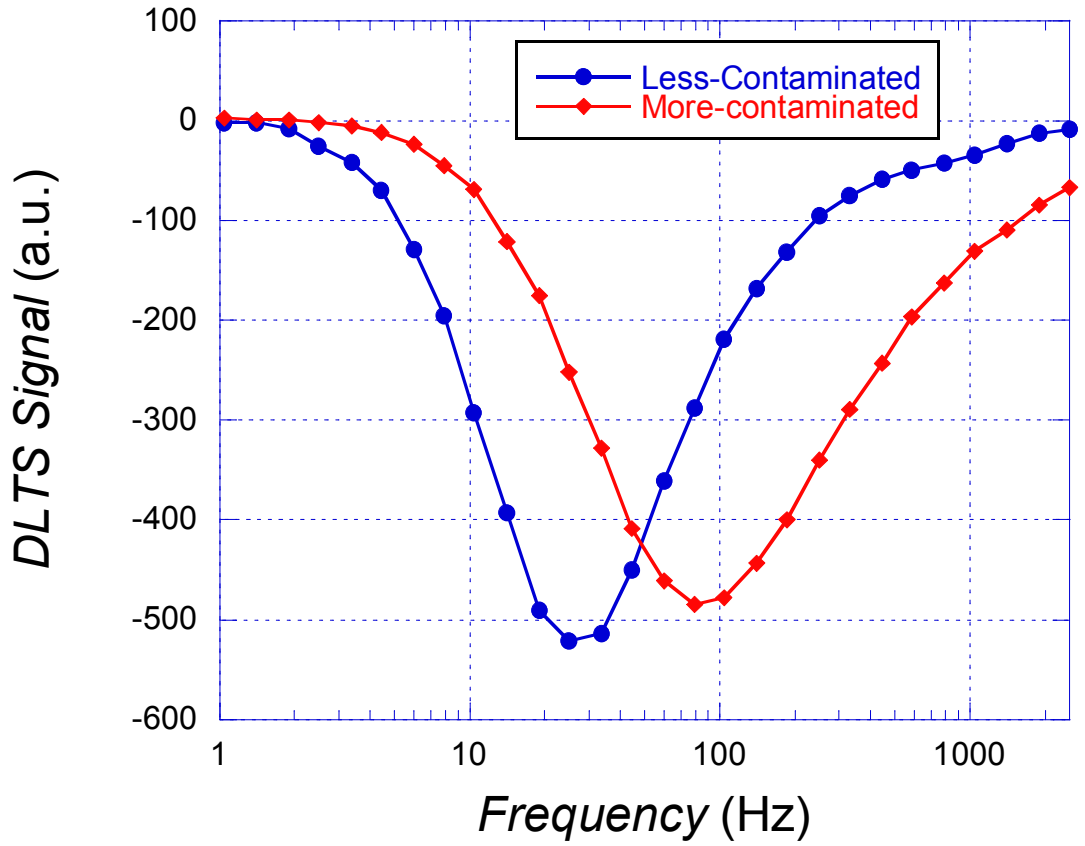

Figure 3.15. The DLTS spectrums for the more- and less-contaminated wafers showing the minority carrier generation peak. The difference in the peaks' frequency indicates different generation lifetime as predicted by our theory.

In order to demonstrate the accuracy of our proposed technique and to compare it with the classic pulsed MOS-C (aside from the less time it takes) a *t*-test with a sample size of 22 was performed. The MOS-C test structures were chosen in random on each wafer, DLTS measurement were done on them, and the corresponding t_f parameters were recorded. Classic pulsed MOS-C measurements were subsequently performed on test structures of each wafer chosen the same way. The pulsed MOS-C was done pulsing the devices from -5 to +5 V.

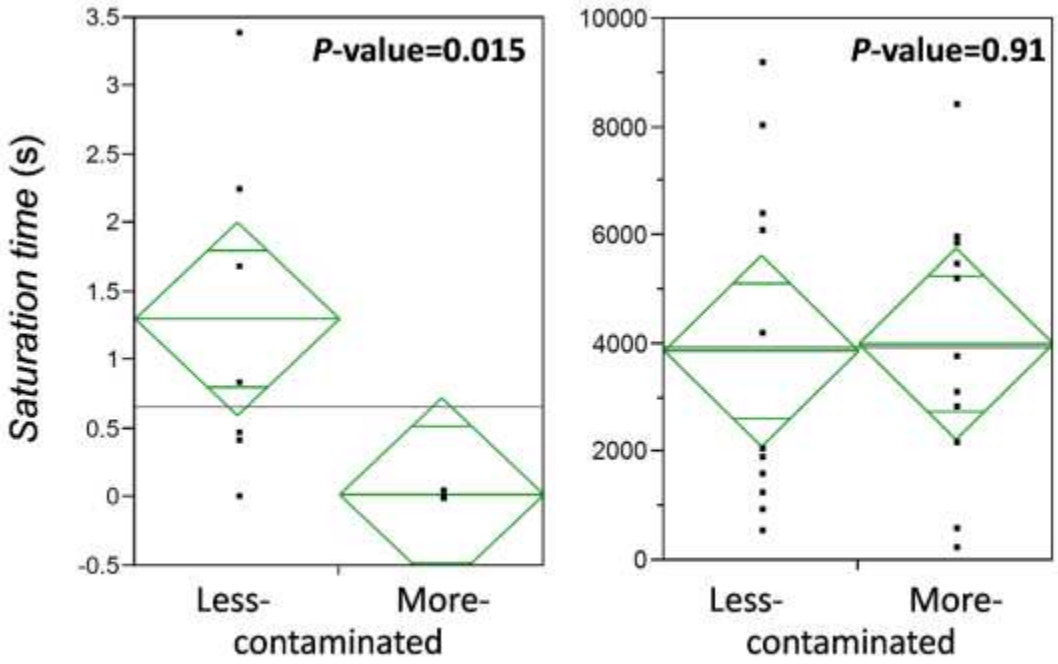

Figure 3.16. Diamond mean comparison plots of the t-test between saturation times of two wafers. Right: saturation times obtained through the classic pulsed MOS-C technique pulsed from -5 to +5 V at room temperature. The t-test does not distinguish any variation of lifetime. Left: The saturation times obtained through the application of frequency-scan DLTS using the same pulse biases at room temperature. The t-test clearly distinguishes between the lifetimes, as one would expect.

The results, analyzed with JMP software [3.20], are showed in Fig. 3.16. As it is expected, the saturation times obtained from classic pulsed MOS-C could not reveal any variation of the lifetimes, *i.e.* the mean values of the t_f for the more- and less- contaminated

wafers could not be declared different due to the test *p*-value of 0.91. In fact, and as it was shown in the preceding section, even when the lifetime is extracted from the whole recorded *C-t* transient (instead of using t_f alone), the parameter still would not be able to reveal any difference between the ultraclean wafers like the ones used in this study. On the other hand, this obviously is not the case when the t_f parameter is extracted from the abrupt *C-t* transients using DLTS. Referring to these saturation times, as depicted in Fig. 3.16, one can effectively distinguish the difference between the samples in terms of their lifetime. The t-test performed between the mean t_f of the more- and less- contaminated wafers declares them to be different with a small *p*-value of 0.015. This comparison clearly shows the superior detection power of our proposed technique over the classis pulsed MOS-C technique.

3.5 Conclusion

We have shown that the DLTS technique can be used as a fast mean to screen carrier generation lifetime using the pulsed MOS-C measurement principals. The measured *C-t* saturation times are showed to be of good accuracy when compared to the modified pulsed-MOS technique for the same structures capable of distinguishing between impurity concentrations as low as 10^{10} cm^{-3}. It is worth noting that screening the generation lifetime with this technique can be done in minutes while other techniques such as modified pulsed MOS-C may take up to several hours before accurate results are obtained. This advantage makes this technique suitable for screening purposes where the duration of such tests is of great importance [3.21].

3.6 References

[3.1] J. S. Kang, and D. K. Schroder, "The Pulsed MIS Capacitor," *Phys. Stat. Sol. (a)* vol. 89, pp. 13-43, May 1985.

[3.2] S. Lee, and D. K. Schroder, "Thin p/p$^+$epitaxial layer characterization with pulsed MOS capacitor," *Solid-State Elec.*, vol. 43, pp. 103-111, May 1998.

[3.3] D. K. Schroder, "Carrier lifetimes in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[3.4] S. M. Sze, *Physics of semiconductor devices*, 3rd ed., Wiley-Interscience, Hoboken, N.J., 2007.

[3.5] S. Y. Lee, "*Thin silicon layer characterization with the pulsed MOS capacitor,*" PhD Dissertation, Elect. Eng. Dep., Arizona State Unv., Tempe, 1998.

[3.6] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., John Wiley and Sons, 2006.

[3.7] D. S. Day, M. Y. Tsai, B. G. Streetman, and D. V. Lang, "Deep-leveltransient spectroscopy: System effects and data analysis," *J. Appl. Phys.*, vol. 50, no. 8, pp. 5093–5098, 1979.

[3.8] D. V. Lang, "Deep-level transient spectroscopy: A new method to characterize traps in semiconductors," *J. Appl. Phys.*, vol. 45, no. 7, pp. 3023–3032, Jul. 1974.

[3.9] M. Schultz and N. M. Johnson, "Transient capacitance measurements of hole emission from interface states in MOS structures," *Appl. Phys. Lett.*, vol. 31, no. 9, pp. 622–625, 1977.

[3.10] N. O. Pearce, B. Hamilton, A. R. Peaker, and R. A. Craven, "Application of deep level transient spectroscopy to metal-oxidesemiconductor relaxation transients," *J. Appl. Phys.*, vol. 62, no. 2, pp. 576–581, 1987.

[3.11] K. Yamasaki, M. Yoshida, and T. Sugano, "Deep level transient spectroscopy of bulk traps and interface states in Si MOS diodes," *Jpn. J. Appl. Phys.*, vol. 18, no. 1, pp. 113–122, Jan. 1979.

[3.12] A. Elhami Khorasani, T. L. Alford, D. K. Schroder, "Modified pulsed MOS capacitor for characterization of ultraclean thin p/p+ silicon epitaxial layers," IEEE Trans. Electron Dev., vol. 60, pp. 2592-2597, 2013.

[3.13] D. K. Schroder, *Modular Series on Solid State Devices: Advanced MOS Devices*. Reading, MA, USA: Addison-Wesley, 1987.

[3.14] F. D. Auret, “Considerations for capacitance DLTS measurements using a lock-in amplifier,” *Rev. Sci. Instrum.*, vol. 57, no. 8, pp. 1597–1603, 1986.

[3.15] G. L. Miller, J. V. Ramirez, and D. A. H. Robinson, “A correlation method for semiconductor transient signal measurements,” *J. Appl. Phys.*, vol. 46, no. 6, pp. 2638–2644, 1975.

[3.16] P. M. Henry, J. M. Meese, J. W. Farmer, and C. D. Lamp, “Frequency-scanned deep-level transient spectroscopy,” *J. Appl. Phys.*, vol. 57, no. 2, pp. 628–630, 1985.

[3.17] K. S. Rabbani, “Investigations on field enhanced generation in semiconductors,” *Solid-State Electron.*, vol. 30, no. 6, pp. 607–613, Jun. 1987.

[3.18] P. U. Calzolari, S. Graffi, and C. Morandi, “Field-enhanced carrier generation in MOS capacitors,” *Solid-State Electron.*, vol. 17, no. 10, pp. 1001–1011, Oct. 1974.

[3.19] Khorasani, Arash Elhami. *Carrier Lifetime Measurement*. Diss. ARIZONA STATE UNIVERSITY, 2013.

[3.20] JMP, Version 10. SAS Institute Inc., Cary, NC, 1989-2012.

[3.21] A. Elhami Khorasani, D. K. Schroder, T. L. Alford, “A fast technique to screen carrier generation lifetime using DLTS on MOS capacitors,” IEEE Trans. Electron Dev., vol. 61, no. 9, pp. 3282-3288, 2014.

CHAPTER 4

OPTICALLY EXCITED MOS CAPACITOR TECHNIQUE FOR CARRIER RECOMBINATION LIFETIME MEASUREMENT

4.1 Introduction

Carrier lifetime is one of the few parameters that can give information about the low defect densities in silicon [4.1]. Although impeded by experimental constraints, in theory there is no lower limit to the defect density that can be identified using lifetime measurements. It is for these reasons that the IC community, largely concerned with unipolar MOS devices, uses lifetime measurements as a "process cleanliness monitoring" tool. Being a device parameter, on the other hand, lifetime also has an important role in how a variety of semiconductor devices (such as p-n junction diodes, DRAM, CCD, thyristors, *etc.*) operate. Therefore, it is of great interest to develop techniques that can measure the recombination lifetime is an accurate, swift, and convenient way.

Lifetimes fall into two primary categories: recombination (τ_r) and generation (τ_g) [4.2]. The concept of recombination holds when the density of carriers is above equilibrium (*i.e.*, $n \times p > n_i^2$). The excess carriers in this case will decay through recombination so that their density can decrease to that of the equilibrium value. Generation lifetime, on the other hand, applies when there is a paucity of carriers (*i.e.*, when $n \times p < n_i^2$). This is the case in the space-charge region (scr) of a reverse biased p-n junction or a deep depleted MOS capacitor.

Photoconductance decay (PCD) [4.3] and surface photovoltage (SPV) [4.4] techniques are widely used for characterization of τ_r in bulk wafers. However, since they measure lifetime from the minority carrier's diffusion length, these techniques cannot be applied to thin epitaxial layers where the layer thickness is smaller than the minority carrier diffusion length. Microwave photoconductance decay (μ-PCD) [4.5], time-resolved photoluminescence (TRPL) [4.6], and frequency-dependent SPV [4.7] are among techniques that have been used to study the recombination lifetime in epitaxial layers by trying to confine the excitation inside the epitaxial layer thickness using short excitation wavelengths. Unfortunately, all of these approaches require special test setups (*e.g.,* dipping the wafer in HF solution [4.8] to achieve surface passivation) before they can yield correct results. Such difficulties make these techniques time-consuming and impractical for process screening purposes, especially on fully developed test vehicles.

The pulsed MOS technique has been shown to be an eminently suitable method for measuring the generation lifetime in epitaxial wafers [4.9]. This is because the carrier generation lifetime is measured inside the scr width that can be easily confined within the epi-layer thickness by exploiting the MOS-C gate voltage (V_G). This technique has gained popularity since the MOS test structures are, in fact, building blocks for a variety of other semiconductor devices [4.10] and the extracted lifetime parameters have direct design application. Moreover, MOS capacitors are favorable test structures due to their relatively easy fabrication process. In fact, it is a standard industrial practice to incorporate MOS devices into process control modules (PCM) across the product wafers for test purposes [4.11]. Therefore, it is invaluable to have the ability of also extracting the recombination lifetime using MOS capacitors.

It will be shown that the recombination lifetime can be measured by employing the reverse of the underlying principles behind pulsed MOS method. If the density of minority carriers in the inversion layer is increased over the equilibrium value employing an external excitation, as soon as the excitation is stopped, the device attains equilibrium through the recombination of the excess carriers. The device capacitance transient in this case, as will be demonstrated, is used to extract the recombination lifetime.

4.2 Experimental

4.2.1 Device Fabrication and Measurement

The samples were prepared on p/p^+ Silicon epitaxial wafers. The epi-layer had a thickness of 4 μm and a resistivity of 30 ohm-cm. The heavily doped substrate was 725 μm thick and had a resistivity of 10^{-3} ohm-cm. Intentional Tungsten ($E_T = E_V + 0.41$ eV) contamination at two densities of less than 10^{10} and 3×10^{10} cm^{-3} were introduced into the wafers in order to form two more- and less-contaminated device groups. Device fabrication was done by growing a 20 nm thermal gate oxide followed by the deposition of circular (1 mm in diameter) Aluminum gates. The metal deposition was done in an E-Beam Lesker-75 PVD evaporator. A post-metallization anneal, to reduce interface traps, was done at 450 ^oC for 30 minutes in an ambient of 95% nitrogen and 5% hydrogen. Quasi-steady state technique [4.12] was used to measure the Si/SiO_2 interface trap density after device fabrication employing a Keithley model 82 C-V system. It was deemed that both device groups have the same interface trap densities equal to 10^{10} $cm^{-2}eV^{-1}$.

The doping concentration profile for samples was determined from differential capacitance method [4.12] using the deep-depletion C-V and $1/C^2$-V curves, and was

calculated to be constant throughout the epitaxial layer and equal to 9.8×10^{14} cm^{-3}. The capacitance measurements were done using an Agilent 4284-A LCR meter controlled on a LabView platform. In order to insure that the illumination turning-off transient does not affect the measurements, it was necessary to use a light emitting diode (LED) instead of an incandescent light bulb as the light source. The optical excitation in our experiments was done employing a fast switching LED whose switching frequency could be controlled via a PC platform. The illumination intensity was measured using a model 850009 UVA/B light-meter.

4.2.2 Device TCAD Simulation

TCAD simulations have been performed using the Synopsys TCAD package. 3D MOS capacitor structures were first constructed and meshed with SDeviceEditor [13] and SMesh [4.14], respectively. The electrical behavior of these meshed structures were then simulated using SDevice [4.15].

Since small signal (*i.e.*, capacitance) simulation of the MOS-C devices is sensitive to the mesh refinement inside the depletion region, a dense mesh was used at this location; while, a coarser mesh was used for the rest of the structure. Although it is computationally more favorable to simulate a smaller device, the gate area was defined large enough to avoid narrow gate effects [4.10]. As demonstrated in Fig. 4.1, the MOS-C depletion region extends beyond the gate in the lateral direction. When the gate area becomes comparable with these extended regions, their charge contribution cause unwanted complexities in the analytical solutions.

The device electrical simulation is performed in the 3D domain and by using the cylindrical coordinate system [4.16]. In this approach one exploits the cylindrical symmetry of the structure as a mean to reduce the computational resources needed for simulating a 3D device. Figure 4.1(a) shows the MOS-C cross section that is used as a base plane for an initial 2D simulation. This plane is then rotated along a symmetry axis (which in this case is set to x= 0) to create a 3D cylindrical MOS-C structure, as depicted in Fig. 4.1(b). Although the final results of this simulation are achieved over substantially shorter time frames, they are equivalent to those obtained doing a complete 3D simulation.

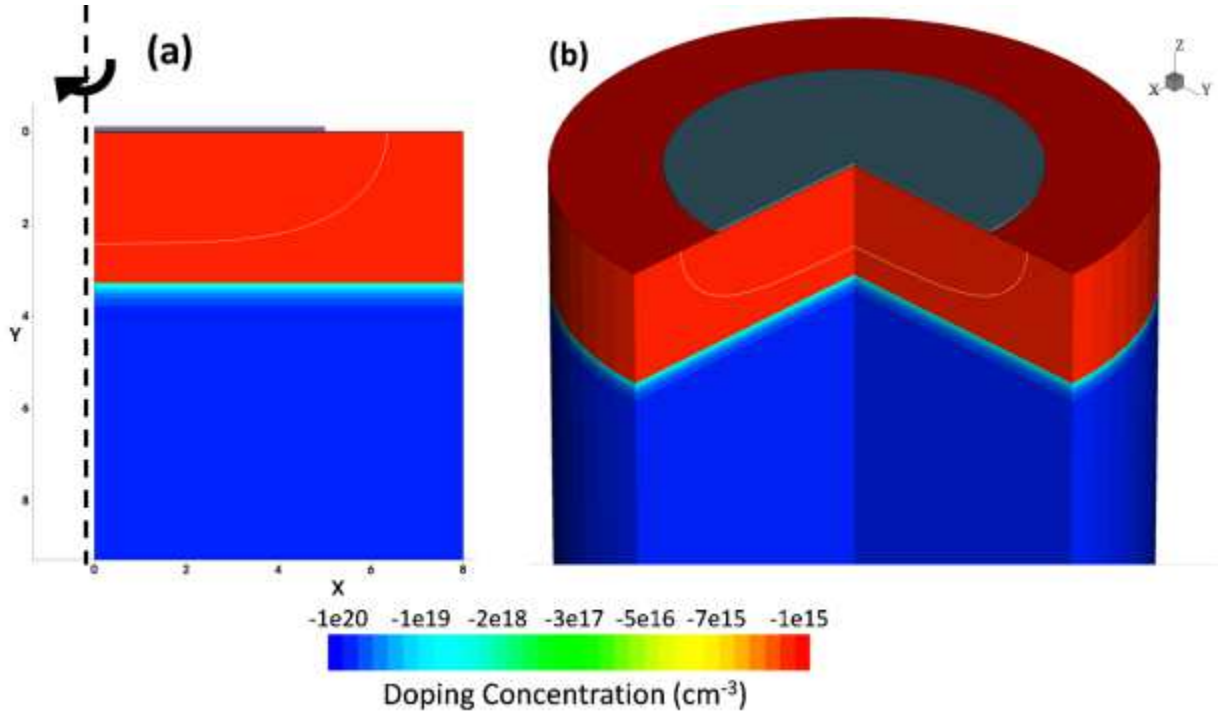

Figure 4.1. (a) A MOS capacitor cross section that is used as a base plane for an initial 2D simulation. (b) The equivalent 3D structure that the final simulation results would represent when cylindrical coordinate system is used.

4.3 Theory

A photon with energy larger than the band-gap of a semiconductor can cause optical generation of electron-hole pairs (ehps). Light can therefore be used as a tool for manipulating the carrier density inside a semiconductor. When this external excitation is stopped, however, the excess optically generated carriers will need to recombine in order for the material to return to the equilibrium (where $n \times p = n_i^2$). Such a generation/recombination cycle is the foundation for our measurement technique and is used to study the carrier lifetime.

The distribution of electrons and holes inside an MOS-C can also be manipulated employing its gate voltage. With the application of an inverting gate bias an inversion layer surrounded by a depletion region will be formed beneath the gate. If the inverted device is then illuminated, the newly added carriers from optical generation will increase both *p* and *n* in silicon causing the inversion layer charge to increase, while decreasing the depletion region width (*i.e.*, the bulk charge). Such a change in the depletion region width can be monitored by measuring the MOS-C capacitance. The MOS-C capacitance, when measured with high ac probing frequency (C_{meas}), is a combination of the oxide capacitance (C_{ox}) in series with the depletion capacitance (C_b) following:

$$C_{meas} = \frac{C_{ox}.C_b}{C_{ox} + C_b} = \frac{C_{ox}}{1 + (K_{ox}W/K_s t_{ox})} \tag{4.1}$$

with K_{ox} and K_s being the dielectric constants for oxide and silicon, and t_{ox} the oxide thickness. Based on (4.1), any change in W would perturb the device measured capacitance given C_{ox} is essentially a constant. Therefore as soon as the illumination starts, the depletion

region will shrink and a rise in capacitance will be observed. In contrast, immediately after the illumination is stopped the optically generated carriers will begin to recombine and the capacitance decreases as W widens. This evolution of the depletion region width is presented using TCAD simulations in Fig. 4.2. The resultant capacitance transient when recorded with respect to time (*i.e.*, the *C-t* plot) can therefore be used to calculate W as a function of time following:

$$W(t) = AK_s\varepsilon_0\left(\frac{1}{C_{meas}(t)} - \frac{1}{C_{ox}}\right) \tag{4.2}$$

with A being the gate area, and ε_0 the vacuum permittivity.

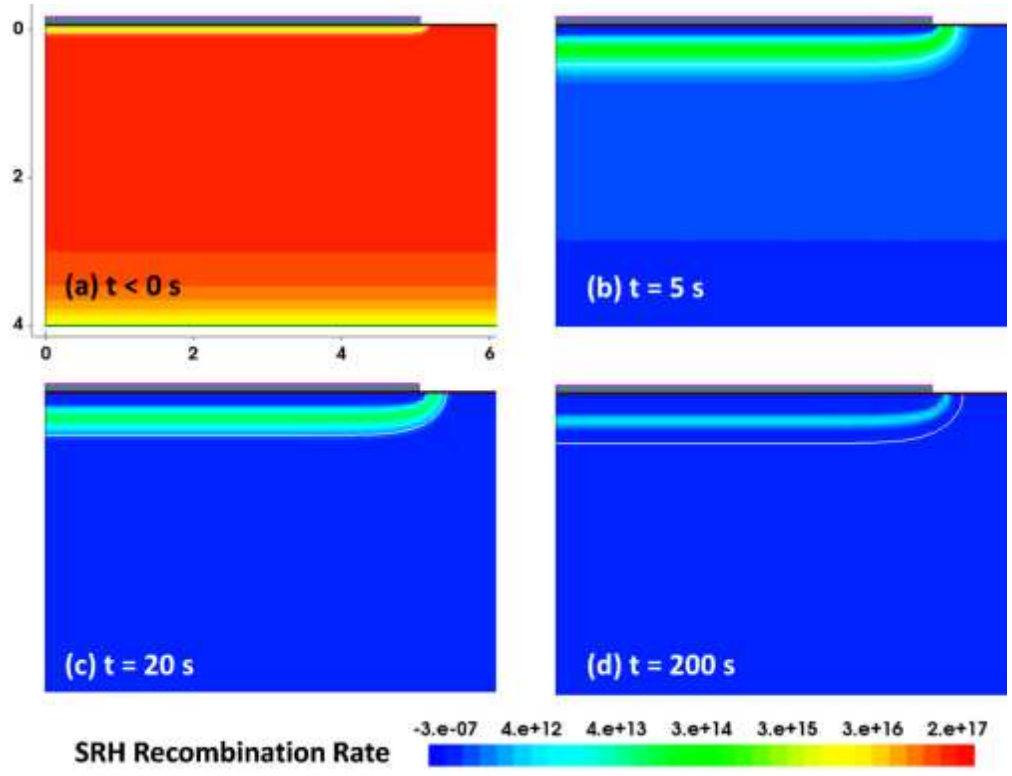

Figure 4.2. The evolution of the depletion region width with time under optical excitation. The simulated structure corresponds to a device with doping concentration of

10^{15} cm^{-3} and the oxide thickness of 20 nm. The illumination is switched off at $t = 0$ and the scr has the minimum width while the illumination is on in (a).

By knowing the depletion region width as a function of time, the electrostatic potential in silicon $\varphi(x)$ (defined by taking φ_i=E_i/q in the bulk as the reference potential) can be calculated at any individual time using [4.12]:

$$\varphi(x,t) = qN_A(W(t) - x)^2/2K_s\varepsilon_0 \tag{4.3}$$

where N_A is the doping concentration, q the electric charge of an electron, and x is the distance from the Si/SiO_2 interface. The electrostatic potential value enables us to calculate the carrier concentration at any location and time within the depletion region using the quasi-Fermi levels for holes (E_{Fp}) and electrons (E_{Fn}), following:

$$p(x,t) = n_i \times \exp(\frac{E_{i,bulk} - E_{Fp}(t) - q\varphi(x,t)}{k_BT}) \tag{4.4}$$

$$n(x,t) = n_i \times \exp\left(\frac{E_{Fn}(t) - E_{i,bulk} + q\varphi(x,t)}{k_BT}\right) \tag{4.5}$$

where k_B is the Boltzman constant, T temperature and n_i the intrinsic carrier concentration.

In general dn/dt = G – R with G and R being the generation and recombination rates. However, when $p.n>n_i^2$ (*i.e.*, immediately after the light pulse is turned off) the recombination dominates the generation and one can write dn/dt = –R [4.17].

Assuming the lifetimes for holes and electrons to be the same and equal to τ_r, the SRH recombination rate can be written as:

$$R = \frac{pn - {n_i}^2}{\tau_r(n + p + 2n_i)} \tag{4.6}$$

Based on (4.6), and as demonstrated by TCAD simulations in Fig. 4.3, the recombination rate will not be uniform throughout the scr if *n* and *p* vary with *x*. This is a key concept that separates the optically excited MOS capacitor technique from other optical and charge-based measurements. The variations of *n* and *p* (due to the application of an inverting gate bias) give rise to a weighted recombination rate in the scr whose centroid (*i.e.,* the location for R_{max}) is where holes and electrons have the same densities. Due to the large magnitude of R_{max}, and as we are only concerned with the scr when measuring the capacitance, to the first order approximation, we can ignore recombination in the rest of the device.

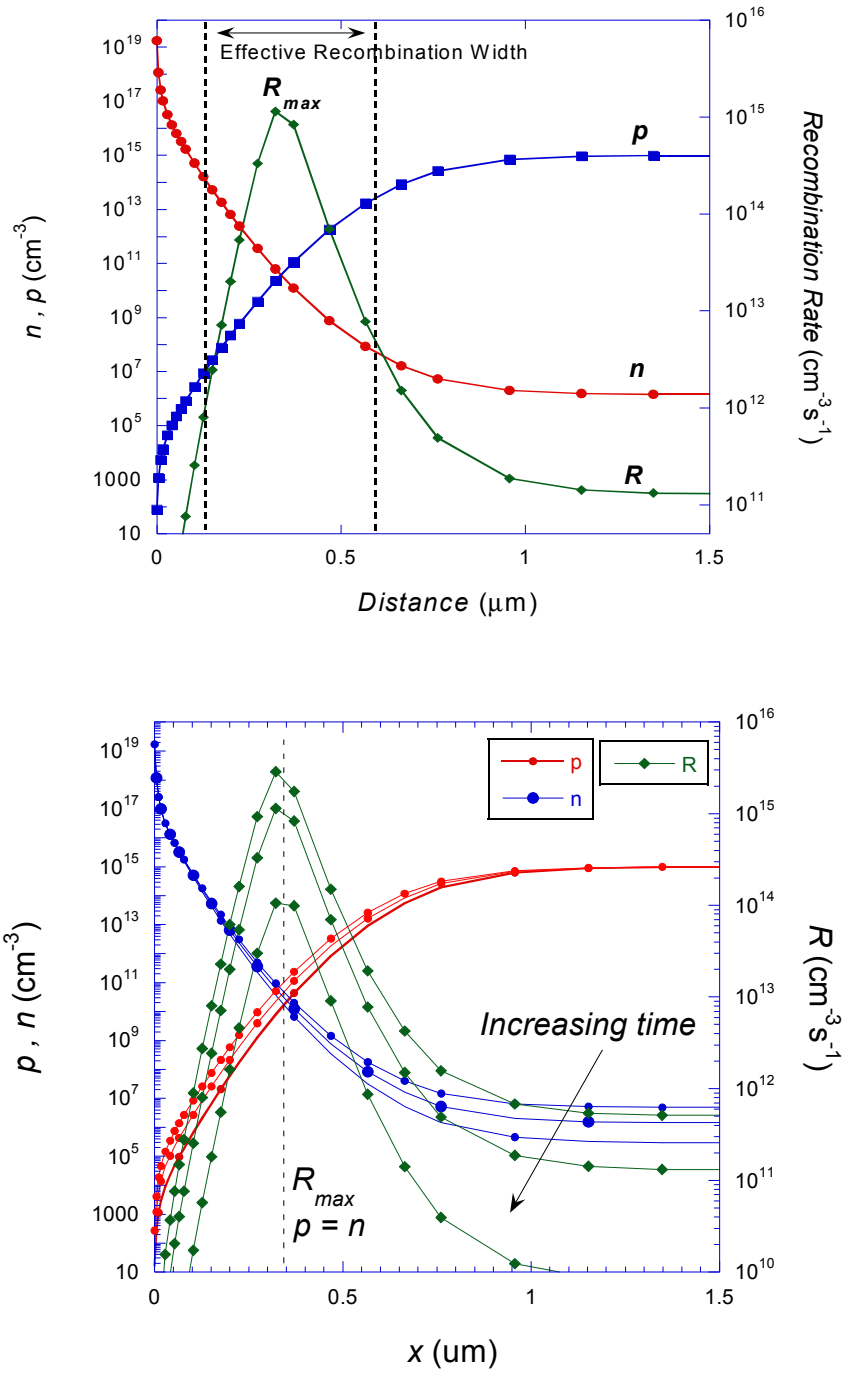

Figure 4.3. Electron/hole density and SRH recombination rate profile drawn with distance from Si/SiO_2 interface during the *C-t* transient in top, and how they change with time in bottom.

Employing the delta-depletion approach [4.18], the charge density (Q_p) that is consumed in the recombination at any time (considering holes for mathematical simplicity) can be written as:

$$Q_p = N_A(W_F - W(t)) \tag{4.7}$$

If one considers the (W_F-W_i) value as the effective recombination width [4.19] with W_F and W_i being the depletion region width at the end and at the beginning of the transient, the charge density Q_p can be correlated with the recombination rate following:

$$\frac{dQ_p}{dt} = \frac{dQ_n}{dt} = -q \int_{W_i}^{W_F} R_{max}.\, dx \tag{4.8}$$

Hence, by having W-t data, one can calculate R_{max} as a function of time.

The location at which $n = p$ can be identified using the steady-state condition. At equilibrium, $n = p = 10^{10}$ cm^{-3} where $\varphi(x) = \varphi_F = (E_F\text{-}E_i)/\text{q}$ with φ_F being a function of the doping concentration. Assuming this location to move proportional to the movement of W as it changes, and considering the E_{Fp} to be equal to E_F throughout the scr [4.20], p can be calculated using (4.4) and (4.5) from knowing the W-t data employing (4.3). With having R_{max} and p for each specific time, (4.9) can be used to extract the lifetime τ_r parameter as the slope of the p vs. R_{max} plot.

$$R(t) = \frac{p(t) + n_i}{2\tau_r} \tag{4.9}$$

Based on the above simple mathematical approach (that is also demonstrated in Fig. 4.4), the capacitance transient can be used to obtain the recombination lifetime inside the epitaxial layer. A key advantage of this measurement is that the location at which one probes the lifetime is within the depletion region width, typically 1-2 μm away from the surface. This is well confined within the thickness of a typical epitaxial layer. No other recombination lifetime measurement technique can attain this local precision without the need of a more complicated setup. Moreover, a positive gate voltage acts to repel holes from the surface minimizing the effect of surface recombination under the gate. Another advantage of our method is that the graphical extraction of lifetime from the slope of p vs. R_{max} plot enables us to check against possible deviations from assumptions that have been made in deriving the formulation. If these assumptions do not hold a linear region should not be observed in the lifetime plot.

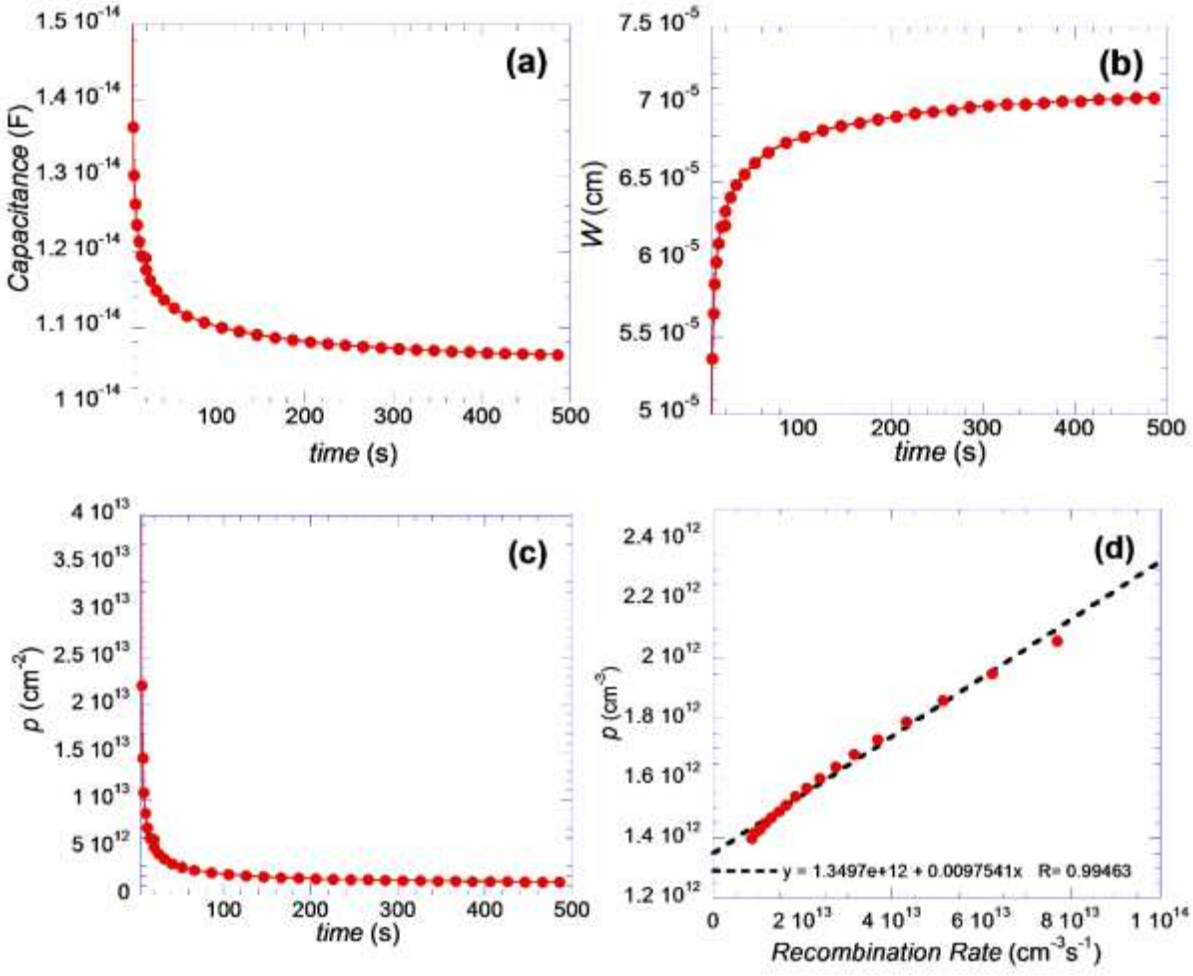

Figure 4.4. (a) *C-t*, (b) *W-t*, (c) *p-t* and (d) the lifetime plot calculated in sequence after one another.

4.4 Results and Discussion

The measured *C-t* plots for a MOS-C device that is being optically excited is demonstrated in Fig. 4.5. In this measurement the MOS gate bias is being held at inversion (5 V) as a light pulse is shone on it. As explained in the preceding section, the light illumination causes an increase in capacitance. The capacitance will immediately start to

decrease when the light is switched off. The turning off *C-t* has an instantaneous drop in its beginning. This phase corresponds to the high level injection and the large number of carriers that are available for recombination at $t = 0$.

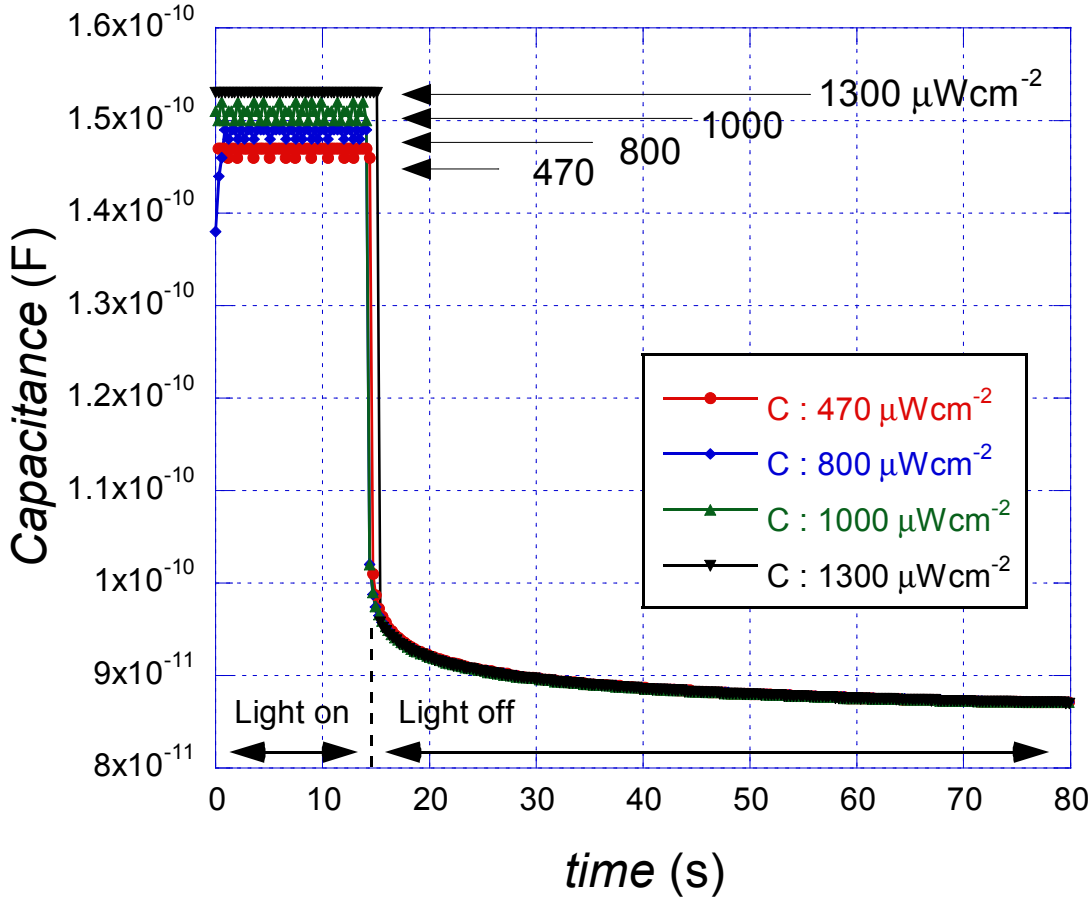

Figure 4.5. Measured *C-t* behavior of a MOS-C with an applied pulse while being biased in the inversion. The use of various light intensity have little effect on the capacitance decay.

Table 4.1
Measured lifetime parameters on more- and less-contaminated wafers

Wafer	Contamination	τ_g (ms)	$L_{n\,eff}$ (μm)	τ_r (μs)
Less-Contaminated	$< 10^{10}$	45	56	150
More-Contaminated	3×10^{10}	25	108	62

The instantaneous drop is then followed by a slow decay where our mathematical model applies. In this region the ehps density has reduced to a point where low level injection (as shown in Fig. 4.3) prevails in an effective recombination width. Such behavior makes the slowly decreasing part of the *C-t* independent of the illumination intensity. In other words, any difference in the light intensity (which results in a different carrier density at t=0) will be resolved in the initial steep portion of the *C-t*. Figure 4.5 shows the *C-t* plots that are measured with different light intensities. Notice that although different light intensities can cause different initial carrier densities (and

hence capacitance values), the *C-t* plots become similar as they enter the slow decaying phase. Therefore, unlike other optical methods, in which it is necessary to know and control the initial density of optically generated ehps, this information is not necessary for our proposed technique.

Figure 4.6 shows the lifetime plots obtained from TCAD simulation of three identical MOS devices that have different predetermined lifetime values of 0.5, 1 and 5 ms. As is

evident from this figure, the extracted lifetimes are in good agreement with their expected values.

In order to demonstrate that our technique measures a lifetime that only represents the epitaxial layer, TCAD simulations have been performed with the lifetime parameter manipulated location- wise throughout the device. Figure 4.7 shows the resultant *C-t* plots when the lifetime in a 0.15 μm band parallel to the surface is set to a different value (5 ms) than the rest of the structure (0.1 ms). Simulations are repeated with the location of this band being changed so that various distances from the surface can be swept. The results, as predicted by our model, shows that the *C-t* is only affected when the lifetime is changed in a certain region well within the epitaxial layer thickness. This location corresponds to the concept of an effective recombination width that was explained in the theory section. On the contrary, the *C-t* remains the same as the lifetime at other locations are changed.

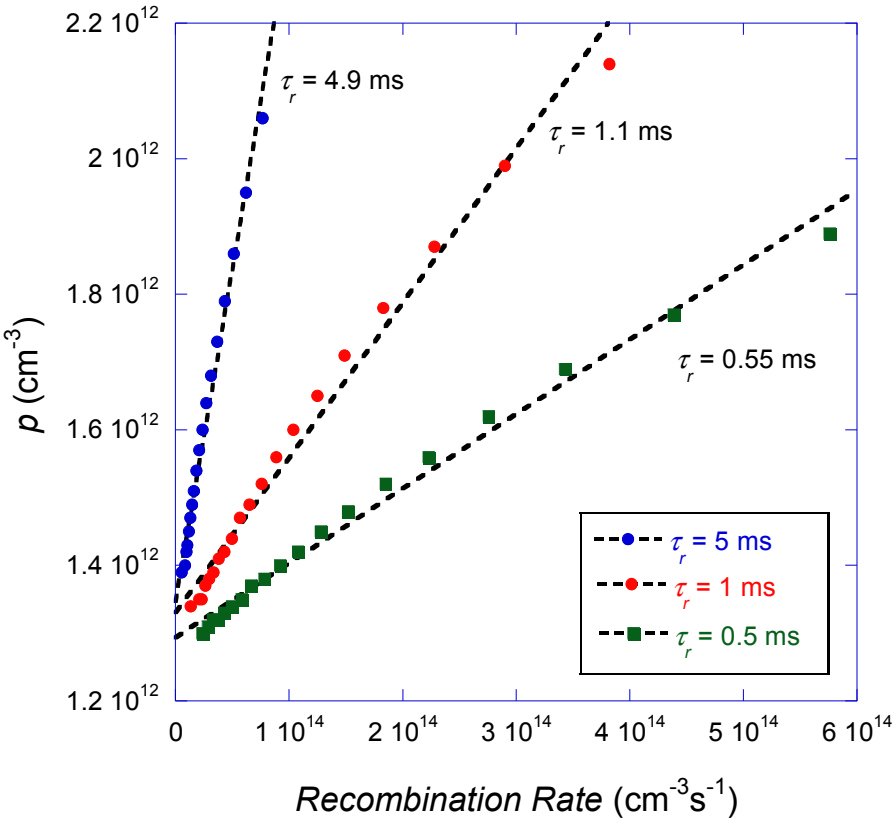

Figure 4.6. Lifetime plots obtained from three simulated MOS-C devices with predetermined lifetime values of 0.5, 1 and 5 ms. As is evident, the extracted lifetimes well reflects the expected values.

Figure 4.8 shows the measured *C-t* transients for a fabricated MOS-C device that is known to have Tungsten contamination. Figure 4.8(a) is obtained using the classic pulsed MOS measurement, in which the device is pulsed for -5 V to +5 V and its capacitance is recorded with respect to time. The *C-t* is then used to calculate the generation lifetime employing a method that is described in details in reference [4.9]. The generation lifetime is deemed to be 19 ms for this device. The optically-excited MOS measurement is also

performed on the same device with its resultant *C-t* plot depicted in figure 4.8(b). The recombination lifetime for this device is calculated to be 60 us.

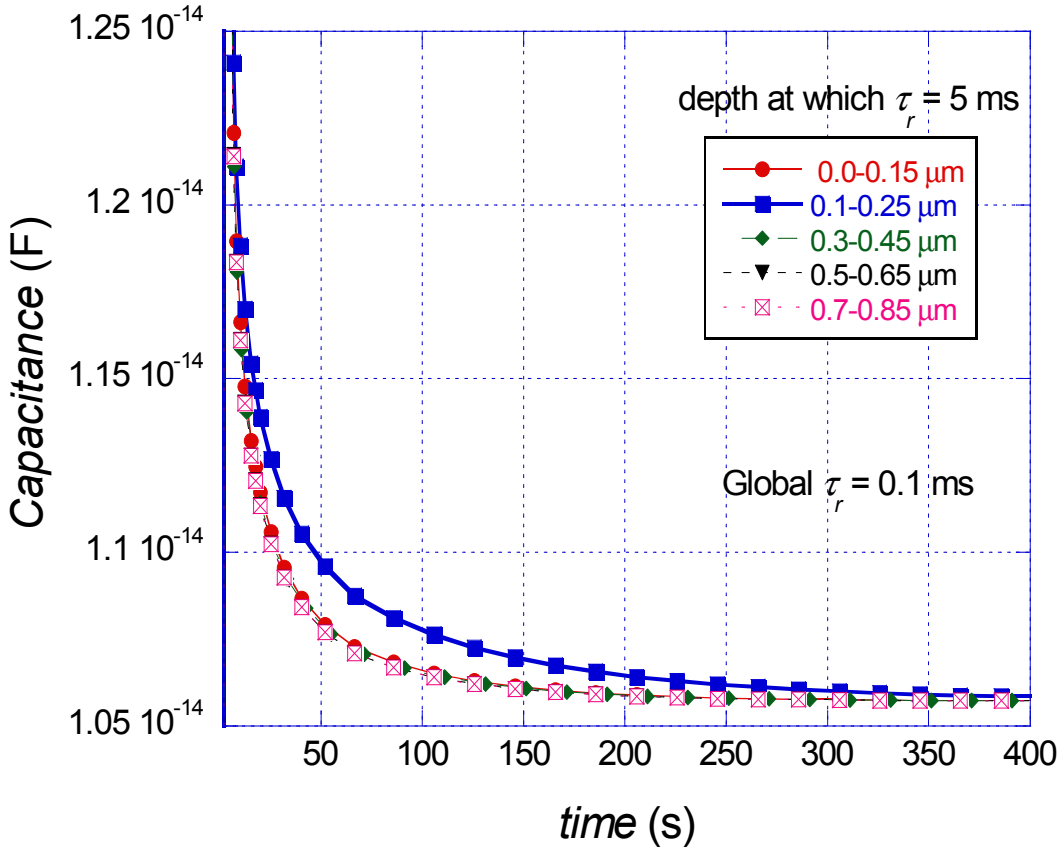

Figure 4.7. C-t transients obtained with TCAD simulations. The C-t plot changes only when the lifetime at a certain region is manipulated.

In order to compare and analyze the two extracted lifetime values, it needs to be pointed out that the recombination and generation lifetimes are related to each other following [4.12]:

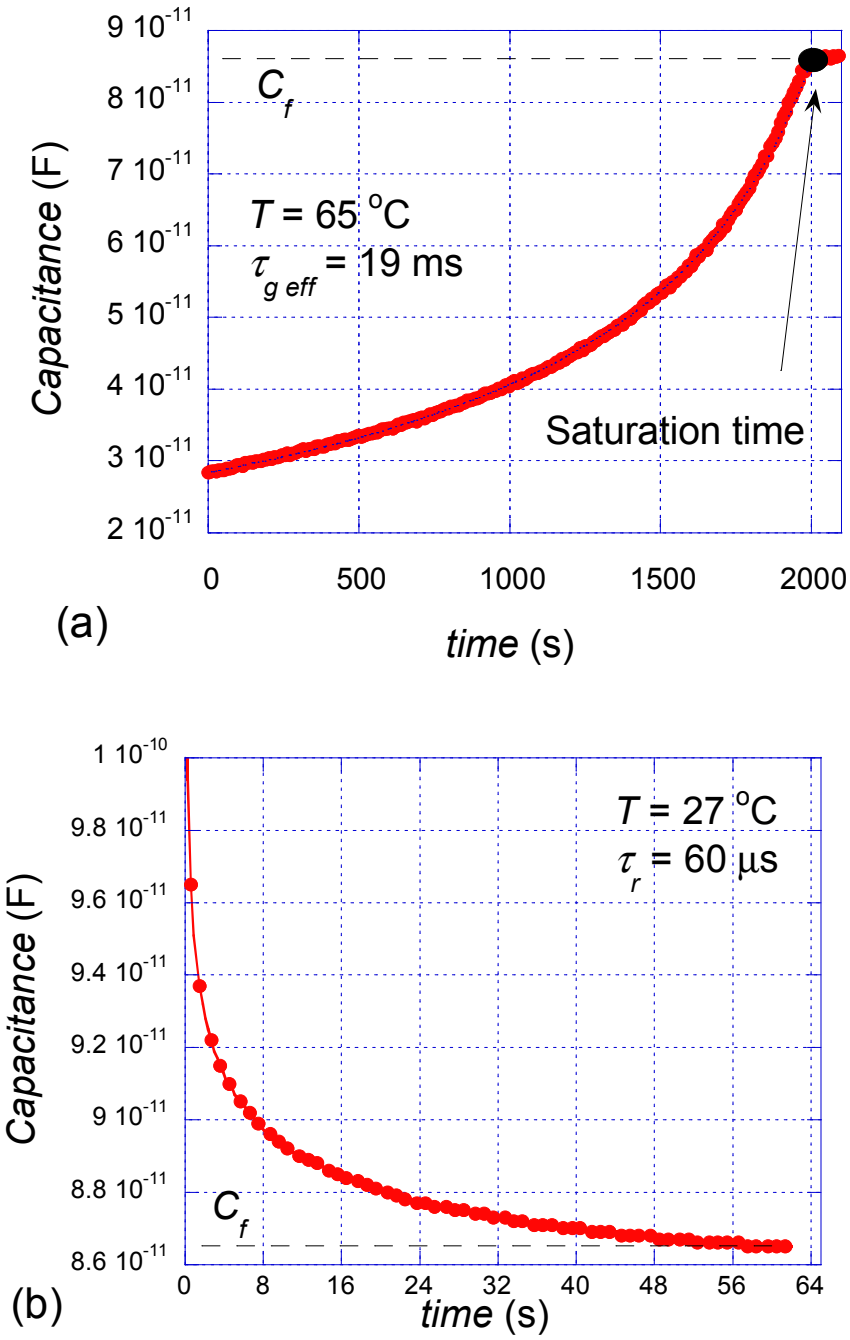

Figure 4.8. Measured *C-t* plots with: (a) pulsed MOS technique and (b) optically-excited MOS technique on samples with Tungsten contamination. The calculated generation and recombination lifetimes are 19 ms and 60 µs, respectively.

$$\frac{\tau_g}{\tau_r} = 2 \times \cosh(\frac{E_T - E_i}{k_B T}) \quad (4.10)$$

where E_T is the energy level for the deep level impurity. Since the metallic impurity in the studied samples is Tungsten with an energy level of E_V + 0.41 eV, based on (4.10), the measured generation lifetime is expected to be ~320 times larger than the recombination lifetime. This ratio is in good agreement with the lifetime values that are extracted for these devices (τ_r =60 µs, and τ_g =19 ms) as depicted in figure 4.8.

Table 4.1 is a summary of the mean measured recombination lifetimes (obtained with the optically excited MOS technique), generation lifetimes, and the effective diffusion length values (obtained with the modified pulsed MOS technique [4.9]) for the more- and less- contaminated samples with 3×10^{10} and $<10^{10}$ cm^{-3} of Tungsten contaminations. The calculated τ_r values reflects the expected difference between the samples following:

$$\tau_r \propto \frac{1}{\sigma_n \vartheta_{th} N_T} \quad (4.11)$$

where σ_n is the minority carrier capture cross section, ϑ_{th} is the carriers thermal velocity, and N_T is the deep-level impurity concentration.

An interesting point here, however, is that the diffusion length values presented in Table I do not follow the defect density variations between the samples. This discrepancy, as already mentioned, is due to the fact that the measured diffusion length in p/p^+ epitaxial layers is an effective value defined through [4.12]:

$$L_{n\,eff} = L_n \frac{1 + (s_i L_n / D_n)\tanh(t_{epi}/L_n)}{tanh(t_{epi}/L_n) + s_i L_n / D_n} \tag{4.12}$$

with D_n being the diffusivity, t_{epi} the epi-layer thickness, and s_i the epi/substrate interface recombination velocity. The dependence of the effective diffusion length on the value of interface recombination velocity makes it a vague parameter. s_i itself depends on a variety of other parameters such as doping concentration, diffusion length in the substrate, and the band gap narrowing of the p^+ region. Such difficulties in interpreting the effective diffusion length render most of the techniques that are based on using it impractical for studying lifetime in the p/p^+ epitaxial layers.

Another important concern when measuring the lifetime is the departure of the extracted parameter from the true bulk lifetime as a result of the surface recombination/generation [4.7], [4.21]. When measuring recombination lifetime in thin epitaxial layers, there are two interfacial recombination velocities that can influence the measurements: Surface recombination at the Si/SiO_2 interface (s_0), and interface recombination at epi/substrate interface. The former can be reduced using a post fabrication anneal (to 1-10 cm/s); while, the latter is usually in the 100 cm/s range even for the high quality epitaxy processes that are available today. If these surface and interface recombinations are not prevented from affecting the measurements (as occurs in most techniques), the extracted lifetime has been shown to become an effective value following [4.7]:

$$\tau_{r\,eff} \approx \left(\frac{1}{\tau_{r\,epi}} + \frac{2s_0}{t_{epi}} + \frac{2s_i}{t_{epi}}\right)^{-1} \tag{4.13}$$

The accuracy of conventional techniques in determining the bulk lifetime is substantially reduced as a result of the surface recombination. This problem is extensively addressed in references [4.5] and [4.7]. As an example, a value of 100 cm/s for s_i in a 4 um thick epi layer will cause a bulk lifetime of 100 μs to be measured 100 times smaller when using microwave photoconductance decay technique [4.7]. This issue however is greatly reduced in our proposed optically excited MOS (OEMOS) technique. In our technique, the epi/substrate recombination plays no role in the measurement. This, as previously addressed, is due to the fact the carrier recombination dynamics is being dominated by what happens at a specific location far away from the interface and well within the depletion region width. Figure 9 shows a comparison between how different interface recombination velocities distort the measured bulk lifetime when classic methods and OEMOS technique are used on the same material.

The role of Si/SiO_2 surface recombination in masking the bulk recombination lifetime is also minimized in the OEMOS technique. There is practically no surface recombination beneath the gate where the surface is heavily inverted and shielded against recombination (due to low hole density). Surface recombination at Si/SiO_2 interface only stays effective in a small area away from the gate that always remain depleted throughout the measurement. Being a small region compared to the volume where bulk recombination takes place, however, this surface component has minimum effect on the extracted lifetime

parameter. Figure 10 shows the extracted lifetime parameter for different values of Si/SiO_2 surface recombination velocities. As the data suggests, the deviation of lifetime is substantially less for OEMOS compared to the classics techniques such as photoconductance decay measurements, where (4.13) holds.

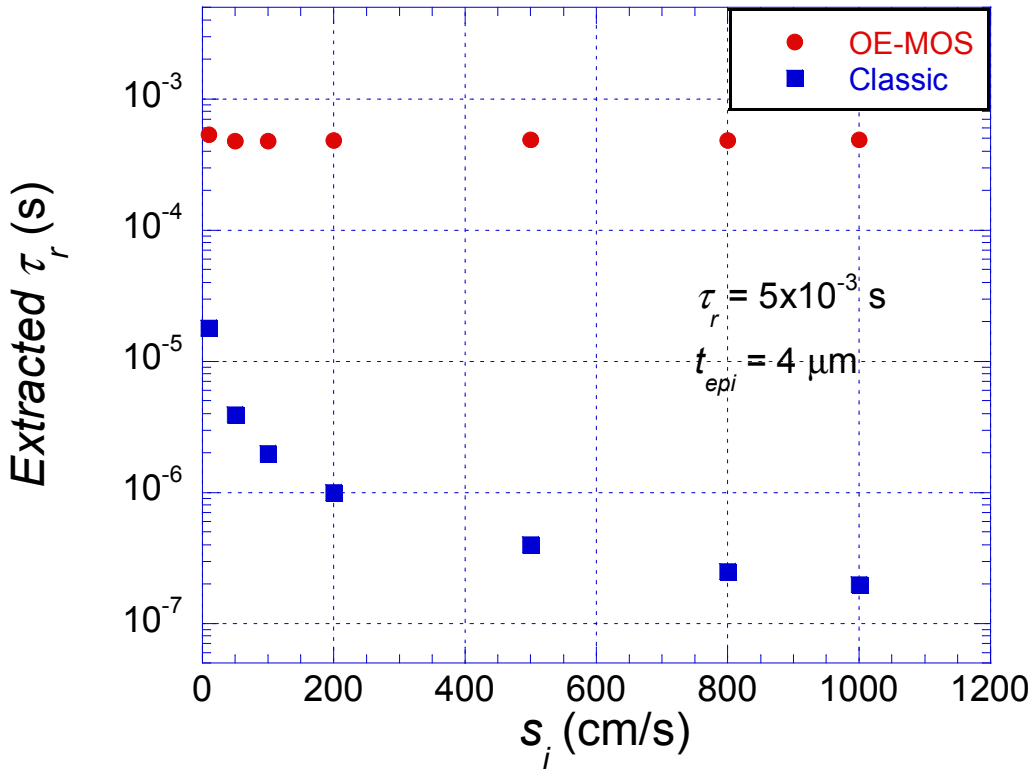

Figure 4.9. The effect of epi/substrate interface recombination velocity (s_i) on the extracted τ_r parameter.

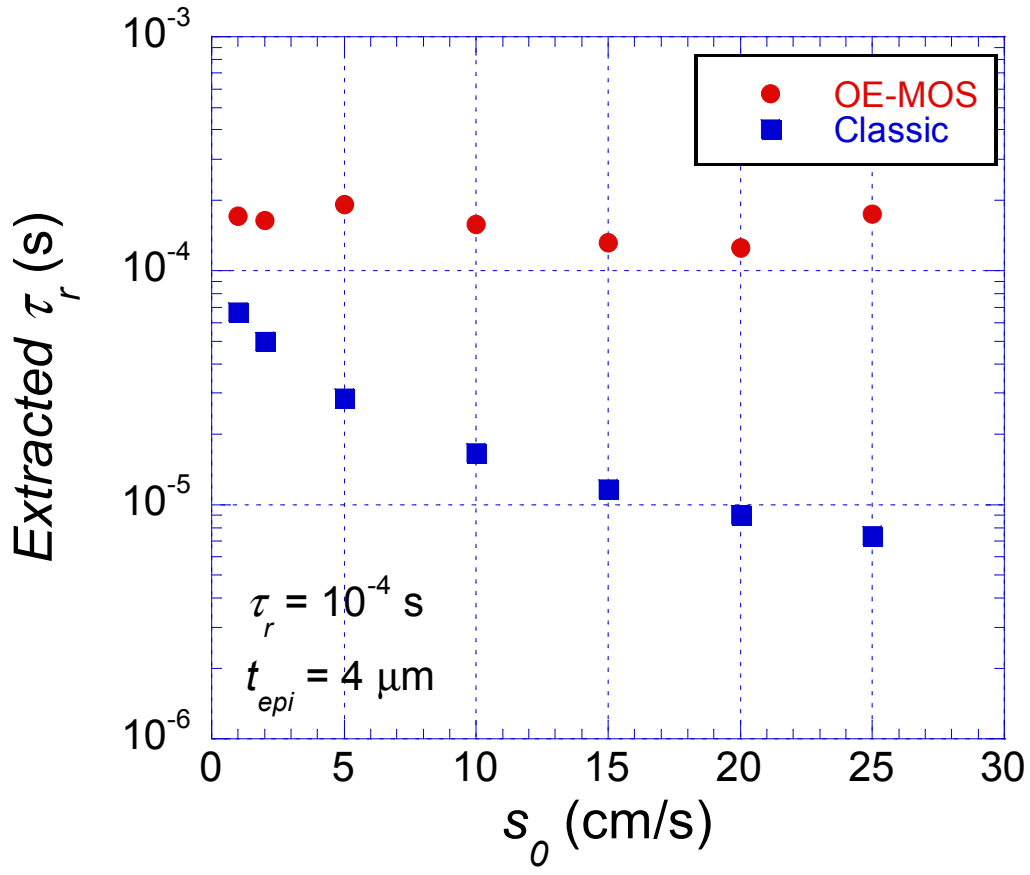

Figure 4.10. The effect of Si/SiO_2 surface recombination velocity (s_0) on the extracted τ_r parameter.

4.5. Conclusion

A new method was presented to characterize the recombination lifetime in p/p^+ silicon epitaxial layers. The technique is based on applying an optical excitation on an inverted MOS capacitor device and monitoring the device capacitance transition after the illumination is removed. The measured *C-t* plot is then exploited mathematically to extract the recombination lifetime parameter. The extracted values are shown to be accurate, both experimentally and with the aid of TCAD simulations. This approach can be a suitable substitute for other optical and electrical methods on silicon as it does not require any surface passivation or complicated test equipment. Most importantly, the effects of surface and interface recombination in masking the bulk lifetime were shown to be minimized compared to the conventional measurement techniques.

4.6 References

[4.1] D. K. Schroder, "Carrier lifetime in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[4.2] D.K. Schroder, "The concept of generation and recombination lifetimes in semiconductors," *IEEE Trans. Electron Dev.,* vol. ED-29, p. 1336, 1982.

[4.3] H. Hashizume, S. Sumie, and Y. Nakai, "Carrier lifetime measurements by microwave photoconductivity decay method," in *Recombination Lifetime Measurements in Silicon*, D. C. Gupta,, F. R. Bacher, and W. M. Hughes, Eds: ASTM, 1998, vol. STP 1340, pp. 47–58.

[4.4] J. Lagowski, P. Edelman, M. Dexter, and W. Henley, "Non-contact mapping of heavy metal contamination for silicon IC fabrication," *Semicond. Sci. Technol.*, vol. 7, pp. A185–A192, 1982.

[4.5] D. K. Schroder, B. D. Choi, S. G. Kang, W. Ohashi, K. Kitahara, G. Opposits, T. Pavelka, and J. Benton, "Silicon epitaxial layer recombination and generation lifetime characterization," *IEEE Trans. Electron Dev.*, vol. 50, pp. 906-912, Apr. 2003.

[4.6] Y. Hayamizu, R. Hoshi, Y. Kitagawara, T. Takenaka, "Novel evaluation methods of silicon epitaxial layer lifetimes by photoluminescence technique and surface charge analysis," *J. Electrochem. Soc. Jpn.*, vol. 63, pp. 505-519, 1995.

[4.7] J. E. Park, D. K. Schroder, S. E. Tan, B. D. Choi, M. Fletcher, A. Buczkowski, and F. Kirscht, "Silicon epitaxial layer lifetime characterization," *J. Electrochem. Soc.*, vol. 148, pp. G411-G419, Jul. 2001.

[4.8] B. Sopori, S. Devayajanam, O. Basnyat, H. Moutinho, B. Nemeth, V. LaSalvia, S. Johnston, J. Binns, J. Appel, "Some challenges in making accurate and reproducible measurements of minority carrier lifetime in high-quality Si wafers," *IEEE PVSC 2014*, pp. 0649-0654, 2014.

[4.9] A. Elhami Khorasani, T. L. Alford, D. K. Schroder, "Modified pulsed MOS capacitor for characterization of ultraclean thin p/p^+ silicon epitaxial layers," *IEEE Trans. Electron Dev.*, vol. 60, pp. 2592-2597, 2013.

[4.10] S. M. Sze, *Physics of semiconductor devices*, 3rd ed., Wiley-Interscience, Hoboken, N.J., 2007.

[4.11] G. S. May, C. J. Spanos, *Fundamentals of semiconductor manufacturing and process control*, Wiley, 2006.

[4.12] D. K. Schroder, *Semiconductor materials and device characterization*, 3rd ed., New York, NY: Wiley, 2006.

[4.13] Sentaurus Device - SYNOPSYS, User manual, ver. H-2013.03, SYNOPSYS, 2013.

[4.14] Sentaurus Mesh Generation Tools – SYNOPSYS, User manual, ver. D-2010.03, SYNOPSYS, 2010.

[4.15] Sentaurus Mesh Generation Tools – SYNOPSYS, User manual, ver. D-2010.03, SYNOPSYS, 2010.

[4.16] Sentaurus Technology Template: Light Triggered Thyristor – SYNOPSYS, 2007.

[4.17] J. Muller, and B. Schiek, "Transient responses of a pulsed MIS capacitor," *Solid-State Elec.*, vol. 13, pp. 1319-1332, 1970.

[4.18] J. S. Kang, and D. K. Schorder, "The pulsed MIS capacitor," Phys. Stat. Sol., vol. 89, pp. 13-43, 1985.

[4.19] K. S. Rabbani, and D. R. Lamb, "On the analysis of pulsed MOS capacitance measurement," *Solid State Elect.*, vol. 21, pp.1171-1173, 1978.

[4.20] D. K. Schroder, *Advanced MOS devices*, Addison-Wesley Longman Publishing Co., Inc., 1987.

[4.21] A. Elhami Khorasani, D. K. Schroder, T. L. Alford, "A fast technique to screen carrier generation lifetime using DLTS on MOS capacitors," IEEE Trans. Electron Dev., vol. 61, pp. 3282-3288, 2014.

CHAPTER 5

REVERSE RECOVERY MEASUREMENT AND LIFETIME STUDY ON THE BODY DIODE OF LD-MOSFET

5.1 Introduction

The body diode of an n-channel power MOSFET is comprised of a lightly n-doped drain (often referred to as the drift region) and a p-doped body. As demonstrated in Fig. 5.1, taking into account the highly doped n^+ drain, the MOSFET body diode resembles a p-i-n structure [5.1]. The drift region of a power MOSFET is fabricated through multiple ion implantations, followed by appropriate annealing to recover the implantation damage and to activate the implanted dopants [5.2]. The drift region has the pivotal role of enhancing the break-down voltage in the power device. Therefore, having the ability to monitor the damage created by ion implantation and the efficiency of anneal treatments are of great importance. The reverse recovery behavior of the body diode in power MOSFETs can be used as a mean to study the carrier lifetime in their drift region, and hence to study their fabrication quality [5.3]. In Laterally-Diffused MOSFET (LDMOS) devices, however, the correct interpretation of the reverse recovery behavior is necessary before the lifetime can be extracted from these measurements. Since there exists a lateral and a vertical p-n junction between the drift and body regions of the LDMOS, the stored charges will be removed/recombined with two different kinetics during the reverse recovery. This makes the lifetime parameter extracted using classic reverse recovery (that accommodates only the one-dimensional current flow as in a conventional planar p-n junction) inaccurate. To gain further insight into this issue, Synopsys mixed-mode TCAD has been employed to

study the device during the reverse recovery transient. Mixed-mode TCAD allows for the simulation of a circuit that combines 2D numerically simulated device structures along with other lumped elements (*e.g.*, voltage supplies, capacitors, or resistors) based on compact SPICE models [5.4]. Based on experimental data and simulation results, we show that by exploiting the reverse bias and gate voltage one not only can extract a correct lifetime without the need of complex mathematics, but can also distinguish between the local lifetimes at the two vertical and lateral junctions.

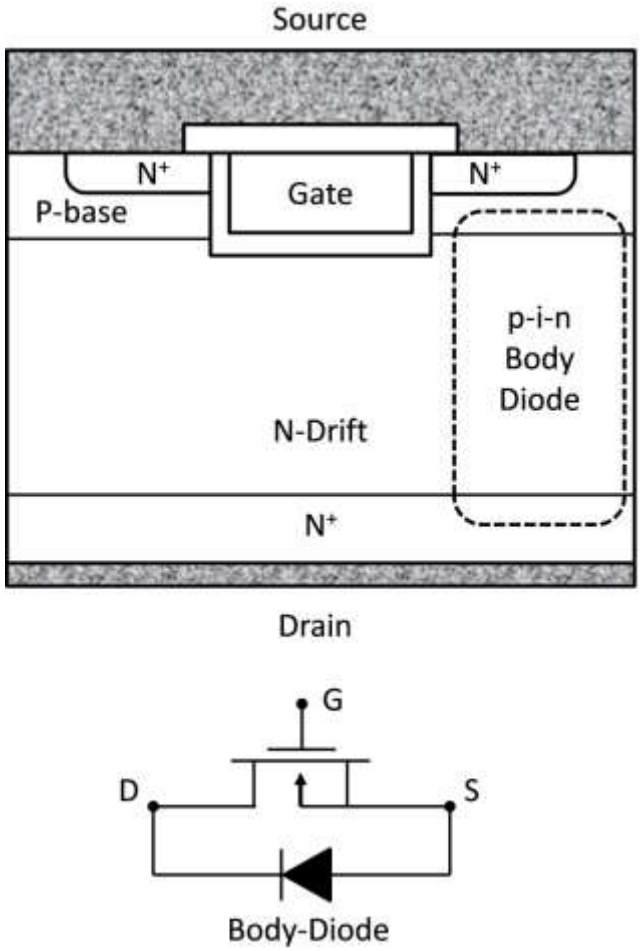

Figure 5.1. The U-MOSFET and its body-diode [5.1]

5.1.2 The Reverse Recovery Behavior

Considering the switching circuit shown in Fig. 5.2, one would expect the diode to start blocking the voltage as soon as the switch is closed and the bias is changed from forward (V_F) to the reverse (V_R). This, in reality, will never happen. Figure 5.3 shows the diode *I-V* characteristic that would be observed instead and in practice. Immediately after the switching, a large reverse current (I_R) will start to flow through the device. This is in contrary to the ideal diode theory where the diode's current is expected to become infinitesimal (*i.e.*, equal to the diode leakage current value) when a reverse bias is applied to its terminals. In real application, the device begins to act as a short (rather than an open) and does not show any of its blocking/rectifying properties when it is suddenly reverse biased after being forward biased. The I_R magnitude can become comparable (and even larger) than the diode's initial forward current (I_F). This phenomenon is called the diode's "*reverse recovery*". It not only can disturb the circuit performance, but in some cases, can destroy the device itself. In fact, what limits the magnitude of the reverse current passing through the diode is the circuitry (R_R) and not the device [5.5]. As it is shown in Fig. 5.2, the magnitude of I_R remains constant for a short period of time, to which we refer to as the "*storage time*" (t_s). After the storage time is passed, the I_R will gradually decay toward a value that corresponds to the diode's leakage current. The total time that it takes for the current to decrease to 10-25% of its maximum value I_R (depending on the standard that is used for the definition) is referred to as the "*reverse recovery time*" (t_{rr}). Figure 5.3 shows different times and their corresponding nomenclature during the reverse recovery behavior of a diode.

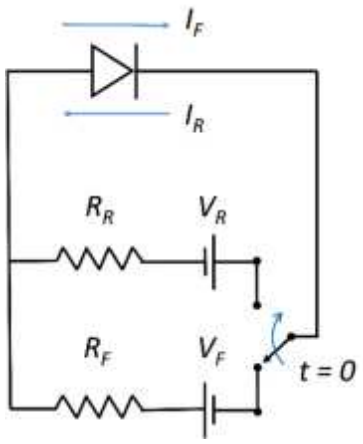

Figure 5.2. Idealized circuit for showing the reverse recovery of diode.

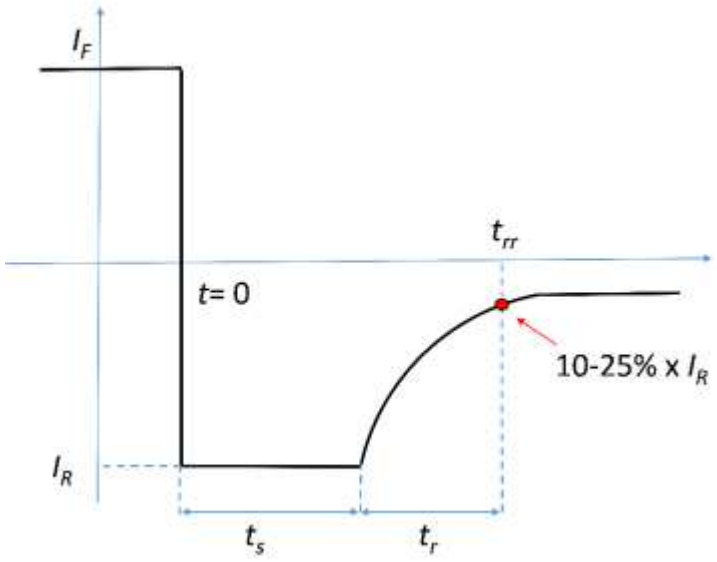

Figure 5.3. Idealized reverse recovery response of a diode when switched from forward to the reverse bias. t_s is the storage time, t_r the recovery time, and t_{rr} is the total reverse recovery time.

Aside from being a property that, by itself, is important for design purposes, the diode reveres recovery can be used to determine the carrier lifetime. In fact, the characterizing carrier lifetime with reverse recovery measurement has been one of the first approaches that were used for this purpose [5.6]. The lifetime determined using reverse recovery, however, is different from the one that is measured with the pulsed MOS in the sense that it corresponds to the carrier recombination (*i.e.*, it is the recombination lifetime τ_r). The lifetime measured using pulsed MOS technique, on the other hand, is the generation lifetime (τ_g).

The relationship between reverse recovery and lifetime was first mentioned by Kingston [5.7] and Kuno [5.8]. When the diode is forward biased, large numbers of majority carriers are injected across the p-n junction and to the opposite side. Once these majority carriers cross the junction, they become minority carriers and will be eliminated through recombination. This process will give rise to a pile up of minority carriers adjacent to the depletion region edges, as demonstrated in Fig. 5.4 for $t < 0$ when the diode is forward biased. In contrast, when the diode is reverse biased, the depletion region acts as a sink for the minority carriers.

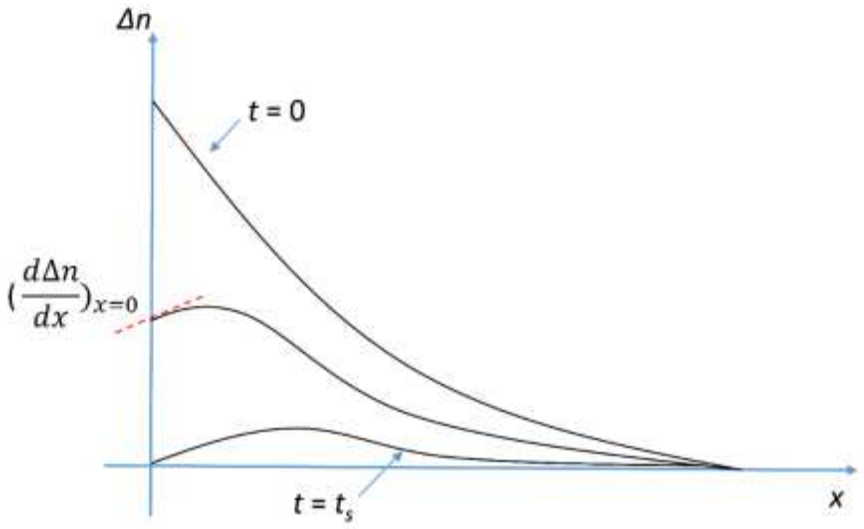

Figure 5.4. Minority carrier distribution next to the scr edge for $t = 0$ when the diode is forward bias and after been reverse biased till $t = t_s$ at the end of the storage time.

At $t = +0$ when the bias has been just switched from forward to the reverse, the minority carrier pile up next to the junction will make the device to act as a short. Before the blocking behavior of the diode can advent (and before the depletion region widen to support the reverse bias), these minority carriers need to be removed. This process will happen through two different mechanisms: First, the recombination; and second, the back injection of carriers in form of a net current. The minority carriers can get recombined through Shockley-Read-Hall mechanism. They can also be pulled out carrier current that for an ideal p^+/n junction is proportional to the slope of minority carriers profile at the depletion region edge (*i.e.*, $d\Delta n/dx|_{x=0}$). The actual magnitude of the reverse current is dictated by the R_R as envisioned in Fig 5.2 and following $I_R \approx V_R/R_R$.

Employing the charge equation for an ideal p^+/n junction, the following relation holds between the current density, the recombination lifetime (τ_r), and the minority carrier charge density (Q_p) in the n-doped side of the junction [5.9]:

$$\frac{dQ_p}{dt} = I(t) - \frac{Q_p}{\tau_r} \tag{5.1}$$

In the steady-state condition (*i.e.*, when $dQ/dt = 0$) and while the diode is being forward biased (*i.e.*, $I(t) = I_F$), (5.1) can be rewritten to give for the total stored charge density as:

$$Q_p = \tau_r . I_F \tag{5.2}$$

During the storage phase (*i.e.*, $0 < t < t_s$), on the other hand, the current is constant and equal to $-I_R$ and (5.1) can be rewritten as:

$$\frac{dQ_p}{dt} = -\left(\frac{Q_p}{\tau_p} + I_R\right) \tag{5.3}$$

Equation (5.3) can then be integrated to find the t_s.

$$\int_{Q_p(0)}^{Q_p(t_s)} \frac{dQ_p}{I_R + Q_p/\tau_r} = \int_0^{t_s} dt = t_s \tag{5.4}$$

The lower limit of the integration in (5.4) is already derived in (5.2). If the upper integration limit (*i.e.*, Q_p at t_s) in (5.4) is assumed to be 0, we would arrive at:

$$t_s = \tau_r \times \ln\left(1 + \frac{I_F}{I_R}\right) \tag{5.5}$$

Equation (5.5), although an approximation, is a key relationship that correlates the storage time with the lifetime in silicon. This relationship forms the basis of how one can characterize the lifetime parameter from reverse recovery measurements.

5.2 Experimental

5.2.1 Transmission Line Pulse (TLP) Technique

Transmission-Line-Pulse measurements has been first proposed in 1985 [5.10] as a tool to study the electrostatic discharge (ESD) phenomenon in the high current regime. Since then, the 100 ns pulse width TLP systems have become a standard method for many applications. Amongst many advantages of the pulsed mode measurements are the diminished self-heating effects (SHE), as well as the reduced risk of device destruction due to excessive energy dissipation. TLP systems are excellent means for generating high quality waves that can range anywhere from below 1 V to above several kilo volts [5.11]. As the pulse rise-times become smaller, however, impedance matched systems are needed for connecting the device to the measurement system in order to avoid the reflection problems.

TLP systems can be categorized based on their characteristic impedance (typically 50 and 100 Ω systems). The pulsed voltage source (V_0) and the characteristic impedance of the transmission line (Z_0) determines the maximum DUT voltage and the maximum short circuit current following:

$$V_{DUT,max} = V_0 \tag{5.6}$$

$$I_{DUT,max} = V_0/Z_0 \tag{5.7}$$

A TLP system, in its simple form, can be thought of as a voltage source being connected to two transmission lines that are separated via a switch, as pictured in Fig. 5.5 [5.12]. The propagation velocity of the wave inside a transmission line (v) is defined as:

$$\vartheta \approx c/\sqrt{\varepsilon_r} \tag{5.8}$$

where c is the speed of light and ε_r is the dielectric constant of the line's insulator. Putting a large resistor (R>>Z) before the first transmission line would create an open end to this line. When a wave hits the open end of a transmission line it will be reflected back. The reflected wave's magnitude depends on the reflection coefficient $\Gamma = (R\text{-}Z)/(R\text{+}Z)$. Given R is very large $\Gamma = 1$ and a voltage wave of the same polarity propagates through the line after hitting its open end adding up to the initial voltage wave. The transmission line is now charged at V_0 voltage.

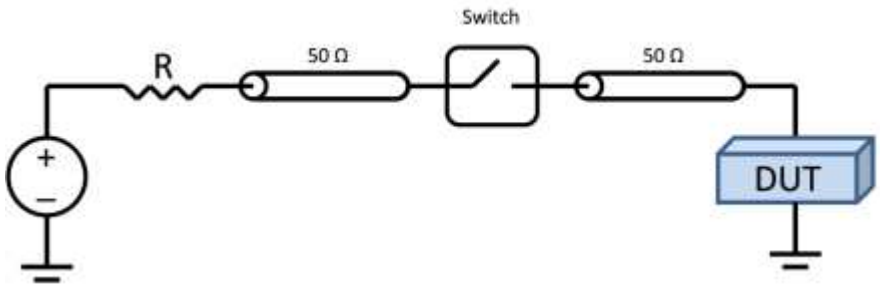

Figure 5.5. Basic structure of a TLP system

Here it will be beneficial to briefly mention how the pulse is being sourced through a TLP system like the one depicted in Fig. 5.5. The process is shown schematically in Fig. 5.6, with L being the length for each of the lines.

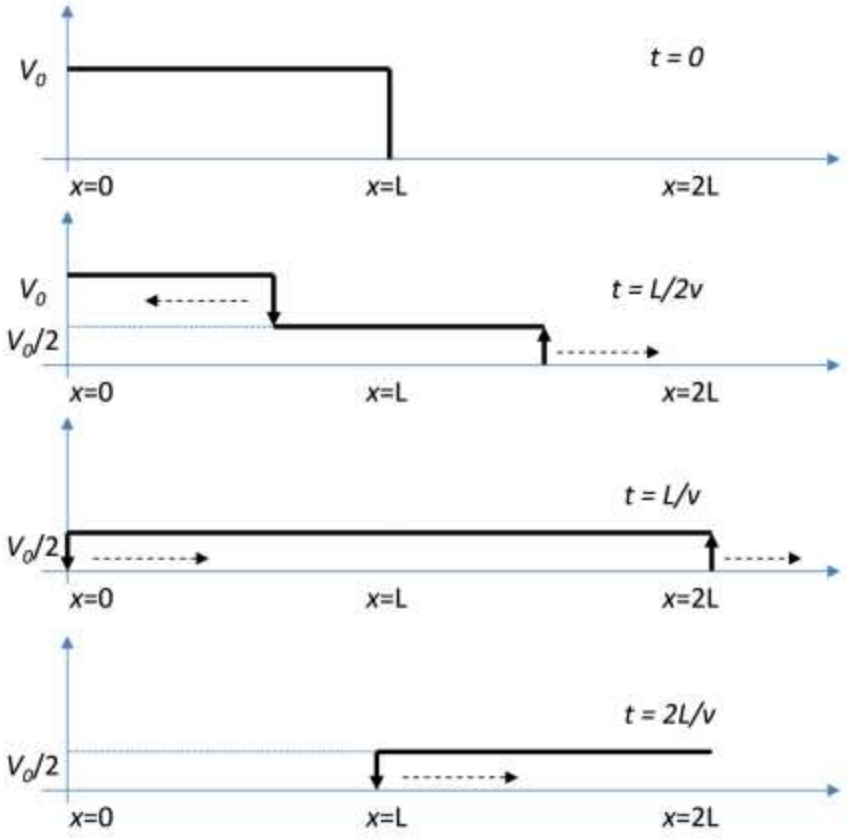

Figure 5.6. Schematic presentation of how the TLP system sources the pulse

At $t = 0$ and when the switch is open the first transmission line is charged by the voltage source to V_0. Immediately after the switch is opened the wave will start to propagate through the second transmission line. At this moment the voltage at the switch will drop down to $V_0/2$. This happens since now the two transmission lines are available to the voltage with the total new impedance being $2Z_0$. The propagation of the wave through the new system happens via two half waves, with amplitudes of $\pm V_0/2$ propagating in opposite

directions, as showed in Figure 5.6. Given the speed of the waves and length of lines, at $t = L/v$ the progressing half-wave will reach the open end of the transmission line and pass through the load (or DUT as depicted in Fig. 5.5). At the same time the recessing half-wave reaches the open end of the line and reflects back. As the reflected half-wave reflects back it will add up to initial wave and cancel it out. It will take $t= 3L/v$ for the cancelation wave to drain through the open end of the line. This whole process will generate a rectangular pulse with the amplitude of $V_0/2$ and the pulse width of $2L/v$. The pulse width (t_p) obtained this way then can be written in terms of:

$$t_p = \frac{2L}{\vartheta} \approx \sqrt{\varepsilon_r}\frac{2L}{c} \tag{5.9}$$

Therefore it can be seen that the pulse width in the TLP can be easily manipulated by modifying the transmission line's length. However, in practice there a maximum pulse width above which the line's dielectric loss would not be negligible anymore and the pulse would lose its uniformity.

In order to monitor the voltage and the current in the DUT using standard TLP systems, current and voltage probes are put as close as possible to the device. The current is often monitor via transformer based current probes; while, the voltage is monitor through voltage dividers and employing 4-point Kelvin measurement setup.

5.2.2 Experimental Setup

A routine way of measuring the reverse recovery behavior is the one mentioned by MIL-STD-750C Method 4031.5 Test Condition A [5.13] which is similar to ideal circuit mentioned in the introduction. The incorporation of this setup at the wafer level testing,

however, is not easily done [5.14]. Instead a Transmission-Line-Pulse (TLP) system has been used along with a wide-band bias-tee to perform the reverse recovery measurements. The block diagrams of this system, as well as the real experimental setup that was used are demonstrated in Figures 5.7 – 5.9. A 2651A Keithley source meter is used to forward bias the diode under tes. The AC-DC coupling is done by a wide-band bias Tee specifically designed for this purpose with a bandwidth of several GHz. The high voltage reverse bias pulses are then generated through a 3010 HPPI TLP system and overlap the DC forward bias driving the DUT into the blocking mode. The current is being monitored by Tektronix current probe CT-1 with a bandwidth of 2GHz on the force side, while the voltage is monitored on the sense side. This is the main difference between this method and that of the MIL-STD-750C standard. Care should be taken that the current probe is only capable of monitoring AC currents. This will result is a recorded current waveforms that needs to be corrected for $I = I_{leakage}$ at $t = t_{final}$.

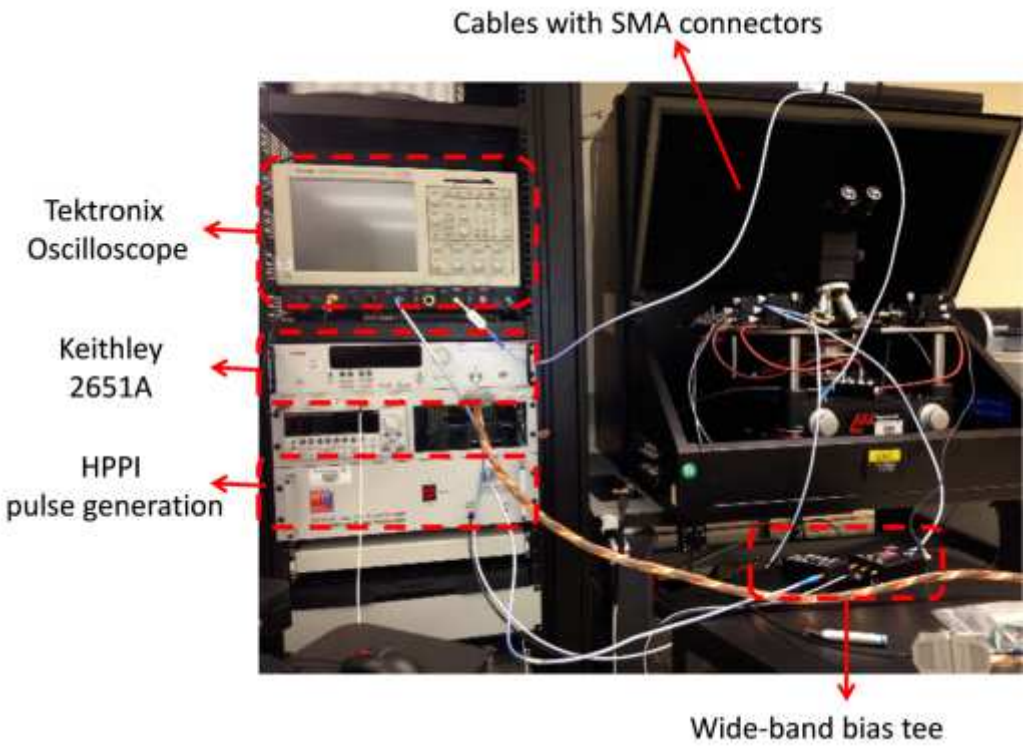

Figure 5.7. The TLP system used for the wafer level testing in this study. From top to bottom in the rack are: The Tektronix digital Oscilloscope, the Keithley 2651A and 2612, and the HPPI high power pulse generator.

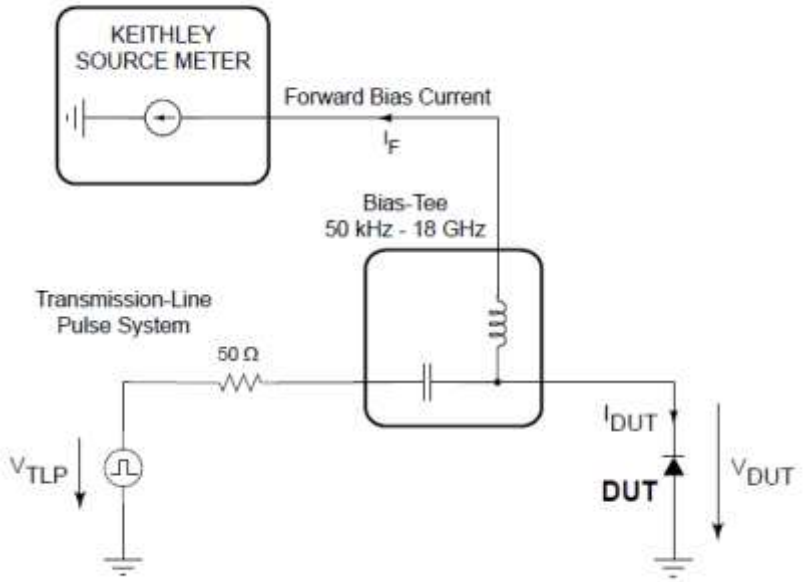

Figure 5.8. Block diagram of the reverse recovery setup

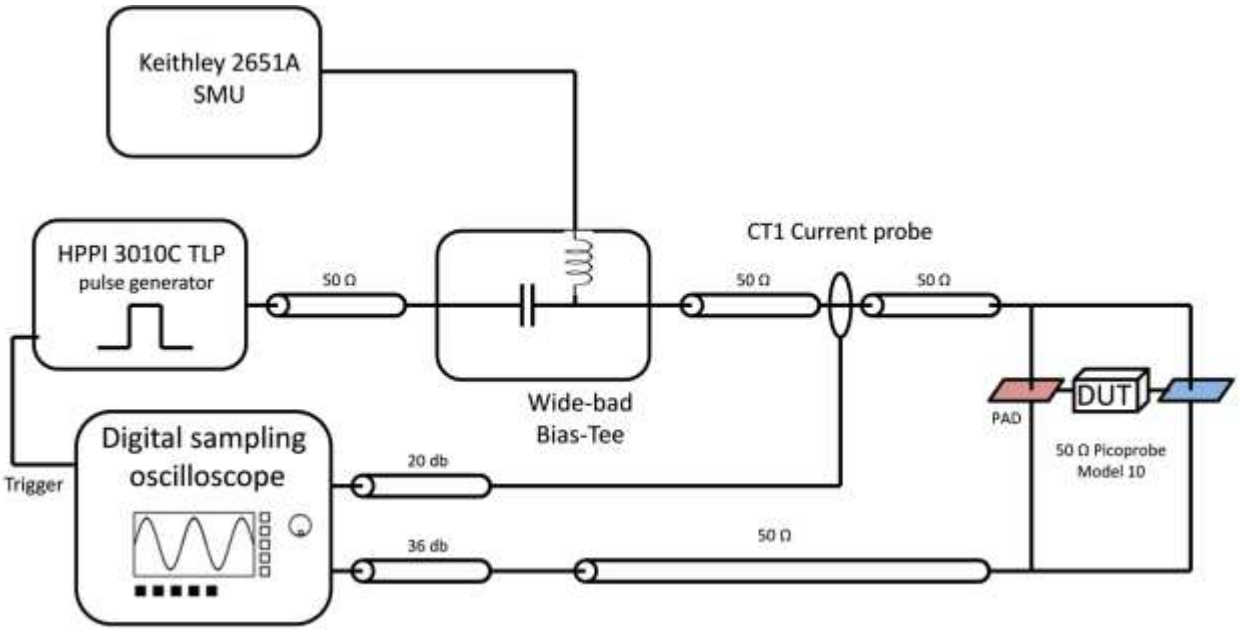

Figure 5.9. Detailed block diagram of 50 ohm reverse recovery TLP system

The calibration of the measurement system is performed using surface-mount devices with known parameters. Figure 5.10 shows the calibration kit that was used in this study. This kit is comprised of Zener diodes with known breakdown voltages, resistors with known resistances and short and open paths all of which were put in a Ground-Signal (GS) configuration. First the open path is used to calibrate for the shunt resistance of the measurement system. Since there is a 36 db attenuator in the system, as shown in Fig. 5.9, a shunt resistance in the order of 50 Ω is present in the system that needs to be accounted for. Then the Zener diodes are used to calibrate the voltage reading (voltage gain factor). The resistors are used in the end to calibrate for the current gain factor.

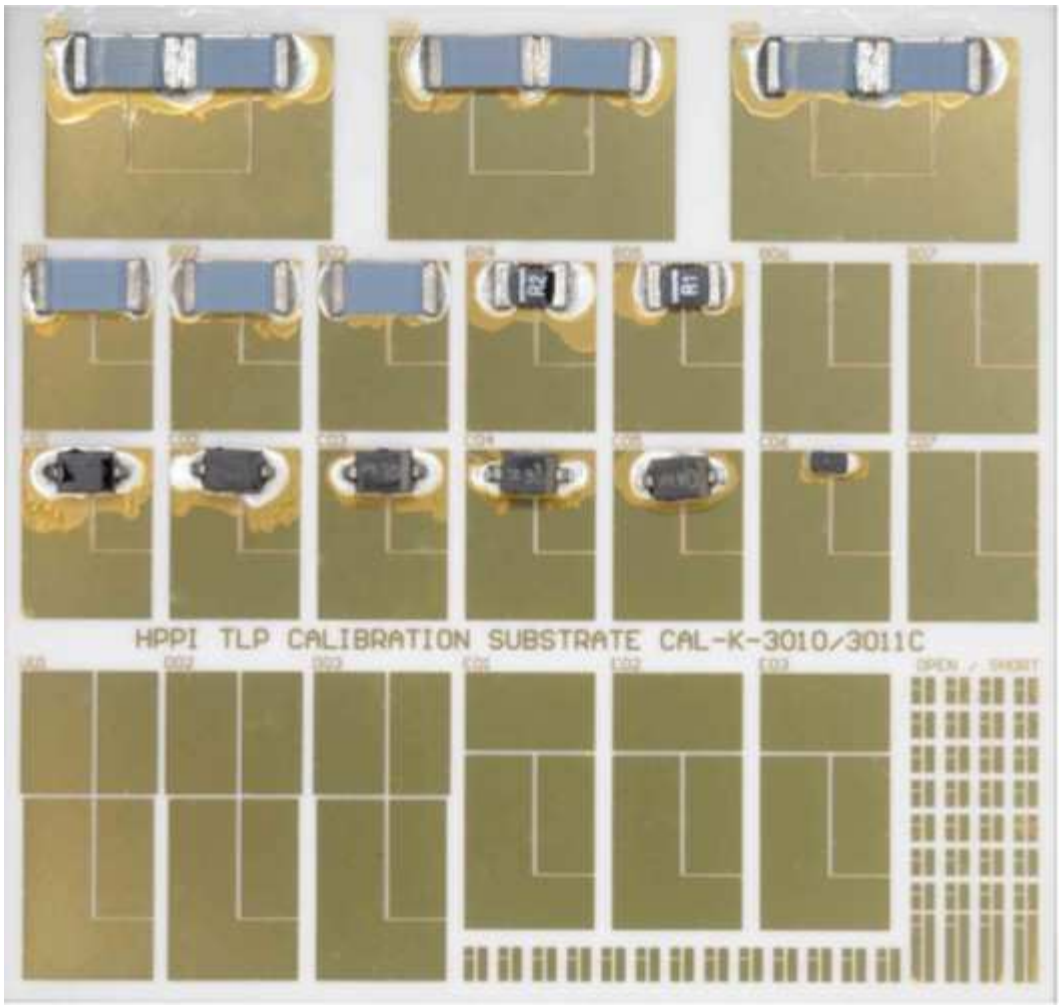

Figure 5.10. TLP Calibration Kit CAL-K-3010/3011C

The wafer-level connection to the DUT was done using 50Ω Pico-probe Model-10 with fixed pitch [5.15]. The probes were comprised of ground and signal (GS configuration) and came with preinstalled resistors or capacitors on the probe tips. Two 10-50/30-125-W-2-L-150 probes were used to force and sense on the LDMOS drain and the source, while a W-2uF probe was used to short its gate during the measurement. It was observed that if the gate was left floating unwanted stray capacitance could have been introduced into the measurements.

A key point when measuring current using the transformer based current probes with a TLP system is the distance between the probe and the DUT. In order to have the best reading, it is necessary to place to probe as close to the DUT as possible. Figure 5.11 shows the effect that the probe distance from DUT has on the measurements.

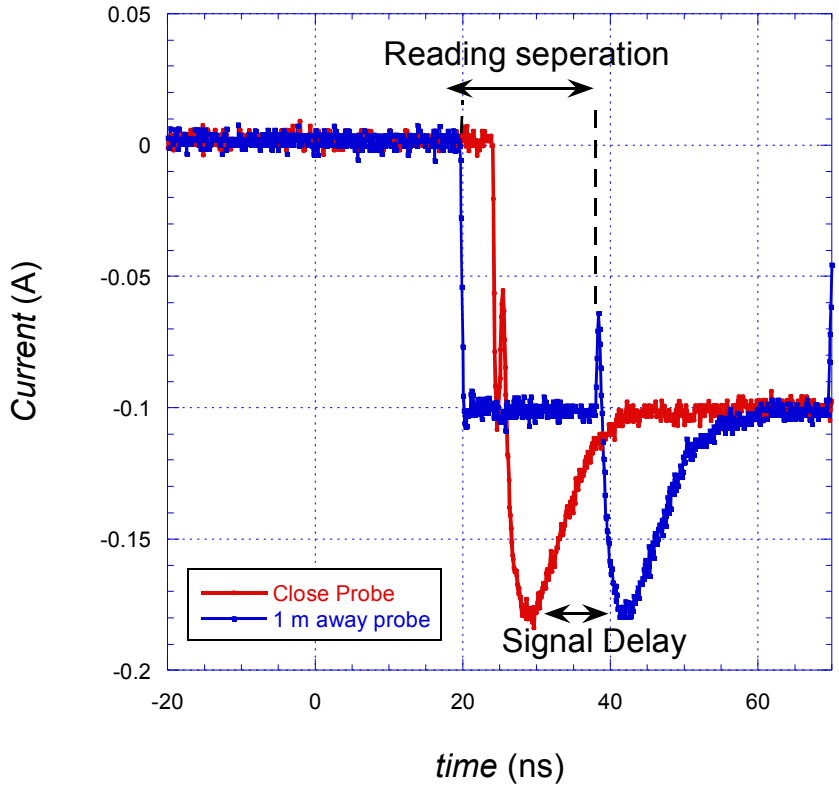

Figure 5.11. The effect of current probe distance from the DUT on the reading. It is necessary to place the probe as close as possible to the DUT.

The observed phenomena can be better explained using a numerical example. Let's assume that the current probe is put 1m away from the DUT. At $t = 0$ the propagating pulse signal with a velocity defined by (5.5) ~ 0.2 m/ns hits the current probe [5.16]. Therefore, the current sensor would show:

$$I(t)_{0<t<10\ ns} = \frac{V_{pulse}}{2 \times 50\Omega} \tag{5.10}$$

after 5 ns the propagating pulse signal reaches the DUT. Since the voltage is being read at the DUT, the voltage waveform will reach V_{DUT} immediately at this time (t = 5 ns). However, it will take another 5 ns for the reflection wave to reach back the current sensor. Hence the current sensor will read the correct current waveform starting at t = 10 ns. This will give rise the behavior being demonstrated in Fig. 5.11.

Figure 5.12 shows the importance of impedance matching. As demonstrated in this figure, in the presence of an impedance mismatching a spike will show up in the voltage waveform as it hits the board/DUT. Therefore, it is necessary to account for the impedance matching specifically when the pulse width are short (<50 ns) and the pulse rise-time can introduce error in the measurements.

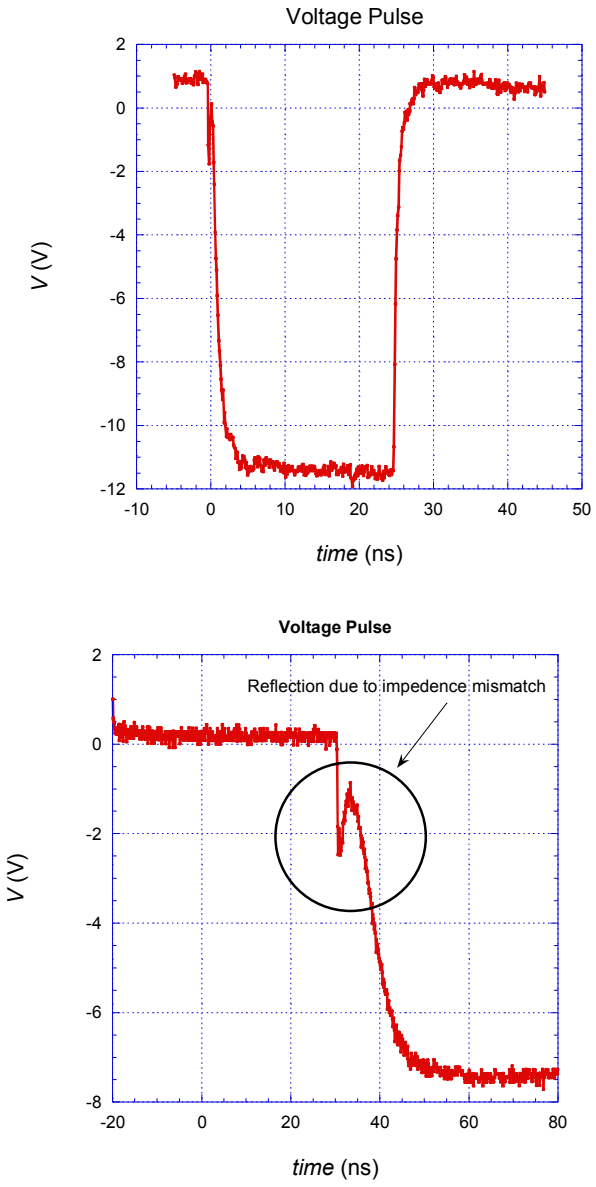

Figure 5.12. Right: Impedance mismatch and the reflection problem when the propagating wave hits the DUT. Left: Impedance matching achieved by using 50 ohm test fixture.

5.3 Theory, Results and Discussion

Let us review the basic equation once again. Taking into account the n-doped side of a planar p-n junction, and assuming the current to be due to the holes, the charge equation can be written as [5.8]:

$$I(t) = \frac{dQ_p}{dt} + \frac{Q_p}{\tau_r} \tag{5.11}$$

where $I(t)$ is the current at time t, Q_p is the charge density, and τ_r is the recombination lifetime. When at $t = 0$ the diode is switched from forward to the reverse bias, the stored charge density due to forward conduction, defined by (5.12), needs to be extracted before the junction can support the electric field and block the reverse bias. The time dependence of this process gives rise to the reverse recovery behavior of the diode. The time it takes for the stored charge at the junction to reduce to zero is referred to as the storage time (t_s). During the storage phase ($0 < t < t_s$) the current remains constant and equal to I_R. Since the stored charge is in part annihilated through recombination, the storage time of the diode is correlated to the carrier's lifetime via (5.13) [5.17]:

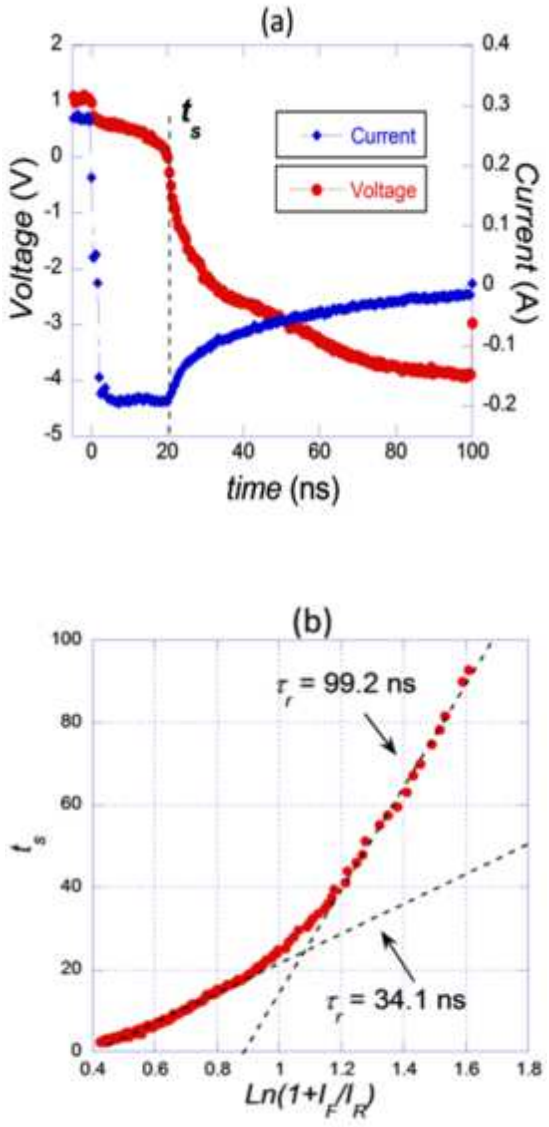

Figure 5.13. (a) Measured reverse recovery behavior of the LDMOS body diode. (b) Plot of t_s vs. ln(1+I_F/I_R) for the same device obtained by repeating measurements for different I_R values.

$$Q = I_F \tau_r \tag{5.12}$$

$$t_s = \tau_r[\ln(1 + I_F/I_R) - \ln(1 + Q_s/I_F\tau_r)] \quad (5.13)$$

with I_F being the forward current, and Q_s the remaining charge at the end of the storage phase.

Based on (5.13) the slope of the $\ln(1+I_F/I_R)$ vs. t_s plot yields τ_r. We will refer to this plot as the lifetime plot. Lifetime has been experimentally extracted this way for a variety of devices [5.1], [5.8], and [5.17].

Figure 5.13(a) shows the measured reverse recovery for the body-diode of a typical high-voltage LDMOS which is integrated into a 180 nm CMOS platform with a structure similar to Fig. 5.14(a). The switching and measurements are done at wafer-level employing an HPPI Transmission-Line-Pulse (TLP) system [5.14] to apply the reverse bias pulse, and a Keithley 2651A for forward biasing the structure. The coupling of the reverse bias pulse and the DC forward bias is done using a wide-band bias-tee, as explain in section 5.2.2. Drain and body contacts are used to bias the device while the gate was grounded. Figure 5.13(b) demonstrates the measured t_s vs. $\ln(1+I_F/I_R)$ for the same device obtained by repeating the reverse recovery measurements at different I_R values. In contrast to what the classic theory suggests, the lifetime plot for LDMOS device is no longer comprised of a single linear region. Instead, as shown in Fig. 13(b), this plot contains two separate linear parts with different slopes that both can be interpreted as separate lifetimes. This phenomenon has not been previously studied.

Such behavior can be addressed from two prospective. First, there are two vertical and lateral p-n junctions through which the stored charge can be extracted, as shown in Fig. 5.14(a). This specific design causes the reverse recovery of LDMOS devices to differ from

other power MOSFETs. And second, the lifetime at the locations corresponding to these two junctions are not

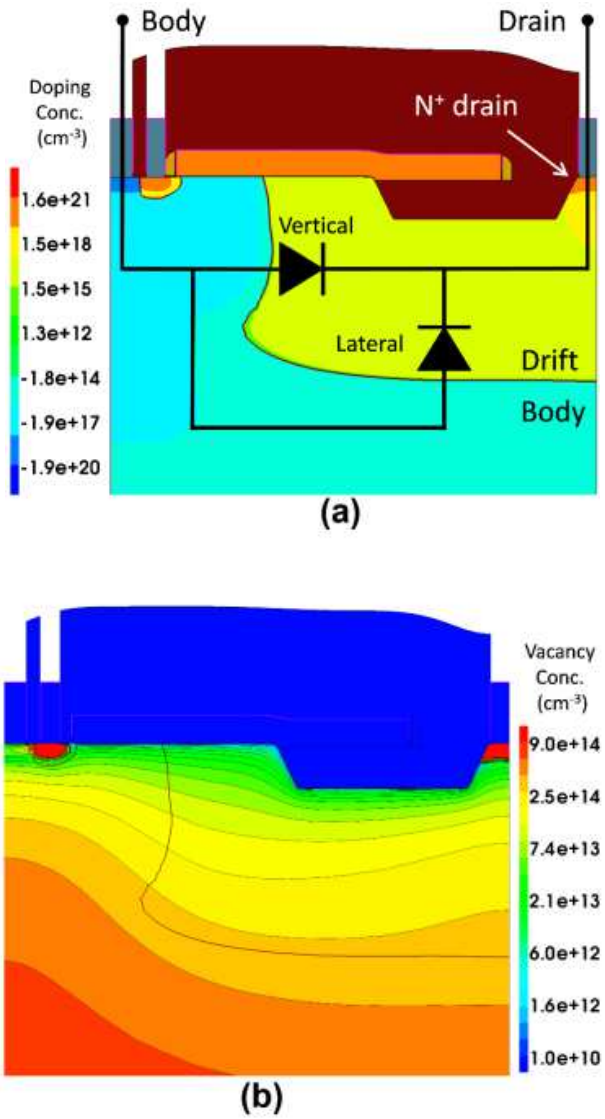

Figure 5.14. (a) The body diode of the LDMOS and the schematic diode equivalent circuit. (b) TCAD simulation of the remaining vacancy concentration in the LDMOS structure after RTA.

the same due to ion implantation. Figure 5.14(b) shows the simulated final vacancy concentration that exists in the silicon after ion implantation/annealing cycles. Since the density of deep-level centers (N_T) are linearly proportional to the radiation induced vacancy concentration [5.18], it is deemed that the lifetime at the lateral junction to be shorter than that of the vertical junction following $\tau_r \sim 1/N_T$ [5.19].

The interpretation of the lifetime plot in Fig. 5.14(b) can now be done if, to the first order approximation, we consider the two vertical and lateral p-n junctions as two diodes that operate in parallel to one another. Therefore, the measured reverse current (I_R) is the sum of the currents that pass through these two diodes, while the V_F and V_R biases are the same for them. In the beginning of the storage phase, as shown in mixed-mode TCAD transient simulations of Fig. 5.15, the current components for two diodes are comparable in magnitude but flow at two perpendicular directions. Hence, the measured reverse recovery is an effective transient corresponding to both diodes working together to drain the stored charges. This becomes problematic since the two diodes have different lifetimes and hence different storage times based on (5.13). Moreover, although the total reverse current I_R is defined by the circuitry (*i.e.,* $I_R=V_R/R_s$), each diode can have an individual reverse current. In other terms, at low I_R values when both diodes have a storage phase (t_s), the lifetime plot no longer corresponds to single (I_R/I_F, t_s) pairs and hence the extracted lifetime becomes a meaningless parameter unless complex mathematical analysis are invoked.

As the V_R is increased, the I_R gets larger than I_F in magnitude, and t_s for the vertical diode becomes 0 (*i.e.*, the vertical diode goes into an overdriven state [5.8]). This happens

since there is a maximum reverse bias (for a constant I_F) beyond which the stored charge can no longer afford the reverse current $I_R=V_R/R_s$ at the junction. As demonstrated in Fig. 5.15, the I_R for the vertical diode increases more rapidly than that for the lateral diode, forcing it into the overdriven state sooner. Therefore the measured storage time will become dominated by the t_s of the lateral diode

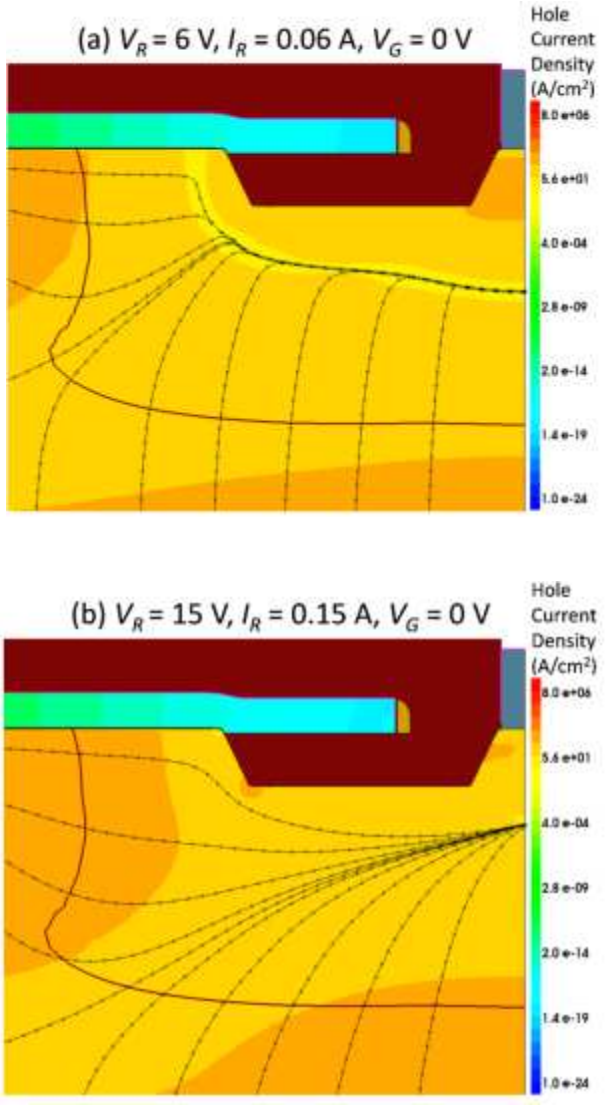

Figure 5.15. TCAD simulation of hole current density contour plots and hole current stream lines for two LDMOS with 6 and 15 V reverse bias. Both devices have spent same

time into their storage time phases and are pulsed from same V_F while the gate was grounded.

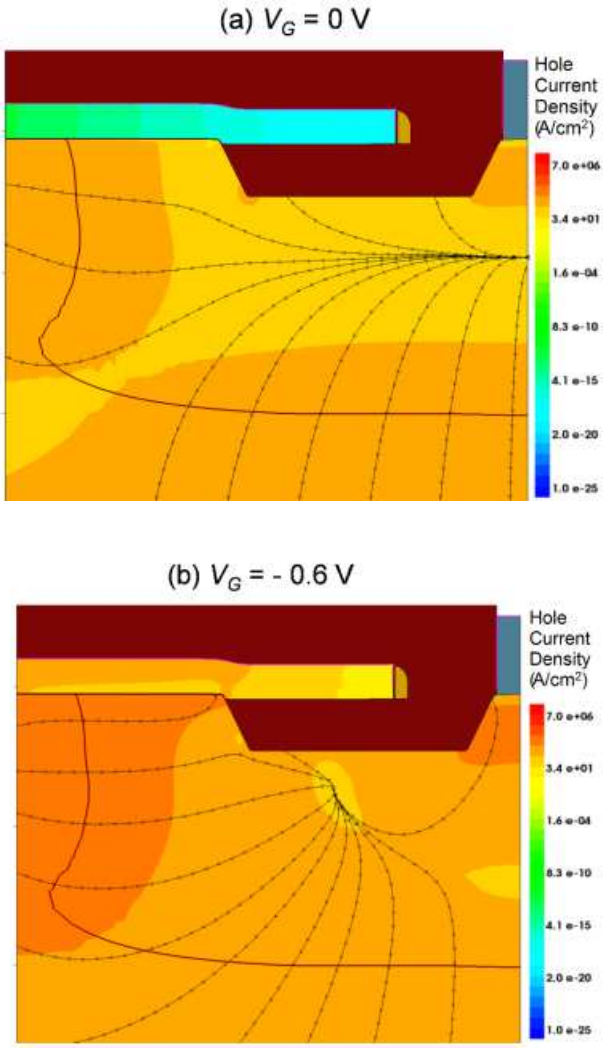

Figure 5.16. TCAD simulation of the hole current density contour plots and hole current stream lines for two LDMOS with 0 and -0.6 V gate biases. Both devices are at equal times during their storage phases and are switched from V_F = -1 V into V_R = 10.5 V.

as V_R increases. As a result of this, the lifetime extracted from high-I_R portion of the lifetime plot corresponds to the bottom of the drift region where the lateral diode is located. It can be seen from Fig. 5.13(b) that this lifetime is shorter in magnitude than the one extracted from the low-I_R region. This is in agreement with the damage profile of Fig. 5.14(b).

Since the current flow in LDMOS devices take place mostly in the lateral direction, the lifetime (*i.e.*, quality monitoring) at the vertical portion of the body-diode is also of interest. Till now the reverse recovery was studied for the case of zero gate bias. Gate bias can be employed as a mean to control the direction at which the current flows out of the drift region and hence to redefine the equivalent circuit of the body-diode. This happens mainly because gate bias can influence the stored charge distribution inside the drift region at t = 0. Since during the storage phase the current is mainly diffusional and controlled by $\mathrm{d}p/\mathrm{d}x$ at the junction, the contribution of the two diodes to reverse current changes under an applied gate bias. A negative gate voltage will act to pull those minority carriers stored deep in the drift region and supply them to the vertical junction. As a result of this, as the gate bias is decreased toward more negative values the reverse current will become dominated by the vertical diode. This makes the role of the lateral diode less pronounced.

The effect of changing the gate voltage on the reverse recovery waveform is demonstrated in Fig. 5.17(a). Figure 5.17(b) shows plots of t_s vs $\ln(1+I_F/I_R)$ corresponding to various V_G values. The two linear parts of this plot, which are observed at zero gate bias, gradually merge into a single straight line as the gate bias is decreased toward more negative values. This, as already stated, is an indication that the involvement of the lateral diode is decreasing till a point that we can neglect it from the analysis. In other terms, under

a negative gate bias one can consider the LDMOS body diode as a single planar p-n junction and apply the classic reverse recovery concepts to its measured reverse recovery. The TCAD simulations in Fig. 5.16 show how the minority carriers' current stream lines and current densities are significantly influenced by the presence of a negative gate bias. As demonstrated in Fig. 5.16(b), the slope of the lifetime graph also increases with decreasing

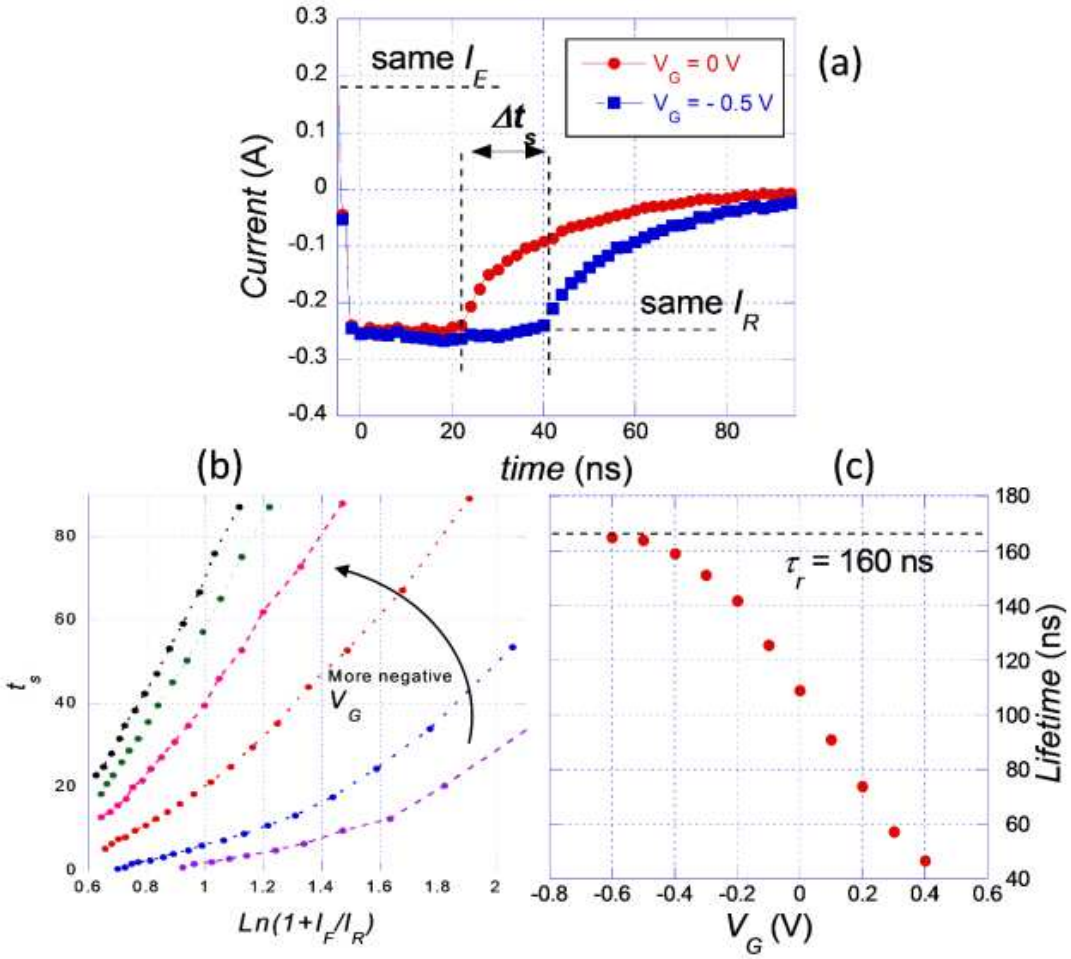

Figure 5.17. The effect of various gate voltages on: (a) the measured reverse recovery waveforms, and (b) on the t_s vs. $\ln(1+I_F/I_R)$ plot. (c) Extracted lifetimes as a function of

gate voltage. The measured lifetime values saturate as gate voltage is decrease toward more negative values.

the gate voltage, which is in agreement with the expectation that the lifetime is higher at the vertical junction. Figure 5.17(c) shows the lifetime values extracted from these measurements. It can be seen that the lifetime increases with the gate bias and begins to saturate. This saturation value can be interpreted as the true lifetime parameter that corresponds to the classic reverse-recovery behavior for the vertical p-n junction between the n-drift region and the p-body. The difference in measured lifetimes for lateral (34 ns) and vertical (160 ns) diodes follows our expectation that the lifetime in the vertical portion of the drift region is higher than the horizontal section due to the damage induced by ion implantation process.

5.5 Conclusion

In this chapter we discussed the reverse recovery of LDMOS devices. It was shown that by manipulating the reverse bias magnitude and gate voltage, the reverse recovery can be used to correctly characterize the body diode of the LDMOS device and to extract carrier lifetime in their drift region.

5.4. References

[5.1] J. Baliga, *Advanced power MOSFET concepts*, Springer, New York, 2010.

[5.2] B. O. Kolbesen, and H. Cerva, "Process-induced defects in silicon: A never ending story," *Electrochem. Soc. Proc*., 1999.

[5.3] B. J. Baliga, and J. P. Walden, "Improving the reverse recovery of power MOSFET integral diodes by electron irradiation," *Solid-State Elect*., vol. 26, pp. 1133-1141, 1983.

[5.4] Synopsys. 2013. Sentaurus Device User Guide. Synopsys Inc.,USA.

[5.5] R. F. Pierret, *Fundamentals of semiconductor devices*, Addison-Wesley Pub. Inc., 2000.

[5.6] D. K. Schroder, "Carrier lifetimes in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[5.7] R. H. Kingston, "Switching time in junction diodes and junction transistors," *Proc. IRE*, vol. 42, pp. 829-834, 1954.

[5.8] H. J. Kuno, "Analysis and characterization of PN junction diode switching," *IEEE Trans. Electron Dev.*, vol. ED11, pp. 8-14, 1964.

[5.9] Y. C. Kao, J. R. Davis, "Correlations between reverse recovery time and lifetime of p-n junction driven by a current ramp," *IEEE Trans. Electron Dev.*, vol. ED-17, pp. 652-657, 1970.

[5.10] T.J. Maloney, N. Khayrana, "Transmission line pulsing technique for circuit modeling of ESD phenomenon," *Proc. EOS/ESD Symp.*, pp. 49-51, 1985.

[5.11] D. Bradley, J. Higgins, M. Key, and S. Majumdar, "A simple laser-triggered spard gap for kilovolt pulses of accurately variable timing," *Opto-Electronics Letters*, vol. 1, pp. 62-64, 1969.

[5.12] W. Simburger, D. Johnsson, and M. Stecher, "High current TLP characterization: An effective tool for development of semiconductor devices and ESD protection solutions," *ARMMS RF & Micro. Soc.*, 2012.

[5.13] MIL-STD-750, www.dscc.dla.mil

[5.14] M. Sauter, W. Simburger, D. johnsson, M. Stecher, "On-wafer measurement of the reverse-recovery time of integrated diodes by transmission-line-pulse (TLP)," *Microelectronics Rel.*, vol. 51, pp. 1309-1314, 2011.

[5.15] http://ggb.com/10/html

[5.16] Advanced TLP/HMM Solution, "How to use pricoprobes and flexible pitch probes," Application Note AN-010, 2014.

[5.17] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., New York, NY, USA: Wiley, 2006.

[5.18] P. Hazdra, J. Vobecky, "Accurate simulation of fast ion irradiated power devices," *Solid-State Elec.*, vol. 37, pp. 127-134, 1994.

[5.19] A. Elhami Khorasani, D. K. Schroder, T. L. Alford, “Optically excited MOS-capacitor for recombination lifetime measurement,” *IEEE Electron Dev. Lett.*, vol. 35, no. 10, pp. 986-988, 2014.

CHAPTER 6

SUMMARY

6.1 Introduction

This chapter discusses several innovative electrical measurement techniques that were proposed for screening semiconductor device and material parameters. The methods were designed with two goals in mind: Short execution time, and accuracy. The economics of semiconductor industry necessitate the availability of such tools for swift investigation of process variations; a capability that enables continuous improvement throughout the technology development progress. The techniques span from oxide reliability to lifetime measurements using various approaches.

In order to be able to come up with a new measurement technique a combination of experimental work and theoretical semiconductor device physics is needed to be employed. TCAD simulations are also powerful tools for studying the more complex phenomenon such as the transient behavior of pulsed devices. These three aspects were had been all used together throughout this work.

6.2 Positive Charge Trapping and Its Application in Differentiating Thick Inter-Layer Dielectrics

Utilizing the positive charge generation/trapping kinetics, a novel fast screening *I-V* measurement technique is proposed that can distinguish between the intrinsic quality of ultra-thick oxides and their corresponding time dependent dielectric break-down (TDDB) time to failure (t_{BD}). We showed that positive charge trapping is a dominant process when

thick oxides are stressed through the ramped voltage test (RVT). Using multiple experimental approaches it was demonstrated that the positive charge trapping is and impact-ionization process and hence can be correlated with the band structure and the bonding states of the oxide under study. Such correlation was then used to predict and model the low electric field TDDB behavior of the same oxides using thermochemical *E* model. The theory was first developed and then confirmed by running constant-voltage stress (CVS) test. It was demonstrated that there is a good agreement between the quality assessment based on our proposed technique and the actual t_{BD} values that are calculated by running CVS-TDDB for three oxides of different qualities.

6.3 A Fast Technique to Screen Carrier Generation Lifetime Using DLTS on MOS Capacitors

It was shown that the DLTS technique can be used as a fast mean to screen carrier generation lifetime using the pulsed MOS-C measurement principals. The measured *C–t* saturation times were showed to be of good accuracy when compared with the modified pulsed-MOS technique for the same structures capable of distinguishing between impurity concentrations as low as 10^{10} cm^{-3}. It is worth noting that screening the generation lifetime with this technique can be done in minutes, while other techniques such as modified pulsed MOS-C may take up to several hours before accurate results are obtained. This advantage makes this technique suitable for screening purposes where the duration of such tests is of great importance.

6.4 Optically Excited MOS-Capacitor for Recombination Lifetime Measurement

A new method was presented to characterize the recombination lifetime in p/p^+ silicon epitaxial layers. The technique is based on applying an optical excitation on an inverted MOS capacitor device and monitoring the device capacitance transition after the illumination is removed. The measured *C-t* plot is then exploited mathematically to extract the recombination lifetime parameter. The extracted values are shown to be accurate, both experimentally and with the aid of TCAD simulations. This approach can be a suitable substitute for other optical and electrical methods on silicon as it does not require any surface passivation or complicated test equipment. Most importantly, the effects of surface and interface recombination in masking the bulk lifetime were shown to be minimized compared to the conventional measurement techniques.

6.5 Reverse Recovery Measurement And Lifetime Study on The Body Diode of LD-MOSFET

The power MOSFET devices are sometimes employed in circuits where the current is flowing through the structure in the third *I-V* quadrant (*e.g.*, the voltage regulator modules, and the H-bridge motor control circuits). A key advantage of these power MOSFETs are their inherent body-diode with rectifying behavior. Like any diode, however, the body diode of the power MOSFET shows reverse recovery behavior when switched from on to the off state. Such behavior results in low efficiency and power decapitations. One of the key goals of the power semiconductor industry has always been to increase the power efficiency of their products. This necessitates the detailed study of the reverse recovery behavior of the Laterally-Diffused-MOSFET as with large application in power circuits.

The transmission line pulse (TLP) technique was used to study the reverse recovery behavior of the body-diode of the LDMOS devices. Experimental results, as well as TCAD simulations were presented. It was shown that by exploiting the reverse bias magnitude and gate voltage, the reverse recovery can be used to correctly characterize the body diode of the LDMOS device and to extract carrier lifetime in their drift region. The extracted lifetime values and trend were showed to be in agreement with the expected values based on ion-implantation damage profiles.

REFERENCES

[1.1] H. Bentarzi, *Transport in metal-oxide-semiconductor structures: Mobile ions effect on the oxide properties*, Springer, 2011.

[1.2] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., Wiley-Interscience, New York, 2006.

[1.3] S. M. Sze, *Physics of semiconductor devices*, 3rd ed. Wiley-Interscience, New York, 2007.

[1.4] J. W. McPherson, "Time dependent dielectric breakdown physics – Models revisited," Microelectronics Reliability, vol 52, pp. 1753-1760, 2012.

[1.5] E. Rosenbaum, J. C. King, and C. Hu, "Accelerated testing of SiO_2 reliability," *IEEE Transactions on Reliability*, vol. 43, pp. 70-80, 1996.

[1.6] M. Shatzkes, and M. Av-Ron, "Impact ionization and positive charge in thin SiO_2 films," *Journal of Applied Physics*, vol. 47, pp. 3192- 3202, 1976.

[1.7] M. Shatzkes, M. Av-Ron, and R. M. Anderson, "On the nature of conduction and switching is SiO_2*," J. Appl. Phys.*, vol. 45, pp. 2065-2078, 1974.

[1.8] J. J. O'Dwyer, "Theory of high field conduction in a dielectric," *J. Appl. Phys.*, vol. 40, pp. 3887-3890,1969.

[1.9] Kashat, and N. Klein, "Current runaway in insulators affected by impact ionization and drift," *J. Appl. Phys.*, vol. 58, pp. 5217-5226, 1977.

[1.10] T. H. DiStefano, and M. Shatzkes, "Dielectric instability and breakdown in wide bandgap insulators," J. Vac. Sci. Technol., vol. 12, pp. 37-46, 1975.

[1.11] N. Klein, and P. Solomon," Current runaway in isulators affected by impact ionization and recombination," *J. Appl. Phys.,* vol. 47, pp. 4364-4372, 1976.

[1.12] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "High-field and current-induced positive charge in thermal SiO_2 layers," J. Appl. Phys., vol. 57, pp. 2830-2839, 1985.

[1.13] Aal, "Fast wafer level reliability assessment of ultra-thick oxides under impact ionization," *IEEE Trans. Dev. Mat. Rel.*, vol. 7, pp. 117-120, 2007.

[1.14] D. R. Young, E. A. Irene, D. J. Dimaria, R. F. Dekeersmaecker, and H. Z. Masoud, "Electron trapping in SiO_2 at 295 and 77 °K," *J. Appl. Phys.*, vol. 50, pp. 6366-6372, 1979.

[1.15] R. Bottini, "Oxide reliability: a new methodology for reliability evaluation at parametric testing," IEEE Proceeding on Integrated Reliability, pp. 142-145, 2006.

[1.16] U. Schwalke, M. Polzl, T. Sekinger, and M. Kerber, "Ultra-thick gate oxides: charge generation and its impact on reliability," *Microelectronics Reliability*, vol. 41, pp. 1007-1010, 2001.

[1.17] T. H. DiStefano, and M. Shatzkes, "Impact ionization model for dielectric instability and breakdown," *Appl. Phys. Lett.*, vol. 25, pp. 685-687, 1974.

[1.18] C. Hu, and Q. Lu, "A unified gate oxide reliability model," *IEEE 37th International Reliability Symposium*, pp. 47-51, 1999.

[1.19] J. W. McPherson, "Time dependent dielectric breakdown physics – Models revisited," Microelectronics Reliability, vol 52, pp. 1753-1760, 2012.

[1.20] Y. Chiang, D. P. Birnie, W. D. Kingery, *Physical Ceramics*, John Wiley & Sons, Inc., 1996.

[1.21] P. Lenahan, J. F. Conley, and B. D. Wallace, "A model of hole trapping in SiO_2 films on silicon," J. Appl. Phys., vol 81, pp. 6822-6824, 1997.

[1.22] R. Degraeve, J. L. Ogier, R. Bellens, P. J. Roussel, G. Groeseneken, and H. E. Maes, "A new model for the field dependence of intrinsic and extrinsic time-dependent dielectric breakdown," IEEE. Trans. Electron Dev., vol. 45, pp. 472-481.

[1.23] C. Kittel, *Introduction to solid-state physics*, 4^{th} ed., Wiley, New York, 1971.

[1.24] 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, pp. 1513-1523, 1998.

[1.25] K. V. Ravi, "Imperfections and impurities in semiconductor silicon," John Wiley & Sons, 1981.

[1.26] D. Fleetwood, "Defects in microelectronic materials and devices," CRC Press, 2009.

[1.27] D. K. Schroder, "Carrier lifetime in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[1.28] D. K. Schroder, B. D. Choi, S. G. Kang, W. Ohashi, K. Kitahara, G. Opposits, T. Pavelka, and J. Benton, "Silicon epitaxial layer recombination and generation lifetime characterization," *IEEE Trans. Electron Dev.,* vol. 50, pp. 906-912, Apr. 2003.

[1.29] D.K. Schroder, "The concept of generation and recombination lifetimes in semiconductors," *IEEE Trans. Electron Dev.,* vol. ED-29, p. 1336, 1982.

[1.30] D. K. Schroder, "effective lifetime in high quality silicon devices," *Solid-St. Electron.*, vol. 27, pp. 247-251, Mar. 1984.

[2.1] H. Bentarzi, *Transport in metal-oxide-semiconductor structures: Mobile ions effect on the oxide properties*, Springer, 2011.

[2.2] E. Rosenbaum, J. C. King, and C. Hu, "Accelerated testing of SiO_2 reliability," *IEEE Transactions on Reliability*, vol. 43, pp. 70-80, 1996.

[2.3] M. Shatzkes, and M. Av-Ron, "Impact ionization and positive charge in thin SiO_2 films," *Journal of Applied Physics*, vol. 47, pp. 3192- 3202, 1976.

[2.4] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "High-field and current-induced positive charge in thermal SiO_2 layers," J. Appl. Phys., vol. 57, pp. 2830-2839, 1985.

[2.5] A. Aal, "Fast wafer level reliability assessment of ultra-thick oxides under impact ionization," *IEEE Trans. Dev. Mat. Rel.*, vol. 7, pp. 117-120, 2007.

[2.6] T. H. Ning, "Thermal reemission of trapped electrons in SiO_2." *J. Appl. Phys.*, vol. 49, pp. 5997-6003, 1978.

[2.7] O. Gunnarsson, and J. E. Han, "The mean free path for electron conduction in metallic fullerenes," Nature, vol. 405, pp. 1027-1030, 2000.

[2.8] J. W. McPherson, 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, pp. 1513-1523, 1998.

[2.9] J. W. McPherson, "Time dependent dielectric breakdown physics – Models revisited," Microelectronics Reliability, vol. 52, pp. 1753-1760, 2012.

[2.10] N. Klein, and P. Solomon, "Current runaway in isulators affected by impact ionization and recombination," *J. Appl. Phys.,* vol. 47, pp. 4364-4372, 1976.

[2.11] U. Schwalke, M. Polzl, T. Sekinger, and M. Kerber, "Ultra-thick gate oxides: charge generation and its impact on reliability," *Microelectronics Reliability*, vol. 41, pp. 1007-1010, 2001.

[2.12] K. Hess*, Advanced theory of semiconductor devices*. Wiley-IEEE Press, 2000.

[2.13] JMP, Version10. SAS Institute Inc., Cary, NC, 1989-2012.

[2.14] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., Wiley-Interscience, New York, 2006.

[2.15] D. C. Montgomery, *Design and analysis of experiments*, 8th ed., John Wiley & Sons, Inc., 2013.

[2.16] D. M. Fleetwood, S. T. Pantelides, and R. D. Schrimpf, *Defects in microelectronic materials and devices*, CRC press, 2008.

[3.1] J. S. Kang, and D. K. Schroder, "The Pulsed MIS Capacitor," *Phys. Stat. Sol. (a)* vol. 89, pp. 13-43, May 1985.

[3.2] S. Lee, and D. K. Schroder, "Thin p/p^+epitaxial layer characterization with pulsed MOS capacitor," *Solid-State Elec.*, vol. 43, pp. 103-111, May 1998.

[3.3] D. K. Schroder, "Carrier lifetimes in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[3.4] S. M. Sze, *Physics of semiconductor devices*, 3rd ed., Wiley-Interscience, Hoboken, N.J., 2007.

[3.5] S. Y. Lee, "*Thin silicon layer characterization with the pulsed MOS capacitor,*" PhD Dissertation, Elect. Eng. Dep., Arizona State Unv., Tempe, 1998.

[3.6] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., John Wiley and Sons, 2006.

[3.7] D. S. Day, M. Y. Tsai, B. G. Streetman, and D. V. Lang, "Deep-leveltransient spectroscopy: System effects and data analysis," *J. Appl. Phys.*, vol. 50, no. 8, pp. 5093–5098, 1979.

[3.8] D. V. Lang, "Deep-level transient spectroscopy: A new method to characterize traps in semiconductors," *J. Appl. Phys.*, vol. 45, no. 7, pp. 3023–3032, Jul. 1974.

[3.9] M. Schultz and N. M. Johnson, "Transient capacitance measurements of hole emission from interface states in MOS structures," *Appl. Phys. Lett.*, vol. 31, no. 9, pp. 622–625, 1977.

[3.10] N. O. Pearce, B. Hamilton, A. R. Peaker, and R. A. Craven, "Application of deep level transient spectroscopy to metal-oxidesemiconductor relaxation transients," *J. Appl. Phys.*, vol. 62, no. 2, pp. 576–581, 1987.

[3.11] K. Yamasaki, M. Yoshida, and T. Sugano, "Deep level transient spectroscopy of bulk traps and interface states in Si MOS diodes," *Jpn. J. Appl. Phys.*, vol. 18, no. 1, pp. 113–122, Jan. 1979.

[3.12] A. Elhami Khorasani, T. L. Alford, D. K. Schroder, "Modified pulsed MOS capacitor for characterization of ultraclean thin p/p+ silicon epitaxial layers," IEEE Trans. Electron Dev., vol. 60, pp. 2592-2597, 2013.

[3.13] D. K. Schroder, *Modular Series on Solid State Devices: Advanced MOS Devices*. Reading, MA, USA: Addison-Wesley, 1987.

[3.14] F. D. Auret, "Considerations for capacitance DLTS measurements using a lock-in amplifier," *Rev. Sci. Instrum.*, vol. 57, no. 8, pp. 1597–1603, 1986.

[3.15] G. L. Miller, J. V. Ramirez, and D. A. H. Robinson, "A correlation method for semiconductor transient signal measurements," *J. Appl. Phys.*, vol. 46, no. 6, pp. 2638–2644, 1975.

[3.16] P. M. Henry, J. M. Meese, J. W. Farmer, and C. D. Lamp, "Frequency-scanned deep-level transient spectroscopy," *J. Appl. Phys.*, vol. 57, no. 2, pp. 628–630, 1985.

[3.17] K. S. Rabbani, "Investigations on field enhanced generation in semiconductors," *Solid-State Electron.*, vol. 30, no. 6, pp. 607–613, Jun. 1987.

[3.18] P. U. Calzolari, S. Graffi, and C. Morandi, "Field-enhanced carrier generation in MOS capacitors," *Solid-State Electron.*, vol. 17, no. 10, pp. 1001–1011, Oct. 1974.

[3.19] Khorasani, Arash Elhami. *Carrier Lifetime Measurement*. Diss. ARIZONA STATE UNIVERSITY, 2013.

[3.20] JMP, Version 10. SAS Institute Inc., Cary, NC, 1989-2012.

[3.21] A. Elhami Khorasani, D. K. Schroder, T. L. Alford, "A fast technique to screen carrier generation lifetime using DLTS on MOS capacitors," IEEE Trans. Electron Dev., vol. 61, no. 9, pp. 3282-3288, 2014.

[4.1] D. K. Schroder, "Carrier lifetime in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[4.2] D.K. Schroder, "The concept of generation and recombination lifetimes in semiconductors," *IEEE Trans. Electron Dev.,* vol. ED-29, p. 1336, 1982.

[4.3] H. Hashizume, S. Sumie, and Y. Nakai, "Carrier lifetime measurements by microwave photoconductivity decay method," in *Recombination Lifetime Measurements in Silicon*, D. C. Gupta,, F. R. Bacher, and W. M. Hughes, Eds: ASTM, 1998, vol. STP 1340, pp. 47–58.

[4.4] J. Lagowski, P. Edelman, M. Dexter, and W. Henley, "Non-contact mapping of heavy metal contamination for silicon IC fabrication," *Semicond. Sci. Technol.*, vol. 7, pp. A185–A192, 1982.

[4.5] D. K. Schroder, B. D. Choi, S. G. Kang, W. Ohashi, K. Kitahara, G. Opposits, T. Pavelka, and J. Benton, "Silicon epitaxial layer recombination and generation lifetime characterization," *IEEE Trans. Electron Dev.*, vol. 50, pp. 906-912, Apr. 2003.

[4.6] Y. Hayamizu, R. Hoshi, Y. Kitagawara, T. Takenaka, "Novel evaluation methods of silicon epitaxial layer lifetimes by photoluminescence technique and surface charge analysis," *J. Electrochem. Soc. Jpn.*, vol. 63, pp. 505-519, 1995.

[4.7] J. E. Park, D. K. Schroder, S. E. Tan, B. D. Choi, M. Fletcher, A. Buczkowski, and F. Kirscht, "Silicon epitaxial layer lifetime characterization," *J. Electrochem. Soc.*, vol. 148, pp. G411-G419, Jul. 2001.

[4.8] B. Sopori, S. Devayajanam, O. Basnyat, H. Moutinho, B. Nemeth, V. LaSalvia, S. Johnston, J. Binns, J. Appel, "Some challenges in making accurate and reproducible measurements of minority carrier lifetime in high-quality Si wafers," *IEEE PVSC 2014*, pp. 0649-0654, 2014.

[4.9] A. Elhami Khorasani, T. L. Alford, D. K. Schroder, "Modified pulsed MOS capacitor for characterization of ultraclean thin p/p^+ silicon epitaxial layers," *IEEE Trans. Electron Dev.*, vol. 60, pp. 2592-2597, 2013.

[4.10] S. M. Sze, *Physics of semiconductor devices*, 3rd ed., Wiley-Interscience, Hoboken, N.J., 2007.

[4.11] G. S. May, C. J. Spanos, *Fundamentals of semiconductor manufacturing and process control*, Wiley, 2006.

[4.12] D. K. Schroder, *Semiconductor materials and device characterization*, 3rd ed., New York, NY: Wiley, 2006.

[4.13] Sentaurus Device - SYNOPSYS, User manual, ver. H-2013.03, SYNOPSYS, 2013.

[4.14] Sentaurus Mesh Generation Tools – SYNOPSYS, User manual, ver. D-2010.03, SYNOPSYS, 2010.

[4.15] Sentaurus Mesh Generation Tools – SYNOPSYS, User manual, ver. D-2010.03, SYNOPSYS, 2010.

[4.16] Sentaurus Technology Template: Light Triggered Thyristor – SYNOPSYS, 2007.

[4.17] J. Muller, and B. Schiek, "Transient responses of a pulsed MIS capacitor," *Solid-State Elec.*, vol. 13, pp. 1319-1332, 1970.

[4.18] J. S. Kang, and D. K. Schorder, "The pulsed MIS capacitor," Phys. Stat. Sol., vol. 89, pp. 13-43, 1985.

[4.19] K. S. Rabbani, and D. R. Lamb, "On the analysis of pulsed MOS capacitance measurement," *Solid State Elect*., vol. 21, pp.1171-1173, 1978.

[4.20] D. K. Schroder, *Advanced MOS devices*, Addison-Wesley Longman Publishing Co., Inc., 1987.

[4.21] A. Elhami Khorasani, D. K. Schroder, T. L. Alford, "A fast technique to screen carrier generation lifetime using DLTS on MOS capacitors," IEEE Trans. Electron Dev., vol. 61, pp. 3282-3288, 2014.

[5.1] J. Baliga, *Advanced power MOSFET concepts*, Springer, New York, 2010.

[5.2] B. O. Kolbesen, and H. Cerva, "Process-induced defects in silicon: A never ending story," *Electrochem. Soc. Proc*., 1999.

[5.3] B. J. Baliga, and J. P. Walden, "Improving the reverse recovery of power MOSFET integral diodes by electron irradiation," *Solid-State Elect*., vol. 26, pp. 1133-1141, 1983.

[5.4] Synopsys. 2013. Sentaurus Device User Guide. Synopsys Inc.,USA.

[5.5] R. F. Pierret, *Fundamentals of semiconductor devices*, Addison-Wesley Pub. Inc., 2000.

[5.6] D. K. Schroder, "Carrier lifetimes in silicon," *IEEE Trans. Electron Dev.*, vol. 44, pp. 160-170, 1997.

[5.7] R. H. Kingston, "Switching time in junction diodes and junction transistors," *Proc. IRE*, vol. 42, pp. 829-834, 1954.

[5.8] H. J. Kuno, "Analysis and characterization of PN junction diode switching," *IEEE Trans. Electron Dev*., vol. ED11, pp. 8-14, 1964.

[5.9] Y. C. Kao, J. R. Davis, "Correlations between reverse recovery time and lifetime of p-n junction driven by a current ramp," *IEEE Trans. Electron Dev.*, vol. ED-17, pp. 652-657, 1970.

[5.10] T.J. Maloney, N. Khayrana, "Transmission line pulsing technique for circuit modeling of ESD phenomenon," *Proc. EOS/ESD Symp.*, pp. 49-51, 1985.

[5.11] D. Bradley, J. Higgins, M. Key, and S. Majumdar, "A simple laser-triggered spard gap for kilovolt pulses of accurately variable timing," *Opto-Electronics Letters*, vol. 1, pp. 62-64, 1969.

[5.12] W. Simburger, D. Johnsson, and M. Stecher, "High current TLP characterization: An effective tool for development of semiconductor devices and ESD protection solutions," *ARMMS RF & Micro. Soc.*, 2012.

[5.13] MIL-STD-750, www.dscc.dla.mil

[5.14] M. Sauter, W. Simburger, D. johnsson, M. Stecher, "On-wafer measurement of the reverse-recovery time of integrated diodes by transmission-line-pulse (TLP)," *Microelectronics Rel*., vol. 51, pp. 1309-1314, 2011.

[5.15] http://ggb.com/10/html

[5.16] Advanced TLP/HMM Solution, "How to use pricoprobes and flexible pitch probes," Application Note AN-010, 2014.

[5.17] D. K. Schroder, *Semiconductor material and device characterization*, 3rd ed., New York, NY, USA: Wiley, 2006.

[5.18] P. Hazdra, J. Vobecky, "Accurate simulation of fast ion irradiated power devices," *Solid-State Elec*., vol. 37, pp. 127-134, 1994.

[5.19] A. Elhami Khorasani, D. K. Schroder, T. L. Alford, "Optically excited MOS-capacitor for recombination lifetime measurement," *IEEE Electron Dev. Lett.*, vol. 35, no. 10, pp. 986-988, 2014.