

# Chapter 3

## Defects in semiconductors

### 3.1 Introduction

A perfect crystal is one in which the arrangement of the atoms within the crystal lattice is periodic, i.e. each atom occupies its designated lattice site and there are no missing atoms. However, in real life situations, it is difficult to obtain a perfect crystal. Imperfections in crystal lattices are due to several reasons. These include the growth technique and growth conditions. During crystal growth, some impurities can be unintentionally introduced into the material, thereby disturbing the orientation of the host atoms. Others can also be intentionally introduced during doping. Post-growth annealing of semiconductor material can also introduce or remove defects. Imperfections in semiconductor material can among others be due to missing atoms, impurities, dopants and others and also atoms occupying the wrong positions which disturb the crystal structure.

These imperfections, also referred to as “*defects*” may introduce electronic states in the bandgap of semiconductor material. The electronic states introduced by these defects can be classified as shallow and deep levels depending on their location relative to the conduction or valence band edge. If a defect transition level is positioned in such a way that it can be thermally ionized at room temperature or at the device operating temperature, it is referred to as a shallow level [1]. With particular reference to ZnO, shallow level defects include hydrogen, aluminium, group I related metals, just to mention a few, which are located just a few tens of milli-electron volts below the bottom of the conduction band minimum. Deep level defects include, the oxygen vacancy, Zn vacancy, which are said to be located within the middle third of the semiconductor band gap and are usually unionized at room temperature. Defects in semiconductors can be divided into two main categories, i.e. point defects and extended defects.

Defects in ZnO are often detrimental to device operation as they tend to modify the electrical and optical properties of devices. They also affect the realization of p-type doping and also the fabrication of UV light emitting devices due to their self-compensation behaviour. In some cases, defects can also be important, especially in the fabrication of devices with high switching speeds.

In this particular chapter, an outline of the defects that can be found in semiconductor material is given. This discussion is going to focus mainly on the type, structure and properties with special reference to n-type ZnO.

### **3.2 Point defects**

Many semiconductors possess large concentrations of defects in the form of point defects (intrinsic and extrinsic). As their name states, “point defects” cause some perturbation of the lattice within a localized area about a lattice site involving only a few nearest neighbours. The formation energy of point defects depends on the growth or annealing conditions [2,3]. Point defects are classified as intrinsic and extrinsic lattice defects. Intrinsic lattice defects are those that are caused by the missing of the host atoms (vacancies), interstitials, and host atoms occupying the wrong lattice sites (antisites). Extrinsic point defects are those due to impurity atoms occupying substitutional or interstitial sites. In II-VI compound semiconductors, the intrinsic lattice point defects of interest are vacancies and host interstitials [4] associated with each of the two sublattices. These point defects affect semiconductor material in many ways as they often control background doping, minority carrier lifetime and luminescence efficiency [1,5]. Point defects can cause device degradation as they also assist the diffusion mechanisms that are involved during growth and processing of the material. As much as doping is concerned, native point defects which act as donors in ZnO may self-compensate the introduced acceptors [6,7]. This poses a challenge in the realization of p-type ZnO. Native point defects have also been believed to play a major role in the high levels of unintentional n-type conductivity observed in ZnO [7]. An understanding of the nature and origins of these point defects in ZnO is essential. The native point defects in ZnO include the oxygen vacancy, zinc vacancy, oxygen antisites, zinc antisites, oxygen interstitials and zinc interstitials. Extrinsic point defects include hydrogen and other impurities introduced during growth e.g. group III related elements. The following sections are going to give a brief outline of some of the intrinsic point defects and hydrogen in ZnO.

### 3.2.1 The Oxygen vacancy

ZnO occurs naturally as n-type, whose conductivity has been attributed to native point defects particularly the oxygen vacancy and some other impurities which act as shallow donors [8, 9, 10,11,12,13]. However, the contribution to the n-type conductivity of ZnO by oxygen vacancies is still an issue under debate. First principles calculations by Janotti *et al.* [3] predicted that the oxygen vacancy  $V_O$ , if isolated is a deep donor and assumes the neutral charge state when the Fermi level is near the conduction band minimum, hence it is not responsible for the n-type conductivity in ZnO even though it has lowest formation energy among all the defects that behave as donors [3, 7]. The electronic structure of the oxygen vacancy comprises of four Zn dangling bonds ( $sp^3$  hybrids) and two electrons. The dangling bonds combine forming a fully symmetric state,  $a_1$  in the band gap and three almost degenerate states in the conduction band [7]. In the neutral state, the  $a_1$  is filled (doubly occupied) leaving the three states in the conduction band empty. Each of the four Zn dangling bonds, contribute 1/2 electron to form the neutral state of the oxygen vacancy. The oxygen vacancy has three possible charge states, i.e. the neutral ( $V_O^0$ ), +1 ( $V_O^+$ ) and +2 ( $V_O^{2+}$ ), with the  $V_O^{2+}$  and  $V_O^0$  being the stable states while the +1 charge is thermodynamically unstable [6]. The atomic structure of the oxygen vacancy in its three charge states is shown in Figure 3.1.

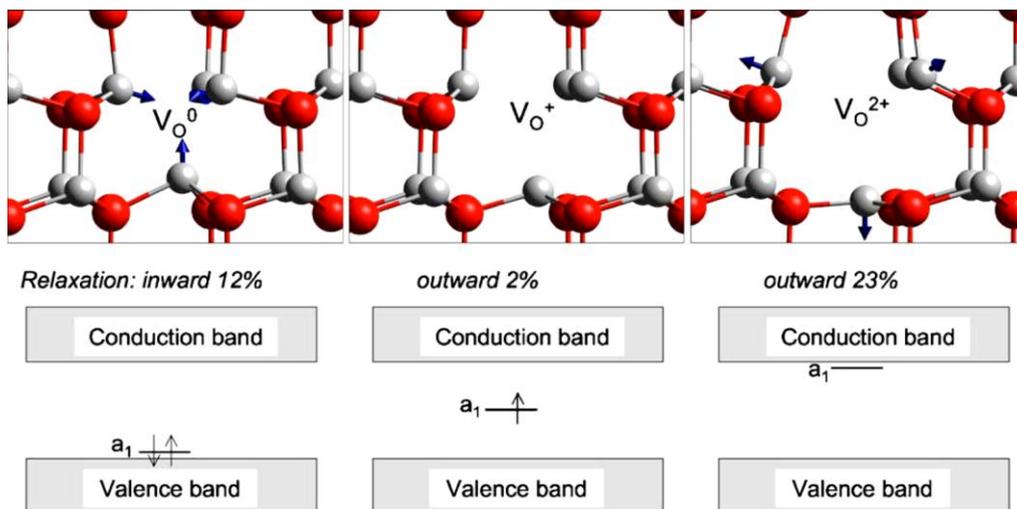


Figure. 3.1: The structure of the local atomic relaxations of the oxygen vacancy and the position of the  $a_1$  state. Redrawn from reference [3]

As indicated in Figure 3.1, the local lattice relaxation of the oxygen vacancy determines the occupation of the  $a_1$  state. In the neutral state, the four Zn atoms strongly relax towards the

vacancy by about 12% of the equilibrium Zn – O bond length. For the +1 state, the Zn atoms relax slightly outwards, while for the +2 charge state, the atoms relax further outwards. The large relaxation of the  $V_{\text{O}}^{2+}$  renders it low formation energy compared to the  $V_{\text{O}}^0$  and  $V_{\text{O}}^+$ . This makes the 2+ charge state to be the energetically the most preferable state for the O vacancy [14]. Also the relaxation difference between the  $V_{\text{O}}^0$  and  $V_{\text{O}}^+$  reduces the formation energy of  $V_{\text{O}}^0$  relative to the  $V_{\text{O}}^+$ . The relaxation differences in the charge states of the oxygen vacancy leads to the negative- $U$  behaviour [15,3], which sees the  $V_{\text{O}}^+$  being thermodynamically unstable.

Under p-type conditions, the Fermi level is near the top of the valence band. The oxygen vacancy is stable in the +2 charge state and has low formation energies; hence it acts as a compensation centre [7], affecting the formation of p-type ZnO. Since the formation energy of  $V_{\text{O}}$  is very low, at high temperature environments it can be easily formed [10] and is stable up to a temperature of 400°C [5]. It is therefore important to ensure that during growth or annealing of p-type ZnO, one must avoid incorporation of the  $V_{\text{O}}^{2+}$ . This can be achieved if the O-chemical potential approaches O-rich conditions and the Fermi level is kept away from the valence band minimum so as to suppress the formation of  $V_{\text{O}}^{2+}$  and reduce self-compensation effects.

### 3.2.2 The Zinc vacancy

Zn vacancies are double acceptors [6]. The acceptor behaviour of the Zn vacancies emanates from the fact that they are partially filled and hence can easily accept additional electrons. The electronic structure of a Zn vacancy is made up of dangling bonds from four oxygen atoms. These O dangling bonds combine to form a doubly charged symmetric state located deep in the valence band, and also introduce three partially occupied states in the bandgap of ZnO that lie very close to the valence band minimum. The Zn vacancy has three possible charge states, i.e.  $V_{\text{Zn}}^0$ ,  $V_{\text{Zn}}^-$  and  $V_{\text{Zn}}^{2-}$ . Zn vacancies can easily be formed in n-type ZnO since their formation energy decreases with increase in Fermi level [1,7], hence  $V_{\text{Zn}}^{2-}$  can occur in high concentrations, acting as compensation centres and is the dominant compensating centre in ZnO [16, 17] .

In p-type ZnO, the formation energy of Zn vacancies is very high and thus can be found in very low concentrations. Formation of the Zn vacancy favours oxygen-rich conditions [7].

### 3.2.3 The Zinc interstitial

The n-type conductivity of ZnO is attributed to the native defects among which the Zn interstitial is suggested to be a candidate. It is a shallow donor estimated to lie at about 30 meV below the minimum of the conduction band [9]. The zinc interstitial occupies the octahedral interstitial site of the wurtzite structure [18,6] since it is stable at this particular position. Stability of the  $Zn_i$  is brought about by the fact that, at the tetrahedral site, it is situated at a distance of approximately  $0.833d_0$  (where  $d_0$  is the Zn – O bond length along the c axis) and has one Zn and one O as the nearest neighbours, while in the octahedral site, it is at a distance of approximately  $1.07d_0$  and has three Zn and three O atoms as nearest neighbours [7]. Based on size considerations, the Zn interstitial is expected to be stable on the octahedral site because of less geometrical constraints. In ZnO, the  $Zn_i$  induces a state with two electrons above the conduction band minimum. The electrons are made available to the conduction band states forming the +2 charge state which is very stable. This behaviour makes the Zn interstitial a shallow donor. However under n-type conditions, the Zn interstitial is said to have high formation energy for Fermi level positions near the conduction band [6,14,1]. This property disqualifies it from being a most likely source of the n-type conductivity in n-type ZnO [14,1] since it exists in very low concentrations. Since the formation energy of the  $Zn_i^{2+}$  decreases when the Fermi level moves closer to the valence band maximum, the  $Zn_i$  becomes a likely source of compensation in p-type ZnO.

### 3.2.4 Negative- $U$ and Metastability

As has been discussed in section 3.2.1, the oxygen vacancy has three charge states, the +2, +1 and the neutral state, with the +2 state having a low formation energy compared to the other two, while the +1 is thermodynamically unstable, a negative- $U$  behaviour is predicted for the oxygen vacancy. This negative- $U$  behaviour is brought about by the fact that the unstable +1 state always has higher energy than the +2 and the neutral state [15,1] for any position of the Fermi level in the band gap. With increase in Fermi level position, a transition is observed to occur from the +2 to the neutral charge state [19], which are the stable states. For n-type material, two electrons of the neutral oxygen vacancy are transferred to the conduction band upon optical excitation, which triggers a large lattice relaxation of the defect resulting in the +2 charge state being stable at energies above the neutral charge state [20]. Another

contributing factor to the negative- $U$  behaviour of the oxygen vacancy is, the positively charged ion is readily ionized as opposed to the neutral state and the electron affinity for  $V_O^{2+}$  is high in energy than the  $V_O^+$  [15]. The negative- $U$  behaviour of the oxygen vacancy has been observed using optical techniques. The positive charge state of the oxygen vacancy is detected using electron paramagnetic resonance, EPR experiments [21,22] and also optically detected Magnetic resonance (ODMR) experiments [23] and Optically Detected Electron Paramagnetic Resonance (ODEPR) experiments [22,4]. This negative- $U$  behaviour of the oxygen vacancy results in defect metastability.

### **Defect Metastability**

This refers to defects that can occur in more than one structural configuration for at least one charge state. Metastable defects have structural transformations between atomic configurations under specific device operation or measurement conditions [24]. These transformations can be significant in a way since they can modify the role of a defect in the optical and electrical behaviour of a material, provided they are reproducible. Controllable metastability in semiconductor material is essential for the fabrication of memory devices [4]. An important aspect of the metastable defects is the reversible cycling that occurs among the different atomic configurations within the semiconductor with changes in device operation conditions, or measurement conditions [24]. The charge state, optical excitation, strain and temperature, just to mention a few, are some of the properties that determine the relative populations of the defect configuration.

Consider a defect with two lattice configurations. This defect configuration can be viewed as two total energy wells whose configuration space is separated by a potential barrier [24]. The stable configuration is the one with the lowest total energy and hence contains the largest population at equilibrium. A coordinate configuration diagram is shown in Figure 3.2.

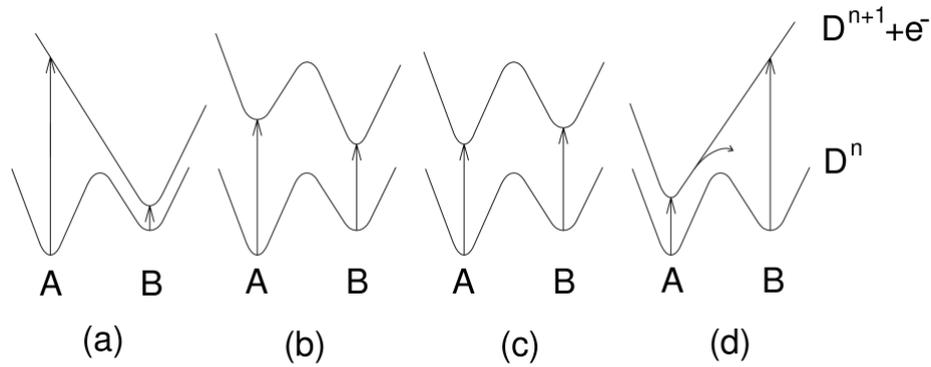


Figure 3.2: Coordinate configuration diagram for a metastable defect with two lattice configurations. Redrawn from reference [4].

As illustrated in Figure 3.2, the lattice configuration marked A for a defect which has a charge state,  $n$  is the stable lattice configuration while B is the metastable configuration. In (a) and (d), large lattice relaxation cases are observed, where the ionized state  $D^{n+1} + e^-$  has a local minimum only in one configuration. This is due to the fact that the large Stokes shifts are apparent, i.e. one of the configurations has very high optical ionization energies compared to the true ionization energy. 3.2 (b) and 3.2 (c) indicate ionized states which both have local minima, where the difference lies in which of A or B has the lowest energy configuration. Such cases are often referred to as small lattice relaxations since large Stokes shifts are absent.

In Figure 3.2(a) and 3.2(b), the stable configuration is not the same for the  $D^n$  and  $D^{n+1}$  charge states. A is the stable configuration in the  $D^n$  state while B is the stable configuration in the  $D^{n+1}$  charge state. This results in this particular defect being *bistable*. Its configuration can be altered by changing the charge state at a temperature at which the potential barriers can be overcome [4]. This is achieved by cooling the defect to below that temperature where it becomes frozen to a particular configuration where its properties can be determined. In Figure 3.2 (c) and (d), A is the stable configuration for both charge states. In a case where a single configuration is the only stable one for the two charge states, optical excitation or electrical injection can be used to establish thermal equilibrium distribution between the two near energy configurations in the ionized state.

DLTS can be used to detect the metastable configurations of a defect. Under equilibrium conditions, i.e. at non-zero temperature, the two configurations are occupied [24]. By measuring the DLTS peak heights, an estimate of the relative concentrations can be obtained.

Adjusting the bias conditions of the sample can set the charge states of the defects [25]. If the temperature of the sample is varied, by cooling from a high temperature under reverse bias, the position of the Fermi level will be altered, where any defect state below the Fermi level will be filled. The same defect can be set in a different state by cooling under zero bias. If a control of the measurement conditions for the defect configurations is successfully managed, the reaction mechanisms, potential barrier heights, electrical and optical properties of individual defect structures can be determined.

### 3.3 Extended defects

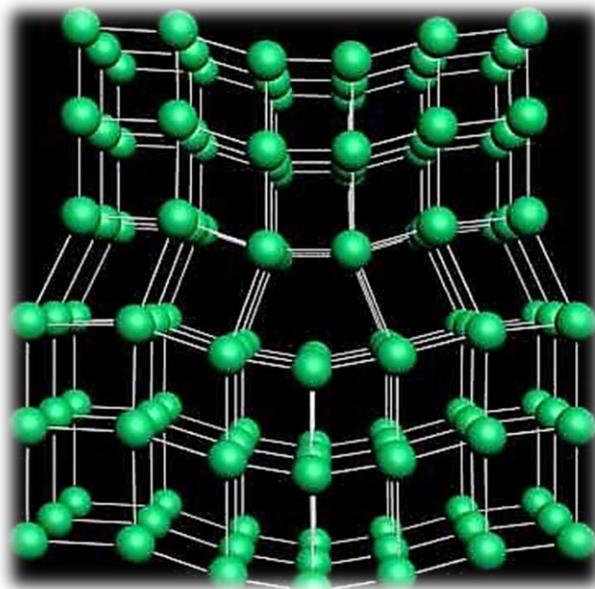
Extended defects in semiconductors are usually identified as volume defects, surface defects, and line defects. Volume defects are viewed to be any volume that differs from the rest of the crystal in structure, composition, orientation and/or state variables [26] e.g. precipitates, second phase grains, and also grains in polycrystalline material. Surface defects are also defined to be free surfaces or interfaces between distinguishable volumes. These are usually a cause of surface conduction in bulk ZnO samples, with a free carrier concentration of about  $(1 - 3) \times 10^{12} \text{cm}^{-2}$  [27, 28]. The effects of surface conduction are observed from the Hall effect measurements where the semiconductor can be treated as having two conducting channels [29, 30]. The first one being dominated by surface electrons whose effect is observed at low temperatures (freeze out region) while the other one is dominated by the bulk electrons, usually dominant at high temperatures, i.e. above the freeze out region. Surface conduction can be reduced by exposing the material to air due to possible adsorption of oxygen and/or changes in surface reconstruction [6]. Line defects are classified into different categories, namely *dislocations* and *disclinations* (also referred to as dislocations of rotation). Disclinations are usually observed in liquid crystals in which the rotational symmetry of crystals is violated.

### 3.3.1 Line defects

As has been mentioned in section 3.3, these types of defects fall into two different categories, namely *dislocations* and *disclinations*. They have three important characteristics [26] which are;

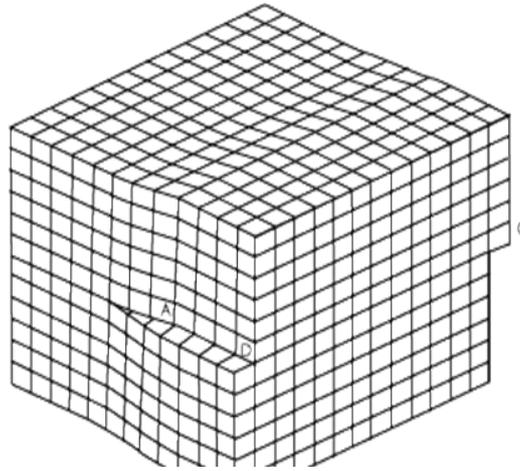
- a. The defect formation energy decreases as the number of dimensions in which the defect extends is reduced. These formation energies are also large and hence easy to avoid or eliminate.
- b. All of them are fundamentally related. Since all line defects are dislocations, a systematic treatment is possible because of their unity.
- c. Dislocations are capable of interacting with all other types of defects and/or can also assist in generating other types of defects.

Dislocations can be explained as linear defects in which some of the atoms of the crystal lattice are misaligned [31]. The two basic types of dislocations are *edge* dislocations and *screw* dislocations. If a plane of atoms in the middle of a crystal is terminated, edge dislocations are formed. In order to have a perfectly ordered crystal structure on either side, adjacent planes bend around the edge of the terminating plane.



*Figure 3.3: Lattice structure illustrating an edge dislocation within a semiconductor.*

Screw dislocations can be viewed as a structure consisting of a helical path that is traced around the linear dislocation by the atomic planes in the crystal lattice. Such a structure is indicated in Figure 3.4.



*Figure 3.4: Lattice structure indicating a screw dislocation*

Dislocations are negatively charged as has been observed from transmission electron microscopy by Muller *et al.* [32]. They affect the electrical properties of materials as they often contain many charges. In crystals, dislocations can also cause lattice distortions or strain and have also been proposed to account for the shear stresses above which single crystals of metals undergo permanent plastic deformation [26]. It has also been proposed that several luminescence peaks are related to extended defects which include dislocations [32]. In ZnO heteroepitaxial layers, threading dislocations with densities of approximately  $10^9 \text{ cm}^{-2}$  are usually observed [6]. These usually run along the *c*-axis in ZnO grown under O-rich conditions whereas for ZnO grown under Zn-rich conditions, it has been observed that the dislocations are inclined to  $20^\circ - 30^\circ$  from the *c* axis [34].

In semiconductors, dislocations can be observed using DLTS by monitoring the emission process of charge carriers. Charge carriers can build up in a line and gradually start repelling accumulation of additional charges. This in turn reduces the effective capture cross section as the amount of accumulated charge increases, causing the emission of carriers to follow a non-exponential process [35].

## References

1. A. Janotti and C. G. Van de Walle, Phys. Rev. Lett. B. **76**, 165202 (2007)
2. C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004)
3. A. Janotti and C. G. Van de Walle, Appl. Phys. Lett. **87**, 122102 (2005)
4. G. D. Watkins, Materials Science Forum, **38-41**, 39 (1989)
5. L. S Vlasenko and G. D. Watkins, Phys. Rev. Lett. B, **71**, 125210 (2005)
6. M. D. McCluskey and S. J. Jokela, Appl. Phys. Rev. **106**, 071101 (2009)
7. A. Janotti and C. G. Van de Walle, Rep. Prog. Phys. **72**, 126501 (2009)
8. Y. Natsume, H. Sakata, Mater. Chem. Phys. **78**, 170 (2002)
9. D.C. Look, J.W. Hemsky, J.R. Sizelove, Phys. Rev. Lett. **82**, 2552 (1999)
10. E. V. Lavrov, Physica B, **404**, 5079 (2009)
11. C. G. Van de Walle, Phys. Rev. Lett. **85**, 5 (2000)
12. F. A. Selim, M. H. Weber, D. Solodovnikov, and K. G. Lynn, Phys. Rev. Lett. **99**, 085502 (2007)
13. D. M. Hofmann *et al.*, Phys. Rev. Lett. **88**, 045504 (2002)
14. F. Oba, M. Choi, A. Togo and I. Tanaka, Sci. Technol. Adv. Mater. **12**, 034302 (2011)
15. C. H. Patterson, Phys. Rev. Lett. B, **74**, 144432 (2006)
16. F. Tuomitso, K. Saarinen, D. C. Look, G. C. Farlow, Phys. Rev. B. **72**, 085206 (2005)
17. F. Tuomitso, V. Ranki, K. Saarinen, D.C. Look, Phys. Rev. Lett. **91**, 205502 (2003)
18. P. Erhart and K. Albe, Appl. Phys. Lett. **88**, 201918 (2006)
19. Kohan, G. Ceder, D. Morgan, C.G. Van de Walle, Phys. Rev. Lett. B, **61**, 22 (2000)
20. D. Pfisterer, J. Sann, D. M. Hofmann, B. Meyer, T. Frank, G. Pensl, R. T. Zaera, J. Zuniga-Perez, C. M. Thomas, and V. Munoz-Sanjose, Phys. Stat. Sol. (c) 3, No. **4**, 997 (2006)
21. C.G. van de Walle, Physica B, **308**, 899 (2001)
22. L. S. Vlasenko, Appl. Magn. **39**, 103 (2010)
23. F. Leiter, H. Zhou, F. Hennecker, F. Hofstaetter, D. M. Hoffman, and B. K. Meyer, Physica B, **308**, 908 (2001)
24. L. J. Benton, Journal of Electronic Materials, **18**, 2 (1989)
25. L. J. Benton and M. Levinson, Defects in Semiconductors II eds. S. Mahajan and J. W. Corbett, (North Holland, New York, 1983) 95.
26. D. B. Holt, B. G. Yacobi, Extended defects in Semiconductors, Electronic properties, Device effects and structures (Cambridge University press 2007) ISBN 978-0-521-81934-3

27. O. Schmidt, P. Kiesel, C. G. Van de Walle, N. M. Johnson, J. Nause, and G. H. Döhler, *Jpn. J. Appl. Phys., Part 1* **44**, 7271 (2005)
28. D. C. Look, H. L. Mosbacker, Y. M. Strzhemechny, and L. J. Brillson, *Superlattices Microstruct.* **38**, 406 (2005)
29. D.C. Look, B. Claflin, H.E. Smith, *Appl. Phys. Lett.* **92**, 122108 (2008)
30. D.C. Look, *Superlattices Microstruct.* **42**, 284 (2007)
31. J. P. Hirth and J. Lothe). *Theory of dislocations*, 2<sup>nd</sup> ed. (Krieger Pub Co. 1992)  
ISBN 0894646176
32. E. Muller, D. Gerthsen, P. Bruckner, F. Scholz, Th. Gruber, A. Waag, *Phys. Rev. Lett. B*, **73**, 245316 (2006)
33. H. Alves, D. Pfisterer, A. Zeuner, T. Riemann, J. Christen, D. M. Hofmann, and B. K. Meyer, *Opt. Mater.* **23**, 33 (2003)
34. A. Setiawan, Z. Vashaei, M. W. Cho, T. Yao, H. Kato, M. Sano, K. Miyamoto, I. Yonenaga, H. J. Ko, *J. Appl. Phys.* **96**, 7 (2004)
35. D.C. Look, Z-Q. Fang, A. Krtshil, and A. Krost, *phys. stat. sol. (c)* 2, No. **3**, 1039– 1046 (2005)

# Chapter 4

## Defect characterization

### 4.1 Introduction

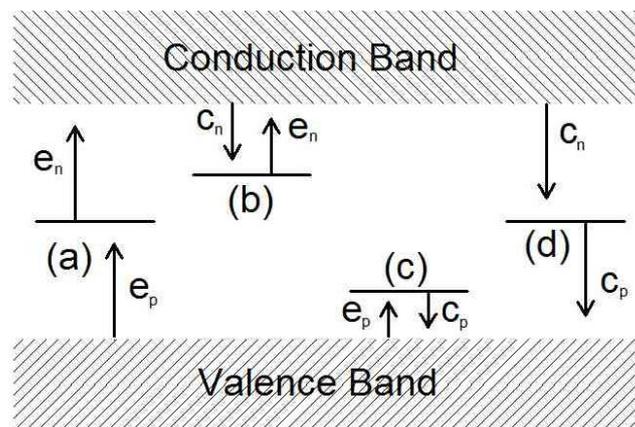
Defects in semiconductors can be detrimental to devices particularly in photovoltaic applications as they tend to reduce the minority carrier lifetime [1] and at times enhance the operation of devices by increasing the switching speeds by acting as recombination centres [2]. It is therefore essential to have a technique for characterizing defects in semiconductors. Electrically active defects in semiconductor crystals have different characteristics, depending on the location of their energy states relative to the conduction or valence band edge. These defects can be classified as shallow level defects and deep level defects. Shallow level defects are those whose energy levels are within a few tens of milli-electron volts from the respective band edges [3] which can be easily characterized by surface sensitive techniques, e.g. the Hall effect technique. Deep level defects typically reside within the middle two thirds of the semiconductor band gap. The energy required to remove an electron or hole from such a trap to the valence or conduction band is much larger than the characteristic energy,  $kT$ . This makes it more difficult to detect these defects with the Hall effect technique. Thus it is essential to have a technique to characterize deep level defects in semiconductors. Deep level transient spectroscopy DLTS is an ideal technique for characterizing electrically active deep level defects in semiconductors.

### 4.2 Emission and capture of carriers from deep levels

Existence of intrinsic defects in semiconductors poses a serious challenge to the semiconductor industry as defects tend to modify the electrical and optical properties and at times influence doping, e.g. in ZnO where it is still a challenge in obtaining p-type material.

Fabrication and development of efficient semiconductor devices requires a good knowledge of the identity, nature and properties of the deep levels that are native and also those that can be introduced in semiconductor material. Based on application, defects can affect device operation and at times can enhance the operation of devices.

The behaviour of these defects as either traps or recombination centres in devices is vital. High concentrations of deep level defects lead to the deterioration of solar cell parameters by increasing saturation current and decreasing the lifetime of the minority carriers and hence the efficiency [1]. On the other hand, deep level defects can improve the switching speeds of semiconductor switches. Deep level defects almost always change their electron/hole occupancy via carrier transitions between the levels and the bands [4]. A deep level can capture an electron from the conduction band. The electron can stay there until it is re-emitted back to the conduction band. Such a defect level is referred to as an electron trap.



*Figure 4.1: Schematic representation of carrier transitions between deep states and the valence- and conduction bands redrawn from ref [4]. Transfer between deep levels is neglected. The effective direction of electron transfer for both electron and hole processes is indicated by the arrows.*

Figure 4.1 illustrates the four most common processes for the transition of electrons and holes between the deep levels and the bands. The transition marked (a) indicates carrier generation, (b) shows electron trapping, (c) indicates hole trapping and (d) shows the recombination process. In the determination of defect state properties, it is usually electron trapping and hole trapping that are used to derive deep state properties [4]. Electron capture and emission most likely occurs for an empty deep level when the electron capture rate  $c_n$  from the conduction band is much larger than the hole capture rate  $c_p$  from the valence band, i.e.  $c_n \gg c_p$ .

The kinetics of charge transfer between the deep levels and the bands which are ultimately used to analyse deep level experimental data, are fully described by the Shockley-Read-Hall

(SRH) model [5, 6]. This model is developed by assuming thermal (or near thermal) equilibrium. Detection of deep level defects is experimentally performed by introducing perturbations to the carrier density or to the occupancy of the deep states involved and monitoring the return to equilibrium.

Let us consider an electron that is captured and emitted by a deep level. The electron capture rate is given by

$$c_n = \sigma_n \langle v_n \rangle n \quad (4.1)$$

where  $\sigma_n$  is the defect's electron capture cross-section,  $n$  is the electron concentration in the conduction band, and  $\langle v_n \rangle$  is the average thermal velocity of free electrons given by

$$\langle v_n \rangle = \sqrt{3kT/m^*} \quad (4.2)$$

where  $m^*$  is the electron effective mass,  $k$  is the Boltzmann constant and  $T$  is the temperature in Kelvin.

As a function of temperature, the electron emission rate is given by

$$e_n(T) = \sigma_n \langle v_n \rangle \frac{g_0}{g_1} N_C \exp\left(-\frac{E_C - E_t}{kT}\right) \quad (4.3)$$

where  $E_C - E_t$  is the energy separation of the deep state from the conduction band,  $g_0$  and  $g_1$  are the degeneracy terms referring to the state before and after electron emission, respectively.  $\sigma_n$  is the electron capture cross section which is assumed to be temperature independent in this case.  $N_C$  is the effective density of states in the conduction band given by,

$$N_C = 2M_C \left(\frac{2\pi m^* kT}{h^2}\right)^{3/2} \quad (4.4)$$

$M_C$  is the number of conduction band minima and  $h$  is Planck's constant.

From equation (4.4), the product  $\langle v_n \rangle N_C$  has a  $T^2$  dependence. Determining  $e_n$  as a function of temperature, an Arrhenius plot of  $\log(e_n/T^2)$  versus  $1/T$  must produce a straight line whose slope will give the energy difference  $E_C - E_t$  while the intercept (at  $1/T = 0$ ) gives the apparent capture cross section  $\sigma_{na}$ , assuming  $\sigma_{na}$  is temperature independent.

If we allow the capture cross-section to vary with temperature, it will take the form [7],

$$\sigma(T) = \sigma_\infty \exp\left(-\frac{\Delta E_\sigma}{kT}\right) \quad (4.5)$$

where  $\sigma_\infty$  is the capture cross-section extrapolated to  $T = \infty$  and  $\Delta E_\sigma$  is the thermal activation energy of the capture cross-section (i.e. the thermal barrier for carrier capture). The possible causes of the temperature dependence of the capture cross-section are the multiphonon capture into deep levels [7]. The temperature dependence of the capture cross-section can be determined from the plot of  $\log \sigma(T)$  versus  $1/T$  where  $\Delta E_\sigma$  can be obtained from the slope

and  $\sigma_\infty$  can be obtained after extrapolation to  $T = \infty$ . For a deep level defect whose capture cross section is temperature dependent, its activation energy  $E_a$  is identified with  $(E_C - E_t) + \Delta E_\sigma$ . The energy level of the trap is not directly obtained and also the identification only holds if  $(E_C - E_t)$  is temperature independent. Thus for a trap whose capture cross-section is temperature dependent, the thermal emission rate is given by,

$$e_n(T) = \sigma_n \langle v_n \rangle \frac{g_0}{g_1} N_C \exp\left(-\frac{(E_C - E_t) + \Delta E_\sigma}{kT}\right) \quad (4.6)$$

where  $\Delta E_\sigma$  is the thermal activation energy of the capture cross-section. Thus the thermal activation energy for emission of an electron to the conduction band is given by

$$\Delta E_a = (E_C - E_t) + \Delta E_\sigma \quad (4.7)$$

Equation 4.7 indicates that the thermal activation energy has two components: (1) the energy difference between the bottom of the conduction band and the trap level, and (2) the thermal activation energy of the capture cross-section, as illustrated in Figure 4.2.

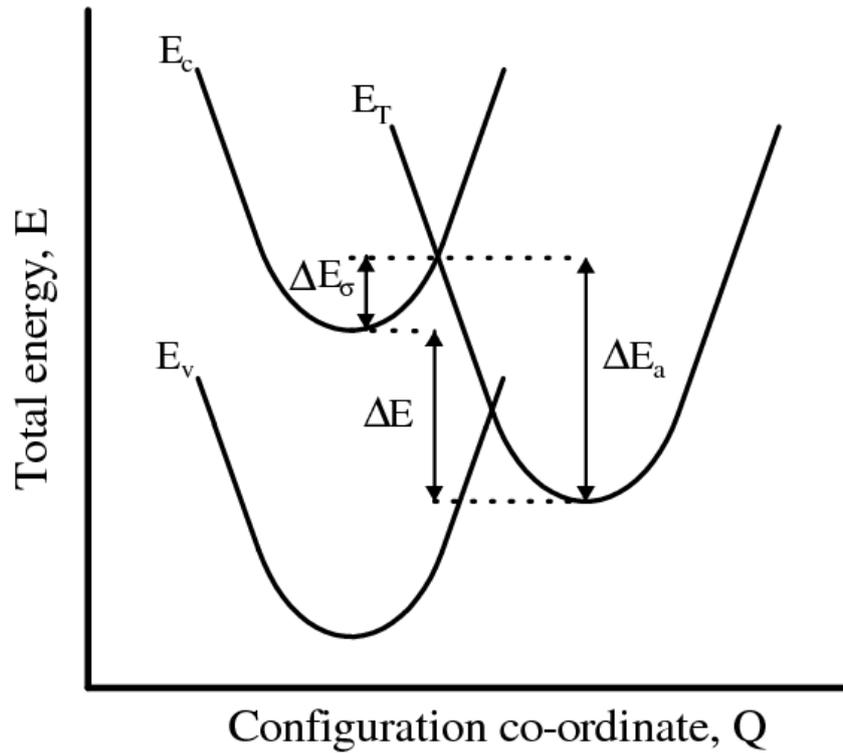


Figure 4.2: Co-ordinate configuration (CC) diagram depicting the energy level of a defect level below the conduction band,  $\Delta E (= E_C - E_t)$ , the thermal activation energy of the capture cross-section,  $\Delta E_\sigma$ , and the total energy an electron requires to escape from the trap level to the conduction band,  $\Delta E_a$ .

The commonly used parameter for characterizing deep levels is the activation energy for thermal emission [2]. The physical parameter  $E_T = E_C - E_t$  is the Gibbs free energy:

$$E_T = \Delta H - T\Delta S \quad (4.8)$$

where  $\Delta H$  and  $\Delta S$  are the changes in enthalpy and entropy due to the change in charge state of the level. Substituting equation (4.8) into (4.6) yields,

$$e_n(T) = \sigma_n \langle v_n \rangle \frac{g_0}{g_1} N_C \exp\left(\frac{\Delta S}{k}\right) \exp\left(-\frac{\Delta H}{kT}\right) \quad (4.9)$$

This implies that, the slope of the Arrhenius plot gives the enthalpy of the deep level, and not the free energy, which can only be determined from optical measurements [8,9]. Hence, one must be careful when comparing energies derived from thermal emission measurements with other methods [4].

### 4.3 Deep level transient spectroscopy, DLTS

DLTS is a high-frequency capacitance thermal scanning technique used for observing a wide variety of traps in semiconductors [10]. This is achieved by probing the depletion region of an ordinary p-n junction, a Schottky diode, or a MOS device structure. The measurement process uses the change in the capacitance transient associated with thermal emission of charge carriers from deep levels to thermal equilibrium after an initial non-equilibrium condition in the space charge region. This technique is capable of revealing the nature of a trap by either a positive peak (majority carrier trap) or a negative peak (minority carrier trap) on a flat baseline as a function of temperature. DLTS is very sensitive, rapid and easy to analyse [10]. It can uniquely determine the position of a peak in temperature using the thermal emission properties of the trap, which one can use to determine the activation energy and the capture cross-section of the trap. In addition, this technique has the ability to identify the concentration profiles of the deep levels.

In this section, the main focus of the discussion is based on how the technique is employed with particular reference to the depletion region of a Schottky diode.

Consider the width of the depletion region of a metal semiconductor contact discussed in Chapter 2, given by,

$$w = \sqrt{\frac{2\varepsilon_s(V_{bi} - V_a)}{qN_D}} \quad (4.10)$$

where  $\varepsilon_s$  is the semiconductor permittivity,  $V_{bi}$  is the built in potential of the junction,  $V_a$  is the externally applied voltage,  $q$  is the electronic charge, and  $N_D$  is the density of the ionized

impurities due to dopants and other defects with levels in the bandgap. The junction capacitance is now given by,

$$C = \frac{\epsilon_s A}{w} = A \sqrt{\frac{q \epsilon_s N_D}{2(V_{bi} - V_a)}} \quad (4.11)$$

where  $A$  is the cross-sectional area of the junction. From equation (4.10) and (4.11), altering the concentration of electrons or holes trapped at deep levels, by thermally or optically stimulated emission results in a change in the junction capacitance. The same change in the junction capacitance with the temperature dependence of  $N_D$  at a constant applied voltage forms the basis of capacitance-based DLTS [2].

During the formation of a metal/semiconductor contact, a region depleted of charge carriers is formed. This region is assumed to be formed within the semiconductor. According to the depletion width approximation, the semiconductor is now divided into two distinct regions, a layer that is entirely depleted of carriers (directly below the metal) and an interior region (bulk) of perfect charge neutrality, and in which no electric field exists. Assume that the semiconductor contains a low concentration of defects that cause a deep electron trap  $E_T$  shown in Figure 4.3. This deep level is assumed to be empty in the depletion region and filled in the bulk of the semiconductor. There exist a region  $\lambda$  within the depletion region that extends from the bulk (depletion region edge), where the deep level intersects the Fermi level. In this region, the deep level is also filled under equilibrium and zero bias conditions, since it also lies below the Fermi level.

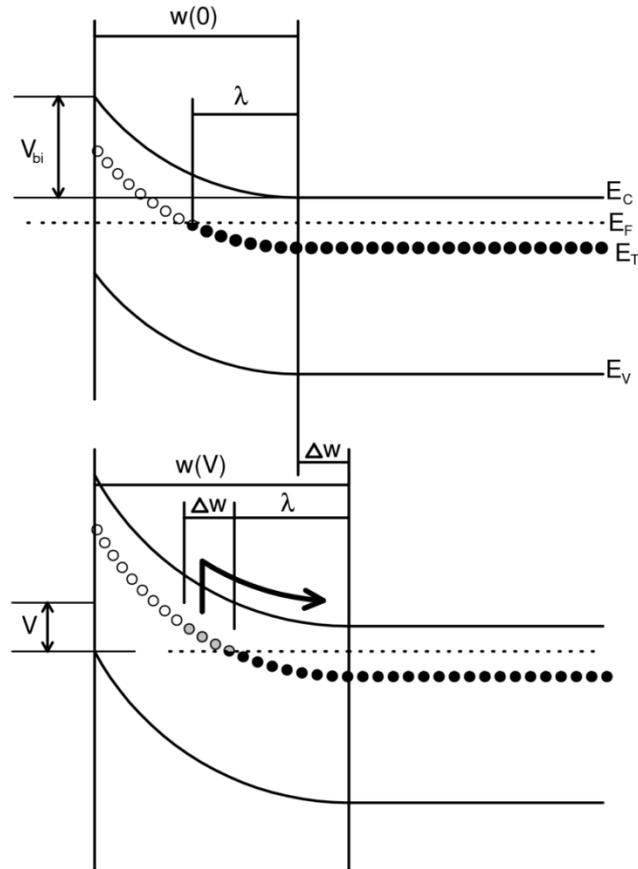
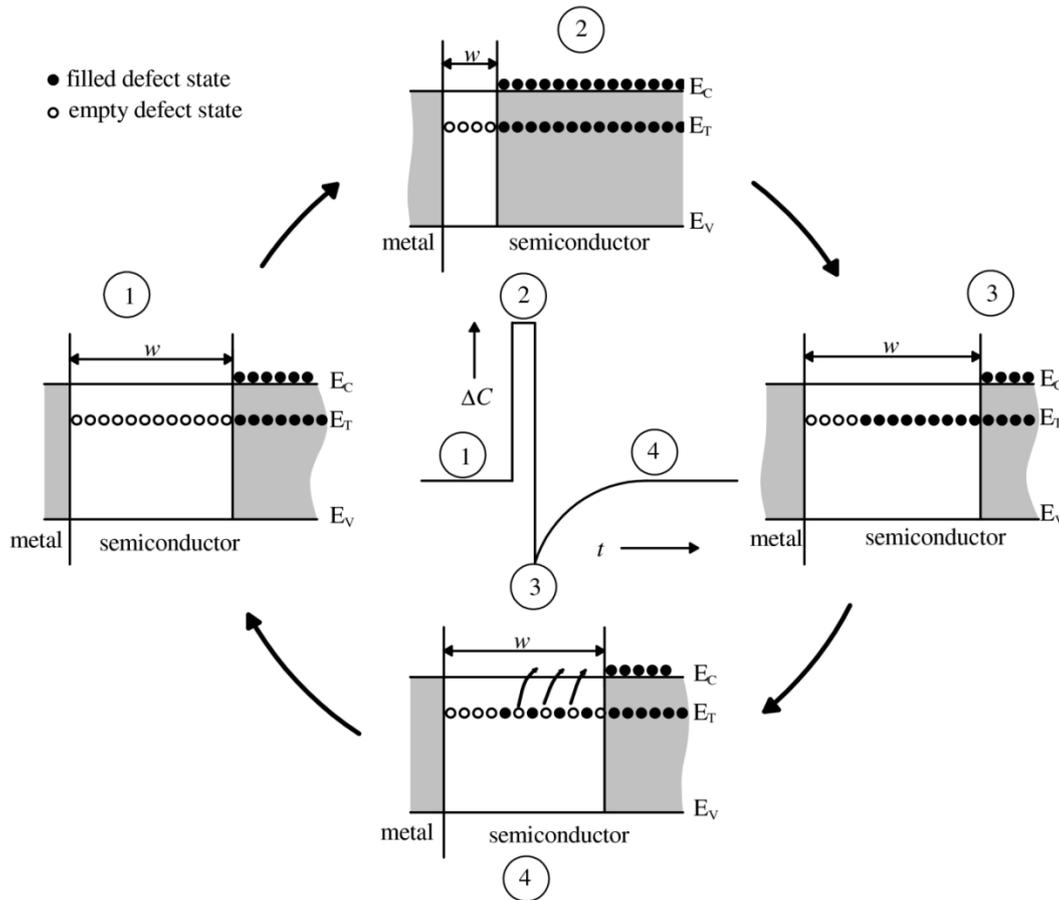


Figure 4.3: Filling (top) and subsequent emission (bottom) of electrons from a deep level in the depletion region, assuming a constant Fermi level. In the top diagram, the filling of the defect during the filling pulse is indicated. It must be noted that because of band bending and defect level depth, the defect level is filled to a depth  $\lambda$  shallower than the depletion region edge. After the filling pulse, the depletion region increases by an amount  $\Delta w$  [11].

Applying a quiescent reverse bias  $V_a$  results in an increase in the depletion region width, which in turn alters the occupancy of the space charge region since some of the deep level traps in the depletion region are now filled. The charge density in the depletion region decreases resulting in a decrease in capacitance [11].

A schematic representation of the capture and emission of carriers is shown in Figure. 4.4.



*Figure. 4.4: Variation of depletion width and trap occupancy of an electron trap in an n-type semiconductor for a DLTS reverse bias and filling pulse. (1) shows the equilibrium condition under reverse bias, while (2) indicates the change after a filling pulse has been applied. (3) and (4) indicate the situation at a time  $t$  after removing the filling pulse. The resulting capacitance transient is shown in the centre. Redrawn from ref [2].*

If the applied voltage is reduced, by applying a positive going pulse, the depletion width is reduced (2), allowing electrons to be trapped at the deep levels. At this stage, it is assumed that the filling pulse width is long enough to allow all the empty levels to completely fill up. This results in an increase in capacitance since the depletion width will be reduced. After removing the pulse (3), the filled states within the depletion region, above the Fermi level start emitting carriers with a characteristic rate into the conduction band where they are instantaneously swept away by the junction electric field (4). This thermal emission of the trapped electrons where the junction electric field sweeps them instantaneously into the conduction band is taken to be the majority carrier capacitance transient [2].

During experiments, the electron emission rate is determined from the time dependence of the capacitance transient. The density of the occupied traps at time  $t$  after removing the filling pulse is given by [2],

$$N(t) = N_T \exp(-e_n t) \quad (4.12)$$

where  $e_n$  is the thermal emission rate and  $N_T$  is the trap concentration assuming all the traps were initially filled. For  $N_T \ll N_D$ , the junction capacitance can be expressed as a time dependent function given by,

$$C(t) = C_0 - \Delta C_0 \exp(-e_n t) \quad (4.13)$$

where  $C_0$  is the equilibrium reverse bias voltage capacitance and  $\Delta C_0$  is a change in capacitance observed immediately after removing the pulse.

### 4.3.1 DLTS Concepts

DLTS is a highly sensitive capacitance or current measurement technique with a good transient response. It is also a spectroscopic technique that is capable of resolving signals due to different traps from each other. Due to the above-mentioned features, amongst others, the DLTS principle of operation lies in the processing of capacitance transients obtained after a repeated pulsing sequence. It is also capable of setting an emission rate window such that the measurement apparatus only responds when it sees a transient whose rate falls within its window. During temperature scanning, since the thermal emission process is strongly dependent on temperature, the peak positions of the observed spectra shift to higher temperatures for higher emission rates and to lower temperatures for lower emission rates allowing a unique determination of these positions by using the rate window thermal scan. Initially, DLTS used a dual-gate average (double boxcar) for precise determination of the emission rate window and for signal averaging, enhancing the signal to noise ratio. This averaging helps in the detection of low concentration traps. Besides the fact that there is no lower limit on the rate window, it is ideal to work with rate windows above  $1s^{-1}$ . This is because the thermal scan time is long, so below the  $1s^{-1}$  rate window, performing signal averaging would require an exceptionally longer time [10]. Figure 4.5 shows how the double boxcar technique is used to select a rate window.

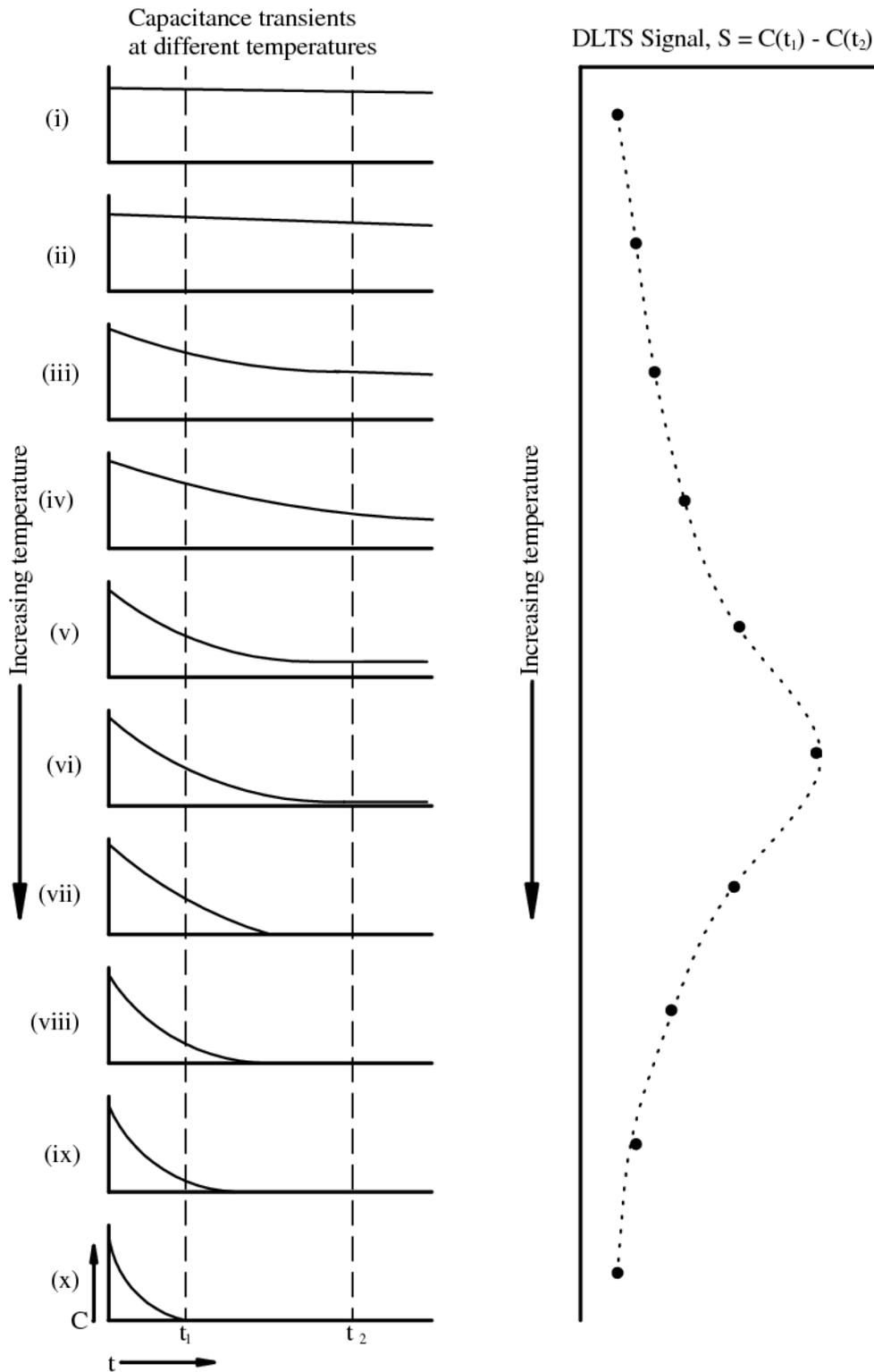


Figure 4.5: The change in the shape of a DLTS transient with increasing temperature (left) and the DLTS signal obtained from the transients as a function of sample temperature (right). Redrawn from ref [11].

Transient signals shown on the left hand side of Figure 4.5 are fed through the double boxcar with pre-set times,  $t_1$  and  $t_2$ . The average difference between the capacitances at  $t_1$ ,  $C(t_1)$  and at  $t_2$ ,  $C(t_2)$  are plotted as a function of temperature as shown on the right hand-side of Figure 4.5. The rate window for the DLTS thermal scan is determined by the values of  $t_1$  and  $t_2$ . As indicated in Figure 4.5, the capacitance difference changes as a function of temperature, i.e. it increases until it reaches a maximum (peak) and then it starts decreasing from either side of the peak. The peak is formed only when the decay constant  $\tau$  falls within the order of  $t_2 - t_1$ . The value of  $\tau$  at the maximum of the  $C(t_2) - C(t_1)$  vs  $T$  for a particular trap,  $\tau_{\max}$  can be related to the gate positions  $t_1$  and  $t_2$ . This is achieved by normalizing the DLTS signal given in Fig 4.5 to obtain  $S(T)$  defined as,

$$S(T) = [C(t_1) - C(t_2)]/\Delta C(0) \quad (4.14)$$

where  $\Delta C(0)$  is the capacitance due to the pulse change at  $t = 0$ . For exponential transients,

$$S(T) = [\exp(-t_1/\tau)] - [\exp(-t_2/\tau)] \quad (4.15)$$

which can also be written as

$$S(T) = \exp(-t_1/\tau)[1 - \exp(-\Delta t/\tau)] \quad (4.16)$$

where  $\Delta t = t_2 - t_1$ . Thus the relationship between  $\tau_{\max}$ ,  $t_1$  and  $t_2$  can be obtained by differentiating  $S(T)$  with respect to  $\tau$  and equating the result to zero:

$$\tau_{\max} = \frac{t_1 - t_2}{\ln\left(\frac{t_1}{t_2}\right)} \quad (4.17)$$

Thus the emission rate of the peak of a DLTS spectra obtained from a thermal scan can be precisely and uniquely determined. Performing a thermal scan with many different rate windows enables one to identify the peak positions and the temperatures at which they occur. This information can then be used to calculate the activation enthalpy and apparent capture cross-section as illustrated in Figure. 4.6.

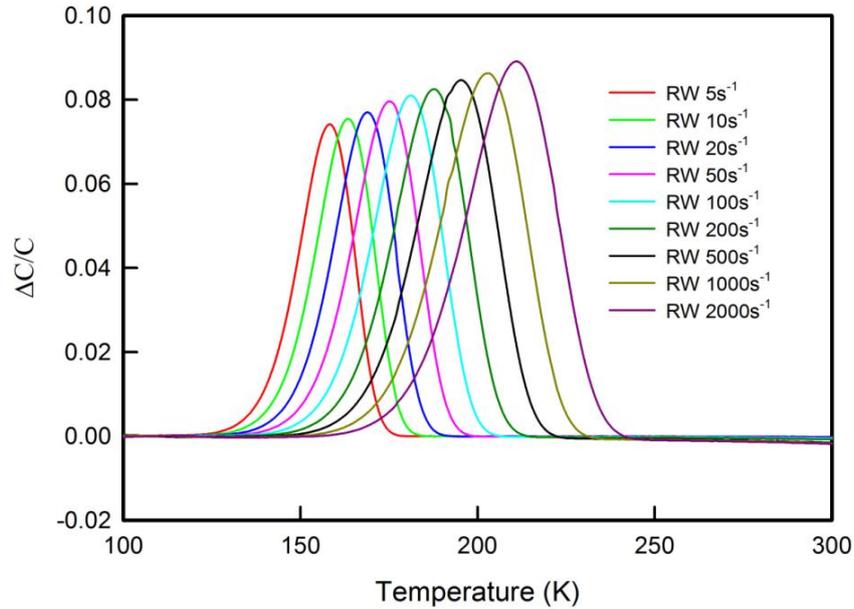


Figure 4.6: (a) Normalized DLTS spectra of the E3 peak in ZnO annealed at 500°C in vacuum to demonstrate how different rate windows can be used to plot an Arrhenius plot shown in Figure 4.6 (b). These spectra were recorded at a quiescent reverse bias of 2.0 V, filling pulse of 0.30 V and pulse width of 2ms.

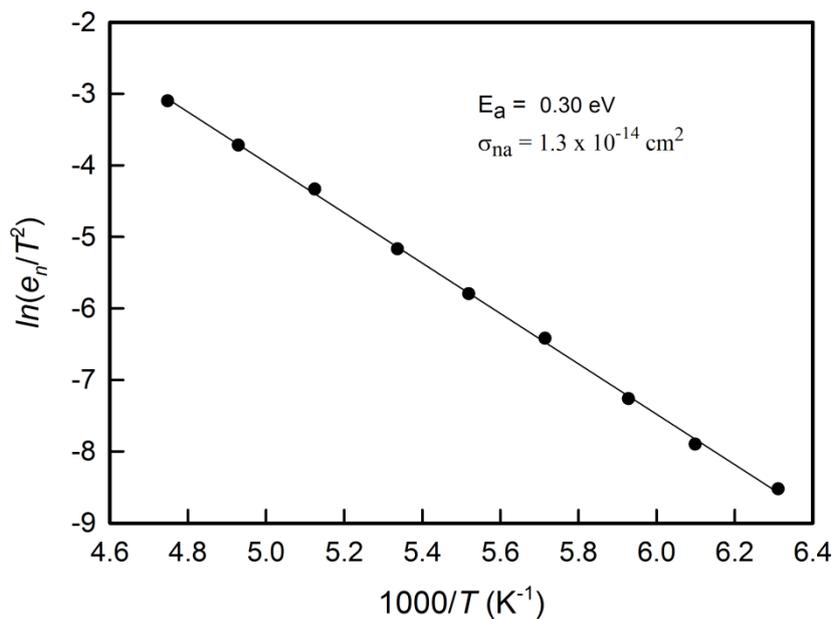


Figure 4.6: (b). Arrhenius plots of the E3 peak obtained using different rate windows indicated in Figure 4.6 (a).

### 4.3.2 Peak height dependence on rate window frequency

As can be observed in Figure 4.6 (a), the heights of the normalized DLTS peaks seem to depend on RW, i.e. the slower the rate window, the smaller the DLTS peak height. This variation has often been attributed to some peculiar characteristics of the traps under investigation, which include broadening of the trap activation energy resulting in non-exponential capacitance transients caused by field dependent electron emission, amphoteric behaviour of the deep levels [12] and temperature dependence of the capture cross-section [13]. Cavalcoli *et al.* [14] suggested and demonstrated the variation of the peak height due to the presence of a coulombic repulsive barrier associated with the defect under investigation. Rockett and Peaker [15] rule out the dependence of the peak height on non-exponential capacitance transients and suggest it to be rather related to the change in shape of the Debye tail of carriers which spill into the depletion region from the neutral bulk region, with both reverse bias and temperature. A study of the effects of varying the reverse bias conditions on the behaviour of the E3 peak is shown in Figure 4.7(a) and (b).

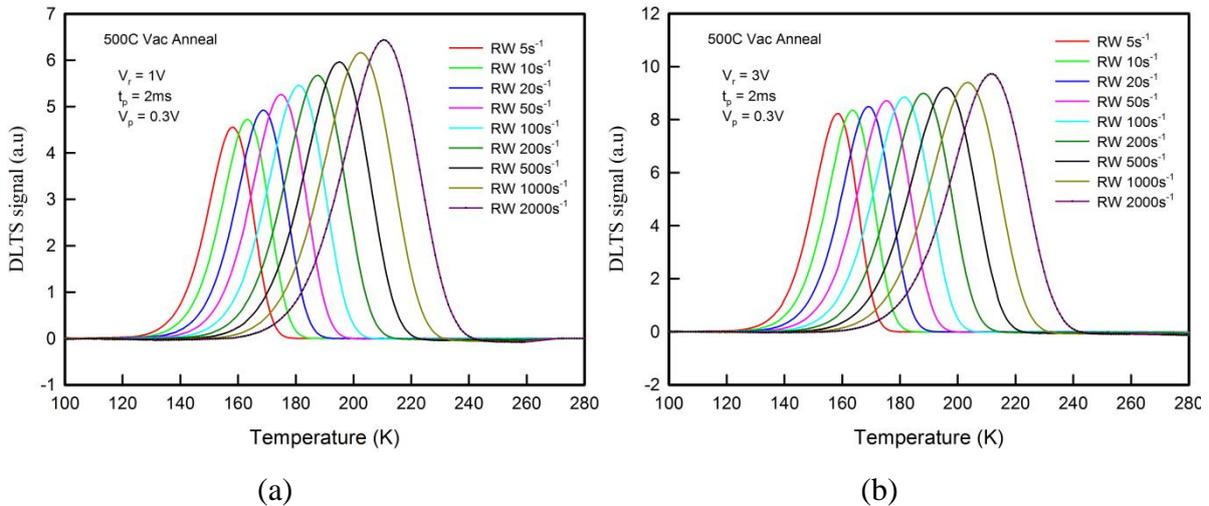


Figure 4.7: DLTS spectra of the E3 peak determined at (a) 1.0 V reverse bias and (b) 3.0 V reverse bias to illustrate the effect of reverse bias conditions on the variation of peak height with rate window.

The peak height variations still show a very strong dependence on emission rates. This indicates that the large peak variations observed in the E3 peak with large and small reverse bias conditions, which suggest field enhanced emission effects are not involved in any important way as has been suggested by Zhao *et al.*[16]. At high temperatures the variation in peak height is expected as the Debye tail penetration into the depletion region is also greater.

Zhao *et al.*[16] however suggest that the variation of the peak height is expected to occur if there is a large temperature dependence of the carrier capture cross section.

### 4.3.3 Temperature dependence of the capture cross-section

DLTS is a good technique for the determination of the trap thermal activation energy. However, the use of this technique is not sufficient since the determination of thermal capture parameters by DLTS is through an indirect process that reduces the accuracy [17]. Since the capture barrier is usually determined from the temperature dependence of the capture cross section and the available free-majority carriers, there is a high uncertainty associated with the actual shape of the filling pulses when the width is reduced. For pulses with short widths, the captured charge becomes difficult to determine and thus the capture coefficient is often underestimated [17].

However DLTS can be used to address this problem. This is achievable if one performs emission DLTS scans using different filling pulse widths. Saturation filling pulse conditions have to be determined first by using a longer filling pulse width. Once these conditions have been obtained, constant width filling pulses have to be selected to assure partial trap recharge. Figure 4.8 shows the normalized DLTS spectra for the E3 peak obtained under different filling pulse widths.

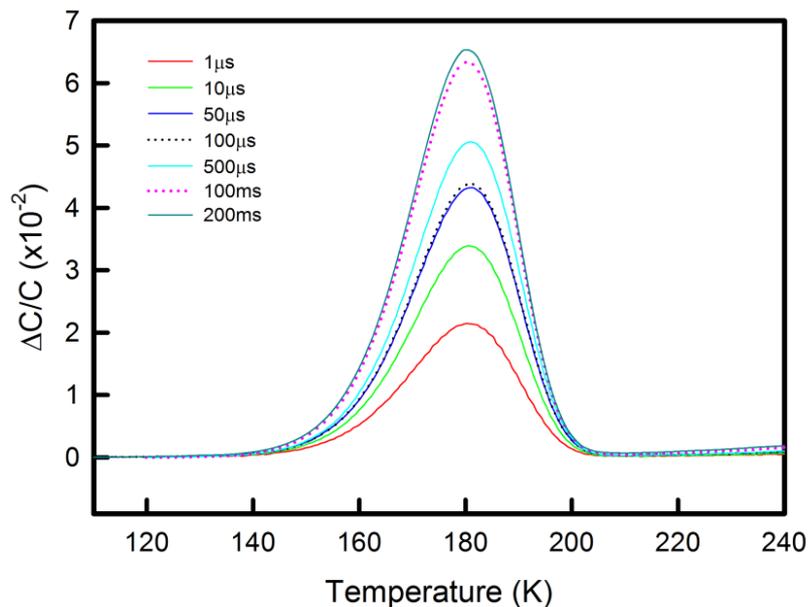


Figure 4.8: Normalized DLTS spectra of the E3 peak measured at different filling pulse widths. The spectra were recorded at a reverse bias of 2.0 V, pulse height of 0 V and rate window of  $100 \text{ s}^{-1}$ .

It can be observed that with short filling pulses, the peak height is smaller, implying partial trap recharge. With longer filling pulses, the peak height increases until it reaches saturation, i.e. implying all traps will be filled.

Since DLTS is capable of setting different rate windows, performing emission DLTS at different sampling times (keeping the ratio of the gate delay times  $t_2/t_1$  constant to ensure self-consistent trap parameters [18]) will provide a set of true peak heights within a given temperature range. The change in the peak heights is related to the trap filling factor [17]. An Arrhenius plot of the DLTS peak heights and temperature positions can be used to determine the capture barrier energy directly.

A capacitance DLTS signal produced by thermal emission of carriers from a trap can be given by,

$$S(T) = C_0(t_f, T)\{\exp[-e_n(T)t_1] - \exp[-e_n(T)t_2]\} \quad (4.18)$$

where  $C_0(t_f, T)$  is the capacitance transient amplitude at temperature  $T$  obtained with a filling pulse width  $t_f$ .  $t_1$  and  $t_2$  are the gate delay times and  $e_n$  is the thermal emission rate. The maximum of the DLTS signals occur at  $T_{\max}$  such that;

$$S(T_{\max}) = C_0(t_f, T_{\max})B \quad (4.19)$$

$B$  is a factor that only depends on the ratio  $t_2/t_1$ . With long filling pulses, it is expected that the trap will be fully charged and the DLTS peaks for various rate windows (keeping the ratio  $t_2/t_1$  constant) will have the same peak height but appear at different  $T_{\max}$ . Under partial filling conditions,  $C_0(t_f, T)$  can now be expressed as

$$C_0(t_f, T) = C_{\text{inf}}[1 - \exp(-c_n t_f)] \quad (4.20)$$

where  $C_{\text{inf}}$  represents the transient capacitance amplitude for a very long charging time (infinite), and  $c_n$  is the capture rate. For a thermally activated capture process,

$$c_n = v_{\text{th}} n_{\text{cap}} \sigma_n = v_{\text{th}} n_{\text{cap}} \sigma_n^{\infty} \exp(-\Delta E_{\sigma}/kT) \quad (4.21)$$

$\sigma_n^{\infty}$  is the high temperature trap electron capture cross section,  $\Delta E_{\sigma}$  is the capture cross-section activation energy, and  $n_{\text{cap}}$  is the electron concentration available for trap capture.

Assuming the thermal velocity  $v_{\text{th}}$  depends on temperature as  $T^{1/2}$  while  $n_{\text{cap}}$  and  $C_{\text{inf}}$  are both temperature independent over the scanned temperature range, we can write,

$$\ln[1 - C_0(t_f, T)/C_{\text{inf}}] = KT^{1/2} \exp(-\Delta E_{\sigma}/kT) \quad (4.22)$$

where  $K$  is the proportionality constant that includes the value of the filling pulse width,  $t_f$  that is kept constant for all the measurement rate windows.

For a saturating filling time (long pulse width), the emission DLTS peak is given by,

$$S(t_{\text{inf}}, T_{\max}) = C_{\text{inf}}(T_{\max})B \quad (4.23)$$

Thus, the capture barrier energy can be obtained as the gradient of the line from an Arrhenius plot of equation (4.22). The true capture cross-section will also be deduced from the intercept, i.e.

$$e^c = n_{\text{cap}} \sigma_n^\infty t_f \sqrt{\frac{3k}{m^*}} \quad (4.24)$$

where  $c$  is the intercept,  $k$  is the Boltzmann constant and  $m^*$  is the electron effective mass.

The other approach to obtain the capture barrier energy and capture cross-section is to use the pulse bias technique by Henry *et al.*[19]. In this particular method, the filling pulse width is related to the capacitance change,  $\Delta C$  as,

$$\Delta C(t_f, T) = \Delta C_{\text{inf}} \left[ 1 - \exp\left(-\frac{t_f}{\tau_c}\right) \right] \quad (4.25)$$

where  $t_f$ ,  $C(t_f, T)$  and  $C_{\text{inf}}$  have the same meaning as has been outlined before.  $\tau_c$  is the capture time constant which is the inverse of the capture velocity,  $nC_n$ , where  $n$  is the carrier concentration and  $C_n$  can be given as a function of the capture cross-section,  $\sigma$  as,

$$C_n = \sigma v_{\text{th}} \quad (4.26)$$

with  $v_{\text{th}}$  having its usual meaning.

A plot of  $\ln \left[ 1 - \frac{\Delta C(t_f, T)}{\Delta C_{\text{inf}}} \right]$  versus  $t_f$  for various temperatures (different rate windows) where the peak of  $\Delta C(t_f, T)$  occurs will yield the values of the capture time constant from the gradient of each line. From each of these values, the capture cross-section can be obtained as a function of temperature as,

$$\sigma(T) = \frac{m(T)}{nv_{\text{th}}} \quad (4.27)$$

where  $m(T)$  is the gradient obtained from each line at a particular temperature. The capture barrier energy and true capture cross-section can be obtained from an Arrhenius plot of equation (4.5),  $\sigma(T) = \sigma_\infty \exp\left(-\frac{\Delta E_\sigma}{kT}\right)$ , i.e.

$$\ln \sigma(T) = \ln \sigma_\infty - \frac{\Delta E_\sigma}{kT} \quad (4.28)$$

A plot of  $\ln \sigma(T)$  versus  $1/T$  gives a straight line whose gradient gives the capture barrier energy and the intercept yields the true capture cross-section.

#### 4.4 Defect Depth profiling

DLTS can be used for the determination of the spatial distribution of a defect within a semiconductor and hence can help predict the defect introduction rate. Assuming that the DLTS signal peak height is directly proportional to the concentration of a deep level, the concentration can be obtained directly from a capacitance change observed using a complete trap filling pulse width. Using this concept, the concentration of a defect is given by [10],

$$N_T = \frac{2\Delta C(0)}{C} N_d \quad (4.29)$$

where  $\Delta C(0)$  is the capacitance change just after removing the complete trap filling pulse,  $C$  is the junction capacitance under quiescent reverse bias conditions and  $N_d$  is the net doping concentration obtained using the CV profiling technique. Determination of the defect concentration using eqn (4.29) usually results in an underestimation of  $N_T$ , particularly for thin films and under low reverse bias conditions. For a better estimation of the trap concentration, we need to consider the region marked  $\lambda$  in Figure 4.3 where the trap level crosses the Fermi level at a  $\lambda$  distance shallower than the depletion region edge. Within this region, the trap is filled and does not contribute to the capacitance transient when a filling pulse has been applied. The width of this particular region is given by [20]

$$\lambda = \left[ \frac{2\varepsilon(E_F - E_T)}{q^2 N_D} \right]^{1/2} \quad (4.30)$$

where  $\varepsilon$  is the semiconductor dielectric constant,  $E_F$  is the Fermi level,  $E_T$  is the trap level and  $q$  is the electronic charge. For one to determine the defect distribution profile, the defect states within the region to be profiled must be filled. The variation of the defect concentration with depth can be obtained using a constant (fixed) reverse bias voltage and changing the filling pulse height, usually referred to as the *fixed bias-variable pulse* method. This technique monitors the incremental change in capacitance  $\delta(\Delta C)$  with a small change in majority carrier filling pulse height,  $\delta V_p$ . The relative incremental change in capacitance with pulse height is given by,

$$\delta \left( \frac{\Delta C}{C} \right) = \left( \frac{\varepsilon}{qw^2 N_D} \right) \frac{N_T(x)}{N_D(x)} \delta V_p \quad (4.31)$$

where  $x$  is the depth below the junction,  $N_D$  is the ionized shallow impurity concentration,  $w$  is the depletion region, corresponding to quiescent reverse bias conditions.  $N_D(x)$  is obtained using the CV profiling technique. Double integration of the Poisson equation based on the

detailed derivation by [20] gives the total signal due to a majority carrier pulse. The corrected expression for the concentration of a deep level defect is given by,

$$N_T = \frac{2\Delta C(0)N_D(x)}{C} \left[ \left( \frac{x - \lambda}{x} \right)^2 - \left( \frac{x_p - \lambda_p}{x} \right)^2 \right]^{-1} \quad (4.32)$$

$x - \lambda$  and  $x_p - \lambda_p$  are the depletion widths before and after applying a filling pulse width, respectively.  $\lambda_p$  is the distance from the depletion region edge to the point where the trap level crosses the Fermi level during the application of the pulse. Using this technique, under low noise measurements, defect concentrations of the order of  $10^{10} \text{ cm}^{-3}$  are detectable.

## 4.5 Laplace DLTS

### 4.5.1 Introduction

Since the discovery of the DLTS technique by Lang in 1974, conventional DLTS has been an important tool in the determination of deep level defect signatures in semiconductors. Having the capability of analysing the temperature-dependent emission transient to produce a sequence of peaks on the temperature scale, ability to perform signal processing using the boxcar and lock-in techniques and excellent sensitivity, DLTS has taken many strides in helping improve device performance efficiency. However, the time constant resolution of this standard DLTS technique is poor for studying fine structure in the emission process [21] as it cannot be used to separate closely spaced transients. Since any variation of time constant in a defect results in additional broadening of the peak, there is need for a technique that can help resolve these time constants. Attempts to overcome this particular limitation have been made by using sophisticated peak deconvolution methods and development of different weighting functions [22]. However the problem with such methods is that extracting multiple closely spaced components decaying exponentially is fundamentally ill posed [21] and so in the presence of noise, there is no unique solution. In DLTS measurements, the signal is often quite small; sensitivity and signal to noise ratio considerations are very important [4]. To address the above mentioned problems of the conventional DLTS technique, Dobaczewski *et al* developed the high resolution Laplace-DLTS technique.

## 4.5.2 Laplace DLTS principles

Laplace-DLTS has a high sensitivity and a sufficiently high energy resolution (of the order of a few meV) to reveal information on the impurity's local environment [23]. In DLTS there are two main categories of transient processing that are classified as analog and digital signal processing. Analog signal processing is performed in real-time, i.e. as the temperature is ramped; one or two capacitance transients are extracted at a time. A filter is then used to produce an output that is proportional to the amount of signal that is observed at a particular time constant range. In digital signal processing, the analog transient output of the capacitance meter is digitized [21]. This is usually performed with the sample held at a fixed temperature and averaging many digitized transients to reduce the noise level. The acquired wave form is then processed using some algorithm. High resolution Laplace-DLTS uses this basis. In identifying the type of algorithm, a method of moments was applied by Ikossi-Anastasiou and Roenker [24], while Nolte and Haller [25] used an approximation the Gaver-Stehfest algorithm to effect a Laplace transform. Even though they achieved a substantial increase in resolution; in the presence of experimental noise levels, the approach appeared to be unstable. In the Nolte and Haller approach, the ultimate theoretical limit of time constant separation in the presence of noise is considered.

To give a quantitative description of nonexponentiality observed in the capacitance transients, one needs to assume that they consist of a spectrum of emission rates, [21]

$$f(t) = \int_0^{\infty} F(s)e^{-st} dt \quad (4.33)$$

where  $f(t)$  is the recorded transient, and  $F(s)$  is the spectral density function. The mathematical representation of the capacitance transients shown in eqn (4.33) is a Laplace transform of the true spectral density function. To determine the spectrum of the real emission rates present in the transient, an inverse Laplace transform has to be performed using numerical methods [4]. A spectrum of delta-like peaks is expected to be produced for multi-, mono-exponential transients, or broad spectrum with no fine structure for continuous distribution [21]. It has to be noted that eqn (4.33) does not have a general solution for any given function  $f(t)$ . According to Lerch's theorem [26], for an analytical multi-exponential function, such a solution exists and is unique. However, the number of possible solutions becomes infinite if the function is superimposed with noise. Numerical methods used in the

Laplace DLTS system seeks to find a spectral function with the least possible number of peaks, consistent with the data and experimental noise [4].

The Laplace DLTS system considered in this study consists of three different software procedures used for numerical calculations, based on the Tikhonov regularization method [27]. The three numerical routines are the CONTIN [28], FTIKREG[29] and FLOG[30]. The three differ in the way the criteria for finding the regularization parameters are defined. During the experiment, sample excitation parameters are set using the Laplace card. In some cases, the card is used to trigger an external pulse generator, which supplies the biasing and pulsing conditions to the sample. The Laplace software then acquires the capacitance transient which will then be converted into LDLTS spectra using numerical algorithms. In DLTS, not all states are expected to produce ideal exponential decays. Some conditions e.g. the Poole Frenkel effect result in an increased emission with increasing field, thus it is important that any algorithm must provide accurate amplitude and rate information in relation to each transient component. The parallel use of the three different software packages increases the level of confidence in the spectra obtained [4].

## Summary

Properties of the DLTS technique which include among others, its high sensitivity, high resolution and ability to uniquely determine the position of a peak in temperature using thermal emission properties of a trap have been exploited to study deep level defects in ZnO. Capacitance based DLTS has been used to monitor the change in the capacitance within the depletion region of fabricated Schottky contacts leading to the determination of “defect signatures”. This technique has been successfully employed to study the temperature dependence of the capture cross-section of deep level defects. Laplace DLTS principles have also been employed to separate closely spaced energy levels in broad peaks.

## References

1. M.A. Green, Solar Cells, Prentice-hall Inc., Englewood Cliffs, NJ, 1982
2. F. D. Auret, P. N. K. Deenapanray, Critical Reviews in Solid State and Materials Sciences, **29**, 1 (2004)
3. H. J. Queisser, Science, **281**, 945 (1998)
4. L. Dobaczewski, A. R. Peaker, K. Bonde Nielsen, J. Appl. Phys. **96**, 9 (2004)
5. Shockley and W. T. Read, Jr., Phys. Rev. **87**, 835 (1952)
6. S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1969)
7. C. H. Henry and D. V. Lang, Phys. Rev. B, **15**, 989 (1977)
8. G. L. Miller, D. V. Lang, and L. C. Kimerling, Capacitance Transient Spectroscopy. Ann. Rev. Mater. Sci. **7**, 377 (1977)
9. C. O. Almbladh and G. J. Rees, Statistical Mechanics of Band States and Impurity States in Semiconductors, J. Phys. C: Solid State Phys. **14**, 4575 (1981)
10. D. V. Lang J. Appl. Phys. **45**, 3023 (1974)
11. W. E. Meyer PhD Thesis, University of Pretoria 2007
12. K. Knobloch, H. Alexander, Mat Sci. Eng. B, **42**, 63 (1996)
13. W. Szkielko, O. Breitenstein, and R. Pickenhein, Cryst. Res. Technol. **16**, 197 (1981)
14. D. Cavalcoli, A. Cavallini, E. Gombia, Phys. Rev. B, **56**, 23 (1997)
15. P.I. Rockett, A.R. Peaker, Electronic Lett. **17**, 22 (1981)
16. J. H. Zhao, T.E Schlesinger, and A.G Milnes J. Appl. Phys. **62**, 7 (1987)
17. J. Criado, A. Gomez, E. Munoz, and E. Calleja, Appl. Phys. Lett. **52**, 8 (1988)
18. J. Criado, A. Gomez, E. Munoz, and E. Calleja, Appl. Phys. Lett. **49**, 26 (1986)
19. C. H. Henry, H. Kukimoto, G. L. Miller, and F. R. Merrit, Phys. Rev. B **7**, 2499 (1973)
20. Y. Zohta, M. O. Watanabe, J. Appl. Phys. **53**, 3, 1809 (1982)
21. L. Dobaczewski, P. Kaczor, I. D. Hawkins and A. R. Peaker, J. Appl. Phys. **76**, 194 (1994)
22. A. A. Istratov, J. Appl. Phys. **82**, 2965 (1997)
23. <http://www.laplacedlts.eu/Dodatki/WordPress/laplacedlts/>
24. K. Ikossi-Anastasiou and K.P. Roenker, J. Appl. Phys. **61**, 182 (1987)
25. D. D. Nolte, and E.E. Haller, J. Appl. Phys. **62**, 900 (1987)
26. G. A. Korn, and T. M. Korn, in Mathematical Handbook , McGraw-Hill, New York, 1968
27. C. Eiche, D. Maier, M. Schneider, D. Sinerius, J. Weese, K. W. Benz, and J. Honerkamp, J. Phys. Condens. Matter, **4**, 6131 (1992)

28. S. W. Provencher, Compt. Phys. Comm. **27**, 213 (1982)
29. J. Weese, Compt. Phys. Comm. **69**, 99 (1991)
30. A. Matulis, Z. Kancleris, Semiconductors Physics Institute, Vilnius, Lithuania