Chapter 3

Defects in Semiconductors

3.1   Introduction

In an ideal crystal lattice, each atom is at its designated position and deviations from this perfect structure are called imperfections or defects. These defects may introduce electronic energy states into the semiconductor band gap, which can be placed into two categories: shallow levels and deep levels. Shallow levels are located near their related band edges (valence band for acceptors and conduction band for donors) i.e. ~0.1 eV from the band edge, thus these levels are thermally ionized at room temperature. The ionization energy of a shallow level can be approximately described by a modified hydrogenic model [1]. For example, a shallow donor resembles a hydrogen atom with a positive nucleus binding an electron. Impurity elements which are used as dopants in semiconductors normally create these shallow levels which are ionized at room temperature and provide free carriers to form p-type or n-type semiconductor. Deep levels are those defects positioned deeper in the band gap than the dopant levels and are found to bind the carriers much more strongly into highly compact, localized states. The deep levels have higher ionization energies, therefore contribute very little to the free charge carriers. Defects with deep levels in the band gap are often referred to as, 'traps', 'recombination centers', or 'generation centers'. Deep levels are important in semiconductors since they modify the properties of the semiconductors and therefore, those of the devices fabricated thereon. Deep levels are desirable in some applications, e.g. in fast switching devices, where they can be exploited as recombination centers which quickly remove minority carriers, enhancing the device's switching speed thereby increasing efficiency [2,3]. Deep levels may also be a nuisance if present in semiconductors that are used for photovoltaic applications since they reduce the cells' efficiency by allowing created electron-hole to recombine. Thus deep level study is of paramount importance in the

semiconductor device industry so that those deep levels which are useful can be deliberately added and those that are deleterious can be reduced or eliminated. This chapter will outline the common properties of deep levels (i.e. structure, charge states, formation and migration mechanisms) in silicon and germanium.

3.2   Primary Defects

Defects in semiconductors can be divided into two main categories; point defects and extended defects. Point defects are not extended in space in any dimension and this implies that the perturbation of the lattice is localized about a lattice site and involves only a few nearest neighbours. There are two kinds of point defects of great interest in semiconductor crystals, intrinsic (e.g. vacancies or self-interstitial) and extrinsic point defects (e.g. impurity atoms occupying substitutional or interstitial lattice sites. Small agglomerations of several point defects like divacancies, vacancy-impurity complexes, vacancy-donor etc are also generally considered as point defects. Extended defects are extended in nature (such as, grain boundaries, dislocations or stacking faults). The discussion in this section is focussed more on point defects which are more relevant to the work covered in this thesis.
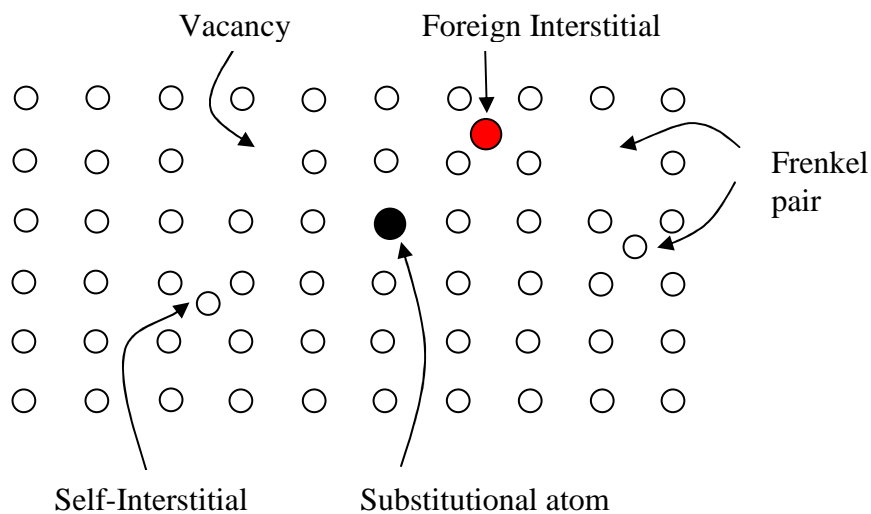

Fig. 3-1. A diagram showing the vacancy, self-interstitial, substitutional defects.

3.2.1   Vacancy Defect

If an atom is removed from its regular lattice site the empty lattice site is called a vacancy defect (V), and is shown in Fig. 3-1. The vacancy in some semiconductors (e.g. in Ge and Si) can have up to five charge states, V^{++}, V^{+}, V^{0}, V^{-} and $V^{=}$. In order

to form a vacancy by removing an atom from its lattice site four bonds are broken in a diamond crystal structure as shown in Fig. 3-2a [4,5]. The broken bonds can form new bonds depending on the charge state (i.e. which is just the number of electrons occupying the dangling bonds) of the vacancy (Fig. 3-2b and c). This causes small inward and outward displacement of neighboring atoms, which either preserves the local symmetry (relaxation) or alters it (distortion). The amplitude of the displacements depends on the charge state of the defect. Another geometric configuration for describing a vacancy is the 'split-vacancy'. The 'split-vacancy' results when an atom resides at the bond center between the empty sites [6], Fig. 3-2d. The 'split-vacancy' is often important primarily to help describe the transition state in vacancy migration [7]. The lattice relaxation depends on the charge state of the point defect (Jahn-Teller effect). Jahn-Teller effect is simply a geometrical distortion which occurs when the electronic state is degenerate, in which case the nuclear state is unstable. Atomic displacements always exist which by lowering the symmetry, split the degenerate level.
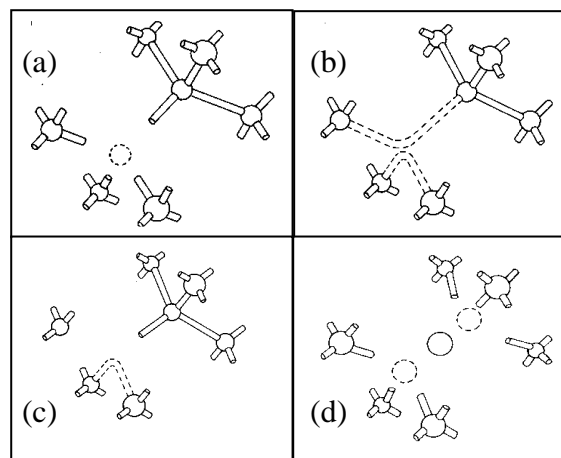

Fig. 3-2. The vacancy configuration in diamond lattice. (a) Four bonds are broken in order to create the vacancy. (b) When there is one electron per dangling bond (i.e., for the neutral vacancy V^0) they form two new bonds leading to local distortion. (c) When an electron is missing (i.e., for the positive vacancy V^+) one of these two bonds is weakened since it contains only one electron. The distortion is thus different from that in the case of V^0. (d) The 'split-vacancy configuration, redrawn from ref. 4.

It should be noted there is general agreement in defect modeling studies that the vacancy formation energy in germanium (1.7 eV – 2.5 eV) is significantly smaller than in silicon (~4.0 eV), for all charge states. When two neighboring atoms are removed and also when two migrating vacancies meet and combine a divacancy is formed. A divacancy can also exist in four different charge states in Si [8].

3.2.2 Interstitial Defect

Interstitials are atoms, which occupy a site in the crystal structure, which is not a regular lattice site as shown in Fig. 3-1. An interstitial defect can be of the same species as the atoms of the lattice (self-interstitial) or of a different nature (an interstitial impurity). Interstitials are generally high-energy configurations. Once again, the introduction of an interstitial induces a relaxation and distortion of the lattice, which surrounds it. The type of configuration the interstitial assumes depends on its ability to make bonds with its neighbors and therefore can change with its charge state. A nearby interstitial defect and vacancy defect is called a Frenkel pair.

3.3 Secondary Defects

The isolated lattice vacancy (V) and self-interstitial (I) are primary defects produced after high-energy particle irradiation in semiconductors. The primary defects are mobile at low temperatures, e.g. in Si the vacancy becomes mobile above 150 K - 200 K and the interstitial is mobile even at 4.2 K [8]. Therefore, deep level transient spectroscopy measurements of room temperature irradiated silicon will not reveal an isolated vacancy or silicon interstitials. The V and I which survive the recombination of simple defects can diffuse into the semiconductor and interact with other intrinsic and extrinsic defects giving rise to complex room temperature stable defects, some of which are depicted in the schematic diagram shown in Fig. 3-3. For example, in silicon when a vacancy becomes mobile, it can be trapped by an oxygen atom to form a V-O complex (A-center), or the doping impurity (e.g. P) to form a V-P complex (*E*-center), or by another vacancy to form divacancies. Other complex defects are via the mobile interstitial (I). It is expected that room temperature defect evolution in silicon shown in Fig. 3-3 should hold similarly for germanium.
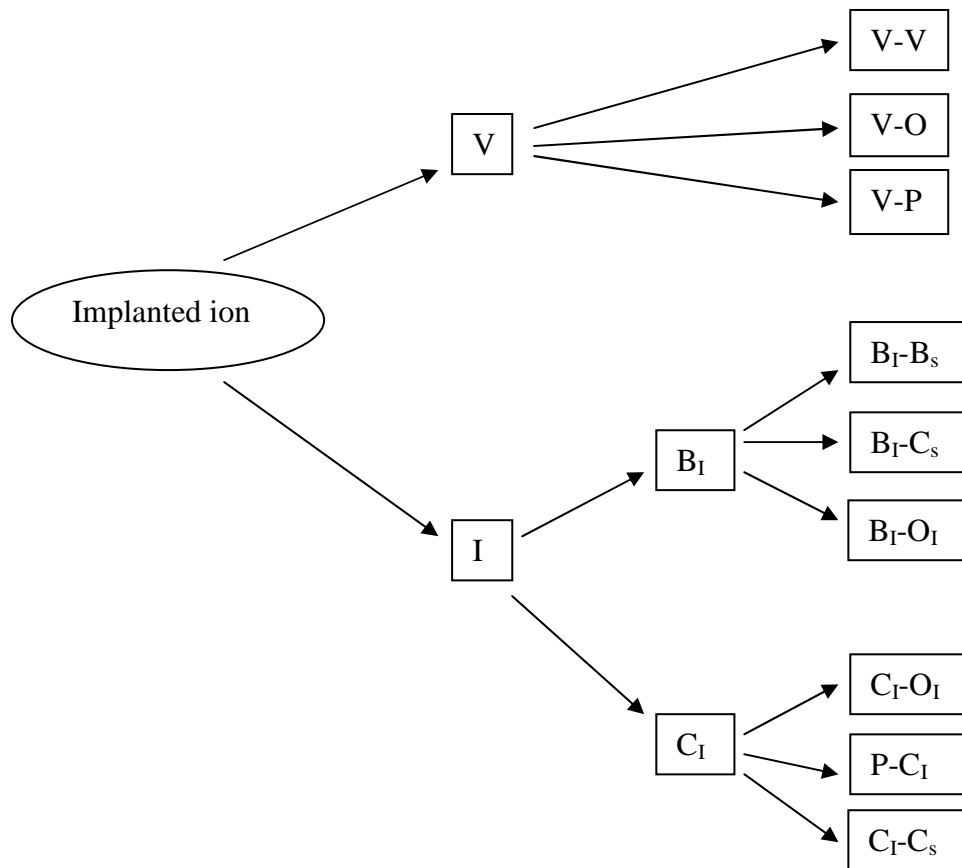
Implanted ion → V → V-V, V-O, V-P

Implanted ion → I → B_I → B_I-B_s, B_I-C_s, B_I-O_I

I → C_I → C_I-O_I, P-C_I, C_I-C_s

Fig. 3-3. Schematic of room temperature defect evolution in crystalline silicon.

3.3.1 The Divacancy

The divacancy is formed by the removal of two neighbouring atoms. Generally divacancies can be created in semiconductors by particle irradiation either as a primary defect, when collision cascade is dense enough, or as a secondary defect by pairing of single vacancies diffusing randomly. The vacancy in germanium is negatively charged in a broad interval of Fermi level position in the band gap, therefore the formation of divacancies by pairing of single vacancies is suppressed by the coulombic repulsion of the vacancies [9]. The divacancy in Ge has not yet been identified by experimental techniques. Using density functional theory (DFT) cluster calculations, Janke *et al* [10] estimated the energy barrier for migration and dissociation of the divacancy. The dissociation energy consisting of the binding energy between two vacancies and the migration of a single vacancy was found to be between (1.5-1.7) eV whereas the migration barrier of the divacancy was found to be 1.1 eV. This corresponds approximately to a thermal stability of 420 K. The

divacancy in silicon is well known, and can appear in four charge states, V_2^+, V_2^0, V_2^- and $V_2^=$.

3.3.2 The *E*-center

The *E*-center can be described as a vacancy trapped next to a substitutional donor atom. The *E*-center can be formed as a primary defect or when the impurity atom captures a mobile vacancy. As to the formation of an *E*-center, although local atomic strain effect may need to be considered, the key role is played by the coulombic interaction between a positively charged antimony and a negatively charged vacancy [11]. Let's consider the *E*-center in phosphorus-doped silicon, V-P. In the neutral charge state i.e. PV^0, two of the three silicon atoms surrounding the vacancy pull together to form an electron pair, leaving an unpaired electron in the orbital of the third silicon atom, while two electrons with antiparallel spins are accommodated by the phosphorus atom. When the Fermi level is above the *E*-center an extra electron is accepted and becomes negatively charged, PV^-. The formation of *E*-center in silicon removes two electrons in the conduction band by converting a positively charged P donor atom to a negatively charged V-P center. The formation of the *E*-center is regarded as an intermediate step in dopant diffusion in silicon and germanium [12].

3.3.3 The *A*-center

The *A*-center (V-O) may be regarded as a vacancy trapped next to an oxygen atom in an interstitial position. Similar to the *E*-center the *A*-center can be formed as a primary defect or when an oxygen impurity traps a mobile vacancy. The *A*-center competes for the vacancies with the *E*-center and its concentration is dependent on the relative O impurity concentration in the sample. The *A*-center has also been found to be an efficient recombination center [13], and therefore can be used as to control minority carrier lifetime in silicon for fast switching device application.

3.3.4 Other Complex Defects

The mobile Si interstitial (I) diffuses and will replace either carbon (C) or group III impurities e.g. boron (B) (depending on the relative concentration of the two species) through the Watkins replacement [14,15] to form interstitial carbon (C_I) or interstitial B_I respectively. The C_I or B_I are mobile at room temperature and will eventually form

defect complexes with other impurities {e.g. interstitial boron – substitutional boron (B_I-B_s), interstitial boron – interstitial oxygen (B_I-O_I), interstitial boron – substitutional carbon (B_I-C_s), interstitial carbon – interstitial oxygen (C_I-O_I) or interstitial carbon – substitutional carbon (C_I-C_s)} as shown in Fig. 3-3. For a particular defect with a large concentration, it tends to aggregate as the temperature increases from room temperature. In case of a divacancy, when mobile or after dissociating, it can form trivacancies, quadrivacancies, pentavacancies, and higher order defects. This behavior should also be true for self-interstitials and for any type of extrinsic defects.

Most of the primary and secondary defects discussed in the previous sections are electrically active and introduce deep levels in the semiconductor band gap. A deep level may act as a minority carrier trap, majority carrier trap or recombination centre depending on its position in the band gap and on relative capture cross-section of minority and majority carriers. A majority carrier trap is an electron trap in n-type semiconductor or a hole trap in p-type semiconductor. Conversely a minority carrier is a hole trap in n-type semiconductor or an electron in p-type semiconductor. If a majority- or minority- carrier lives a mean lifetime in the captured state and is thermally ejected to the band from which it came, the center may be regarded as majority carrier trap or minority carrier trap respectively. From defect spectroscopy measurements such as deep level transient spectroscopy (DLTS) it is possible to extract the defect properties such as the concentration, energy level, and capture cross-section of defect level. The capture cross-sections, $\sigma_{majority}$ and $\sigma_{minority}$ can now be used to deduce whether the defect will act as minority carrier trap, majority carrier trap or a recombination centre.

Recombination centers are deep levels with approximately equal capture cross-sections for both electrons and holes and these centers are normally located near the middle of the band gap. After capturing a majority carrier, if the majority carrier stays trapped at the center long enough for the trap to capture a minority carrier, then recombination takes place and the center is acting as a recombination center. Most of the defect spectroscopy techniques measure the defect concentration, energy level, and capture cross section for majority and minority carrier traps, e.g. when using DLTS, all the detected defects are in a situation where they behave as majority or

minority carrier traps. Hence it is difficult to say which of the detected defects will be a recombination center. There have been attempts to improve on the technique used to distinguish traps from recombination centers. Markvart *et al* [16] developed an improved version of DLTS, known as recombination DLTS that can be used to identify defects that act as recombination centers. These centers act as "stepping stones" for carriers and contribute to the current-voltage characteristics of rectifying junctions at a recombination rate, U given by [3,17]

$$U = \frac{\sigma_p \sigma_n v_{th} (pn - n_i^2) N_T}{\sigma_n [n - n_i \exp\{(E_T - E_i)/kT\}] + \sigma_p [n + n_i \exp\{-(E_T - E_i)\}]} \qquad (3.1)$$

where E_i, E_T, N_T, n_i, n, p, σ_n, and σ_p are the intrinsic Fermi level, defect level, defect concentration, intrinsic carrier density, electron concentration, hole concentration, electron capture cross section and hole capture cross section respectively. From equation (3.1) it is clear that the recombination rate is higher for larger E_T (i.e. most efficient recombination centers are those close to the middle of the band gap and with similar capture cross-sections σ_n and σ_p).

3.4   Theory of Displacement of Atoms in Solids

When passing through matter high-energy particles are decelerated and in the process transfer energy to the material. The transferred energy can modify the structure and properties of the materials. In case of crystalline semiconductors, particle-induced materials modification will occur as long as the projectile particles can transfer energy, E, larger than the displacement energy, E_d, to the lattice atoms [3]. The capacity of a solid to slow a projectile is called the stopping power, and is defined as the amount of energy lost per unit length of trajectory in the solid. The stopping power depends on the type and energy of the projectile and on the mass of the target material.

3.4.1   Energy-Loss Mechanisms

When a particle enters a target material there are two main mechanisms that cause an energy loss: (a) elastic collisions with the nuclei of the target material (nuclear stopping) and (b) inelastic collisions with bound or free electrons (electronic

stopping). In electronic stopping, the term inelastic is used to signify that the collisions may result both in the excitations of bound electrons of the medium and in the excitations of the electron cloud of the ion. The relative effect of the two mechanisms depends on the mass of the target material as well as the energy, mass and charge of the incident particle. Fig. 3.4 is a "universal" diagram showing the nuclear stopping $(d\varepsilon/d\rho)_n$ and electronic stopping $(d\varepsilon/d\rho)_e$ in terms of Thomas-Fermi (TF) ion energy, ε and path length, ρ as a function of $\varepsilon^{1/2}$ (which is proportional to the velocity of the implanted ion) [3,18]. The parameters ε and ρ are dimensionless quantities which can be expressed in terms of laboratory energy E and distance x, respectively as;

$$\varepsilon = \frac{4\pi\varepsilon_0 a M_2}{Z_1 Z_2 e^2 (M_1 + M_2)} E \qquad (3.2)$$

and

$$\rho = N \pi a^2 \frac{4M_1 M_2}{(M_1 + M_2)} x \qquad (3.3)$$

where M_1 and M_2 are the mass numbers of the incident and target atom respectively, Z_1 and Z_2 are their atomic numbers, e is the electronic charge, N is the concentration of atoms, ε_o is the permittivity of free space and a is the screening radius. The screening radius is normally expressed by [19]

$$a = \frac{0.8853 a_0}{(Z_1^{2/3} + Z_2^{2/3})^{1/2}} \qquad (3.4)$$

where $a_0 = 0.529$ Å is the Bohr radius. The universal curve shown in Fig. 3-4 enables the approximation of nuclear stopping and associated quantities (e.g. damage production and sputtering) for all particle-target combinations using a single curve. The ion bombardment analysis can be divided into two distinct regimes.

    (a) *Nuclear microanalysis regime*: High-energy ions are slowed down mainly by electronic stopping. The contribution from the nuclear stopping tends to be

small at high energies because fast ions have only short time to interact with the target nuclei.

(b) *Ion implantation regime*: When the ion has slowed down sufficiently, the collisions with the nuclei become more and more probable, and the nuclear stopping finally dominate the slowing down process. In this regime nuclear stopping reaches a maximum value (ε_1) around $\varepsilon^{1/2}$ = 0.6 and decreases thereafter as shown in Fig. 3-4 and this corresponds to process associated with low energy ions e.g. ion beam etching and sputtering.
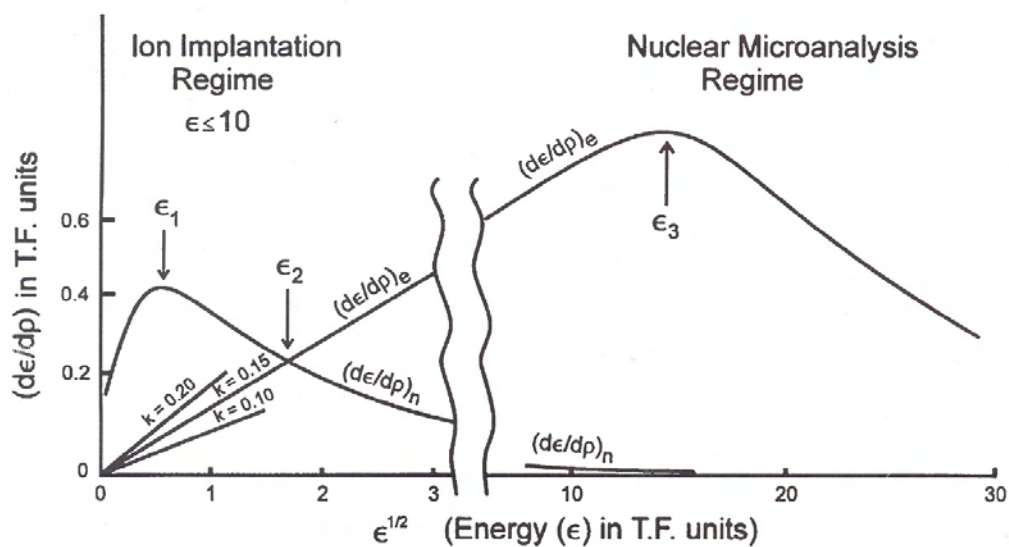

Fig. 3.4. Nuclear and electronic stopping versus the reduced energy of an implanted ion redrawn from ref. 3.

It is important to note that in contrast to the nuclear stopping $\left(d\varepsilon / d\rho \right)_n$ which depends only on ε (i.e. independent of the type of incident particle and target atoms), electronic stopping $\left(d\varepsilon / d\rho \right)_e$ can be expressed by [3]

$$\left(d\varepsilon / d\rho \right)_e = k\,\varepsilon^{1/2}, \quad \text{for} \quad \varepsilon^{1/2} < 14 \tag{3.5}$$

where k is a function of $M_1, M_2,\ Z_1$ and Z_2, thus electron stopping does not exhibit true ε-scaling and therefore may not be described by a universal curve.
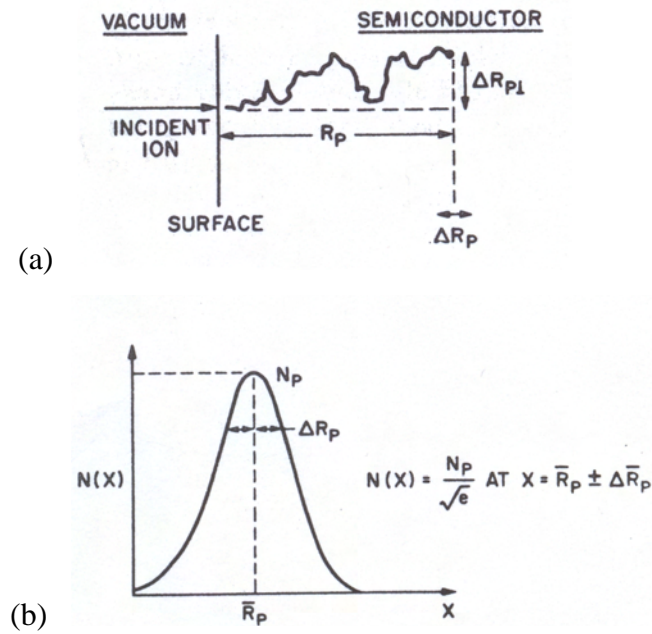
VACUUM   SEMICONDUCTOR

$\Delta R_{P\perp}$

INCIDENT ION R_P

SURFACE ΔR_P

(a)

N_P

ΔR_P

$N(X)$ $N(X) = \dfrac{N_P}{\sqrt{e}}$ AT $X = \overline{R}_P \pm \Delta \overline{R}_P$

(b) $\overline{R}_P$ X

Fig. 3.5. (a) Basic parameters for an implanted ion. R is the total path length, R_p is the projected range, ΔR_p and ΔR_{pL} are the projected standard deviations in the directions parallel and perpendicular to the incident beam, respectively. (b) N(x) is the number of ions per cm^3 at depth x.

The process of ion stopping is a statistical process and so some ions will undergo many collisions, stopping in a distance shorter than the average value. The range R of the projectile is related to its mean track length and for a projectile with initial energy ε_0 the range can be written as

$$R = \int_{0}^{0} \frac{1}{\left(d \ / d \ \right)_{total}} d \tag{3.6}$$

where

$$\left(d \ / d \ \right)_{total} = \left(d \ / d \ \right)_{n} + \left(d \ / d \ \right)_{e} \tag{3.7}$$

If $M_1 \geq M_2$ the projected range R_{p} is obtained by multiplying the range R with the projection factor $\sim (1 + M_2/3M_1)^{-1}$. The average or projected range R_{p} is defined as the

projection of R on the incident ion beam direction, a projected standard deviation or straggle ΔR_p is the statistical fluctuation along incident ion direction if the spatial distribution of the implanted ions is approximately Gaussian as depicted in Fig. 3-5. The implanted ions may also be scattered along the direction perpendicular to the incident direction and the statistical fluctuation along this direction is the projected lateral straggle ΔR_{pL}. This lateral penetration of ions may limit dimensions in some devices [20]. The ion concentration profile in the solid is related to the projected range R_p, standard deviation ΔR_p and ion dose (fluence) Φ, (assuming Gaussian approximation) by

$$N(x) = \frac{\Phi}{\Delta R_p \sqrt{2}} \exp\left[-\frac{(x - R_p)^2}{2\Delta R_p^2}\right] \qquad (3.8)$$

The profile given by equation (3.8) is often called Lindhard-Scharff-Schiott (LSS) [21] profiles for implants in semiconductors. The assumption behind the purely Gaussian stopping distribution described by equation (3.8) is that the implantation takes place into amorphous material, which ignores the effects of channeling related to the single crystal nature of the target.

Computer based simulation of implantation profiles are now possible using Monte Carlo-based techniques such as Transport of Ions in Matter (TRIM) [22] and Stopping Range of Ions in Matter (SRIM) [23]. These simulation codes should be treated with caution when used for crystalline material since they assume an amorphous material and ignore the effects of channeling, diffusion effects during and after ion implantation and also do not account for annihilation of vacancies and interstitials and thus overestimates the concentration of vacancies and interstitials produced by the implantation.

3.4.2   Defect Production by Irradiation

It is important to describe the dynamics of collision because it is fundamental to defect production in semiconductors. Assume that an incident particle with kinetic energy E and mass M_1 strikes a target atom of mass M_2.
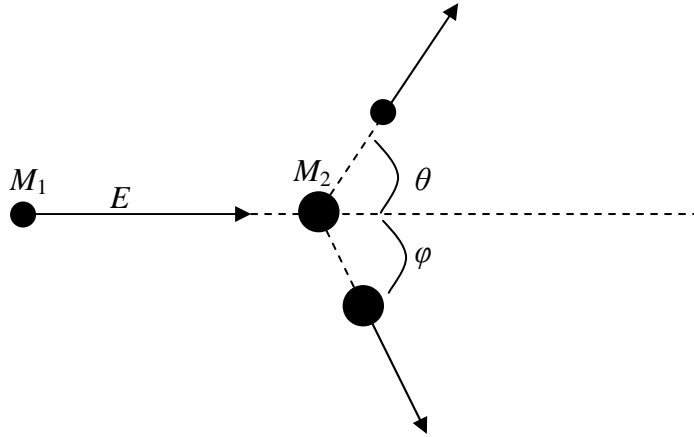
Fig. 3-6. Illustration of the collision between an incident particle of mass M_1 and energy E with a target atom of mass M_2.

The kinetic energy T transmitted to the target atom depends directly on the angular deflection θ of the incident particle (Fig. 3-6). For purely elastic collision, (i.e. when momentum and kinetic energy is conserved) the energy transferred is given by

$$T = 2E \frac{M_1}{M_2} \frac{1 - \eta(\theta)}{\left(1 + M_2 / M_1\right)^2} \tag{3.9}$$

where E is the energy of the projectile and M_1 and M_2 have previously been defined and $\eta(\theta)$ is a function implicitly given by

$$\cos \theta = \frac{1 + (M_2 / M_1)\eta}{\sqrt{1 + 2(M_2.M_1)\eta + \left(M_2 / M_1\right)^2}} \tag{3.10}$$

In the non-relativistic limit the maximum energy T_{max} is transferred for $\theta = 0$, i.e., for $\eta(\theta) = -1$, therefore

$$T_{max} = \frac{4M_1 M_2}{\left(M_1 + M_2\right)^2} E \tag{3.11}$$

In the case where the mass of projectile is approximately equal to the target, the expression of the transferred energy will simplify to

$$T_{\max} = E \qquad (3.12)$$

For a neutron irradiation, $M_1 \ll M_2$ the energy transferred is now given by

$$T_{\max} = \frac{2M_1}{M_2} E \qquad (3.13)$$

In the case of electron irradiation, $M_1 \ll M_2$ relativistic corrections are required and the energy transferred in now written as

$$T_{\max} = \frac{2M_1}{M_2} E \left(2 + \frac{E}{M_1 c^2} \right) \qquad (3.14)$$

where c is the speed of light. Equation (3.14) can be approximately represented by

$$T_{\max} = \frac{2148}{Z} E^2 \qquad (3.15)$$

In this case the Z is the atomic number of the target atom, E is in MeV and $T_{\max}$ is in eV. In order for an atom to be permanently displaced from its lattice position, the energy that it receives must be greater than the displacement energy. The minimum energy necessary to displace an atom from its lattice position is called the threshold energy (T_d). Normally the threshold energy is assumed to be isotropic (i.e. independent of the direction in which the atom is displaced in the lattice.

Table. 3.1. The threshold energies and corresponding minimum incident electron energies for atom displacement in silicon and germanium.

Material	T_d (eV)[a]	$E_{\min}$ (keV)[b]
silicon	21.0	370
germanium	27.5	424

[a]Experimental values obtained from ref. 24.

[b]Values calculated based on equation (3.14).

It should be pointed out that threshold energies for most materials is generally greater than the formation energy of Frenkel pairs, because defect formation is a complex multi-body collision process (e.g. a recoil atom can bounce back to its lattice position or kick back another atom to its lattice position) [25]. Table. 3.1 summarizes the threshold energies and the minimum electron projectile energy E_{min} necessary to displace an atom in silicon and germanium. If an incident particle has energy much greater than the threshold energy (T_d), it will transfer energy to the target atom. The displaced atom may in turn collide and displace other atoms, creating cluster damage, until it eventually comes to rest, usually in an interstitial site and sometimes in a vacancy site. Light energetic particles (such as Si, He, Ar, neutrons, electrons and protons) tend to leave tracks of relative small defect concentrations. These ions initially slow down mainly by electron stopping process with little displacement damage until eventually nuclear stopping becomes dominant at the end of their range.
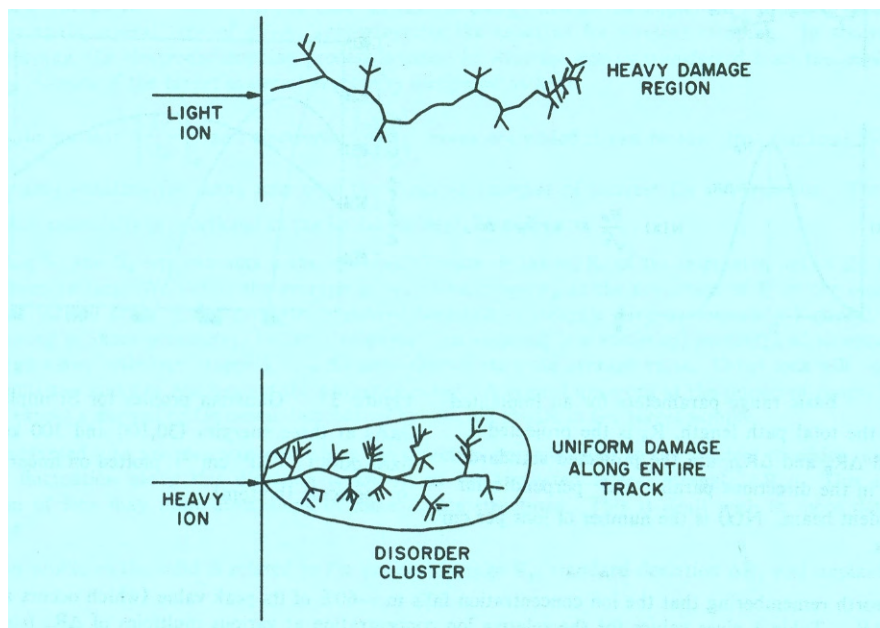

Fig. 3-5. Schematic of ion track in a solid, and associated damage for a light ion (top) and a heavy ion (bottom), redrawn from ref. 20.

Therefore there is generally little lattice damage along the track except for the end of the range. Heavy ions by contrast may create damage clusters along their track. A comparison of lattice damage by light and heavy ions is depicted in Fig. 3-5. The heavy ions may undergo relatively higher degree of nuclear stopping than light ions even right from the surface. The volume of the crystal in which the ion energy is

deposited is usually larger than the volume in which the lattice damage occurs. When the damaged areas start to overlap with increasing ion dose, an amorphous layer can result, which implies that all the nuclei have been displaced from their lattice position and the long range order which describes a crystal is no longer present.

The number of displaced atoms ($N_{disp.}$) by irradiating ion can be estimated by [26]

$$N_{disp.} \approx \frac{E_n}{2T_d} \tag{3.16}$$

where E_n is the total energy deposited in primary and secondary nuclear collisions. The expression in (3.16) should be used with caution, because it overestimates the number of displaced atoms since it overlooks the effects of (i) ion channeling and (ii) vacancy-interstitial recombination.

3.4.3 Defect Annealing Mechanisms

Defect characterization techniques such as deep level transient spectroscopy (DLTS) and photoluminescence (PL) cannot be used to probe the defect structure. The only way to correlate the defect structure obtained from electron paramagnetic resonance (EPR) to the DLTS and PL measurements is by thermal annealing studies. The characteristic temperature at which a defect disappears is the defect's annealing temperature. This annealing temperature is the parameter that allows for results obtained by different techniques to be compared.

The defects annealing mechanism can be classified into two main categories:

(a) Diffusion, as the temperature is increased defects migrate to sinks (e.g. by moving to the surfaces or grain boundaries) or they are subsequently trapped by other defects or impurities (e.g. direct recombination of the interstitial with a vacancy, complex formation or hydrogen passivation) to form new defects. The mean distance between the interstitial and the vacancy depends on the energy deposited by the irradiating particles, thus the annihilation process of the interstitial and a vacancy is mainly related to the irradiation energy [27]. The ability of a defect to migrate through the crystal is determined by the thermal energy of the crystal and its charge state.

(b) Dissociation, which is the breaking-up of the complex defects.

Each of the process i.e. defect migration, recombination, and complex formation is characterized by migration enthalpy or activation energy (E_a). It should be noted that for the simple defects the migration energy for vacancies in a solid is much higher than that for interstitials. At a fixed annealing temperature (T_a) the annealing kinetics can be deduced by monitoring the decrease in defect concentration with time. The defect annealing kinetics can provide information on the defect distribution, annealing mechanism and hence their identity. Consider an irradiated sample with a given defect level with concentration N_T, the number of defects which anneal per unit time is proportional to the number of defects $N_T(t)$ present at time t and thus can be written as

$$\frac{dN_T}{dt} = -Kf(N_T) \tag{3.17}$$

where K is the rate constant and if $f(N_T) = N_T$, then the annealing kinetics is said to be of first order and if $f(N_T) = N_T^2$ then it is of the second order. Solving the differential in equation (3.17) gives annealing kinetics for first order as

$$N_T(t) = N_T(0)e^{-Kt} \tag{3.18}$$

where $N_T(0)$ is the initial defect concentration at $t = 0$. The rate constant K given in equation (3.17) and (3.18) has the form

$$K = K_0 e^{-\frac{E_a}{kT}} \tag{3.19}$$

where k is Boltzmann constant, K_0 is a pre-exponential constant (which contains the vibrational frequency associated with the process) and E_a is the associated activation energy (which might be migration energy, dissociation energy etc depending on the process). Experimentally the activation energy can be obtained from isothermal annealing studies. The variation of N_T versus time is measured at a constant temperature T_1 and a plot of $\ln(N_T)$ vs. t is straight line for first order kinetics which will give a rate constant K_1 from the gradient. If the process is repeated for other constant temperatures T_2 and T_3, then rate constants K_2 and K_3 are obtained respectively. From equation (3.19) a graph of $\ln(K)$ vs. $1/T$ is an Arrhenius plot and

will yield the activation energy E_a for the defect annealing from the gradient and K_0 from the vertical axes intercept.

It is interesting to note that most simple defects anneal out at between 200˚C and 400˚C in silicon, while higher order defects are introduced at higher temperatures between 350˚C and 500˚C. The structure of these higher order defects is critically dependent on the irradiation condition (i.e. irradiation ion energy and ion mass). It has also been shown that the formation of a silicide phase at a metal-Si interface during thermal annealing injects vacancies into the substrate. Thus the silicidation technique can be used to remove interstitial – related defects in processed p-type Si [28].

References

[1]   G.L. Miller, D.V. Lang and L.C. Kimerling, *Ann. Rev. Mater. Sci.* **7** (1977) 377.

[2]   A. Hallen and M. Bakowski, *Solid-State Electron*, **32** (1989) 1033.

[3]   F.D. Auret and P.N.K. Deenapanray, *Crit. Rev. in Sol. State and Mater. Sci.*, **29** (2004) 1.

[4]   M. Lannoo and J. Bourngoin, *Point Defects in Semiconductors I, Theoretical Aspect* Springer series in solid state science 22, (1981).

[5]   W. Fank, *Inst Phys. Conf. Ser.* **22**, (1975) 23.

[6]   A. Antonelli, E. Kaxiras and D.J. Chadi, *Phys. Rev. Lett*. **81** (1998), 2088.

[7]   S.A. Centoni, B. Sadigh, G.H. Gilmer, T.J. Lenosky, T.D. de la Rubia and C.B Musgrave, *Phys. Rev. B* **72** (2005) 195206.

[8]   G.D. Watkins, *Mater. Sci. Semicond. Processing* **3** (2000) 227.

[9]   J. Coutinho, R. Jones, V.J.B. Torres, M. Barroso, S. Oberg and P.R. Briddon, *J. Phys.: Condens. Matter.* **17** (2005) L521-7.

[10]   C. Janke, R. Jones, S. Oberg and P.R Briddon, *Phys. Rev B* **75** (2007) 195208.

[11]   A. Mesli, L. Dobaczewski, K. Bonde Nielsen, V.L. Kolkovsky, M. Christian Petersen, and A. Nylandsted Larsen, *Physical Review B* **78** (2008) 165202.

[12]   A.N. Larsen and M. Mesli, *Physica B*, **401-402** (2007) 85-90.

[13]   V. Rianeri, G. Fallica and S. Libertino, *J. Appl. Phys*. **79** (1996) 9012.

[14]   G.D. Watkins, in: Hulin (Ed.), *Radiation Damage in Semiconductors*, Dunod, Paris, 1964, p97.

[15]   G.D. Watkins, *Phys. Rev. B* **12** (1975) 5824.

[16]   T. Markvart, D.P. Parton, J.W. Peters and A.F.W. Willoughby, *Materials Science Forum* **143-147** (1994) 1381.

[17]   W. Schockley and W.T. Read, *Phys. Rev*. **87** (1952) 835.

[18]   J. Lindhard, V. Nielsen, M. Scharff, and P.V. Thompsen, *Kgl. Dan. Vid. Selsk, Mat. Fys. Medd.* **33** (1963) 10.

[19]   J. Lindhard and M. Scharff, *Phys. Rev.* **124** (1961*)* 128.

[20]   S.J. Pearton, *Solid State Phenomena*, Vol. **1-2** (1988) 247.

[21]   J. Lindhard, M. Scharff, and H. Schiott, *Kgl. Dan. Vid. Selsk, Mat. Fys. Medd.* **33** (1963) 14.

[22]   J.P. Biersack and L.G. Haggmark, *Nucl. Instrum. Methods* **174** (1980) 257.

[23]   J.F. Ziegler *"The Stopping and Range of Ions in Matter"* Vol 2-6, Pergamon Press, 1977-1985

[24]   M. Lannoo, J. Bourngoin, *Point Defects in Semiconductors II, Experimtal Aspect* Springer series in solid state science 35, (1983).

[25]   H. H. Andersen, *Appl. Phys.* **18**, (1979) 131.

[26]   G.H. Kinchin and R.S. Pease, *Rep. Prog. Phys.* **18** (1955) 1.

[27]   A. Mesli, L. Dobaczewski, K. Bonde Nielsen, V.L. Kolkovsky, M. Christian Petersen and A. Nylandsted Larsen, *Physical Review B* **78** (2008) 165202.

[28]   Dong-Zhi Chi and S. Ashok, *Mat. Res. Soc. Symp. Proc.* 442 (1983).

Chapter 4

DLTS and Laplace-DLTS Aspects

4.1 Introduction

Deep level defects can be detrimental to or enhance the operation of devices fabricated on semiconductors as discussed in the earlier chapter. Therefore, it is essential to develop a sensitive experimental tool for characterizing the deep level defects in a semiconductor. The deep-level transient spectroscopy (DLTS) is a high-frequency (1 MHz) transient capacitance technique, which has proved to be a very useful tool to probe the defects close to the semiconductor surface since it was first discovered by Lang in 1974 [1,2]. Recently, high-resolution Laplace-DLTS (LDLTS) [3-4], which greatly enhances the resolution and spectroscopic nature of capacitance based defect characterization tools has been developed. In this chapter the emission and capture of carriers form deep level defects is discussed in section 4.2, the DLTS and LDLTS theory is presented in sections 4.3 and 4.4 respectively. The electric field effect on the deep levels is discussed in section 4.5.

4.2 Emission and Capture of Carriers from Deep Levels

The fabrication and development of efficient semiconductor devices require prior knowledge of the properties of deep levels as carrier traps or generation-recombination centers. As mentioned in the previous chapter, the generation-recombination centers can reduce minority carrier lifetime and diffusion length which may, for example, limit bipolar transistor performance [5], reduce the efficiency of photovoltaic cells [6] or increase the switching speed in semiconductor switches. A defect level can be defined as an electron trap if, when it captures an electron from the conduction band the electron stays there until it is re-emitted back to the conduction band. This may occur for an empty 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. $c_n \gg c_p$. Conversely, a recombination center is one for which c_n and c_p are almost similar, i.e. $c_n \approx c_p$. According to Shockley and Read [7] a deep level almost always changes its electron occupancy via carrier transitions between the level and the bands. Figure 4-1 shows the four common processes, neglecting transfer between the deep levels. The figure shows a trap which may exist in either two states, negative or neutral. Similar treatment is possible for other pairs of two states, e.g. neutral and positive or negative and double negative. If a trap is neutral it may capture an electron from the conduction band Fig. 4-1 process (a) or it may capture an electron from the valence band process (d) leaving behind a hole (hole emission). Processes (b) and (c) are electron emission and hole capture respectively.
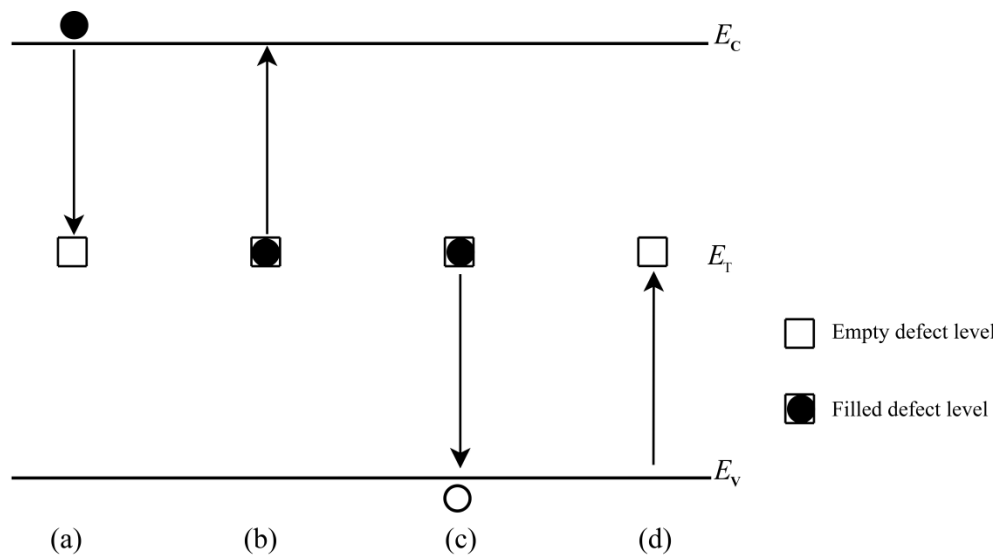

Fig. 4-1. Schematic representation of transitions of carriers between deep states (E_T) and the valence (E_V)- and conduction (E_C)- bands, neglecting the transfer between deep levels (a) electron capture, (b) electron emission, (c) hole capture, (d) hole emission. The arrow indicates the transition of the electron in the process (redrawn from ref. 7).

The kinetics which governs the charge transfer between the deep level and the bands are fully described by the Shockley-Read-Hall [7,8] (SRH) model. The model is developed assuming thermal (or near thermal) equilibrium and studying of deep level by DLTS uses the perturbation of the occupancy of the levels and then monitoring the return to equilibrium. The electron capture rate is given by

$$c_n = \sigma_n \langle v_n \rangle n \tag{4.1}$$

where σ_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 which is given by

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

where m^* is the effective mass of the electron, k is the Boltzmann constant, and T is the temperature in Kelvin. Similarly, the hole capture rate is given by

$$c_p = \sigma_p \langle v_p \rangle p \tag{4.3}$$

where σ_p, is the defect's hole capture cross-section, p is the hole concentration and $\langle v_p \rangle$ is the thermal velocity of the holes. A similar expression to that of $\langle v_n \rangle$ in equation (4.2) can be written for $\langle v_p \rangle$. The thermal emission rate e_n, of electrons from traps to the conduction band is proportional to the Boltzmann factor exp($-E_T/kT$), and can be expressed as a function of temperature by [5,6,9]

$$e_n(T) = \frac{\sigma_n \langle v_n \rangle N_C}{g} \exp\left[-\frac{E_T}{kT} \right] \tag{4.4}$$

where E_T, is the energy level below the conduction band minimum (also referred to as the defect activation energy if one assumes that σ is temperature independent), g is the degeneracy of the defect level, T is the temperature in Kelvin, N_C is the effective density of states in the conduction band given by

$$N_C = 2M_C \left(\frac{2\,\pi m^* kT}{h^2} \right)^{3/2} \tag{4.5}$$

here M_c is the number of conduction-band minima, h is Planck's constant.

An analogous expression can be written for hole emission rate e_p, to the valence band. In equation (4.4), the terms $\langle v_n \rangle$ is proportional to $T^{1/2}$, and N_C is proportional to $T^{3/2}$, while σ_n may or may not be temperature dependent, thus the product $\langle v_n \rangle N_\mathrm{C}$ has T^2 dependence. If e_n is measured as a function of temperature and ignoring the temperature dependence of σ_n, an Arrhenius plot of $\log(e_n/T^2)$ against $1/T$ is a straight line which yields E_T from the slope and σ_{na}, the apparent capture cross-section (from intercept at $T^{-1} = 0$). The parameters E_T and σ_{na} are often referred to as the '*defect signature*'. If a temperature-dependent capture cross-section is assumed then it usually takes the form [4]

$$\sigma_n(T) = \sigma_\infty \exp\left[\frac{\Delta E}{kT}\right] \tag{4.6}$$

where σ_∞, is the capture cross-section extrapolated to $T = \infty$ and ΔE_σ is the thermal activation energy of the capture cross-section (i.e. thermal barrier for carrier capture). The cascade capture into shallow levels and multiphonon capture into deep levels are some of the possible contributing factors to capture cross-section temperature dependence [10]. The temperature dependence of the capture cross-section may be determined from the plot of $\log(\sigma_\mathrm{n})$ vs. $1/T$, where ΔE_σ, is extracted from the slope and σ_∞ after extrapolation to $T = \infty$. Thus, the modified activation energy for a deep level which exhibits a temperature-dependent capture cross-section can be written as,

$$\Delta E_a = E_T + \Delta E \tag{4.7}$$

The modified activation energy has two components; (i) the energy difference between the trap level and the bottom of the conduction band E_T, and (ii) the thermal activation energy of the capture ΔE_σ, as depicted in Fig. 4-2. A more general expression of the thermal emission rate is now given by,

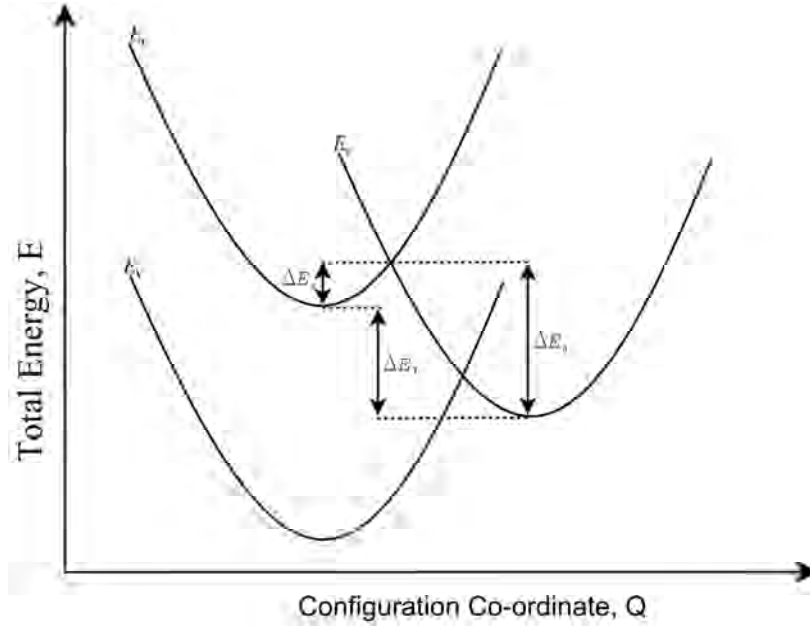$$e_n(T) = \frac{\sigma_n \langle v_n \rangle N_C}{g} \exp\left[-\frac{E_T + \Delta E}{kT}\right] \tag{4.8}$$

Fig. 4-2. Configuration co-ordinate (CC) diagram depicting the energy level of the defect below the conduction band ΔE_T (=E_T), the thermal activation energy of the capture cross-section ΔE_σ, and the total energy an electron requires to escape from the trap level to the conduction ΔE_a, redrawn from ref. 6.

In this study, the defect characterization is based on E_T (the position of the defect level from the band edge) and σ the apparent capture cross-section, therefore great care should be taken when these thermal emission measurements are compared with results from other techniques, e.g. optical measurements.

The physical meaning of E_T is that it is the Gibbs free energy change for the ionization of the state given by [11]

$$E_T = \Delta H - T\Delta S \tag{4.9}$$

where ΔH and ΔS are the changes in enthalpy and entropy due to the change in charge state of the level. Substituting equation (4.9) into (4.4) yields

$$e_n(T) = \frac{_n\langle v_n \rangle N_C}{g} \exp\left[-\frac{\Delta S}{k}\right] \exp\left[-\frac{\Delta H}{kT}\right] \tag{4.10}$$

Therefore, the Arrhenius plot yields the activation enthalpy of the deep level, and not the free energy, which can only be determined from optical measurements [6,9].

4.3 Deep Level Transient Spectroscopy (DLTS)

The conventional deep level transient spectroscopy (DLTS) is a powerful high frequency (MHz range) capacitance transient thermal scanning technique that is used to probe the space-charge region of a p-n junction, Schottky diode or MOS device structure. This technique is based on the transient capacitance change associated with the thermal emission of charge carriers from a trap level to thermal equilibrium after an initial non-equilibrium condition in the space-charge region. The DLTS technique offers the following characterization features:

(a) High sensitivity and good resolution.

(b) Straight forward, easy analysis of spectra and rapid scanning.

(c) Capability of measuring over a wide range of depths in the forbidden gap and detection of very shallow levels.

(d) Spectroscopic nature (i.e. signals due to different traps can be resolved from one another) [1].

A DLTS scan reveals each trap by a positive or negative peak on a flat base-line plotted as a function of temperature. The sign of each peak indicates whether it is due to a majority- or minority- carrier trap and positions of the peaks are simply and uniquely determined by the instrument rate-window and the thermal emission properties of the respective trap [1]. It is also possible to extract the thermal emission rate, activation energy, concentration profile and capture cross-section of each trap from the DLTS measurements. Since all the experimental work on defect levels studies in this study was based on the depletion region formed by the Schottky diode on a semiconductor, the discussion below will be confined to capacitance transient within the space-charge region of a Schottky barrier diode.

4.3.1 Capacitance Transient Processing

The relationship between capacitance and depletion width for a Schottky barrier diode has been dealt with in detail in chapter 2. The depletion width W is given by

$$W = \sqrt{\frac{2\varepsilon_s}{qN_D}\left(V_{bi} - V_a\right)} \qquad (4.11)$$

where, N_D is the density of ionized impurities due to dopants and other defects with levels in the band gap, q is the electronic charge, ε_s is the permittivity of the semiconductor, V_{bi} is built-in potential, and V_a is an externally applied voltage. The corresponding junction capacitance is

$$C = A\sqrt{\frac{q\,\varepsilon_s N_D}{2(V_{bi} - V_a)}} = \frac{A\,\varepsilon_s}{W} \qquad (4.12)$$

where A, is the area of the junction. The capacitance of the depletion region depends on the applied bias voltage and the dopant concentration as shown in equations (4.11) and (4.12). It is the sensitivity of the capacitance to the change in charge in the depletion region that is exploited and forms the basis of DLTS.

In the derivation of the depletion width equation, the depletion approximation has been used, which assumes that the semiconductor can be divided into two distinct regions, i.e. the bulk region which is electrical neutral and the space charge region which is depleted of charge carriers. In a real junction, as depicted in Fig. 4-3, there is a region λ, which lies between the truly depleted region and the bulk region. This region is defined as the distance between the depletion region edge and the point where the deep levels E_T crosses the Fermi level E_F. Fig. 4-3 also illustrates the energy band diagrams and space charge for a metal-n-type semiconductor with deep donors for (a) unbiased junction and (b) after applying a bias voltage V_a. In equilibrium, and under zero bias, deep levels in the region λ are below the Fermi level and therefore filled with carriers (a). After applying a reverse bias V_a the depletion region increases, the space charge region is altered, thus decreasing the capacitance of the depletion region. The deep levels N_T and shallow dopants N_D contribute to the charge density ρ in the depletion region and only the shallow dopants N_D are the source of the charge density in the region λ.
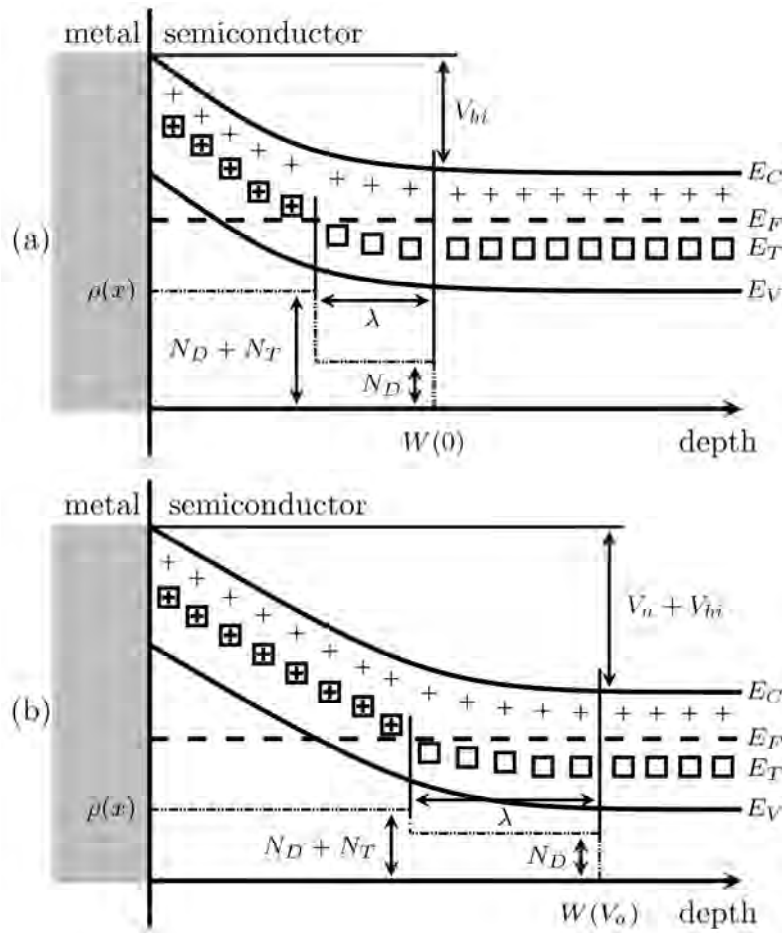
Fig. 4-3. Energy band diagram, the region λ, and space charge for an n-type metal-semiconductor junction with deep donor levels for (a) an unbiased and (b) after applying a quiescent reverse bias of V_a. For each condition the corresponding charge density ρ distribution is also shown, after ref. 14.

If the concentration of holes and electrons trapped at the deep levels is altered, (by say thermally stimulated emission of the carriers to the conduction or valence band), then this change can be monitored by measuring the junction capacitance or capacitance transient at constant applied bias voltage [6,12]. To ensure that the deep levels are filled with charge carriers after thermal stimulated emission, continuous refilling of the deep levels is achieved by application of a repetitive voltage filling pulse superimposed on a constant reverse bias voltage. The variation of the depletion region width and capacitance after the application of a voltage bias and a filling pulse sequence for majority and minority carrier traps in n-type semiconductor (e.g. n-type Ge) is depicted in Fig. 4-4 and Fig. 4-5 respectively. For simplicity, the bending of the bands due to the electric field in the space charge region has not been indicated. Also

the lambda effect has been ignored, so it is assumed that the defect levels in the depletion region are above the Fermi level and those deeper than the depletion region are beneath the Fermi level.
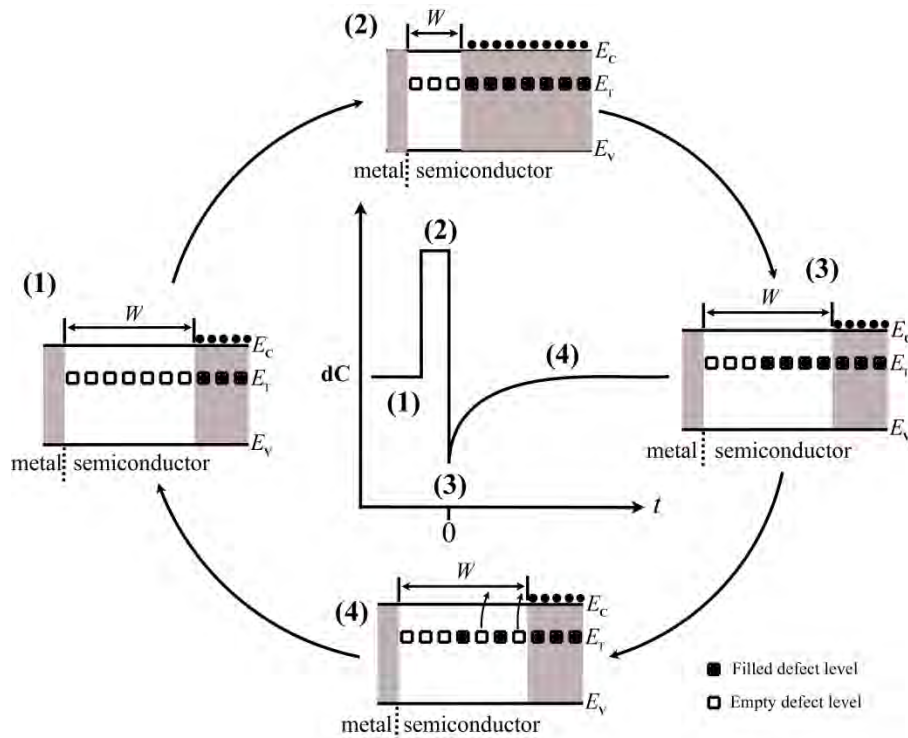

Fig. 4-4. Variation of the depletion region width and capacitance after the application of a voltage bias and a filling pulse sequence for a majority carrier (electron) trap in n-type semiconductor, after ref. 6.

The capacitance transient resulting from the pulse sequence is shown in the center. Under a quiescent reverse bias V and steady state, Fig. 4-4 part (1), the deep levels under the Fermi level are assumed filled and those above are empty as governed by the Fermi distribution function. The empty deep levels in the band gap are indicated by empty squares. After applying a majority carrier filling pulse in part (2), the depletion width is reduced, trapping electrons in those levels that are now below the Fermi level, symbolized by the solid squares. It is assumed here that the pulse width t_p is long enough to allow the complete filling of the trap levels. There is a corresponding step increase in capacitance because of the reduced depletion width. Immediately after the pulse is removed and the quiescent reverse bias V restored, part (3), the filled states lie within the depletion region, above the Fermi level, therefore the levels will start emitting the trapped carriers with a characteristic rate to the

conduction band where they are instantaneously swept away by the junction electric field, part (4).
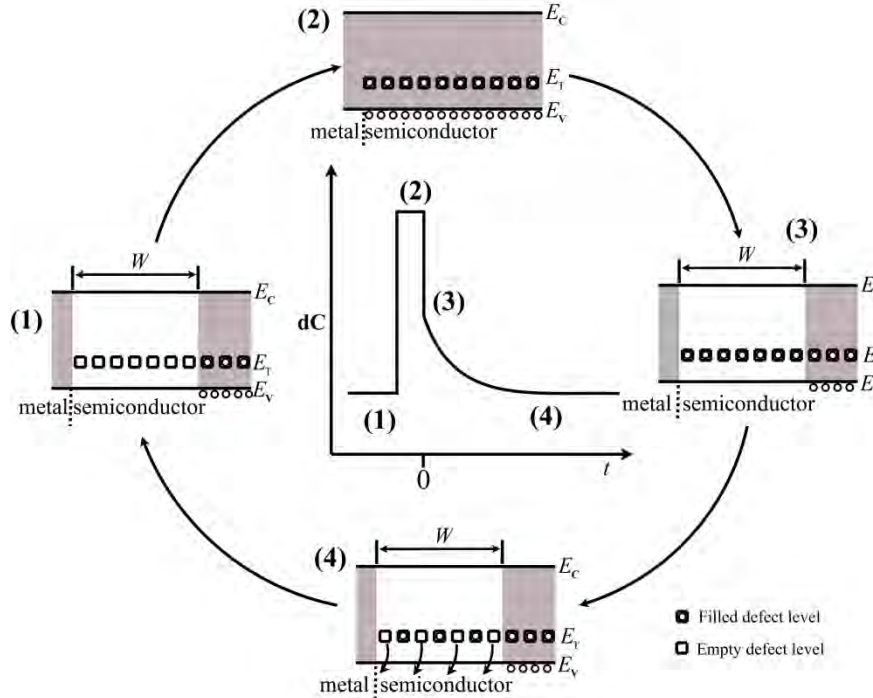

Fig. 4-5. Variation of the depletion region width and capacitance after the application of a voltage bias and a filling pulse sequence for a minority carrier (hole) trap in n-type semiconductor.

The capacitance variation as the trapped carriers are emitted to the conduction band is the so-called majority capacitance transient. The emission rate can be determined from the time dependence of the capacitance transient. The density of occupied defect levels at time t after removing the filling pulse is given by [6]

$$N(t) = N_T \exp(-e_n t) \qquad (4.13)$$

where e_n is the electron thermal emission rate and N_T is the trap concentration junction capacitance, if $N_T \ll N_D$, can then be given by an exponential time varying function as

$$C(t) = C_o - \Delta C_o \exp(-e_n t) \qquad (4.14)$$

here C_o is the equilibrium reverse bias capacitance and ΔC_o the change in capacitance immediately after the removal of the pulse, i.e. at $t = 0$ as shown in Figs. 4-4 and 4-5. When a large enough pulse is applied such that the junction is forward biased, (minority carrier injecting pulse), Fig. 4.5 part (2) and part (3), minority carriers (holes in this case) are trapped. After removing the pulse and the quiescent reverse bias V restored, part (4), the trapped carriers are emitted to the valence band giving rise to the minority carrier capacitance transient, which is of opposite sign to the majority carrier transient.

4.3.2   DLTS Principles

The utility of DLTS is in the processing of the capacitance transients obtained after repeated pulsing sequence discussed in section 4.3.1. The basic idea of DLTS method can be represented by the illustration in Fig. 4-6. The system response occurs only when the emission rate of the trap falls within the 'rate window'. For a given rate window, the system response is shifted to higher temperatures for a trap with higher emission rate and hence the system can resolve signals from different traps as a function of temperature, as shown in Fig. 4-6.
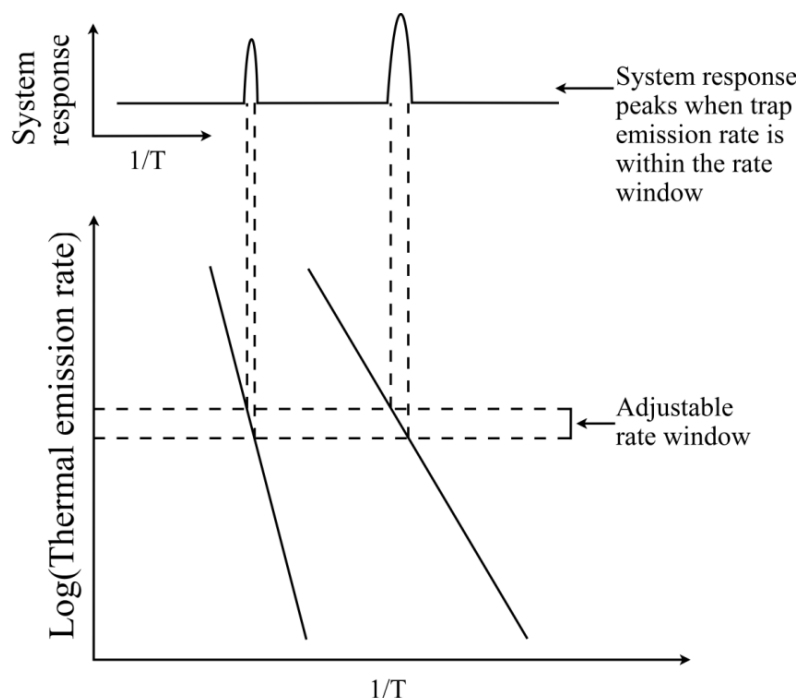

Fig. 4-6. A Schematic illustration of how a rate window produces a peak in its response when the emission rate of the input signal matches the rate selected by the window, redrawn from ref. 1).

In consequence, DLTS has the ability to set up an emission 'rate window' so that the measurement system gives an output only when a transient with a rate within this narrow window occurs. Since the emission rate is strongly temperature dependent, a thermal scan can reveal the presence of different traps at characteristic temperature when their emission rates coincide with the window. Most early DLTS systems used the dual-gated (double boxcar) signal filter for determining the rate window and averaging transients to enhance the signal-to-noise ratio (SNR) of the output, enabling low concentration defects to detected [1,9]. As the temperature is scanned, the filter takes samples of the transients at preset times t_1 and t_2 and produces an output proportional to their average difference, as shown in Fig. 4-7(a). Thus, the rate window is determined by the values of t_1 and t_2. The output signal changes from a small response as the decay time constant ($\tau = 1/e$) moves into the range detectable by the filter, to a maximum, and then drops as the decay time constant again falls outside the filter detectable range, giving rise to a DLTS spectrum depicted in Fig. 4-7(b). The early analog filter design by Lang [1] had an intrinsic dc rejection mechanism, which would give a zero output on the filter when no defect is detected. The normalized DLTS signal $S(T)$ is defined by

$$S(T) = \frac{C(t_1) - C(t_2)}{\Delta C(0)} \tag{4.15}$$

where $C(t_1)$ is the capacitance at t_1, $C(t_2)$ is the capacitance at t_2, and $\Delta C(0)$ is the capacitance change due to the filling pulse at $t = 0$ as described in figures 4-4 and 4-5. The position of the peak on the temperature axis depends on the rate window, e.g. a smaller rate window will shift a defect peak to lower temperatures. If DLTS spectra is produced by using different rate windows, then a series of spectra are produced as shown in Fig. 4-8(a). The emission rate at a maximum peak height is a uniquely defined and can be calculated via the expression for the time constant, (τ_{max}) at maximum peak height

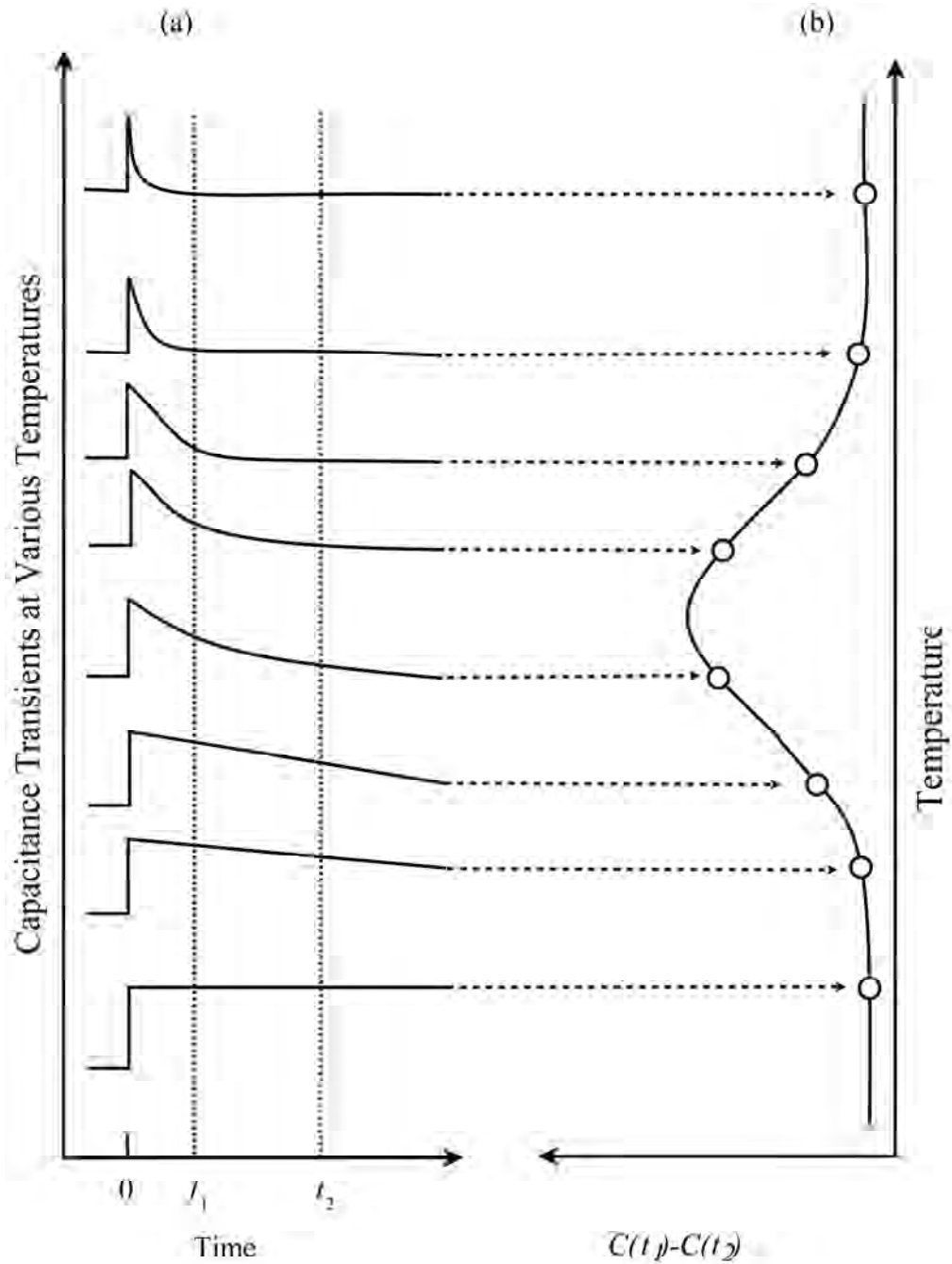$$\tau_{\text{max}} = \frac{t_1 - t_2}{\ln\left(\dfrac{t_1}{t_2}\right)} \tag{4.16}$$

Fig. 4-7. A schematic diagram which shows how a rate window concept can produce a DLTS spectrum. Part (a) shows the capacitance transients at different temperature which after processing with a double boxcar, and thereby resulting in the spectra shown on part (b), after ref. 1.
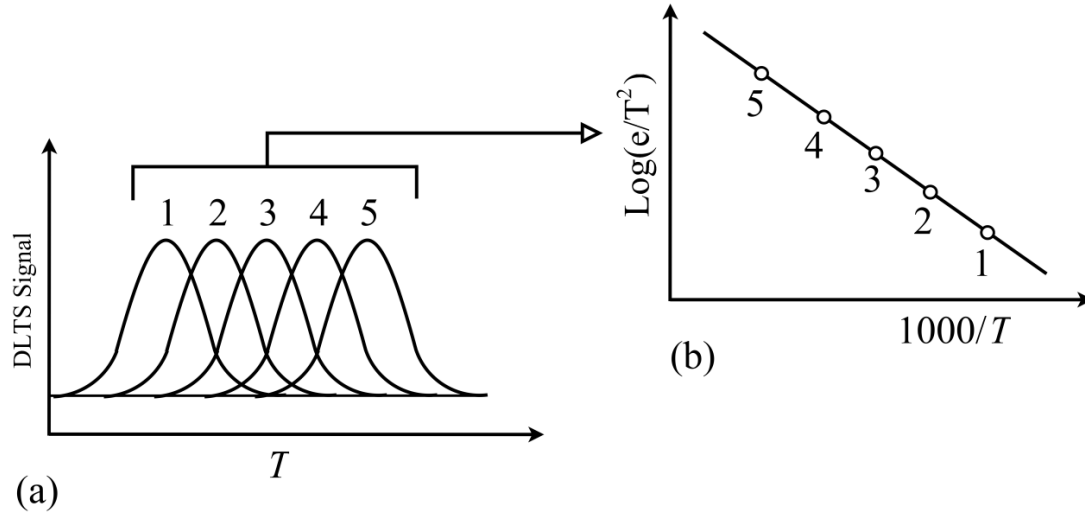
Fig. 4-8. A diagram showing the (a) DLTS spectra at various rate windows and (b) the Arrhenius plots obtained from the spectra, after ref. 1.

For each spectrum, the temperature at the maximum peak height can be measured and the emission rate e, for which the DLTS system shows a maximum response, is calculated via equation (4.16). These points are then used to plot a semi-log graph of $\log(e/T^2)$ vs. $1000/T$, (Arrhenius plot), shown on Fig. 4-8(b) from which the defect activation energy E_T and apparent capture cross-section σ, are extracted.

An alternative to the double boxcar weighting function used in the original Lang's work [1] is to use a lock-in amplifier (LIA) [13]. In a LIA set-up the rate window is set by altering the frequency of the filter. The LIA response to this transient is the integral product of the capacitance signal and the weighting function $w(t)$ given by

$$S(\ ) = \frac{1}{}\int_0 C(t)w(t)dt \tag{4.17}$$

where $w(t) = \sin\left(\frac{2\ t}{}\right)$ is a sine wave of fixed frequency. The LIA gives the same result to that of the double boxcar method. For an exponential transient, using a sine wave weighting function the DLTS signal reaches a maximum when $\ = \left(\frac{1}{0.423}\right)$.

4.3.3 Defect Depth Profiling

The DLTS peak height is directly proportional to the concentration of a deep level, therefore the concentration can be obtained directly from the capacitance change corresponding to completely filling the trap with a saturation minority carrier pulse (in case of a minority carrier trap) or the largest possible majority-carrier pulse (in case of a majority-carrier trap). DLTS enables one to determine the defect spatial distribution within the semiconductor and thereby other parameters such as the introduction rate.

The concentration of deep levels N_T is often calculated by using simple expression [1,14]

$$N_T = \frac{2\Delta C(0)}{C} N_D \tag{4.18}$$

where N_D is the concentration of shallow impurities, C is the junction capacitance under quiescent reverse-biased conditions, and $\Delta C(0)$ is the capacitance change due to the pulse at t = 0 (i.e. just after removing the filling pulse). Equation (4.18) is only applicable if the minority carrier pulse or majority carrier pulse is large and long enough to completely fill the trap and when $\Delta C(0) \ll C$. The correct pulse for defect concentration determination can be checked by making several scans with increasing larger and longer pulses until the defect peak no longer increases in size. As has been pointed out in the past [1,14] using equation (4.16) sometimes results in significant underestimation of N_T especially for thin films and at low reverse bias voltages. In order to determine the corrected expression for N_T one has to consider the region λ (the so called λ effect), where the defect level crosses the Fermi level a distance λ shallower than the depletion region edge as shown in Fig. 4-3(b). The traps in this region are occupied and do not contribute to the change in capacitance when a filling pulse is applied. The width of the transition is given by [14]

$$= \left[\frac{2\ (E_F - E_T)}{q^2 N_D} \right]^{1/2} \tag{4.19}$$

where ε is the semiconductor dielectric constant, E_F is the Fermi level and q is the electronic charge. To obtain the deep level distribution profile, the deep levels in the region to be profiled must be filled with carriers. The depth profiling technique used in this study uses a fixed bias voltage and a variable filling pulse (fixed bias-variable pulse method) [15]. In the fixed bias-variable pulse method the incremental change in capacitance $\delta(\Delta C)$ is monitored as the majority carrier pulse V_p is changed by a small amount δV_p. The relative incremental change in capacitance due to the pulse increment can be expressed by [14]

$$\Delta\left(\frac{\Delta C}{C}\right) = \left(\frac{}{qw^2 N_D}\right)\frac{N_T(x)}{N_D(x)}\Delta V_p \tag{4.20}$$

where x is the depth below the junction, N_D and w are the ionized shallow impurity concentration and depletion region width, respectively, corresponding to quiescent reverse biased conditions. The shallow impurity profile $N_D(x)$ is obtained from C-V measurements. The total signal due to the majority carrier pulse, is then determined by double integration of the Poisson equation according to a detailed derivation by Zohta and Watanabe [14], to give the corrected deep level concentration expression as

$$N_T = \frac{2\Delta C(0)N_D(x)}{C}\left[\left(\frac{x-}{x}\right)^2 - \left(\frac{x_p - {}_{p}}{x}\right)^2\right]^{-1} \tag{4.21}$$

here $x - \lambda$ and $x_p - \lambda_p$ are the depletion region widths before and after applying a filling pulse respectively, and λ is distance from where the deep levels cross the Fermi level to the depletion region edge and λ_p is the value of λ during the pulse. In the limit and low noise measurements, values of 10^{-5}-10^{-6} for $\Delta C/C$ can be achieved and if the shallow dopants concentration is $N_D \approx 10^{16}$ cm^{-3}, a low defect concentration of the order of 10^{10} cm^{-3} is detectable.

4.4 Laplace-DLTS

Since its discovery almost 35 years ago, conventional deep level transient spectroscopy (DLTS) has been a valuable tool in identifying deep level states in

semiconductors, thereby improving device efficiency. Unfortunately this technique has limitations in the details of information on the identity of defects it can measure due to its poor emission rate resolution. The double boxcar or lock-in-amplifier filter used in DLTS exhibit good sensitivity but has poor time constant resolution. Due to this poor time constant resolution, DLTS cannot be used to separate closely spaced transients, and thus its inability to study defect fine structure. Over the past 25 years there has been much effort applied to improve DLTS resolution by developing different weighting functions [16]. These higher order filters showed an improvement of resolution by a factor of up to 3 but at the expense of noise performance.

In 1990, Dobaczewski *et al* [3,4] developed an improved high-resolution version of DLTS, called Laplace-DLTS (LDLTS). This new concept is an isothermal DLTS technique and employs a regularized inverse Laplace transform instead of the conventional boxcar analysis. This results in an order of magnitude improvement in emission rate resolution in the studies of the thermal emission of carriers from deep states. Consequently, LDLTS can separate closely spaced transients (with emission rates differing by a factor greater than 2) when a number of defects with similar emission characteristics are present, thereby overcoming the major deficiency of DLTS. Apart from the remarkable sensitivity (ability to measure very low concentrations of defects), LDLTS can probe very narrow regions of the semiconductors (e.g. regions of shallow implants) and can also be used to study selectively the active regions of devices.

4.4.1   Laplace-DLTS Principles

Generally, in DLTS there are two main classes of transient processing methods, which are analog and digital signal processing. Analog signal processing is a real-time process which involves extracting the capacitance transients as the temperature is ramped. An Analog filter will then produce an output proportional to the signal input at a particular time constant range. In order to increase the resolution and sensitivity of DLTS, several different filters have been investigated which include, boxcar [1], lock-in amplifier [13], exponential [17] and multiple boxcar [18]. The digital signal processing involves digitizing the transient output of the capacitance meter, normally done with sample held at a fixed temperature and averaging many of these digitized transient to reduce noise. If the transient is digitized, then it is much easier to apply

signal processing tasks, even complex ones. The concept of digitizing capacitance at constant temperature and extracting the time constant is the basis of high resolution Laplace-DLTS. The extraction of all accessible time constants from the transients is achieved by numerical algorithm. There is a well-documented problem associated with the extraction and separation of multiple, closely spaced time constants. The problem is finding a suitable choice of algorithm to use in the extraction of the time constants. The problem is due to several factors, including that (i) the exponential decay transient baseline is not known with any degree of precision; hence this becomes an unknown variable in the analysis, (ii) both polarities of the transient may be present simultaneously and (iii) exponential transients are not ideal (due to dependence of the emission rate on electric field or due to inhomogeneous strain which produces a continuum of emission rates for a particular defect). Several algorithms that have been developed in an attempt to solve the transient extraction problem have been reviewed [4,6]. The techniques considered include, "a method of moments" by Ikossi-Anastasiou *et al* [19], "a Gaver-Stehfest approximation algorithm to effect a Laplace transform" by Nolte *et al* [20], and "Tikhonov regularization method" to separate the constituent exponentials in a photo-induced current transient spectroscopy (PICTS) signal by Eiche *et al* [21]. The Tikhonov regularization method uses a similar approach to the technique employed with the Laplace-DLTS work discussed here. To develop an algorithm for transient processing, assume that the recorded transient *f(t)* is a non-exponential transient, which is composed of a superposition of exponential transients and is given by

$$f(t) = \int_0^\infty F(s)e^{-st}ds \tag{4.22}$$

where $F(s)$ is the spectral density function. The function *f(t)* given in equation (4.22) is the Laplace transform of the true spectral density $F(s)$. Thus, to find the true spectrum of emission rates in the transient, an inverse Laplace transform for the function *f(t)*, should be performed using some numerical method. For such a procedure, (assuming that all the decay transients are exponential and have the same sign), a spectrum of delta-like peaks is produced for multi-, or mono-exponential transients. All the numerical methods used in LDLTS attempt to find a spectral function with the least possible number of peaks, which is consistent with the data and

experimental noise [4]. Although the problem has been described in a general way, it should be noted that equation (4.22) does not have a general solution for any given function $f(t)$. For an analytical multi-exponential function such a solution exists and, according to Lerch's theorem [22] it is unique. However, if noise is superimposed on this function the number of solutions can be infinite. Therefore, the problem is to find the best estimate for $F(s)$, and according to the prior knowledge about the system being investigated and its boundary conditions, to exclude unphysical solutions and choose only the simplest one, i.e. the one that reveals the least amount of detail or information that was not already known or expected [3,6].
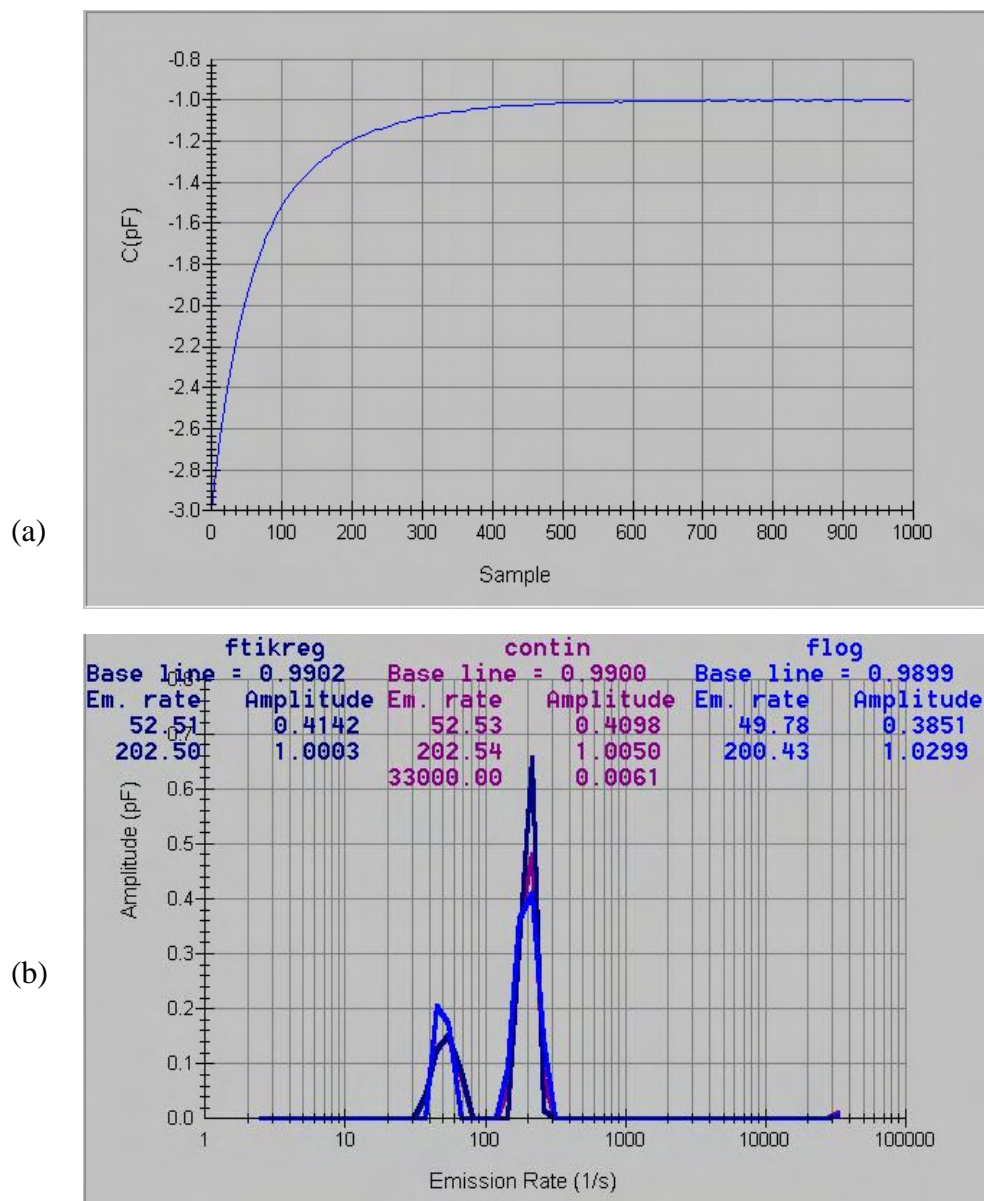
(a)

(b)

Fig. 4-9. (a) A majority-carrier capacitance transient and (b) the corresponding spectra obtained from the transient with use of the three numerical routines.

For Laplace transform inversion in the Laplace DLTS system used here, three numerical routines CONTIN [23], FTIKREG [24] and FLOG (which was specifically developed for Laplace DLTS by Matulis [25]), all of them based on the Tikhonov regularization method, however they differ in the way the principle for finding the regularization parameters are defined. The Laplace card sets the sample excitation parameters (or in some cases is used to trigger the external pulse generator, which supplies the biasing and pulsing conditions to the sample). The Laplace software then acquires the capacitance transient, shown in Fig. 4-9, graph (a) before the transient is converted into the LDLTS spectra depicted in graph (b) using the three numerical routines, i.e. CONTIN, FTIKREG and FLOG and from the spectra, the emission rates and magnitude of the signal can be measured.
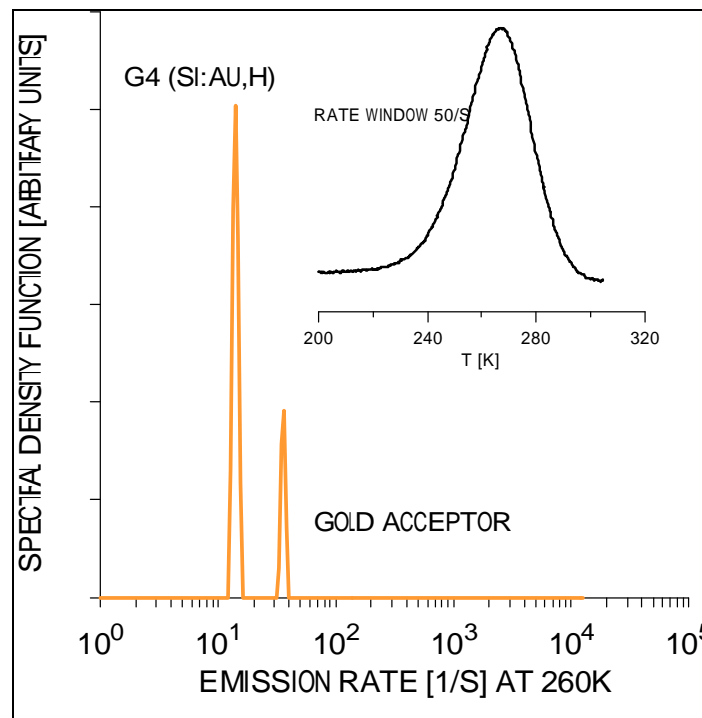

Fig. 4-10. DLTS and LDLTS spectra of hydrogenated silicon containing gold. The conventional DLTS spectrum is shown as an insert at the top of the figure. The broad peak centered at 260 K is attributed to electron emission from the gold acceptor G4. The main spectra were obtained by the Laplace technique and clearly separate the gold-acceptor level and the gold-hydrogen level G4 (redrawn from ref. 26).

The parallel use of three different numerical routines increases the confidence level in the spectra obtained. The evolution of LDLTS in past few years has enabled the theoretical limit of DLTS to be achieved. For relatively shallow states that emit at low temperature, the reduction in line width is remarkable and an increase in resolution of at least two orders of magnitudes is readily achievable but its sensitivity is about an order of magnitude less than that of DLTS. For example, if a sample with a trap concentration approximately 1% of the shallow dopant concentration and quiescent capacitance of about 10 pF is studied by LDLTS then the signal-to-noise ratio (SNR) of 1000 is readily achievable. This is necessary to separate defects with similar emission rates with time constant of 2 i.e. $(e_1/e_2) \approx 2$. An illustration showing a comparison of LDLTS (full figure) and DLTS (inset) spectra obtained from the same (Si:Au,H) sample is depicted in Fig. 4-10 [26]. DLTS shows a broad featureless spectrum of DLTS whilst the LDLTS resolves the broad DLTS spectrum into two peaks.

Finally, to completely characterize defect levels in the semiconductor by DLTS and LDLTS, both the majority and minority carrier traps should be identified. If the depletion layer is formed from a p-n junction then both majority and minority carriers can easily be injected into the depletion region by applying a correct filling pulse. However, when a Schottky diode is used to form the depletion region, a forward bias filling pulse will not always inject minority carriers into the depletion region, therefore minority carrier traps may not be observed. This is true for Si, but not entirely true for all semiconductors e.g. Ge. For Ge, metal-semiconductor Schottky barrier diodes with large barrier heights (in relation to the band gap) can be formed. It has been shown that for a high barrier height, an inversion layer with high concentration of minority carriers can be formed near the semiconductor surface [27,28]. When a forward filling bias is applied to such a diode it results in a flux of holes from the inversion layer to the semiconductor bulk. Thus the minority carrier traps can be filled, making them visible to the DLTS and LDLTS techniques. The minority carrier trap study using Schottky barrier diodes on Si is achievable if traps are filled by optical means. Throughout this work, the Schottky barrier diodes (SBDs) have been used because they are easy to fabricate, easy to use and quality diodes are possible.

4.5 Electric Field Effect

The DLTS and LDLTS techniques employed for defect characterization probe defects in the reverse-biased depletion region. Therefore the emission process takes place in the presence of a high electric field. For an n-type semiconductor, the magnitude of the electric field in the depletion region is given by

$$|E| = \sqrt{\frac{2qN_D}{\varepsilon}(V_{bi} - V_a)} - \frac{qN_D x}{\varepsilon} \qquad (4.21)$$

where q is the electronic charge, N_D is the shallow dopant concentration, V_{bi} the built-in-potential, V_a is the applied bias voltage and ε the permittivity of the semiconductor.
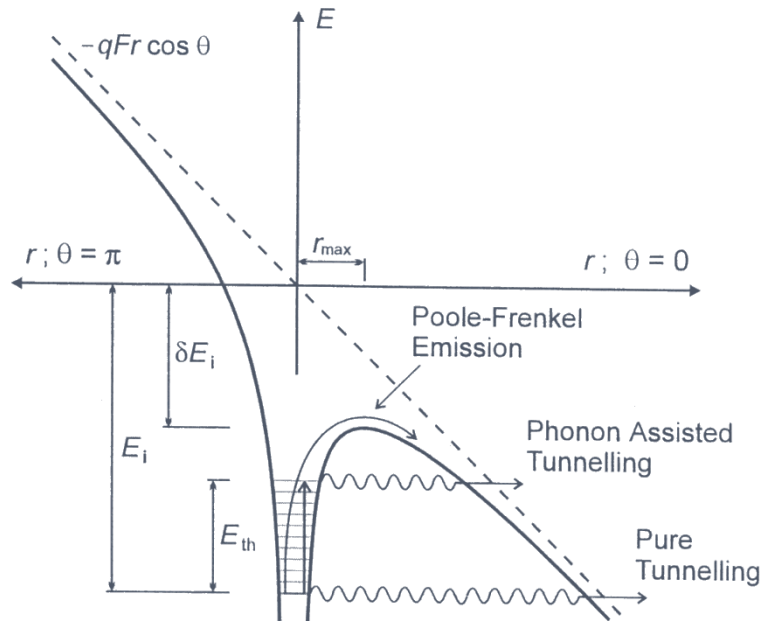

Fig.4-11. The Coulombic well and the three mechanisms of field enhanced emission; Poole-Frenkel emission, phonon assisted tunneling and pure tunneling where E_{th} is the position of the virtual level above the deep level, E_i is the ionization energy of the deep level and δE_i is the change in the ionization energy due to the electric field, redrawn from ref. 30.

The electric field in the depletion region can reach average values of 10^5-10^7 V/m, depending on the doping density and the bias voltage. Such high electric fields can influence the shape of defect potentials and may, therefore, enhance carrier emission

65

rates from potential wells, by the Poole-Frenkel effect [29,30,31] or phonon-assisted tunneling [6,32]. A potential well may be considered as a trap with significant spatial extent and can be described by different models, e.g. a Coulombic, square or Gaussian potential well. For a Coulombic potential well, the mechanisms of field-enhanced emission are depicted in Fig. 4-11. The Poole-Frenkel effect is a mechanism in which a carrier is thermally emitted over the top of the barrier, which has been lowered by the application of an electric field. The emission rate in the presence of the electric field F is now given by

$$e(F) = e(0)\exp\left(\frac{q^3}{kT}F^{1/2}\right)$$ (4.22)

where $e(0)$ is the zero field emission rate, q is the electron charge, T is the temperature in Kelvin and ε is the permittivity of the semiconductor. If a plot of $\log[e(F)]$ vs. $F^{1/2}$ is a linear plot, then it is experimental evidence of a charge leaving a trap of opposite sign. This implies a donor type defect in n-type material and acceptor type trap in p-type material. The Poole-Frenkel mechanism is dominant if the potential well has some appreciable spatial extent. The other mechanisms shown in Fig. 4-10 are phonon-assisted tunneling and pure tunneling. In the phonon-assisted tunneling process, a charge carrier absorbs thermal energy and is excited to a virtual state at E_{th} above the deep level. The electron will then be able to tunnel through the barrier from this virtual level to the conduction band. For any of the field-enhanced emission mechanisms, the electric field is spatially varying and so emission occurring from different positions in the depletion will be affected to a different extent, giving rise to non-exponential capacitance transients.

References

[1]   D. V. Lang, *J. Appl. Phys*. **45** (1974) 3023.

[2]   D. V. Lang, *J. Appl. Phys*. **45** (1974) 3014.

[3]   L. Dobaczewski, P. Kaczor, I.D. Hawkins and A.R. Peaker, *J. Appl. Phys*. **76** (1994) 194.

[4]   L. Dobaczewski, A.R. Peaker and K.B. Nielsen, *J. Appl. Phys*. **96** (2004) 4689.

[5]   P. Blood and J.W. Orton, *Rep. Prog. Phys*. **41** (1978) 11.

[6]   F.D. Auret and P.N.K. Deenapanray, *Crit. Rev. Sol. Stat Mater. Sci*. **29** (2004) 1.

[7]   W. Shockley and W.R. Read, JR, *Physical Rev*. **87** (1952) 835.

[8]   R.N. Hall, *Phys. Rev*. **83** (1951) 228.

[9]   G.L. Miller, D.V. Lang and L.C. Kimerling, *Ann. Rev. Mater. Sci*. **7** (1977) 377.

[10]   C.H. Henry and D.V. *Lang, Phys Rev. B* **15** (1977) 989.

[11]   V.P. Markevich, I.D. Hawkins, A.R. Peaker, K.V. Emtsev, V.V. Litvinov, L.I. Murin and L. Dobaczewski, *Phys. Rev. B* **70** (2004) 235213.

[12]   L.C. Kimerling, *J. Appl. Phys*. **45** (1974) 1839.

[13]   F.D. Auret, *Rev. Sci. Instrum*. **57** (1986) 1597.

[14]   Y. Zohta and M.O. Watanabe, *J. Appl. Phys*. **53 No.3** (1982) 1809.

[15]   D.V. Lang, In "Thermally Stimulated Relaxation of Solids" (P. Braunlich, ed.) pp93-133, Springer-Verlag, Berlin, 1979.

[16]   A.A. Istratov,.*J. Appl. Phys. 82* (1997) 2965.

[17]   J.A. Borsuck and R.M. Swanson, IEEE Trans. Electron Devices **27** (1980) 2217.

[18]   C.R. Crowell and S. Aliphani, *Solid State Electron* **24** (1981) 25.

[19]   K. Ikossi-Anastasiou and K.P. Roenker, *J. Appl. Phys*. **61** (1987) 182.

[20]   D.D. Nolte and E.E. Haller, *J. Appl. Phys*. **62** (1987) 900.

[21]   C. Eiche, D. Maier, M. Schneider, D. Sinerius, J. Weese, K.W. Benz, and Honerkamp, *J. Phys. Condens. Matter*. **4** (1992) 6131.

[22]   G.A. Korn and T.M. Korn, in mathematical handbook , (McGraw-Hill, New York) (1968).

[23]   S.W..Provencher, *Comput. Phys. Commun*. **27** (1982) 213.

[24]   J. Weese, *Comput. Phys. Commun.* **69** (1991) 99; ibid. **77** (1993) 429.

[25]   A. Matulis, Z. Kancleris, Semiconductors Physics Institute, Vilnius, Lithuania.

[26]   P. Deixler, J. Terry, I.D. Hawkins, J.H. Evans-Freeman, A.R. Peaker, L. Rubaldo, D.K. Maude, J.C. Portal, L. Dobaczewski, K.Bonde Nielsen, A. Nylandsted Larsen, and A. Mesli, *Appl. Phys. Lett.* **73** (1998) 3126.

[27]   V.P. Markevich, A.R. Peaker, V.V. Litvinov, V.V. Emstev, L.I. Murin, *J. Appl. Phys.* **95** (2004) 4078.

[28]   E.H. Rhoderick, *Metal-Semiconductor Contacts* (Clarendon Press, Oxford, 1978).

[29]   J. Frenkel, *Phys. Rev.* **54** (1938) 647.

[30]   S.D. Ganichev, E. Ziemann, W. Pretll, I.N. Yassievich, A.A. Istratov, E.R Weber, *Phys. Rev. B* **61** (2000) 10361.

[31]   N. Zangenberg, J. Goubet, A.N. Larsen, *Nuclear Instruments and Methods in Physics Research B* **186** (2002) 71.

[32]   S. Makram-Ebeid and M. Lannoo, *Phys. Rev. B* **25** (1982) 6406.