

9 Resonant Tunneling in Quantum Structures

Mihail Ion Lepsa

Peter Grünberg Institute (PGI-9)

JARA-Fundamentals for Information Technology

Research Center Jülich, 52425

Contents

1. Introduction	2
2. Resonant tunneling in semiconductor heterostructures	2
2.1 Resonant tunneling process	2
2.2 Semiconductor heterostructures	3
2.3 Physics of double barrier resonant tunneling	4
3. Molecular beam epitaxy growth of semiconductor heterostructures	10
3.1 Molecular beam epitaxy	10
3.2 MBE growth	15
4. Resonant tunneling diode: processing and electrical characterization	20
References	22

1. Introduction

Resonant tunneling in semiconductor structures have attracted much interest for both fundamental research and applications in high speed electronics. Based on a pure quantum process, the realization of the specific nanostructures was possible only after the developing of the molecular beam epitaxy (MBE), an advanced technique for the growth of high quality semiconductor thin films. Here, we give first a brief physical description of the resonant tunneling considering mainly the semiconductor layered systems. Then, basics of the MBE growth of III-V compound semiconductors with specific exemplification are presented. Finally, the processing and the electrical characterization of a double barrier resonant tunneling diode are shortly considered.

2. Resonant tunneling in semiconductor heterostructures

2.1 Resonant tunneling process

Tunneling is a quantum process in which a particle penetrates into and traverses a barrier region where its potential energy exceeds its initial, kinetic plus potential, energy. According to classical physics, conservation of the energy would not permit this to occur, particle being totally reflected by the barrier. It is possible in quantum physics, where the particle wave function, and hence its probability density, is spread out over both the classically allowed and classically forbidden parts of the path. Wave functions penetrate from allowed regions into adjoining forbidden region, for a limited distance, and this allows non-classical particle motions that can cross a sufficiently narrow barrier. In a simple picture, this is illustrated in Fig. 1a, where the particle, e.g. electrons, is described by plane waves which partly are reflected and partly are transmitted by the potential barrier. Tunneling process is present in many processes in the fields of atomic, nuclear and solid state physics.

In general, in solids, two categories of tunneling phenomena can be involved. Thus, one can speak about the elementary tunneling processes in atomic field, with “microscopic barrier”, invoked to explain certain transport phenomena in the bulk semiconductors and also in molecular crystals. On the other hand, there are the tunneling processes through macroscopic potential barriers or junctions (the interface of different materials). Concerning this aspect, it is interesting to point out that, most of the people in the field consider only the last category as referring to “tunneling phenomena in solids”. An explanation could be that the junction tunneling

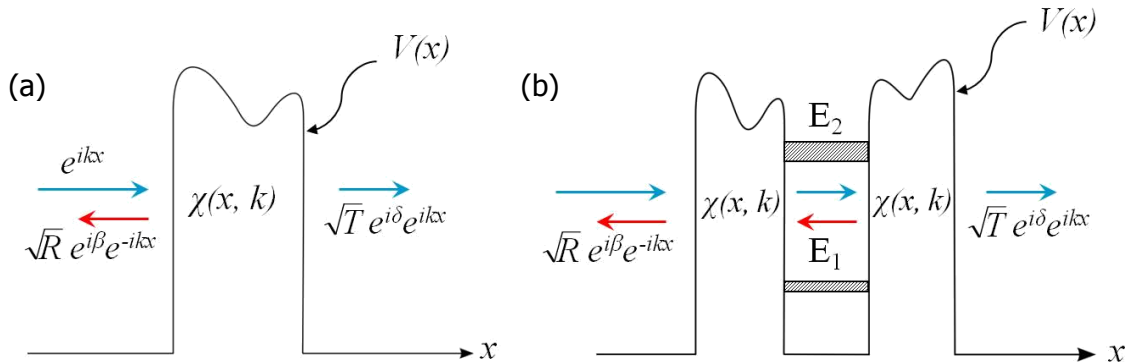


Fig. 1. Illustration of the tunneling (a) and resonant tunneling (b) processes.

processes have great importance for the electronic devices. In their turn, the junction tunneling processes can be also divided into two classes according to either the tunneling being in heterojunction systems, e.g. metal-insulator-metal, metal-insulator-semiconductor, superconductor-insulator-superconductor, Schottky junctions, etc., or in the same material (semiconductor) like interband tunneling.

Resonant tunneling is distinguished from the simple tunneling process by the presence within the classically forbidden region of quasibound, or metastable, states of the tunneling particle. The most typical situation is that one of the double barrier resonant tunnelling, which is involved in both nuclear and solid state phenomena (see Fig. 1b). In this case, two thin finite potential barriers clad a quantum well. Incident particles to the structure have unity tunnelling probabilities at energies corresponding to the quasi-states from the quantum well. The extension of the double barrier resonant tunneling problem to the more general case of a periodic rectangular potential (Kronig-Penney) offers a simple model of the behaviour of electrons in a crystal lattice. The electron energy spectrum along the periodic potential consists of continuous *bands* separated by forbidden gaps.

2.2 Semiconductor heterostructures

The advance of the epitaxial growth techniques, namely Molecular Beam Epitaxy (MBE) and Metal Organic Vapour Phase Epitaxy (MOVPE), made possible to grow very pure and thin semiconductor layers (especially from III-V compounds) with accuracy down to one atomic layer. As a consequence, the resonant tunneling process could be observed in structures consisting of stacked thin semiconductor layers.

The interface between two dissimilar crystalline semiconductors is called heterojunction. As opposed to a homojunction, the semiconducting materials have unequal band gaps. The band gap discontinuity is distributed between the valence and conduction bands. As seen in Fig. 2a, the corresponding band “offsets”, ΔE_V and ΔE_C , are abrupt discontinuities in the band edges at the heterojunction interface. Additionally, electrostatic band bending occurs due to charge redistribution and the adjustment of the Fermi level near the interface. The band bending changes (increasing usually) with doping, but the band offsets is presumably to remain constant. According to the relative ordering of the band edge energies we can distinguish several types of band alignments. The most common (and generally considered to be the “normal”) alignment is the *straddling* configuration, in which the bandgaps of the two materials entirely overlap (see Fig. 2b and c). Another situation is when the bandgaps are partial overlapped, i.e. the conduction band of the smaller-gap material might lie above that of the larger-gap material, or its valence band might lie below that of the larger-gap material. Such a band alignment is called *staggered* and is illustrated in Fig. 2d. The staggering might become so extreme that the bandgaps cease to overlap. This situation is known as a *broken gap* alignment, being illustrated in Fig. 2e. Another nomenclature, employed usually in describing superlattices, divides the band alignments in two main categories: type I and type II alignments. If the extrema of both the conduction and valence bands lie both in the same layers, the alignment is referred as type I, whereas if the band extrema are in different layers the alignment is type II. Comparing the two notations, one can easily observe that the type I alignments corresponds to the straddling case and the type II alignments include both staggered and broken-gap situations.

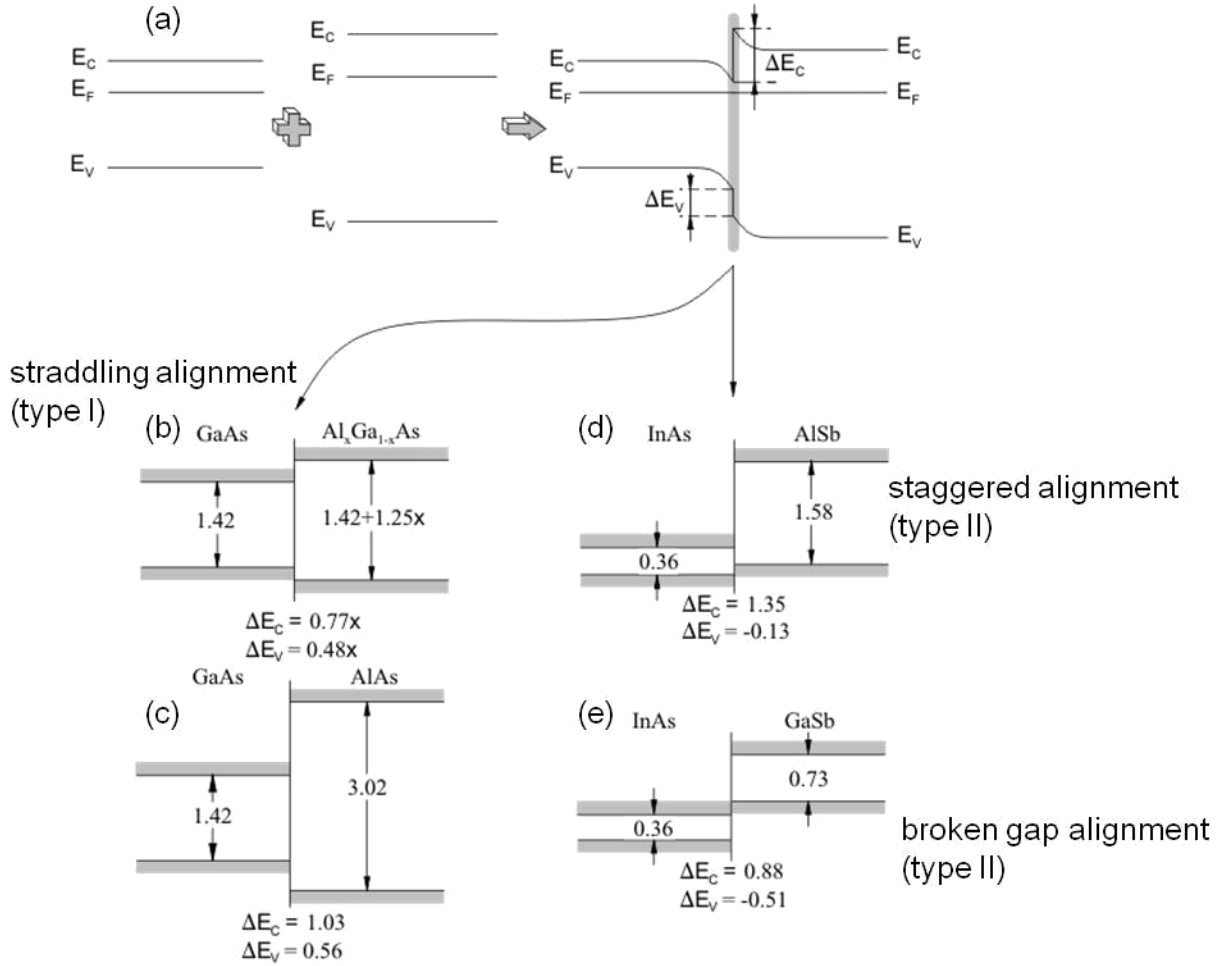


Fig. 2. Conduction and valence-band offsets in a semiconductor heterojunction (a) and experimentally determined band alignments for lattice-matched or nearly lattice-matched heterojunctions: (b) and (c) show straddling alignments, (d) is typical for staggered alignments and (e) corresponds to broken-gap alignment.

The combination of multiple semiconductor heterojunctions forms a heterostructure. By choosing specific semiconductor materials and layer thicknesses, potential barriers and/or quantum wells for the carriers can be obtained. The double barrier resonant tunneling (DBRT) heterostructure consists of two layers of semiconductor material, such as AlAs, embedded in another semiconductor having a smaller band gap, such as GaAs. The band diagram of this layer sequence represents a quantum well between two potential barriers (see Fig. 3). Two electron reservoir layers clad the basic structure resulting in the DBRT diode.

2.3 Physics of double barrier resonant tunneling

The DBRT devices operate on physical principles that are shared by all resonant-tunneling structures. As seen, the process of resonant tunneling implies the existence of metastable states in the potential barrier region. Thus, one can say that the fundamental requirement for this process to occur is quantization by spatial confinement. For the DBRT structures, the presence of the

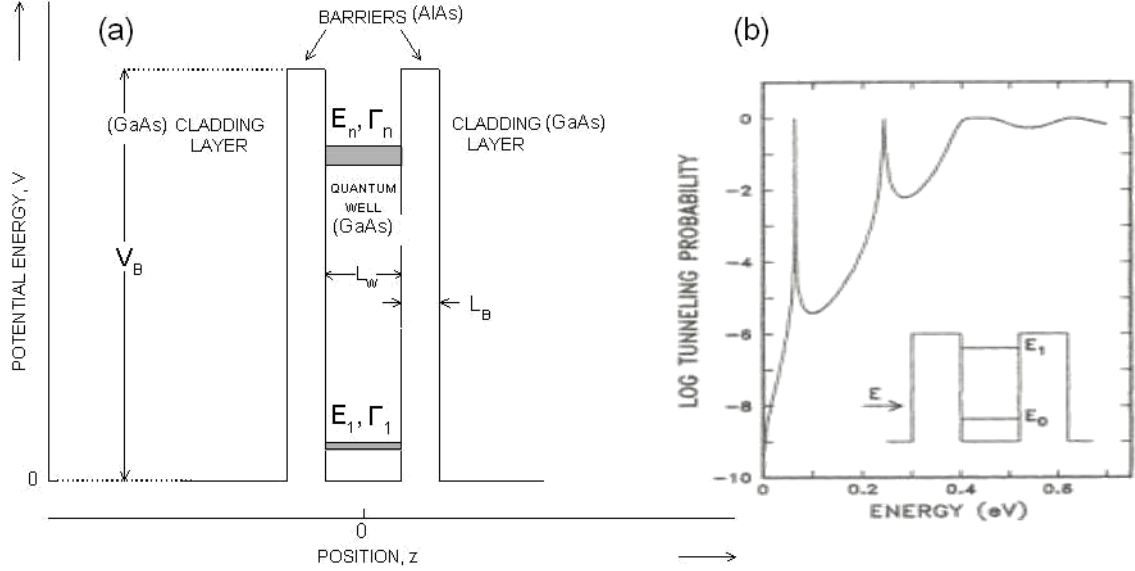


Fig. 3. (a) Potential energy schematic diagram of a symmetric double barrier quantum well heterostructure. (b) Transmission probability through the symmetric DBRT structure as function of the energy of the incident carriers.

quantum well between the two barriers ensures this condition. Each quasi-stationary state from the quantum well is characterized by an energy E_n and a lifetime τ_n . In the real case, when different scattering processes are involved, the condition for spatial quantization is $E_{n+1} - E_n \gg \hbar/\tau_s$ where τ_s is the scattering time for an electron in the two dimensional subband associated with the n^{th} quasi-level. This is similar to the condition that applies to other types of quantization in solids. For example, magnetic quantization requires $\omega_c \tau_s \gg 1$, where ω_c is the cyclotron angular frequency.

The spatial quantization ensures that the quasi-two dimensional states exist in the quantum well, but it does not specify the nature of the resonant tunneling that occurs through these states. The nature of the resonant tunneling is usually defined by the relative magnitudes of τ_n and τ_s . If $\tau_n \ll \tau_s$ the resonant tunneling through the n^{th} state is said coherent, since under this condition the electron does not scatter during the process and thus the phase of the wave function at each point in space is continuous in time. If $\tau_n \gg \tau_s$, the resonant tunneling is said to be incoherent, or sequential. In this case the wave function phase is partially randomized by the scattering events that occur in the double barrier structure.

The basic principles of resonant tunneling are readily illustrated if one neglects both the complicating effects of compositional grading in the heterojunction and the electrostatic band bending. This means that we assume that the heterojunctions of the double barrier structure are perfectly abrupt and no externally voltage is applied. Such a situation is shown in Fig. 3. The potential energy difference, or band offset V_B , at the interfaces is some fraction of the difference in the bandgaps between the two materials. For thin barriers the band gaps for both quantum well and barrier materials are usually taken at the minimum Γ -point of the Brillouin zone.

a) Electronic states in a double barrier structure

The energies corresponding to the quasi-stationary states of the DBRT structure shown in Fig. 2.1 can be found solving the one dimensional effective-mass Schrödinger equation:

$$\left[-\frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m^*(z)} \frac{\partial}{\partial z} + V(z) \right] \psi(z) = E \psi(z) \quad (1)$$

where $\psi(z)$ is the envelope function, $m^*(z)$ is the effective mass of the electron and $V(z)$ is the potential energy at the conduction band minimum. Considering the description of similar kind of states given in, an electron placed in the well in a quasi-stationary state moves inside the well for a considerable period of time τ_n , called the lifetime of the almost stationary state. The energy of this state is not discrete but broadened, whose “width” is related to the lifetime by $\Gamma_n \sim \hbar/\tau_n$ (uncertainty principle for energy). In these conditions we have to look for solutions of the Schrödinger equation which represent outgoing waves outside the barrier region (at infinity). This corresponds to an electron finally leaving the quantum well by tunneling. Solving Schrödinger equation with such boundary conditions, a set of complex values of the energy are obtained, which can be written in the form $\overline{E}_n = E_n - i\Gamma_n/2$. The physical significance of the complex energy values is in connection with the limited time spent by the electron inside the quantum well during the tunneling process. If we write the time factor of the wave function of the quasi-stationary state in the form $e^{-(i/\hbar)\overline{E}_n t} = e^{-(i/\hbar)E_n t} e^{-\Gamma_n/(2\hbar) \frac{t}{\tau_n}}$, then the probability of finding the electron “inside the system” (quantum well) decreases with time as $e^{-(\Gamma_n/\hbar)t}$. Thus Γ_n determines the lifetime τ_n of the state.

b) Coherent resonant tunneling

The interaction of the electron with the double barrier structure during the resonant tunneling process is described entirely by the transmission probability $T(E)$. This quantity is the ratio of the outgoing probability current density on one side of the structure to the incoming probability current density on the opposite side. Within the effective-mass model, $T(E)$ is found assuming the same trial envelope functions as in the case of the quasi-bound states except that one (left) cladding layer has an inward-going wave in addition to the outward wave. The most handy way to solve for $T(E)$, under these conditions, is by considering the transmission (T) matrix approach. Appropriate for multiple heterostructures the procedure consists, in principal, in the following: for each interface of the structure, using the specific boundary conditions, the amplitudes of the waves on both sides are related to each other by a T-matrix; after this the T-matrix of the whole structure, defined in the same way, is obtained as a product of the previous elementary matrices; finally the transmission and reflection probabilities can be calculated using the elements of the total T-matrix. Transmission probability through a symmetric DBRT structure as function of the energy of the incident carriers is illustrated in Fig.3b. As seen, it has the same form as the optical transmission of the optical transmission of a Fabry-Perot resonator, with sharp peaks at longitudinal energies corresponding to each quasi-bound level in the quantum well. Physically the behavior of $T(E)$ is the result of the interference of the coherent electron wave function with itself, in the quantum well, by way of multiple reflections on the two barriers. For off-resonant

energies the interference is destructive and the $T(E)$ is very small, but for energies equal to the quantum well quasi-levels, the interference is constructive resulting in the cancellation of the reflected wave outside the structure and enhancing of the transmitted one.

During resonant tunneling through a double barrier structure the electrons reside in one of the decaying quasistable states of the quantum well. This suggests that the process can be also described in the frame of the quantum scattering theory. In such an approach, at energies near the resonances, the transmission probability $T(E)$ can be expressed in the Breit-Wigner form permitting a direct correlation of the life time of the quasibound states, τ_n , with the full energetic width at half maximum of the transmission probability, Γ_n ($\tau_n = \hbar/\Gamma_n$):

$$T_{coh}^{rez} = \frac{\left(\frac{\Gamma_n}{2}\right)^2}{\epsilon - E_n \pm i\left(\frac{\Gamma_n}{2}\right)} \quad (2)$$

c) Incoherent resonant tunneling

As in any physical system, inelastic scattering that occurs in the double barrier structure changes the phase of wave function and its coherence is lost. This should have an important effect on the transmission probability, since the sharp resonances that are characteristic for typical quasi-bound levels requires a multiple-pass self-interference of the wave function in quantum well. The presence of inelastic collision phenomena into the quantum well can be taken into account assuming that the scattering potential has a small constant imaginary part $i\Gamma_s/2$ which absorbs part of the incident electron flux and causes the breakdown of the unitarity of the total coherent probability (transmission plus reflexion). The “lost electrons” into the quantum well will tunnel out, forward or backward relative to the sense of incident flux, in a separate incoherent process. The probability of each of these events is correlated with the relative transparency of corresponding barrier. For symmetric barriers these probabilities are equal. In this picture the coherent transmission probability T_{coh} is expressed as:

$$T_{coh}^{res} = \frac{\left(\frac{\Gamma_n}{2}\right)^2}{\epsilon - E_n \pm i\left(\frac{\Gamma_T}{2}\right)} \quad (3)$$

where $\Gamma_T = \Gamma_n + \Gamma_s$. We observe that this expression for T_{coh} is similar with Eq. (2) for the fully coherent case in which at the denominator Γ_n is replaced by the greater quantity Γ_T . Alongside with the T_{coh} an incoherent transmission probability T_{incoh} can be calculated:

$$T_{incoh}^{res} = \frac{\frac{\Gamma_n \Gamma_s}{4}}{\epsilon - E_n \pm i\left(\frac{\Gamma_T}{2}\right)} \quad (4)$$

Using Eqs. (3) and (4) the total transmission probability in the presence of inelastic scattering is given by:

$$T^{res} = T_{coh}^{res} + T_{incoh}^{res} = \frac{\frac{1}{4} \Gamma_n \Gamma_T}{\hbar(E - E_n)^2 + \left(\frac{1}{2} \Gamma_T\right)^2} \quad (5)$$

For any finite value of Γ_s , a given transmission resonance is broader than in the coherent limit, and the peak value of the transmission is less than unity. Γ_s can be related to the scattering time through the uncertainty principle, $\Gamma_s \approx \hbar/\tau_s$.

d) Current density

The physical principles underlying the general shape of the I-V curve of a double barrier structure can be understood by referring to the diagram in Fig. 3, in which we use the simple picture introduced in. One assumes that the structure has a quantum well containing only one quasi-bound state energy level E_1 . If the total energy and momentum of the electrons in the plane of the heterojunctions are conserved, then only those electrons having E_z on the cathode side approximately equal to E_1 can traverse the structure with any significant probability. As the bias voltage, V_{bias} , is increased from zero, the quasi-bound level drops relative to the electron energy in the cathode side. The number of electrons in the cathode Fermi sphere that have $E_z = E_1$ increases and it will determine a corresponding increasing of the current. This eventually approaches a peak at a voltage close to that which aligns the resonant level with the conduction band edge in the cathode side. At higher voltages there are no electrons with $E_z = E_1$, so that the current decreases precipitously and a negative differential resistance (NDR) occurs in the I-V curve.

To calculate the current density one usually considers that on the cathode and anode sides of the double barrier structure there are two independent ‘electron reservoirs’, in the thermodynamic sense, within which the electrons are all in equilibrium with a constant temperature and Fermi

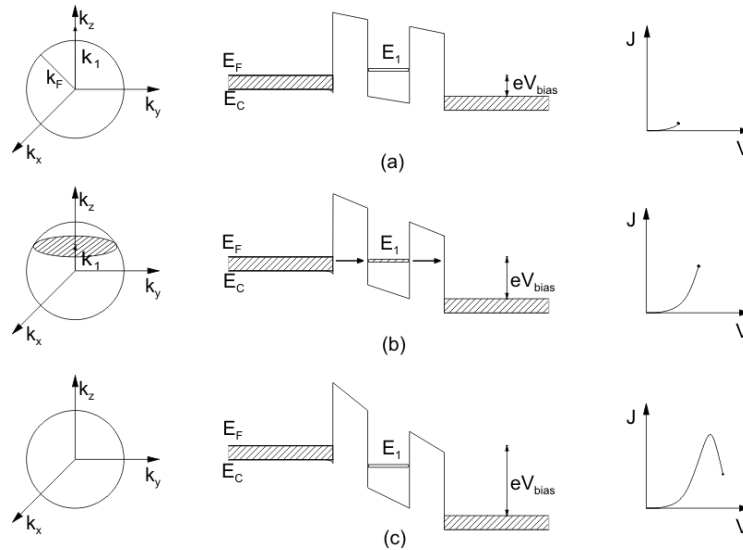


Fig. 3. Fermi sphere, band-edge profile, and I-V curve for a double –barrier structure for different bias conditions.

level, and which are so large that the current flow into or out of the structure represents a negligible perturbation. The electrons inside the two reservoirs are treated in the independent-electron approximation. Also, the corresponding left and right wave functions do not overlap, so no terms which account for the non-orthogonality of these will appear in the expression of the density current, as proposed in. Quantitatively, the net current through the double barrier structure is taken to be the difference between the current from the left to right and that from right to left. The electron effective mass in the electrodes is supposed to be the same, as usually happens in practice. Within this picture each of the two components of the current density is obtained by summing the expectation values of current operator impinging on the structure, multiplied by the corresponding transmission probability T_{tot} , over the occupied electronic states. Finally, the general accepted expression for the static current density has the following form:

$$J = \frac{em^*k_B T}{2\pi^2\hbar^3} \int_{-\infty}^{\infty} dE_z T_{tot}(E_z) \ln \left\{ \frac{1 + \exp \left[\frac{-(E_F - E_z)/k_B T}{1} \right]}{1 + \exp \left[\frac{-(E_F - E_z - eV_{bias})/k_B T}{1} \right]} \right\} \quad (6)$$

with k_B the Boltzmann's constant, T the absolute temperature, $\hbar$ the reduced Planck constant and E_F the Fermi level in the left reservoir.

e) Band bending under bias conditions

An external bias voltage applied across a double barrier structure modifies the equilibrium band profile in the active region. As we have foreshadowed in the previous section, the analysis of the current density implies the determination of the modified band profile, or band bending, to a satisfactory degree of accuracy. The band bending affects the transmission probability in several ways. In a first approximation, it has a direct effect on the quasi-bound levels, decreasing their energy relative to the conduction band edge of the cathode side layer by an amount $1/2(eV_w)$, where V_w is the component of the total voltage that occurs across the double barrier structure, as shown in Fig. 4. Physically, this shift represents the decrease in the potential energy at the centre of the quantum well and is unrelated to any Stark effect. We have seen that it is this effect which gives the qualitative explanation of the I-V characteristic. The second effect of the band profile modification is a change in the peak transmission. Thus, the peak of the first resonance of a symmetric double barrier structure under bias will fall below 1.0 because the transparency of the barriers becomes asymmetric.

A simple way to estimate the band bending is by treating the double barrier structure as an insulator, modeling the device as a semiconductor-insulator-semiconductor (SIS) diode. Unless the current density is very high, one can then assume that the diode is in a condition of quasi-equilibrium. Under this condition the electron concentration on each side of the structure is characterized by a uniform quasi Fermi level. The band bending is obtained by solving Poisson's equation on each side and connecting the solutions with a uniform electric field across the double barrier structure.

The real DBRT diodes have always interposed between the active region and the high doped contact layers some low doped layers and undoped spacer layers. Thus, as shown in Fig. 4, a positively charged depletion layer forms in the anode side and a negatively charged accumulation layer on the cathode side. The overall band bending is then complicated by the fact that a significant amount of electrons can occupy the quantum well and that both the quantum

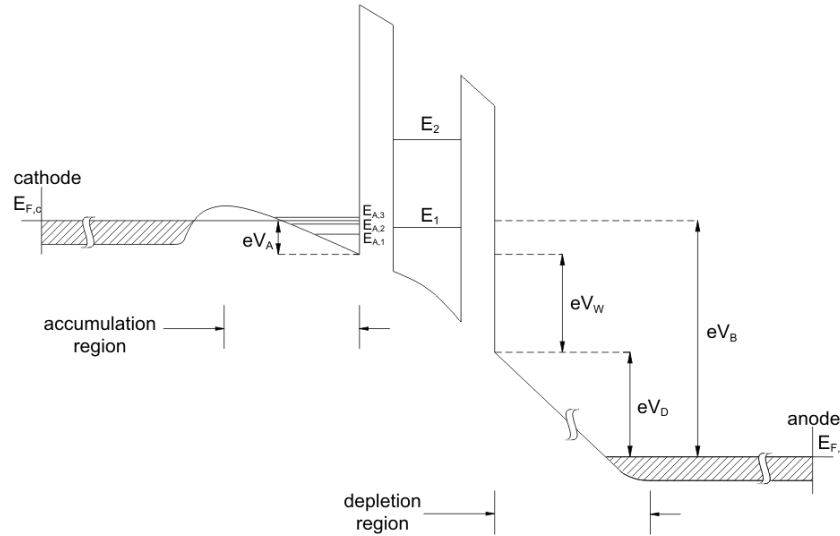


Fig. 4. Conduction band bending diagram of DBRT diode under the bias voltage

well charge storage and some fraction of the charge in the accumulation layer resides in quasi-two-dimensional states. To deal properly with these effects one must carry out self consistent calculations of Poisson and Schrödinger equations. Such calculations indicate that the charge build up in the quantum well can have a profound effect on the I-V curve resulting in an intrinsic bistability. Also, they show that the electron states in the accumulation layer consist of a continuum at energies above the neutral band edge in the cathode side and quasibound states at lower energies. The presence of the spatial quantization in the accumulation layer has been confirmed experimentally in some DBRT devices by the presence of undulations in the I-V curve at voltages below the first peak.

3. Molecular beam epitaxy growth of semiconductor heterostructures

3.1 Molecular beam epitaxy

The study of the DBRT structure in different semiconductor material systems was possible by the development of the molecular beam epitaxy growth technique. Molecular beam epitaxy (MBE) is a refined form of physical vapour deposition for epitaxial growth of high quality semiconductor, metal and insulator thin films. The epitaxial growth refers to the deposition of a crystalline film on a crystalline substrate, the film being in registry with the substrate. The term *epitaxy* comes from the Greek roots epi, meaning "above", and taxis, meaning "in ordered manner". It can be translated "to arrange upon". Using MBE high quality semiconductor layers are grown regarding purity, crystal phase, control of layer thickness and doping.

The main characteristic features of the MBE are:

- precise controlled **atomic or molecular thermal beams** react with a clean heated crystalline surface (substrate);
- **ultra high vacuum** (UHV) conditions;

- the beams are obtained by **effusion from solid (sublimation) or liquid (evaporation)** ultra pure material sources at high local temperature;
- **low growth rate** ($1\mu\text{m/h} \sim 1 \text{ monolayer/s}$) which permit the surface migration of the impinging species resulting in the growth of **atomically flat surfaces**;
- the growth governed mainly by the **kinetics of the surface processes**, under conditions far from thermodynamic equilibrium;
- ultra rapid shutters in front of the beam sources allowing nearly atomically abrupt transitions from one material to another and therefore, the obtaining of **abrupt interfaces**.

Historically, the basic concepts of the MBE growth process have been developed in 1958 by K. G. Günther at Siemens Research Laboratories (Erlangen) within so called '*three temperature method*'. The foundation of the MBE was done in the mid-sixties, when the first semiconductor films were grown by J.R. Arthur and A.Y. Cho at Bell Laboratories while studying the interaction of Ga atoms and As_2 molecules with crystalline GaAs in UHV conditions.

In the following, some of the characteristics of the MBE will be discussed.

a) Vapor pressure

It is well known that in a closed volume, a liquid or solid element is in equilibrium with the above gas phase (see Fig. 5a). The corresponding pressure of the gas is called the saturated or equilibrium vapor pressure, p_{eq} . Increasing the temperature, the vaporization will be stimulated and p_{eq} will increase. Each element has a characteristic phase diagram which describes the coexistence of different physical states of the element. This is shown in Fig. 5b. Only at one point, the so called triple point (T) the three different physical states coexist. The MBE method makes use of the vaporization of solid or liquid materials in UHV conditions. The vapors are directed to a crystalline substrate as atom (molecular) beams where they condense in the solid state. In Fig. 6, the vapor pressure of different elements as function of temperature is shown. Depending on the element, the vaporization results in molecules or atoms and a certain p_{eq} is reached by sublimation or evaporation at different temperatures. For example, a p_{eq} of 10^{-3} torr is reached at 270°C for As by sublimation of As_4 molecules, at 1082°C for Ga and 1007°C for Al, both by evaporation of atoms.

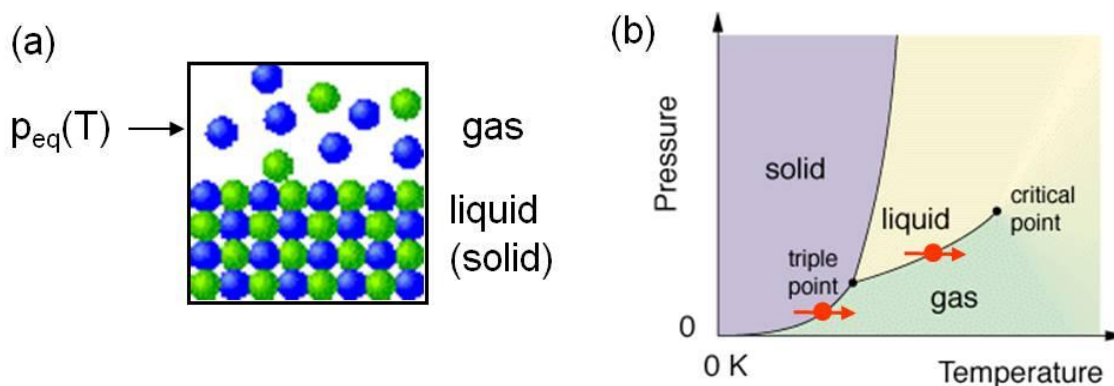


Fig. 5 (a) Illustration of the vaporization of a liquid or solid material in a closed volume resulting in the saturated vapor pressure of the gas phase. (b) Phase diagram of a single element.

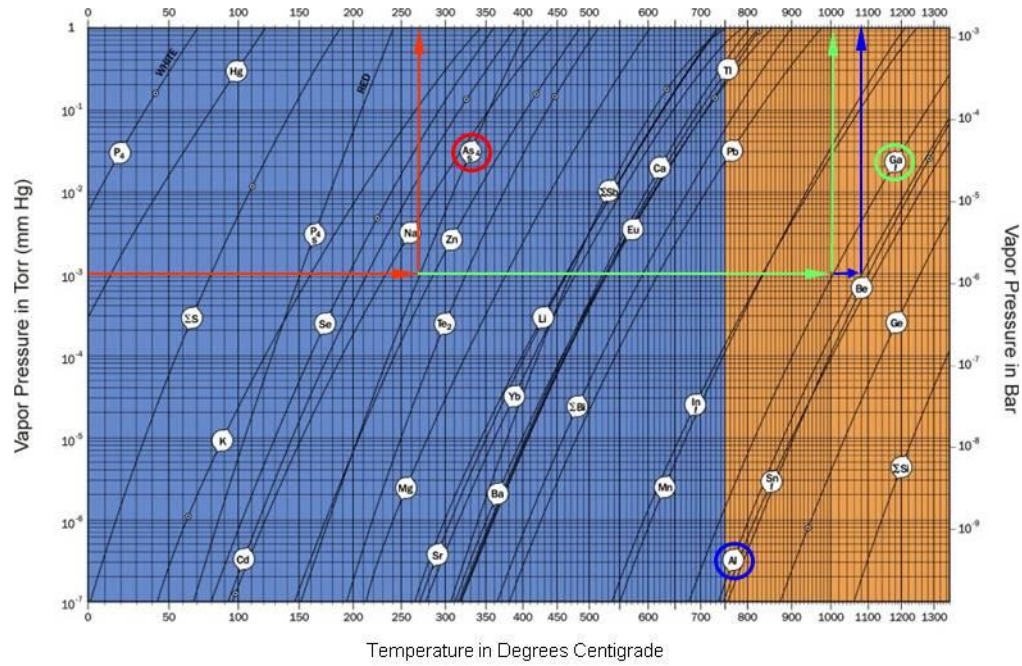


Fig. 6. Vapor pressure of different elements as function of temperature. The arrows indicate the necessary temperatures to reach a vapour pressure of 10^{-3} torr for As_4 (red, 270°C, sublimation), Ga (green, 1082°C, evaporation) and Al (blue, 1007°C, evaporation).

b) Molecular (atomic) beams

The molecular (atomic) beams necessary for MBE are obtained in effusion cells. The effusion cells exploit the evaporation process of condensed materials as molecular flux in vacuum. The understanding of the properties of real effusion cells is complicated and not straightforward, so, easier models are needed. The molecular (atomic) rate of evaporation per unit area from a liquid free surface, A_e , in equilibrium with its vapor was first derived by Hertz and later modified by Knudsen obtaining:

$$\frac{dN_e}{A_e dt} = \alpha_v (p_{eq} - p) \sqrt{\frac{N_A}{2\pi k_B M T}} \quad (7)$$

where p_{eq} is the equilibrium vapor pressure at the absolute temperature T , p is the ambient hydrostatic pressure acting upon the evaporant in the condensed phase, M is the molar mass of the evaporating species, and N_A and k_B are the Avogadro constant and Boltzmann constant, respectively. The evaporation coefficient α_v was introduced by Knudsen to account for the fraction of evaporant vapor molecules that are reflected by the condensed phase surface. α_v is dependent on the microscopic status of the surface and is strongly unpredictable. The maximum evaporation rate is obtained for $\alpha_v=1$ and $p=0$. Langmuir has shown that the Hertz-Knudsen equation applies also to sublimation from solid free surfaces. Phase transitions of this type, which constitutes vaporization from free surfaces, is commonly referred to as Langmuir or free evaporation. The free evaporation is isotropic. To force $\alpha_v=1$, Knudsen invented a technique, in

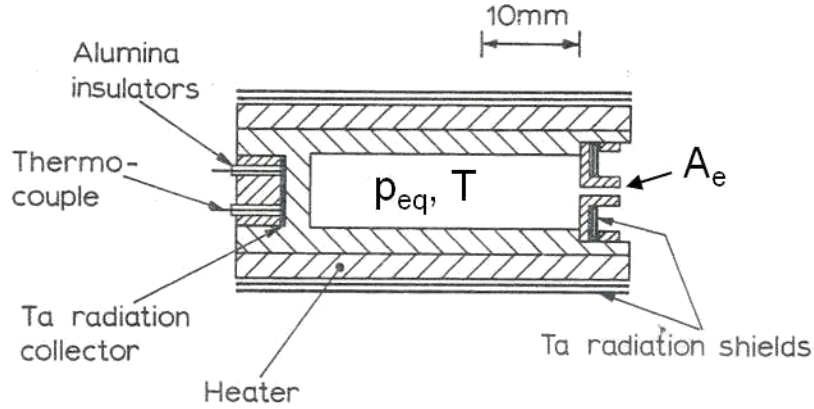


Fig. 7. The ideal effusion source used by Knudsen to study atomic beams of Hg.

which the evaporation occurs as effusion from an isothermal enclosure with a sufficiently small orifice (Knudsen cell, see Fig. 7). Under these conditions, the orifice represents an evaporation source with the evaporant pressure p_{eq} but without the ability to reflect vapor molecules. If A_e is the orifice area, the effusion rate, Γ_e , from the Knudsen cell is given by

$$\Gamma_e = \frac{dN_e}{dt} = A_e (p_{eq} - p_v) \sqrt{\frac{N_A}{2\pi k_B MT}} \quad (8)$$

with p_v the pressure in the reservoir where the molecules effuse from the cell orifice. In UHV environment, it is reasonable to consider $p_v = 0$. Then, the Knudsen effusion equation can be written in the most used form:

$$\Gamma_e = 8.33 \times 10^{22} \frac{p_{eq} A_e}{\sqrt{MT}} \quad (9)$$

where all the quantities are in SI units.

The Knudsen equation can be used to determine the impingement rate on a substrate from ideal effusion sources. In Fig. 8 are shown two situations, when the source orifice is oriented axially and non-axially with respect to the substrate. The impingement rate in the central point A is:

$$I_A = \frac{\Gamma_e}{\pi r_A^2} = 2.653 \times 10^{22} \frac{p_{eq} A_e}{r_A^2 \sqrt{MT}} \quad (10)$$

where r_A is the distance from the orifice to the substrate. The flux at the edge point B of the substrate can be calculated as:

$$I_B = \frac{\Gamma_e \cos^2 \vartheta}{\pi r_B^2} \cdot \frac{r_A^2}{r_B^2} = I_A \cos^4 \vartheta \quad (11)$$

This is called the cosine law of effusion. Proceeding in an analogous way, the impingement rate from a source which is tilted by an angle ϕ from the perpendicular to the substrate can be calculated:

$$I'_B = I_A \frac{r_A^2}{r_B^2} \cos \vartheta \cos(\vartheta + \phi) \quad (12)$$

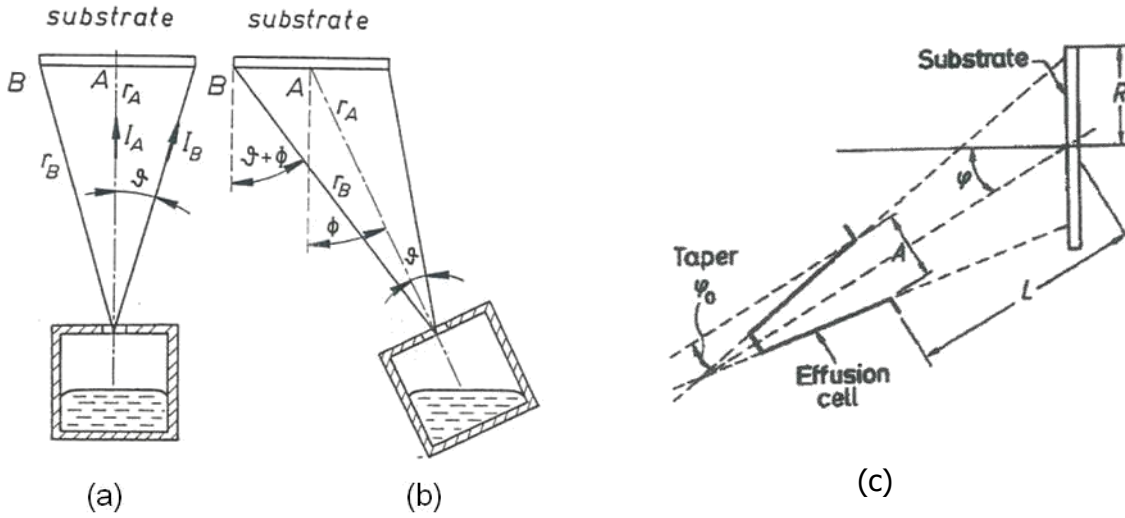


Fig. 8. Molecular flux on a substrate oriented axially (a) and non-axially (b) with respect to the orifice of the ideal effusion cel. (c) Molecular flux from a conical, Langmuir type non-equilibrium effusion cell tilted against the perpendicular substrate axis.

c) Ultra high vacuum

Vacuum conditions are necessary in MBE for two reasons. First, the beam nature of the mass flow towards the substrate has to be preserve. This means that the highest admissible value of the residual gas pressure in the vacuum reactor should assure the condition that the mean free path of the molecules (atoms) of the reactant beam penetrating the environment of the residual gas has to be larger than the distance from the outlet orifice of the beam source to the substrate crystal surface. In standard MBE chambers, this distance is $\sim 0.2\text{m}$ and a maximum residual gas pressure of 5.7×10^{-4} Torr can be estimated. It is evident that the beam nature of the mass transport is preserved even in high vacuum (HV) conditions. However, the low growth rates typical for the MBE technique (about $1\mu\text{m/h}$, or 1 monolayer (ML)/s) combined with the important requirement of negligible unintentional impurity incorporation in the crystallized epilayer lead to much more rigorous limitations for the total pressure of the residual gas in the MBE reactor. To get an estimation of this pressure we consider that the residual gas consists only of nitrogen molecules. The number of molecules which strike the unit area of the substrate can be expressed as:

$$w_{N_2} = 8.33 \times 10^{22} \frac{P_{N_2}}{\sqrt{MT}} \quad (13)$$

where all the quantities are in SI units.

At room temperature, $T=300\text{K}$, and a residual gas pressure of 10^{-6} Torr, considering molar mass of nitrogen molecule 28×10^{-3} kg/mol, $w_{N_2} = 3.82 \times 10^{14}$ molecules/ cm^2s or 7.64×10^{14} atoms/ cm^2s .

This corresponds approximately to 1 ML nitrogen atoms per second, which is also the typical growth rate in semiconductor epitaxy. These means that the background pressure has to be significantly lower than 10^{-6} Torr. An acceptable value of the background doping for the growth of III-V semiconductor layers is of the order of 10^{13} atoms/ cm^3 . On the other hand, the atom density of semiconductors is about 10^{22} atoms/ cm^3 , hence nine orders of magnitude higher than

the background doping level. So, we can conclude that the background pressure of the residual gas should be nine orders of magnitude lower than 10^{-6} Torr if the sticking coefficient of nitrogen atoms is 1, i.e. 10^{-15} Torr. This pressure is technically unachievable. However, the sticking coefficient is in general less than one so achievable residual gas pressures of 10^{-12} Torr should be enough to obtain low background doping. These are UHV conditions.

3.2 MBE growth

a) MBE system

A typical MBE system is presented schematically in Fig. 9a. The chamber is from stainless steel. Inside, UHV is maintained at a background pressure lower than 10^{-10} Torr using ion getter, titanium sublimation and cryo- pumps. Effusion cells are mounted symmetrically on a source flange. In Fig. 9b, a standard effusion cell is shown. The heating element is from tungsten wire being surrounded by several tantalum foils that serve as a radiation shield to improve the power efficiency of the source and to reduce the heating of the surrounding cryo-panel (see below). A precise temperature control of each cell assures the desired constant vapor pressure and corresponding beam flux. The material is heated in pyrolytic boron nitride (PBN) crucibles (see Fig. 9c). Rapid shutters mounted in front of the effusion cells switch the atomic or molecular beams on and off. Each cell is surrounded by a cryo-shield continually cooled with liquid nitrogen at a temperature of 77K. A second cryo-panel forms an inner shield, so that the main MBE area is surrounded by cold walls operating as additional pumping units and minimizing spurious fluxes of atoms and molecules from the walls of the chamber. In the middle of the chamber, the rotatable substrate holder contains also the heater of the substrate. The pressure of the specific material can be measured with a Bayert-Alpert ion gauge, which can be positioned in the same place as the substrate holder. This pressure is often referred as the beam equivalent pressure (BEP) since the absolute value of the ion gauge reading is dependent on the geometry of the beam with respect to the ion gauge entrance slit and other properties, like the ionization probability of the material. The materials loaded in the effusion cells have very high purity. To grow III-V compound semiconductor layers, the effusion cells should contain materials like Ga,

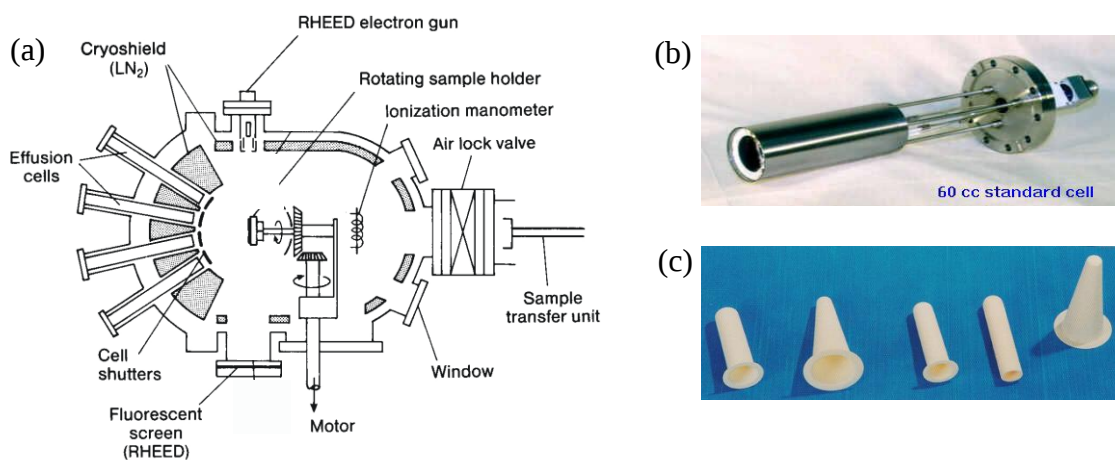


Fig. 9. Schematic illustration of an MBE system (a) and pictures of an effusion cell (b) and PBN crucibles (c).

As, In, Al for the layer growth and Si and Be for intentional n-type and p-type doping, respectively. The UHV environment in the MBE is also ideal for many in-situ characterization tools, like reflection high energy electron diffraction (RHEED). RHEED system provides real time information about the state of the surface during the growth and can be used to determine the growth rate of different layers.

b) Growth of GaAs

The growth of GaAs is a good example to illustrate the MBE of III-V semiconductors. Surface processes involved in general in the MBE growth are schematically illustrated in Fig. 10. These are:

- physical and chemical adsorption of the constituent atoms or molecules impinging the substrate surface;
- surface migration and dissociation of the adsorbed molecules;
- incorporation of the constituent atoms into the crystal lattice of the substrate(grown layer);
- thermal desorption of species not incorporated into the crystal lattice.

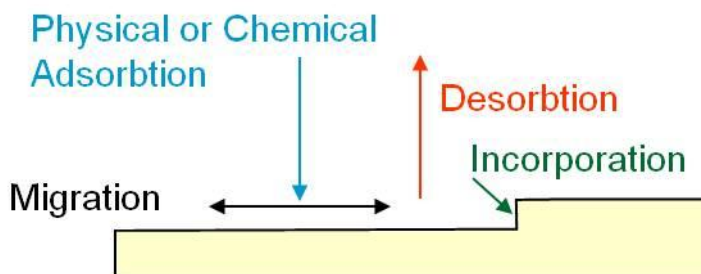


Fig. 10. Surface processes involved in MBE growth.

GaAs is generally prepared in MBE by the evaporation of elemental, atomic Ga, and the sublimation of elemental arsenic which can either be dimeric (As_2) or tetrameric (As_4). As_2 is obtained from As_4 by cracking. The growth is done usually on heated GaAs (100) substrates. The sticking coefficient of different species impinging the substrate is defined as the ratio between the adsorption rate to the impinging rate. Ga has the sticking coefficient unity for the used growth temperatures. The incorporation of the arsenic depends of the involved species. When GaAs is grown from As_2 and Ga, As_2 is first adsorbed into a mobile, weakly bound precursor state (see Fig. 11a). Dissociation of adsorbed As_2 molecules can occur only when these encounter paired Ga lattice sites while migrating on the surface. In the absence of free surface Ga adatoms, As_2 molecules have a measurable surface lifetime and desorb. The sticking coefficient of As_2 is proportional with the Ga flux I_{Ga} and tends to unity when the surface is completely covered with one monolayer of Ga. Additionally, this increases with increasing substrate temperature. In this way, stoichiometric GaAs is grown when $I_{\text{As}_2}/I_{\text{Ga}} \geq 1$. For the growth of GaAs with Ga and As_4 , the process is more complex (see Fig. 11b). From the mobile precursor state, pairs of As_4 molecules react on adjacent Ga sites resulting in the incorporation of four As atoms and desorption of one As_4 molecule (second-order reaction). The sticking coefficient is always less than or equal to 0.5. Stoichiometric GaAs is obtained when $I_{\text{As}_2}/I_{\text{Ga}} \gg 1$. Finally, the important conclusion of the presented models is that the growth of GaAs is kinetically controlled

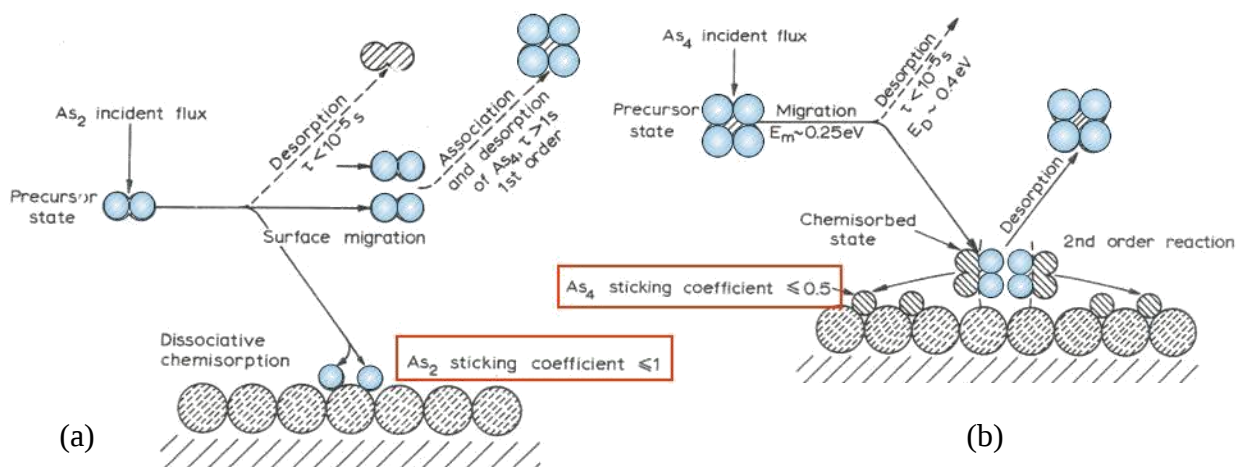


Fig. 11. Models for the growth of GaAs with As_2 (a) and As_4 (b).

by adsorption of group V element, while the growth rate is determined only by the group III element flux. This is available also for the growth of other III-V compound semiconductors.

c) RHEED

RHEED is one of the most important in-situ analysis tool for physical vapor deposition. As already mentioned, RHEED is used in MBE to analyze the surface of the substrate/grown film obtaining fundamental information about surface geometry and chemistry, both in static and dynamic conditions (during growth) and to monitor the growth rate. The technique employs a high energy electron beam (several tens of KeV), directed on the sample surface at grazing incidence (a few degrees); the diffraction pattern is imaged on a symmetrically placed fluorescent screen (see Fig. 9a). Thanks to the grazing incidence, the electron beam is scattered in the first few atomic layers, thus giving a surface-sensitive diffraction pattern. Besides, the grazing geometry avoids any interference of the RHEED apparatus with the molecular beams, making the technique suitable for real time growth analysis.

A qualitative explanation of the origin of RHEED patterns can be seen in Fig. 12a. If electrons interact only with the first atomic layer of a perfectly flat and ordered surface, the three-dimensional reciprocal lattice points degenerate into parallel infinite rods. In the resulting Ewald construction, the intersection of the rods with the Ewald sphere (having a radius much larger than the inter-rod spacing for typical RHEED energies) consists therefore of a series of points placed on a half circle. In reality, thermal vibrations and lattice imperfections cause the reciprocal lattice rods to have a finite thickness, while the Ewald sphere itself has some finite thickness, due to divergence and dispersion of the electron beam. Therefore, even diffraction from a perfectly flat surface results in a diffraction pattern consisting in a series of streaks with modulated intensity, rather than points (Fig. 12b). If the surface is not flat, many electrons will be transmitted through surface asperities and scattered in different directions, resulting in a RHEED pattern constituted by many spotty features. Therefore, a first important information provided by RHEED regards the flatness of a surface. Furthermore, it is evident that diffraction from an amorphous surface (such as an oxide on top of a semiconductor) gives no diffraction pattern at all, and only a diffuse background will result. This is important, for example, for evaluating oxide desorption when a new substrate is initially heated up prior to growth in the MBE chamber, exposing the

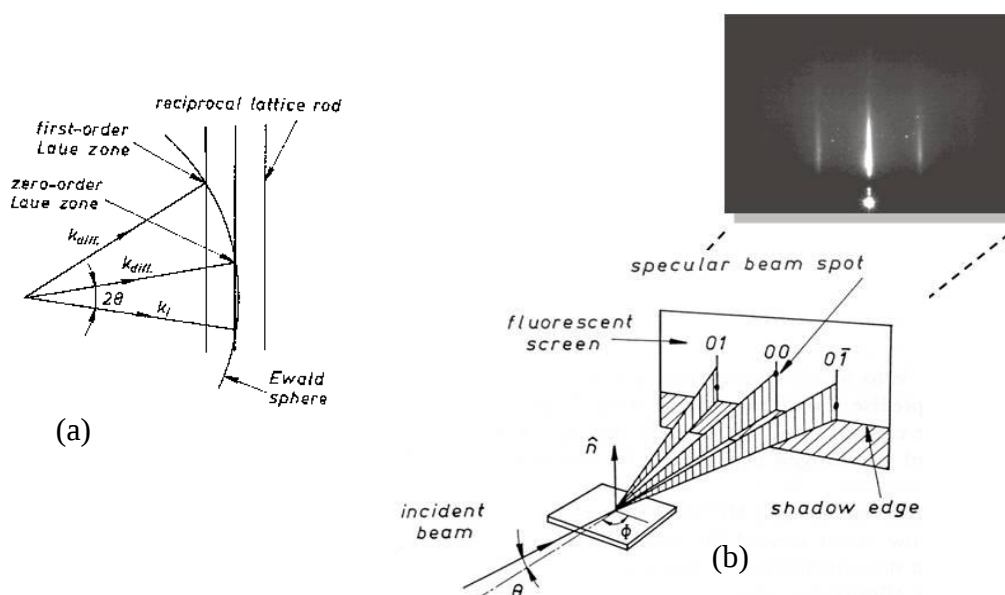


Fig. 12. a) Ewald sphere construction for a two-dimensional reciprocal lattice (side view).
b) RHEED geometry and formation of a diffraction pattern.

underlying, crystalline semiconductor surface. The RHEED diffraction pattern consists of main streaks with the bulk lattice periodicity (due to penetration of the beam in the first atomic layers), intercalated by weaker, surface reconstruction-related lines.

Besides the structural information, RHEED can also provide information about the growth rate of atomic layer deposition. On a perfect laterally flat surface and under the assumption of a two dimensional growth mode, the surface condition varies during the growth periodically. It starts with a complete filled flat surface layer then it continues with half a monolayer coverage ("rough

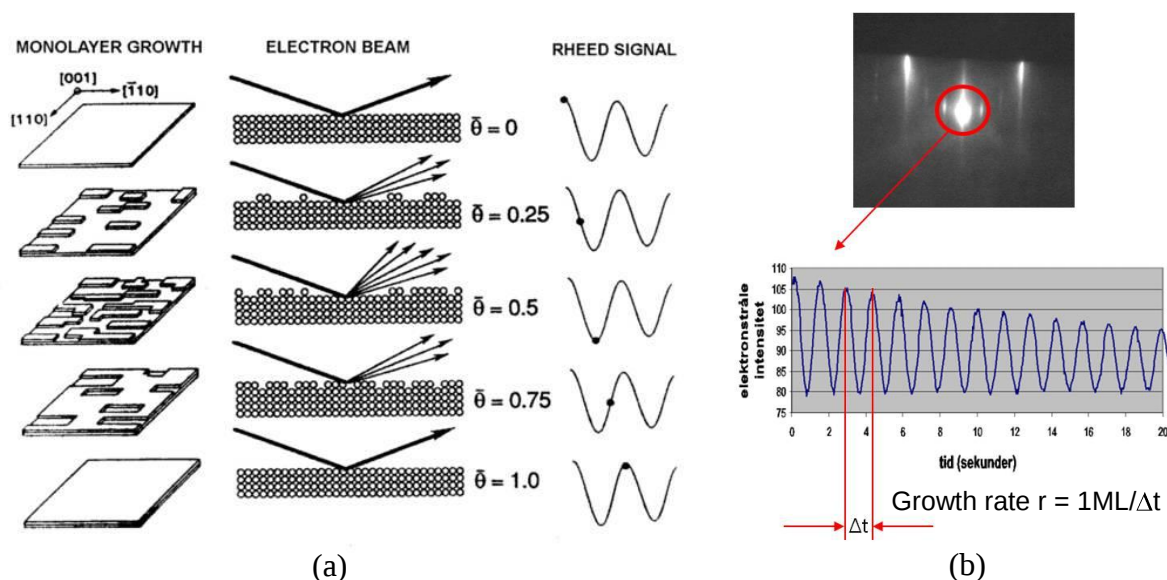


Fig. 13. a) Schematic illustration of the appearance of the RHEED intensity oscillation during the growth of one monolayer. b) Evaluation of the growth rate from RHEED oscillations.

surface”) and finally back to a complete surface coverage (see Fig. 13a) after a monolayer growth. This periodical change in the surface conditions can be seen in the RHEED intensity of a specular diffraction spot and can directly be used to calculate the growth rate as the inverse of the RHEED oscillation frequency (see Fig. 13b).

c) Growth modes.

The adsorbed atoms move on the surface of the substrate and interact also each other. Depending of the interaction with the substrate and between atoms themselves, three possible modes of crystal growth may be distinguished. These modes are illustrated in Fig. 14. The layer-by-layer mode, or Franck van der Merwe mode, is happened when the atoms are more bound to the substrate than to each other (see Fig. 14a). The island mode, or Volmer-Weber mode, displays the opposite characteristics. Small cluster are nucleated directly on the substrate and then growth into islands of the condensed phase (see Fig. 14b). This growth mode appears when the crystal lattice constant of the deposited material is significant different from the substrate crystal lattice constant (lattice mismatched material systems). An intermediate case is the layer plus island mode, or Stranski-Krastanov. After forming a first monolayer, or a few monolayers, subsequent layer growth is unfavorable and islands are formed on top of this ‘intermediate’ wetting layer. The real surface of the substrate is not ideally flat, monoatomic steps can be found on the surface, which are correlated with the surface orientation. On a crystal surface with cubic symmetry the step width is depending on the surface misalignment with respect to a certain crystal orientation. The steps change the surface potential and have a strong influence on the growth. An adsorbed atom on such a step can move downwards or upwards to the edges of the step. In the case of the downward movement, the atom cannot fall onto the next step because it has to overcome the so called Schwöbel barrier (see Fig. 15a). However, atoms moving in the upward direction can be incorporated into the crystal phase at the upper step edge. The relation between the step width, l , and the diffusion length, λ , depending on temperature, determines the growth mode: 2-D growth at lower temperatures and the step flow growth at higher temperatures. These are illustrated in Fig. 15b together with the corresponding RHEED intensity profiles. In the step flow growth mode no oscillations are observed.

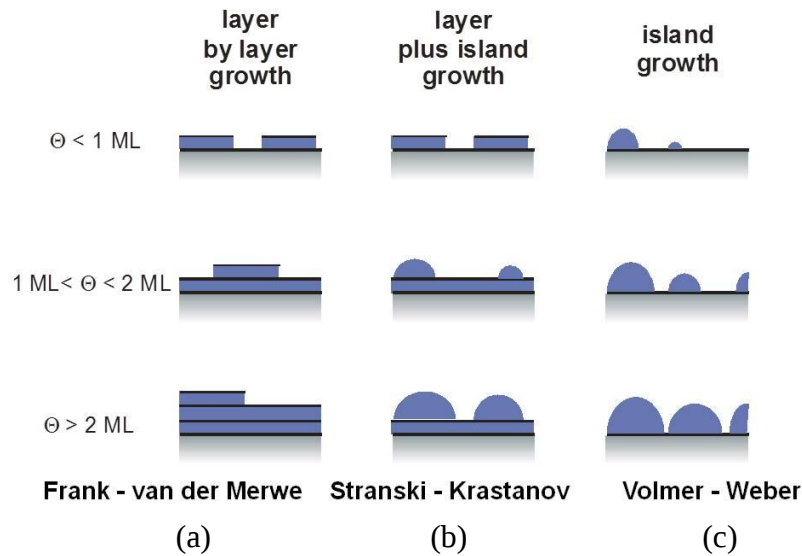


Fig. 14. Schematic illustration of the crystal growth modes.

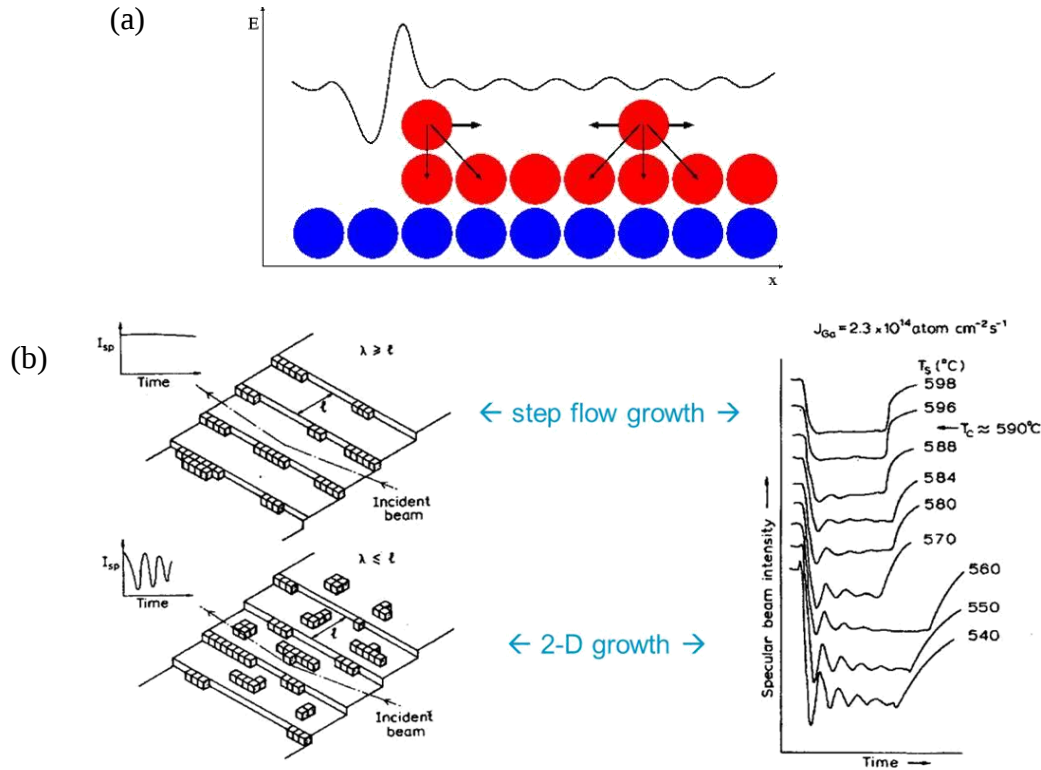


Fig. 15. a) Surface potential at step edge referred as the Schwöbel barrier. (a) Illustration of the step growth mode and 2-D growth mode together with the corresponding RHEED intensity profiles.

4. Resonant tunneling diode: processing and electrical characterization

a) RTD processing

A typical GaAs/AlAs DBRT structure grown by MBE is shown in Fig. 16. The transmission electron microscopy (TEM) micrograph from the inset illustrates the quality of the active region composed from a GaAs quantum well and two AlAs barriers: very good control of the thickness and sharp interfaces. The active region is cladded by spacer, low doped and contact (highly doped) layers. For the electrical characterization, the structure needs to be processed as a device. The processing includes lithographic, etching and metallization processes. The main steps for the fabrication of the resonant tunneling diode (RTD) diode are illustrated in Fig. 17. First, mesas are defined to obtain the active area of the device. This is realized by optical lithography and wet chemical etching in $\text{H}_3\text{PO}_4:\text{H}_2\text{O}_2:\text{H}_2\text{O}$ solution or dry etching using a CH_4/H_2 plasma. The area of the mesas should be on the order of tens of μm^2 . Next, contacts are processed by lift-off technique using a specific metallization, e.g. Ge/Ni/Au evaporated layers. The contacts have to be annealed in a rapid thermal process oven to obtain the ohmic behavior. Finally, the contacts are isolated with SiO_2 and measurements pads are defined using again the lift-off technique this time with Au metallization. An scanning electron microscope (SEM) picture of a processed RTD is shown in Fig. 16a.

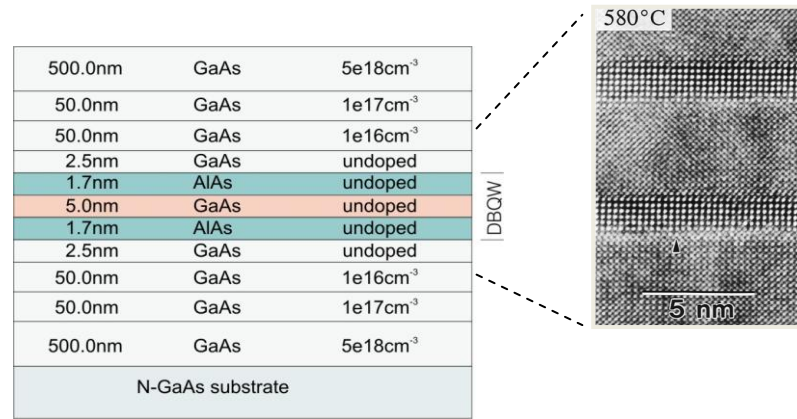


Fig. 16. DBRT structure based on GaAs/AlAs material system. The inset show a TEM picture of the active region.

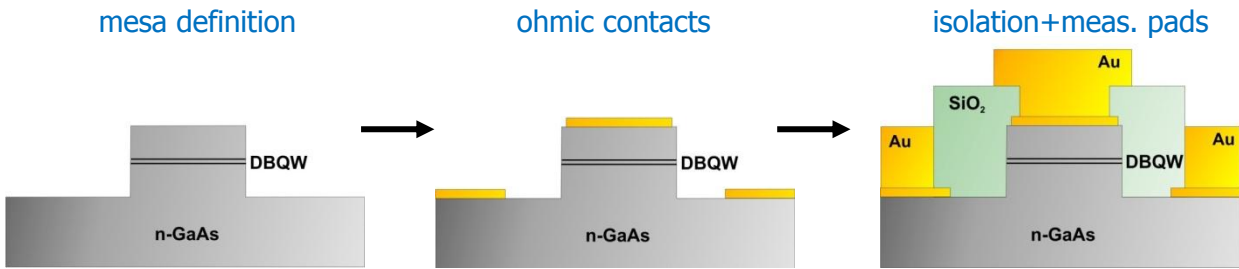


Fig. 17. Process flow of an RTD.

b) DC electronic transport

DC electrical measurements on a RTD can be done using for example a probe station and a semiconductor parameter analyzer (see Fig. 18a). A typical measured current-voltage (I-V) characteristics is shown in Fig. 18b. One can observe some characteristic features and parameters. Thus, the I-V curve presents a negative differential resistance (NDR) region as have been explained in section 2.3d. This is symmetric relative to the bias polarity, which is expected considering that the device structure is symmetric (see Fig.16). On the I-V characteristics, peak and valley current densities, J_p and J_v , and peak and valley voltages, V_p , V_v , can be defined. A parameter used to assess the quality of a RTD is the peak to valley current density ratio $PVR=J_p/J_v$. The presence of the NDR region in the I-V characteristics and the fact that the electronic transport is based on resonant tunneling which is a very fast process make the RTD suitable for high speed electronics, i.e. oscillators, logic circuits, etc. In this sense, good RTDs should have a high J_p , a low J_v , and a large PVR . Related to the power consumption, the characteristic voltages should be as small as possible.

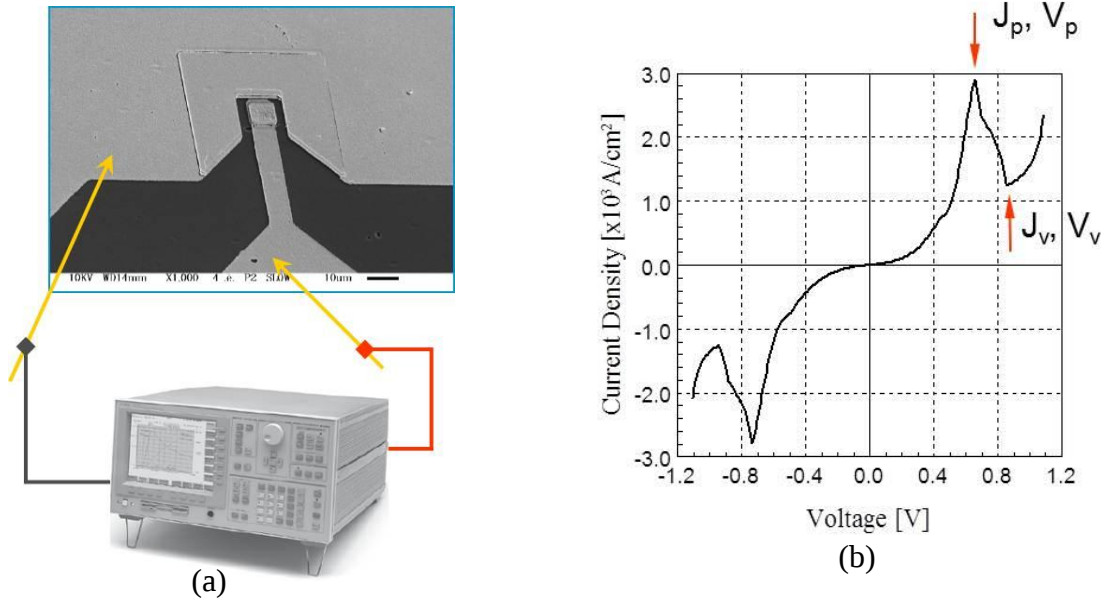


Fig. 18. (a) Schematic illustration of the measurement setup: the diode shown in the SEM picture is connected with the needle probes to an semiconductor parameter analyzer. (b) I-V characteristics of an RTDs.

References

- L. D. Landau and E. M. Lifshiz, “*Quantum Mechanics*”, 3rd ed., section 134, Pergamon, New York, (1977).
- D. Bohm, “*Quantum Theory*”, p. 283, Prentice-Hall, Engelwood Cliffs, N.Y., 1951.
- E. O. Kane, “Basic Concepts of Tunneling”, in “*Tunneling Phenomena in Solids*”, eds. E. Burnstein and S. Lundqvist, Plenum Press, New York, 1969.
- E. Merzbacher, “*Quantum Mechanics*”, p. 100, John Wiley & Sons, New York, 1970.
- W. R. Frensley, “Quantum Transport”, in *Heterostructures and Quantum Devices*, eds. N. G. Einspruch and W. R. Frensley, p. 273, Academic Press, London, 1994.
- H. Mizuta and T. Tanoue, *Physics and Applications of Resonant Tunneling Diodes*, Cambridge University Press, New York, 1995
- E.H. Parker, ed., *The Technology and Physics of Molecular Beam Epitaxy*, Plenum press, New York, 1985.
- M.A. Herman and H. Sitter, *Molecular Beam Epitaxy, Fundamentals and Current Status*, Springer-Verlag, Berlin, 1996.
- R.A. Stradling and P.C. Klepstein, eds., *Growth and Characterisation of Semiconductors*, Adam Hilger, Bristol, 1990.