III-V Molecular Beam Epitaxy Structures for Electronic and Optoelectronic Applications

Thesis by

John Stephen Smith

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
1986

(Submitted April 4, 1986)

To my parents

Acknowledgments

I would like to express my appreciation for the guidance and encouragement Dr. Yariv has given me during my work in his quantum electronics group. I would also like to thank Dr. Shlomo Margalit, Dr. Zev Rav-Noy, and Pamela Derry, who collaborated in some of the work reported in this thesis, and Dr. L. C. Chiu, Dr. Christoph Harder, Dr. Kerry Vahalla, Joel Paslaski, and Anders Larson, with whom I have worked in other areas.

A special thanks to Dr. A. Y. Cho, who taught me the "black art" of MBE, and to Randy Bartman, for many valuable discussions.

I am also grateful to the Fannie and John Hertz foundation for their financial support.

Abstract

Molecular beam epitaxy (MBE), a method for depositing epitaxial semiconductor layers with an extreme degree of control, has previously been limited largely to featureless substrates. This thesis describes a procedure for high quality MBE growth over finely patterned GaAs substrates, which is suitable for device fabrication requiring lateral definition of small ($\sim 1-2$ micron) dimension. The results of series of experiments determining the characteristics of MBE growth over patterned substrates are presented. Temperature and flux ratio dependence of faceting during MBE growth over patterned substrates is shown for temperatures ranging from 580° C to 700° C and for As/Ga flux ratios from 1.4:1 to 4:1, and surface diffusion lengths for gallium are measured. The (811) and (411) facets are shown to have growth rates which are at a local maximum and minimum, respectively, for facets tilted in the $[0\bar{1}1]$ direction, and implications with respect to other experiments on the atomic growth kinetics on (100) facets are discussed. The material grown over facets tilted in the [011] direction is shown to be of high quality, with electroluminescence equal to that over the (100) plane.

This method was used for the fabrication of an index guided laser structure, with an optical guide similar to that of the liquid phase grown channeled substrate planer laser structure, as well as tightly packed arrays of these lasers. Yields of individual lasers exceeded 90 percent, and thresholds were uniform to 10 percent. Other potential applications for this technique include a vertical structure high electron mobility transistor, formation of "quantum wire" structures, and methods of directly contacting quantum well layers for device applications and research purposes.

Certain aspects of transport of hot electrons over quantum wells are discussed, with application to several types of devices, including a new type of infrared detector structure, a solid state electron multipler, and a proposed quantum well base transistor.

-TABLE OF CONTENTS-

Chapter I

INTRODUCTION

Electronics is almost entirely dominated by one material, silicon, despite the fact that essentially similar phenomena occur in a wide variety of semiconductors. This can largely be traced to the existence of a native oxide of silicon which is an excellent dielectric material. Even after a great deal of research, no good substitute for this native oxide has been found for use with other materials. Despite this handicap, it is likely that materials such as gallium arsenide will begin to displace silicon in certain applications because of other properties which are superior to those of silicon. Among these properties are higher mobility and saturation velocity for electrons, which means that equivalent electronic devices will be faster than when fabricated in silicon, a direct bandgap, which makes possible efficient generation and detection of light, and the ability to fabricate structures where the bandgap of the material changes from one part of the material to another. This last ability comes from the use of ternary, quaternary or higher order compound semiconductors where the alloy composition is varied. This technique, termed bandgap engineering, allows an additional degree of freedom in the design of structures, making possible a great number of new types of devices. One example is the high electron mobility transistor (HEMT),[1] which is a field effect transistor where the carrier electrons are separated from their parent donor atoms by an band gap change. This separation reduces the interaction of the electrons with the charged donor atoms, greatly increasing their mobility within the channel formed. Because of the vast head start that silicon based technology has, it will not be displaced as the primary electronic technology in the foreseeable future. Technology based on other semiconductor materials will supplement silicon based technology in applications which take advantage of the speed or other properties. Examples of these applications are ultra high speed processing units, or "supercomputers," demultiplexers, high speed front ends, and optoelectronic circuits, in addition to the microwave applications in which GaAs technology is already becoming dominant, and semiconductor lasers, where a direct bandgap is required.

The speed and complexity of electronic devices has been increasing at an exponential rate over the last two decades. This rapid expansion in the capability and reduction in cost of electronics can be traced in large part to one technique: photolithography. This process allows simple optical reduction to produce patterns of a polymeric material on a two-dimensional surface with definition of a micron or so in extent. This relatively cheap, fast process far exceeds the capability of any other technique, mechanical processes being far cruder, and techniques such as directed beams being far more expensive, and intrinsically slow due to being basically serial processes. Design of semiconductor devices consists largely of using this patterned polymer to "mask" the surface of the wafer during a subsequent process, after which the photoresist is removed. The process of depositing and patterning photoresist, masked surface treatment, and removal is repeated many times until the desired devices are formed. Much of device-related research involves development of variations of this basic technique, increasing the accuracy of the pattern transfer, forming the structure of new devices, fabricating a geometry which allows study of an interesting effect, or improving the characteristics of existing devices. Even purely theoretical research is often driven by what is possible in the laboratory, because analysis of phenomena or predictions of effects are not very "interesting" or useful if there is no way in which they can be observed. An example of this is the rapid development of the work on superlattices[2] and quantum well structures[3] only after a way of fabricating such structures was found using MBE growth techniques, which allow the deposition of layers as thin as individual atomic planes.

This thesis discusses a new method which makes the technique which made these extremely thin layered structures possible, molecular beam epitaxy (MBE), compatible with photolithography. This allows interesting new structures and devices to be formed, several of which are discussed. In addition, growth over patterned surfaces yields information about the MBE growth mechanism itself.

In chapter two, molecular beam epitaxy on featureless GaAs substrates is discussed. MBE is a technique for growing very thin epitaxial layers on single crystal substrates in an ultrahigh vacuum apparatus.[4] MBE is particularly sensitive to contamination, and the method requires exacting experimental technique.

Chapter three discusses a technique for growing epilayers by MBE on wafers which have previously-formed structures, such as grown layers or etched patterns. Such structures would be entirely removed by the cleaning procedures applied to flat substrates, and it was necessary to find modified procedures which would still provide a surface suitable for MBE, yet not etch away the structure which is on the surface of the wafer. Fabrication of new types of devices by this method was proposed by Tsang and Cho in 1977;[5] however, the technique discussed is the first which allows the realization of device quality material over surfaces with small scale features.

Chapter four presents the results of a series of experiments which were carried out investigating faceting and structures formed during growth by MBE over etched ridges over a full range of temperatures and arsenic flux ratios. A simple model of the growth process is developed, and implications for the use of the technique for device fabrication are discussed.

Chapter five describes a laser structure which was suggested by the properties of regrowth by MBE over ridges, which I will refer to as a 411 laser, because it relies on the enhanced growth rate over the [4$\bar{1}$1] facet. This simple structure provides an index guided geometry, comparable to that of the channeled substrate planar (CSP) structure grown by liquid phase epitaxy (LPE). A similar structure which uses the dependence of the growth on the angle of incidence of the molecular beams to form a buried guide using a specially modified MBE system is also discussed.

In Chapter six, an array of lasers of the 411 type is discussed. The sharper propagation of surface features during MBE growth as compared to LPE growth allows the formation of arrays which are much more closely packed, allowing the formation of a true index guided array, which has proven difficult to achieve by other techniques.

Chapter seven describes several other applications for the regrowth technique, including a vertical geometry field effect transistor (VFET) in which the channel is formed by the regrowth process. It is shown that this geometry allows the use of high electron mobility transistor (HEMT) techniques to increase the performance of the device, and avoids the contact problems that the HEMT structure usually entails. Other applications include the formation of quantum wire structures, and a method for supplying electrical contact to quantum well layers.

Chapter eight discusses transport of hot electrons over quantum wells, and several devices which use the low capture probability which occurs under some conditions. Applications include a new type of infrared detector which uses an interband absorption transition in a superlattice. This detector, while not as sensitive as the band-to-band detectors currently under development, has potential application in the near infrared bands for inexpensive infrared imaging. Other applications include a transistor structure which uses a quantum well as its base region, and a electron multiplier which utilizes the fact that under some conditions, electrons propagating through a heavily doped multiple quantum well structure can scatter electrons which were trapped in the quantum wells, resulting in a net current gain.

References for chapter 1:

[1] H. Morkoç and P. M. Solomon, IEEE Spectrum, vol. 21, pp. 28-35, Feb. (1984).

[2] L. Esaki, and R. Tsu, IBM J. Res. Dev. 14, 61 (1970).

[3] A. Y. Cho and J. R. Authur, Jr., in Progress in Solid State Chemistry, edited by J. O. McCaldin and G. A. Somorjai, Vol. 10, pp. 157–191, Pergamon, New York, (1975).

[4] W. T. Tsang in Semiconductors and Semimetals, vol. 22, part A, (1985).

[5] W. T. Tsang and A. Y. Cho, Appl. Phys. Lett., Vol. 30, No. 6, (1977).

Appendix A

Directional dependence of some GaAs properties

Extensive references are made in this thesis to features and orientations on GaAs surfaces. The following is a summary of the notation used for referring to these directions. The crystal symmetry group of GaAs is the face centered cubic (FCC) zincblende structure. Directions are referred to by giving ordered triplets of integers $[lmn]$ in square brackets, and a bar indicating negative numbers, which correspond to vectors $la\hat{x}+ma\hat{y}+na\hat{z}$ where $\hat{x}$, $\hat{y}$ and $\hat{z}$ are the conventional lattice vectors shown in figure A.1, and a is the conventional lattice constant. In this case of a cubic conventional lattice, they form a simple cartesian coordinate system. Planes are referred to by ordered triplets of integers (lmn), perpendicular to the directions $[lmn]$. In addition to the translation symmetries of the FCC lattice, the symmetry of the basis allows a three-fold rotation symmetry about the [111] direction which means that the $[lmn]$,$[nlm]$ and $[mnl]$ directions are equivalent. In addition, the crystal is symmetric with respect to reflection in the (110), (011) and (101) planes, and the $[lmn]$ and $[\bar{l}\bar{m}n]$ directions are equivalent, as are the $[lmn]$ and $[l\bar{m}\bar{n}]$ directions, and the $[lmn]$ and $[\bar{l}m\bar{n}]$ directions. However, the presence of a feature such as a lattice termination breaks some of the symmetries. For example, references to (100) substrates, (010) substrates, and $(00\bar{1})$ substrates are all equivalent, but the [011] direction on a (100) substrate is equivalent to the $[\bar{1}\bar{1}0]$ and [110] directions on a $(00\bar{1})$ substrate, and is not equivalent to the [011] direction on a (010) substrate.

The symmetries of the GaAs crystal and the lack of convention with respect to the notation in the presence of a crystal surface makes it somewhat difficult to follow references to crystal directions in the literature. References to particular directions in this thesis are with respect to a (100) wafer surface. Notations can be converted into one which is more familiar by applying a symmetry transformation to all of the referenced directions and planes. For example, a paper which refers to directions on wafers with a $(00\bar{1})$ surface may be converted into the convention used here by applying the transformation $\{lnm\} \rightarrow \{\bar{m}\bar{l}n\}$ to all of the triplets.

GaAs cleaves easily and cleanly along the orthogonal (011) and $(0\bar{1}1)$ planes, which led to the use of (100) wafers for growth of laser structures, allowing the use of the (011) or $(0\bar{1}1)$ cleavage planes to form the mirrors of the feedback cavities. Most of the work on GaAs has continued to use (100) wafers. The (100) surface of GaAs consists of alternating simple square two dimensional lattices of Ga and As, as shown in figure A.2, with the three dimensional FCC lattice indicated. If the lattice is terminated on an arsenic plane, the dangling bonds are oriented along the [011] direction, or if it is terminated on a gallium plane, the dangling bonds are oriented along the $[0\bar{1}1]$ direction.

When a GaAs surface has its surface oxide removed by heating in a vacuum, a strong HEED diffraction pattern appears along the $[0\bar{1}1]$ and $[0\bar{1}\bar{1}]$ azimuths, consisting of a series of lines corresponding to the square surface lattice which has a periodicity of $a/\sqrt{2}$ where a is the lattice constant of GaAs. As the heating continues, and the surface starts to lose arsenic, the lines shorten, and those along the $[0\bar{1}1]$ azimuth start to become arrowhead shaped. When growth of GaAs commences in the usual arsenic stabilized regime, $\frac{1}{4}$ integral orders are seen in the $[0\bar{1}1]$ azimuth, corresponding to a period of $4(a/\sqrt{2})$ in the $[0\bar{1}\bar{1}]$ direction, and $\frac{1}{2}$ integral orders appear in the $[0\bar{1}\bar{1}]$ azimuth, corresponding to a period of $2(a/\sqrt{2})$ in the $[0\bar{1}1]$ direction.

During growth of MBE epilayers, it has been observed that defects originating from several sources, including the substrate surface and those associated with the gallium effusion cell, cause a growth enhancement defect on the surface of the wafer which is oval in shape, as shown in figure A.3. The long axis of these defects occurs in the $[0\bar{1}1]$ direction. Another surface defect which is seen in MBE epilayers, ripples with a period of a few microns which are associated with insufficient arsenic overpressure and high aluminum mole fraction growths, also have their crests oriented along the $[0\bar{1}1]$ direction.

The action of most wet chemical etching on GaAs surfaces is nonisotropic, in particular that of the important oxidation-dissolution etches. If a mask has an edge in the $[011]$ direction, the mask is undercut by these etches leaving an oblique angle overhang as shown in figure A.4, and this is referred to as the "dovetail" direction. Edges in the $[0\bar{1}1]$ direction are less undercut, and there is no overhanging edge. The $[0\bar{1}1]$ direction is thus referred to as the "vee-groove" direction. Reference [1] is a report on an extensive study of the crystallographic properties of these etches.

Table A.1 is a summary of these various orientation dependent properties.

References for appendix A

[1] D. W. Shaw, J. Electochem. Soc., 874, April (1981).

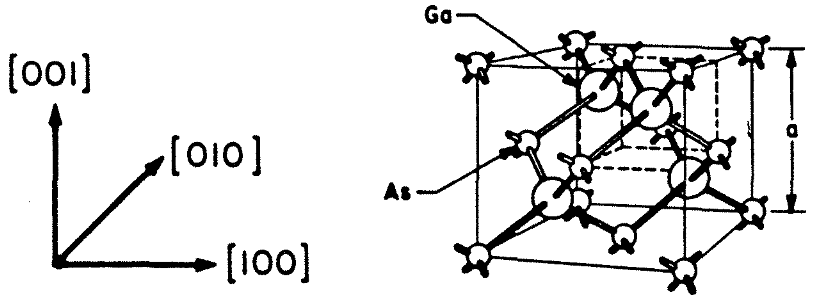
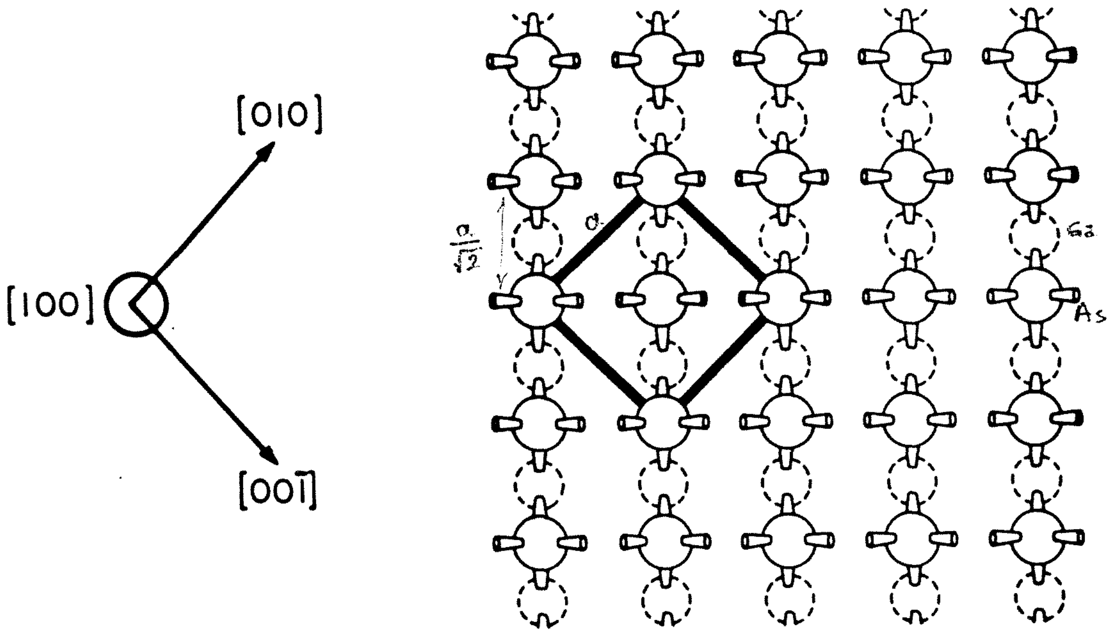


Figure A.1) Crystal structure of GaAs with conventional cell indicated.



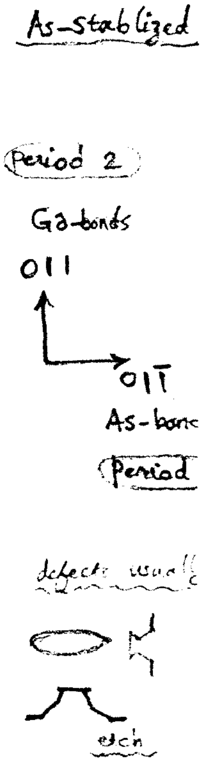


Figure A.2) (100) surface of GaAs showing orientation of conventional unit cell

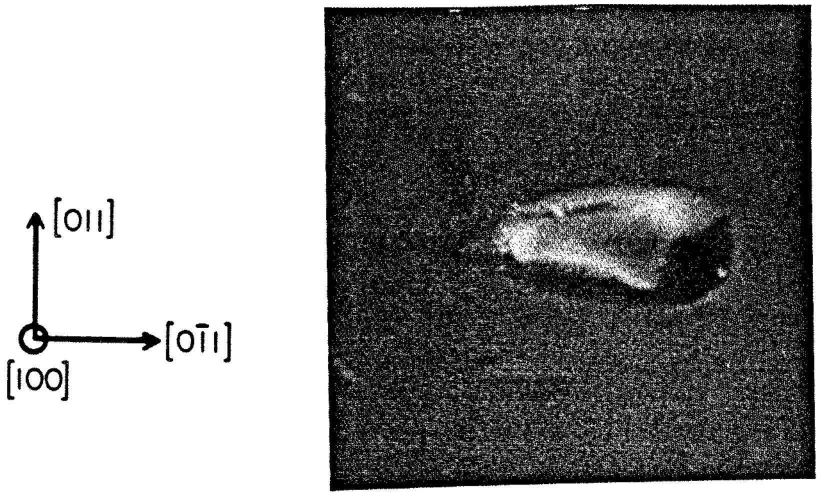


Figure A.3) Orientation of oval defects on (100) surface.

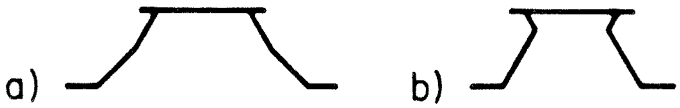


Figure A.4) Etch profiles on a (100) face. a) $[0\bar{1}1]$ ridge. b) [011] ridge.

	011	$0\bar{1}1$
As uppermost bonds	perpendicular	parallel
Ga uppermost bonds	parallel	perpendicular
HEED before growth		arrowhead shape
HEED (As stabilized)	$\frac{1}{2}$ integral orders	$\frac{1}{4}$ integral orders
HEED (Ga stabilized)	$\frac{1}{4}$ integral orders	$\frac{1}{2}$ integral orders
Oval defects	short dimension	long dimension
surface ripples	parallel to crests	perpendicular to crests
H_2O_2/acid etches	dovetail	vee-groove

Table A.1) Some of the direction dependent phenomena on (100) substrates.

Chapter II

Molecular Beam Epitaxy

2.1 Introduction

Molecular beam epitaxy (MBE)[1] is one of several techniques for growing high quality layers lattice-matched onto single crystal substrates. ③ As the name implies, the material is supplied to the surface of the wafer as beams of molecules from thermal cells in a vacuum chamber. This contrasts with liquid phase epitaxy (LPE)[2] in which the component materials are supplied from a metallic solvent bath, chemical vapor deposition (CVD),[3] in which the materials are supplied by a gas which reacts on the surface of the wafer to deposit the material, and metal organic CVD (MOCVD)[4] which is a special case of CVD where some of the reacting species are metal-carrying organic molecules.

One of the main virtues of the MBE technique is that since the materials are transported to the surface of the substrate by a collisionless beam, it is possible to modulate the incorporation of a material by simply imposing a shutter between the source of the beam and the growing layer. A typical growth rate during an MBE deposition is one atomic layer per second, or approximately one micron per hour, which means that a reasonably fast shutter action results in an atomically abrupt transition in the composition of the growing layer, if there are no significant diffusion or surface segregation effects. This makes possible the growth of structures which are bandgap engineered to the greatest possible precision. This allows, for example, the fabrication of devices in which a confining layer is thin enough so that electrons within it exhibit quantum size effects. These quantum effects are used in devices such as quantum well lasers and high electron mobility devices. Materials which are essentially synthetic crystals can be made by growing periodic structures with very thin layers, in some cases individual atomic layers,[5] deposited alternately on the substrate.

The differences between MBE and a simple vacuum evaporation deposition are the degree of elimination of contaminants in the source materials, on the substrate crystal, and in the vacuum chamber, and the processing of the single crystal substrate which provides a nearly ideal terminated lattice. The growth is carried out at an elevated substrate temperature, allowing the deposited material to crystalize into the lattice provided by the underlying substrate. In the case of compound semiconductors, it would seem an impossible task to supply the constituent materials at precisely the right ratio to grow a stoichiometric material. In this case, use is made of the fact that of one of the elements involved is usually much more volatile than the other at the substrate growth temperature. (The group V element in the case of III–V semiconductors). This allows an excess of that material to be supplied to the surface. That which is not immediately incorporated in the growing semiconductor is then re-evaporated. This makes it possible to grow a stoichiometric semiconductor material without requiring a degree of control over the beam fluxes and uniformity which would be impossible to achieve. For the group III materials, the sticking coefficient, the ratio of the atoms incorporated in the growth to those incident on the surface, is generally near unity, but re-evaporation must be taken into account at higher substrate temperatures.

It is a common misconception that the ultra high vacuum conditions under which MBE is carried out are intrinsically cleaner than the gas or liquid environments used by the other growth techniques. This is not the case, as can be seen by the fact that if we take the MBE chamber and fill it with a flowing high purity gas, the total rate of transport of contaminants to the surface of the wafer is reduced because of the short mean free path and reduced mobility of the impurities. In a vacuum chamber, molecules which are desorbed from the walls cross the chamber with only a very small chance of colliding with other molecules. Thus the only protection of the wafer from contamination is a low rate of desorption from the walls of the chamber due to special processing. The ultra high vacuum techniques of very clean handling of fixturing, special cleaning techniques, avoidance of all materials with significant vapor pressures or porosity, and high temperature baking, could all be applied to apparatus for LPE or CVD deposition, but they are in fact not necessary. This is due to the inhibition of the motion of impurities due to the presence of the flowing buffer gas, usually hydrogen, which can be purified by passage through a heated palladium metal membrane to levels unattainable for any other material. MBE growth of high quality materials puts stringent requirements on the absence of impurities within the growth chamber.

While a wide variety of III-V and II-VI semiconductor materials have been grown by MBE, the majority of work has been on GaAs and the ternary semiconductor alloy $Ga_xAl_{1-x}As$ because of the fortuity that GaAs and AlAs have nearly the same lattice constant, and thus the alloy is lattice matched over its full range. In most other compound semiconductor material systems, it is necessary to use quaternary alloys where one of the degrees of freedom is used to stay at lattice match. This seems to be violated in the case of strain-layer superlattices; however, it is still necessary to match the average lattice constant to that of the substrate material. This necessity to maintain the lattice constant by control of the material fluxes makes the process much more difficult, particularly because of the tendency of source fluxes to drift as the material in the cells is depleted. There is an additional complication in the case of alloys containing more than one group V compound, which is that the sticking coefficients are very dependent on the surface conditions, making it difficult to control the alloy ratio.

2.2 Equipment

The MBE system used in this research is a commercially manufactured Riber 2300 with an added liquid nitrogen control system and an automated data acquisition and control system which is described in appendix B. Figure 2.1 is a schematic diagram of the growth chamber. The essential features of the apparatus are:

1. Vacuum chamber. The main chamber is fabricated from welded stainless steel. All of the seals are silver plated copper rings which are compressed between machined knife edges. All motions within the chamber are coupled from the outside by magnetic fields or deformable bellows, eliminating sliding or rotating seals. The only materials used at temperatures above room temperature are, molybdenum, tantalum, tungsten, rhenium, silica ceramics, pyrolithic boron nitride, and sapphire. The only additional materials used in low temperature locations within the chamber are 302 stainless steel, silver plated copper, and glass, and titanium in the titanium sublimation pump and the ion pump.
2. Substrate manipulator. Mechanism for holding the substrate which is mounted on a molybdenum block, with a resistive heating coil and thermocouple for maintaining the substrate at an elevated temperature, and provision for rotation of the sample about its own axis for increased uniformity.
3. Infrared pyrometer. Provides additional information about the temperature of the substrate. Because of the provision for rotation of the substrate during the growth, the thermal contact between the substrate thermocouple and the substrate holder is necessarily weak, contributing to lags and offsets from the true temperature. The infrared pyrometer provides a more responsive temperature readout, with the restriction that it can only be used intermittently because of contamination of its window when it is used.

4. Knudsen cells for each of the constituent and donor materials. These consist of pyrolitic boron nitride crucibles which are heated by tantalum coils and are provided with tungsten-rhenium thermocouples for temperature measurement and control. Two types of arsenic effusion cells can be used. In a simple Knudsen effusion cell, the arsenic is evaporated as As_4. The As_4 must then be cracked at the substrate surface into As_2 before it can be incorporated into the growing layer. If a two section cell is used, a "cracker cell," with a second section running at a high temperature, the arsenic is cracked within the cell and supplied to the substrate as As_2. All of the work discussed in this thesis used a standard arsenic cell.

5. Tantalum shutters. Provided in front of each of the cells along with a main shutter in front of the substrate to provide abrupt initiation and cessation of the molecular fluxs onto the substrate. This allows the growth of very thin layers with sharp interfaces.

6. Liquid nitrogen shrouds surrounding all of the Knudsen cells and lining the interior of the vacuum chamber. These are necessary to avoid desorption of gasses from the walls during the growth which would cause contamination of the growing layers.

7. A high energy electron diffraction (HEED) gun and phosphor screen. This provides a high energy beam of electrons which is diffracted off the surface of the sample at a low angle $(2-3^\circ)$ and is observed on the phosphor screen. The electrons penetrate only slightly into the crystal, and the diffraction pattern is a sensitive probe of the conditions at the surface of the wafer.

8. Quadrupole mass spectrometer. This is used to diagnose the condition of the vacuum, including the presence of leaks or other contamination.

9. Substrate load lock chamber. The main chamber requires an interval of several days after it is exposed to air during which it is heated to 200° C to strip the walls of contaminants, and each of the cells is baked out at high temperature. Use of a load lock avoids going through this process each time a new substrate is loaded.

10. Titanium sublimation pump and 400 liter/second ion pump. (not shown in figure). Roughing from atmospheric pressure is accomplished with liquid nitrogen cooled sorption pumps, avoiding contamination from mechanical pump oils.

2.3 Surface analysis

One of the advantages of MBE is the vacuum environment, which allows the use of many surface analysis techniques which require a vacuum for operation. Many of the experiments which formerly needed to be done by cleaving in a vacuum in order to obtain a clean surface can be done on MBE grown surfaces, without exposing them to air, by doing the analysis in an attached chamber. These techniques include low energy electron diffraction (LEED), Auger electron spectroscopy, and secondary ion emission spectroscopy (SIMS). These techniques have been applied within the growth chamber as well, but the apparatus does not tolerate the metal molecular beams well.

High energy electron diffraction is often used as a surface diagnostic during MBE deposition. A high energy ($\sim 8 - 15$ Kev) electron beam is directed at the crystal surface at a low angle, and the electrons diffracting off the surface are viewed on a phosphor screen. HEED is usually the only surface analytical tool used inside the growth chamber because the glancing incidence of the beam allows both the electron gun and the observation screen to be well protected from the molecular beams. The energy of the component of the velocity of the electrons into the substrate is on the order of a few hundred electron volts because of the glancing angle of incidence. The diffraction pattern from a crystal with a smooth surface is that of a two dimensional surface, which is a set of lines at the reciprocal lattice of the crystal, due to the rapid attenuation of the beam in the crystal. The diffraction pattern from a smooth amorphous oxide layer is a series of rings, corresponding to the average distance between adjacent atoms. The diffraction pattern of a rough surface is similar to that of bulk diffraction, an array of points, because the electron beam can penetrate the surface projections and diffract off the three dimensional lattice. The presence of these features allows diagnosis of surface conditions during the growth. When a GaAs surface has its surface oxide removed by heating in a vacuum, a strong diffraction pattern appears along the $[0\bar{1}1]$ and $[0\bar{1}\bar{1}]$ azimuths, consisting of a series of lines corresponding to the square surface lattice which has a periodicity of $a/\sqrt{2}$ where a is the lattice constant of GaAs. As the heating continues, and the surface starts to lose arsenic, the lines shorten, and those along the $[0\bar{1}1]$ azimuth start to become arrowhead shaped. When growth of GaAs commences in the usual arsenic stabilized regime, $\frac{1}{4}$ integral orders (diffraction lines with a period $\frac{1}{4}$ that of the ideal surface lattice) are seen in the $[0\bar{1}1]$ azimuth, corresponding to a period of $4(a/\sqrt{2})$ in the $[0\bar{1}\bar{1}]$ direction, and $\frac{1}{2}$ integral orders appear in the $[0\bar{1}\bar{1}]$ azimuth, corresponding to a period of $2(a/\sqrt{2})$ in the $[0\bar{1}1]$ direction. Figures 2.2 and 2.3 show the diffraction pattern of a GaAs substrate in the $[01\bar{1}]$ and $[011]$ azimuths at the start of a growth.

2.4 Substrate cleaning

The method originally used to produce the clean surface required for MBE growth was ion bombardment. While this method produces a surface which is essentially free of impurities, the substrate is left with surface damage which is only partially removed by annealing. The necessity for ion bombardment cleaning of substrates was eliminated by the discovery by Cho and coworkers[6] of a series of chemical treatment steps for GaAs which remove almost all of the carbon contamination of the substrate and leave a thin native oxide which protects the surface until it is introduced into the vacuum chamber. Just before the growth is to start, the substrate temperature is raised to approximately 580°C and the oxide is desorbed, leaving the surface free of contamination.

Cleaning procedures for each of the surfaces involved in the growth include solvent rinses, using multiple solvents, to dissolve as much of the contamination as possible; surface etching steps, which are necessary to remove unknown contamination by lifting off contaminants which may not be soluble, and high vacuum baking, raising the surface to a high temperature to rapidly desorb any remaining contaminants and pump them away.

The substrate cleaning procedure described is a modification of the one described by Cho.[6] In the following, "rinsing" refers to moving the wafer thorough several baths of solvent, changing the solvent each time. In the case of deionized water baths, this was carried out approximately 10 times for each "rinse." It is easy to underestimate the importance of careful experimental technique while cleaning substrates. For example, a single drop of water falling from a wet glove can greatly increase the number of surface defects when the epitaxial layers are grown. New glassware must be thoroughly cleaned, and reserved for the wafer preparation process.

The commercially grown and polished wafers are degreased in boiling trichloroethylene (TCE) and then rinsed thorough acetone, methanol, and deionized water baths. They are then soaked in HCl for two minutes, which removes the native oxide and some organic contaminants. The HCl is removed with a deionized water rinse, followed by a thorough rinse in methanol. The sample is hand polished on a lens tissue which is soaked with a dilute bromine/methanol solution (~ 1 minute in a .5% solution by volume) on a glass optical flat, and re-rinsed in methanol and deionized water. It is essential to avoid contamination of the sample or the solutions with water during the polishing process, as the presence of even small amounts of water while the wafer is in a bromine/methanol solution results in a reaction which rapidly clouds the surface and makes the substrate unusable. After the polishing process, the wafers are handled in baskets which makes it possible to transfer them from one bath to another without letting the bead of solution which remains on the surface roll off. This bead protects the surface of the wafer from contamination by the air, and also serves as a sensitive test for contamination. If the surface of the wafer is contaminated with organics, the bead will roll off the surface easily due to the reduced surface tension. The samples are free etched in 4:1:1 (H_2SO_4, H_2O, H_2O_2) for 5 to 10 minutes, rinsed in deionized water, and soaked in HCl for two minutes. This HCl soak strips the oxide that was left by the 4:1:1 bath and prepares the surface for the growth of the thin protective oxide which can be desorbed just before the growth. The sample is transferred to a deionized water bath where the protective oxide is grown, where it is left for two minutes. If the sample is removed prematurely, the solution will roll off the surface due to the low surface tension of water on an unoxidized GaAs surface, and the surface will be contaminated. Finally, the samples are rinsed in deionized water, blown dry with a nitrogen jet, and mounted for growth.

2.5 Substrate mounting

The samples are mounted for growth by soldering them onto a molybdenum holder. The soldering process is used to provide good thermal contact with the holder, and to avoid stresses on the sample which would be created by clips or other mechanical holders. The indium solder which is used is liquid at the growth temperature, and the wafer is held only by the surface tension of the liquid. The molyblock (sample holder) is first cleaned by etching in concentrated HCl and a bromine methanol solution. It is then etched for a short period in a 50-50 mixture of concentrated nitric acid and hydrochloric acid, and rinsed in flowing deionized water. This etch removes the oxide from the surface of the metal as well as etching the metal itself to provide a clean mounting surface. The mount is then rinsed in acetone, and blown dry with a nitrogen jet. It is necessary leave the substrate in acetone before it is blown dry, because the presence of water will quickly oxidize the surface.

Soldering the wafer onto the holder without leaving gaps is one of the more difficult processes. The clean and freshly etched molyblock is then placed on a hot plate heated to 330°C. After it reaches the melting point of indium, which is 160°C, indium is quickly spread over the surface with a tantalum wire. The sample is then placed on the surface and moved back and forth until the bubbles are removed from the indium. If the temperature of the heater during mounting is too high, or if the molyblock is left for more than a brief time without being covered by indium, the surface will oxidize and the indium will not wet the surface. If the temperature is too low, the mounting process will leave gaps in the indium due to the sample being cold, which freezes and sheers the indium layer. Even small gaps in the indium cause the temperature of the surface during the growth to be low and nonuniform, and affects the properties of the grown layer.

After the sample is placed in the load lock and the chamber is pumped down, it is briefly raised to 250°C. This removes adsorbed water vapor and other lightly adsorbed gases without affecting the wafer's protective oxide coating.

2.6 Growth

After the sample is transferred into the growth chamber, the liquid nitrogen shrouds are filled, and the Knudsen cells are heated to the temperatures corresponding to the desired composition and doping density. The substrate is then heated to 580°C over approximately five minutes, while the surface is monitored with the high energy electron diffraction apparatus. As the substrate temperature approaches 580°C, the protective oxide is evolved in two pulses which are observable on the chamber ionization gauges as shown in figure 2.4. As the oxide desorbs, the HEED pattern changes from a diffuse scattering off the disordered surface of the oxide into the parallel lines generated by diffraction off the two dimensional lattice of the crystal. At the appearance of the diffraction pattern, the arsenic cell shutter is opened to provide a flux of arsenic to replace that which is being lost from the surface, and the heating is continued to the desired growth temperature, whereupon the other shutters corresponding to the desired material composition are opened. The growth then proceeds, with temperatures and shutters controlled to form the desired structure.

2.7 Growth temperature

The optimum substrate temperature for a MBE growth depends on the desired properties of the layers which are to be grown. GaAs will grow single crystal on a clean substrate at temperatures over approximately 350°C, and growth has been reported in excess of 800°C. At high substrate temperatures, arsenic is lost from the surface at a high rate, and the rate of use of arsenic to replace that lost from the surface becomes prohibitive. The photoluminescence of GaAlAs improves for higher substrate temperatures, but interface smoothness decreases. For example, electronic devices such as the high electron mobility transistor (HEMT) are grown at low temperatures, for optimum interface smoothness and high mobilities, and multiple quantum well lasers are grown at as high a temperature as possible to decrease nonradiative recombination in the GaAlAS barrier layers between the quantum wells.

2.8 Flux calibration

To grow material with the desired composition, it is necessary to calibrate the molecular fluxes from each of the Knudsen cells. In a true Knudsen cell, both the gas phase and the liquid or solid phase are present in equilibrium, and the molecular flux from a small opening in the cell at the substrate is

$$F = \frac{PA}{\pi l^2 \sqrt{2\pi MkT}} \tag{2.1}$$

where P is the equilibrium pressure, A is the area of the cell opening, l is the distance between the cell and the substrate, k is Boltzmann's constant, M is the molecular mass, and T is the cell temperature. This allows an estimate of the fluxes from the vapor pressure curves; however, the restriction on the cell opening necessary to approach equilibrium within the cell is rarely met by ordinary MBE cells, and it necessary to measure the actual cell fluxes. The Ga, Al, and As fluxes can be measured by placing a Bayard Alpert gauge in the beam. Since a Bayard Alpert gauge measures the the density of the gas within its cage, the beam flux is

$$F = N\bar{v} = N\left(1/2\sqrt{\frac{m}{2\pi kT}}\right) \tag{2.2}$$

where N is the numerical density measured by the ionization gauge, and T is the temperature of the beam. Since the ionization gauge is internally normalized to read as an equivalent pressure at room temperature, $N = P_g/Ik \cdot 300°\mathrm{K}$ where P_g is the pressure reading from the gauge, and I is the ratio of ionization efficiency of the beam molecules compared to that of nitrogen, which is the gas the gauges are calibrated for. The measurement is somewhat inaccurate in the case of arsenic, because arsenic does not stick well on surfaces even at room temperature, and it therefore reflected off the gauge guard back into the gauge cage. In the case of Ga and Al, the sticking coefficient on growing layers for temperatures less than 620°C is unity, and the fluxes can be calibrated by growing material under known conditions and measuring the thickness of the layers produced. The layer thicknesses can be measured under a microscope by cleaving off a section of the wafer which can then be stained with a selective etch, for example a 30% H_2O_2 solution whose pH has been raised to 7.02 by adding a base, which etches GaAs layers faster than GaAlAs layers, and then looking at the cross section under a scanning electron microscope (SEM) or a phase contrast optical microscope. The flux needed from the dopant cells is too small to be measured by direct techniques, so it is measured by growing layers and then measuring the carrier density using a Hall measurement. The dopant density, combined with the known growth rate, yields the dopant flux. Typical flux values for each of these materials are given in figure 2.5.

2.9 Source preparation

Each of the source materials used for deposition must be decontaminated before loading into the Knudsen cells. Each of the materials is commercially available in purities of 6 9's or higher (99.9999 percent pure) except for beryllium. Arsenic is delivered in small lumps sealed in glass ampules, under an inert gas. The ampule is broken just before loading, and care must be taken to avoid contamination of the arsenic with glass fragments. Gallium is purchased in bottles at 7 9's purity. No cleaning is done, but care must be taken when it is loaded to avoid the trapping of gas under the liquid gallium. Aluminum is obtained in pellet form.

It is etched in concentrated hydrochloric acid for five to ten minutes, and rinsed thoroughly in deionized water before loading. Only small amounts of aluminum can be loaded into the cell (5-10 grams) at a time because its wets the cell walls and a large quantity will collapse them when it contracts as it cools. Tin is used as an n-type dopant. It is etched in concentrated hydrochloric acid and rinsed in deionized water before loading. Tin can be used to obtain very high doping levels, $\sim 10^{19}\text{cm}^{-3}$, but it accumulates on the surface during the growth,[7] and doping transitions are limited to a rate of approximately one order of magnitude for every 1000 angstroms. Silicon is used for n-type doping where abrupt interfaces are required. High purity silicon is available in ingot form, which is fragmented and an interior piece is loaded into the cell. Beryllium, which is the only satisfactory p-type dopant, is not readily available in high purities because of its toxicity. A batch of single crystal beryllium was grown in Germany a few years ago, and only small amounts of it are still available. The pieces now available are so small that no cleaning procedure is usually attempted except for the high temperature bake.

In the next chapter, a technique is discussed for extending this growth procedure to substrates which have previously formed structures on their surface and cannot be cleaned by the established procedures, which heavily etch the surface before it is mounted in the chamber for growth.

References for chapter 2

[1] W. T. Tsang, SEMICONDUCTORS AND SEMIMETALS, vol. 22, part A, (1985).

[2] H. C. Casey, Jr., and M. B. Panish, Heterostructure lasers, pp. 109, Academic Press, New York, (1978).

[3] R. A. Burmeister, Jr. G. P. Pighini, and P. E. Greene, Trans. AIME 245, 587 (1969).

[4] R. D. Dupuis, P. D. Dapkus, R. D. Yingling, and L. A. Moudy, Appl. Phys. Lett. 29, 141 (1976).

[5] A. C. Gossard, P. M. Petroff, W. Wiegmann, R. Dingle, and A. Savage, Appl. Phys. Lett. 29, 323, (1976).

[6] A. Y. Cho, H. C. Casey, Jr., C. Radice, and P. W. Foy, *Electron. Lett.* **16**, 72 (1980).

[7] A. Y. Cho, J. of Appl. Phys. Vol 46, No. 4, 1733, (1975).

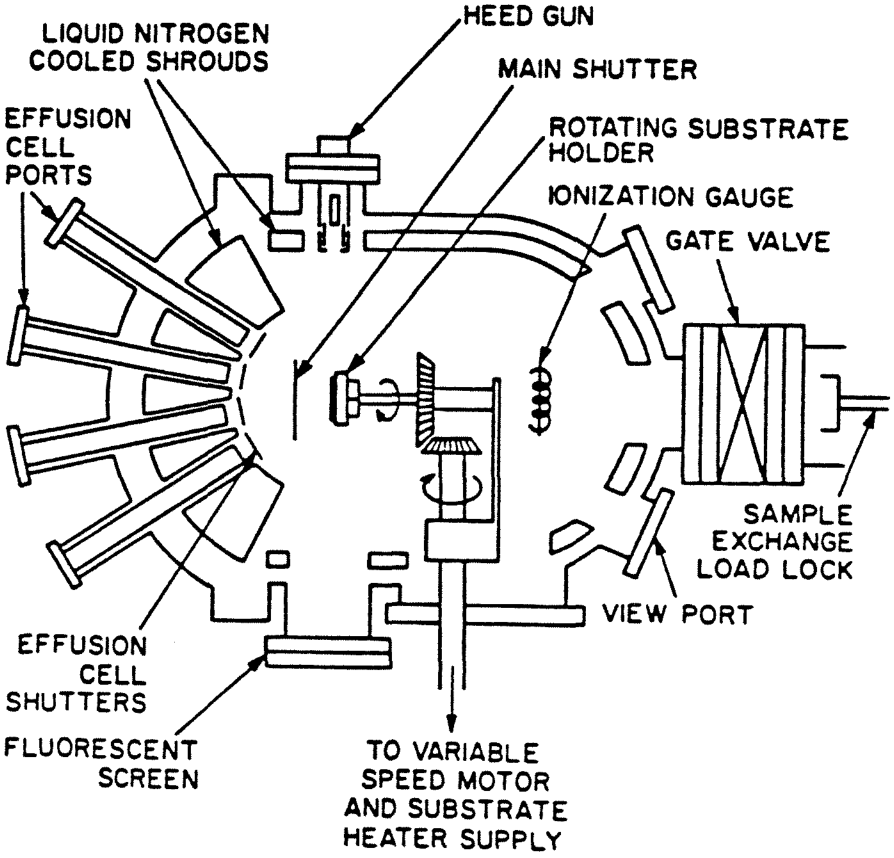


Figure 2.1 Schematic diagram of MBE system.

Figure 2.2 HEED diffraction pattern at $[0\bar{1}1]$ azimuth under arsenic stabilized conditions.

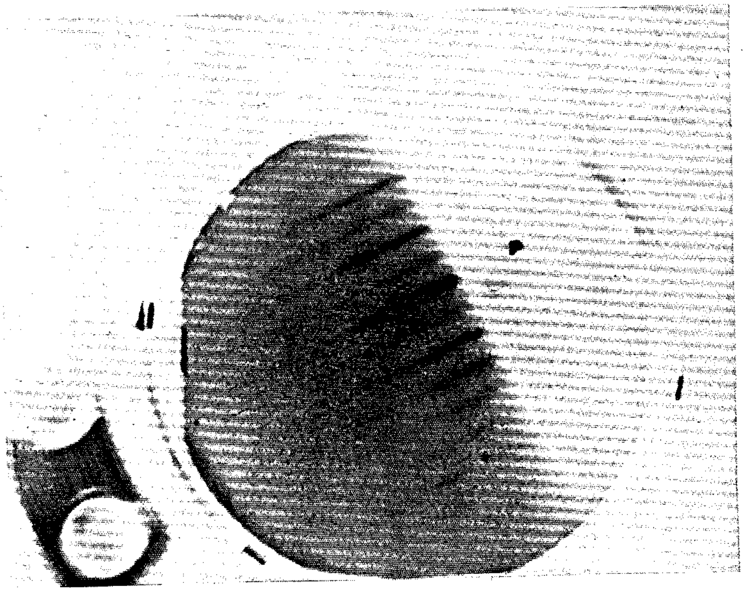


Figure 2.3 HEED diffraction pattern at $[0\bar{1}\bar{1}]$ azimuth under arsenic stabilized conditions.

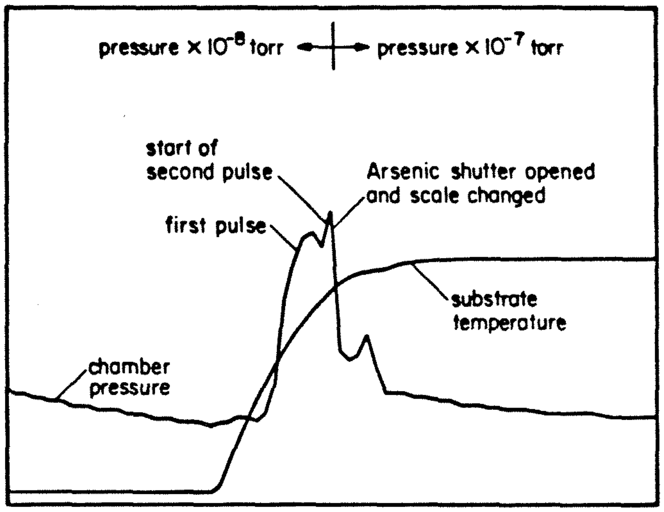


Figure 2.4 Pressure curve at the start of a growth, showing double pulse of oxide desorption.

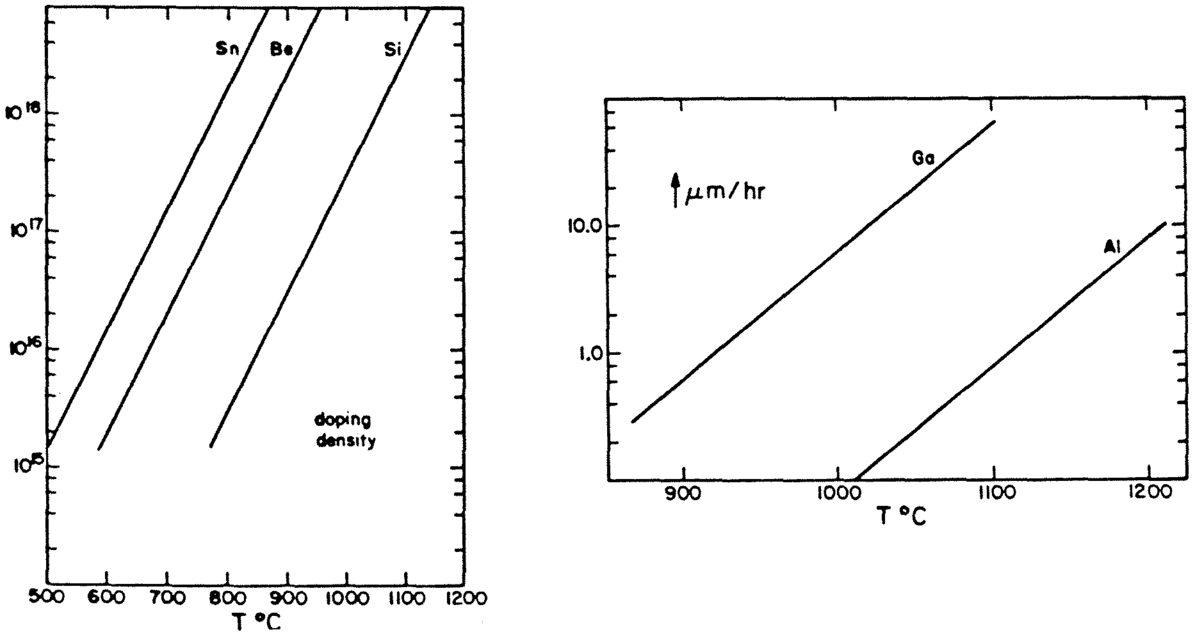


Figure 2.5 Typical flux vs. temperature curves for a Riber 2300 system.

Appendix B

MBE control system

Molecular beam epitaxy is a valuable growth technique for research because of the accuracy to which the growth can be controlled. Alloy composition and doping densities are very accurate and reproducible, limited only by the accuracy to which the temperatures of the cells are constant from growth to growth. Structures with large numbers of very thin layers can be grown simply by opening and closing shutters in front of each of the Knudsen cells. The large number of control operations during a growth, dozens in a simple growth and hundreds in a superlattice growth, makes computer control very desirable. In addition, it is necessary to monitor a large number of parameters affecting the growth conditions, as well as checking the many pieces of equipment both inside and outside vacuum.

Figure B.1 shows the hardware configuration of such a control system. A commercially available system interface, the Hewlett-Packard 3497, was used as the central part of the control apparatus. This unit has both analog and digital control outputs for control of the temperature controllers and the shutter stepper motors, as well as a low offset scanner for the thermocouple readings. An external $6\frac{1}{2}$ digit voltmeter was used to achieve the necessary .1° resolution with the low EMF tungsten-rhenium thermocouples.

As shown in figure B.2, the software is divided into three main parts: the control program, a system allowing the generation and storage of control files containing the list of procedures for growth of particular structures, and a report generation system allowing data taken during the growth to be examined in detail. The MBE control program is an interrupt driven program, with four pseudo-concurrent processes. The base process is a screen control process, graphing any of the monitored variables chosen by the operator. The secondary process is the operator input process, activated only when keys are pressed, allowing the operator to override the control file and to specify temperatures, pressures or other data that he wishes to monitor. The third process is the data acquisition process, activated every five seconds, which records the inputs from each of the sensers. The fourth, highest priority process, is activated whenever a control action is required, for example, the opening or closing of a shutter, or change of a temperature.

The use of separate programs for each phase of the process and communication through standardized files allows complete records of all growths to be kept for later reference, and allows a growth description file to be kept, modified, and reused. The interrupt driven control program allows detailed measurements to be taken during the growth, while still taking control action with an accuracy of one tenth of a second, much less than the time for deposition of one monolayer on the sample.

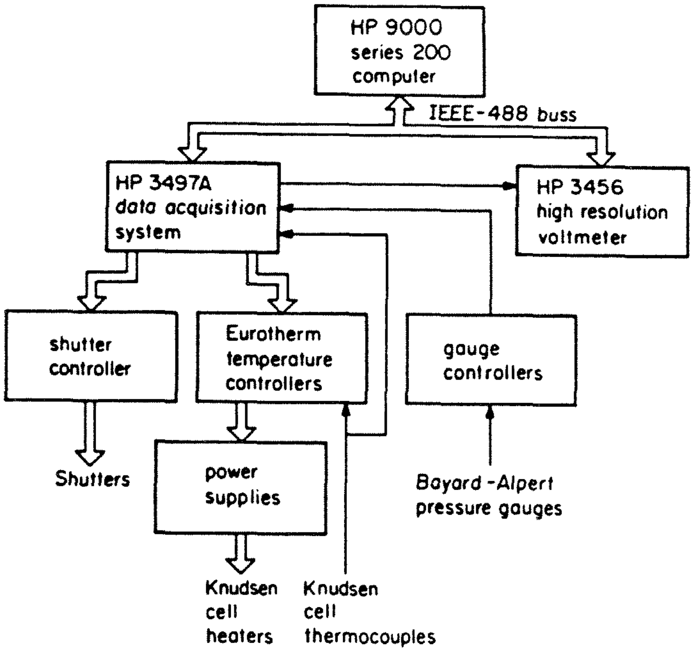


figure B.1) MBE control system hardware configuration.

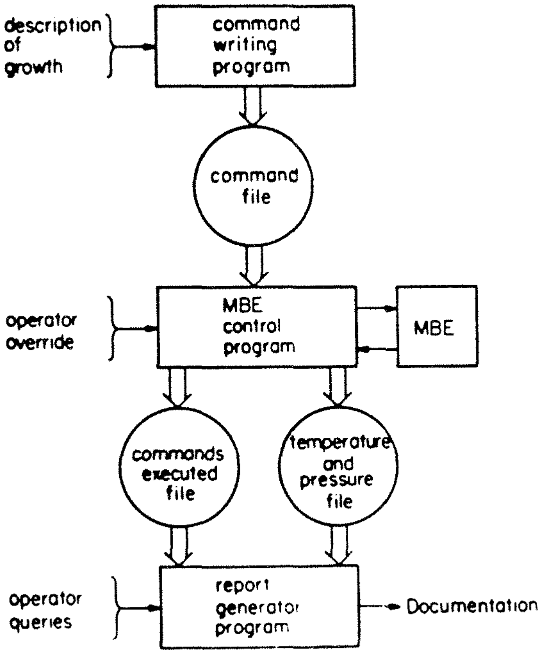


figure B.2) MBE control system software configuration.

Chapter III

MBE growth on patterned GaAs substrates

3.1 Introduction

Much of the success of electronics lies in the essentially parallel way in which devices are fabricated. Large semiconductor wafers are processed in a succession of steps which include application of photoresist, treatment of the partially masked surface, and removal of the photoresist, which are repeated until the desired structures are formed. The processes that are used include wet or dry etching, ion implantation, thermal diffusion of materials into the surface, deposition of metals and dielectric materials by evaporation, sputtering, or chemical vapor deposition (CVD), and epitaxy by CVD, or liquid phase epitaxy (LPE), among others.

Dimensions in the vertical direction can usually be controlled very tightly because the processes applied usually increase in depth, or increase the thickness of the layer they form, with time. Since slow processes can be used and the time that they act can be precisely determined, control over vertical dimensions is only limited to the extent that the conditions during the process are repeatable and uniform. In the case of MBE and metal organic chemical vapor deposition (MOCVD), for example, structures can be grown with a vertical resolution of a few atomic layers.

In the lateral direction, dimensions are limited to the tolerance of the photolithographic process, which is now near the limits imposed by the diffraction of the light used to form the image, on the order of one micron. This asymmetry in control over dimensions in the lateral versus the vertical dimension has resulted in the majority of electronic devices being treated separately in the vertical and lateral directions in both design and analysis, with effects due to the actual three dimensional structure of the device treated separately as edge effects. For example, conventional field effect transistors (FETs), have as their basic structure a channel layer, an insulator or other isolation layer, and a gate, in the vertical direction. The lateral dimension is used to provide contacts for the channel at each end, and the contact for the gate. Bipolar junction transistors (BJT)s have a three layer structure in the vertical direction, emitter, base, and collector, and the most important effect in the lateral dimension is the spreading resistance in the base, a parasitic effect. In order to increase the density of electronic integration much beyond its present level, it is becoming necessary to consider the devices as fully three dimensional both during design and during analysis.

Optoelectronic devices, on the other hand, require three dimensional structure because of the need to control the propagation of light with a waveguide. This has resulted in the development of techniques such as LPE growth over channels or around ridges to achieve the necessary geometries.

3.2 MBE limitations

The molecular beam epitaxy technique has been more limited than other semiconductor processing techniques because of the low tolerance of the growth technique to impurities on the surface. Carbon contamination of the substrate as low as .1 monolayer has been found to degrade epitaxial layer quality.[1] This low tolerance has largely restricted the technique to epitaxial growths on native substrates, with lateral definition only coming in subsequent processing; that is, when MBE layers are used in the design of a device, they must be the first process in the sequence. Methods of protecting the surface after the completion of a growth enabling the reinitiation of growth after exposure to air have been developed, including deposition of pure As layers[2] and growth of thin InAs layers[3] which can be thermally removed later, but these methods only protect areas which are left untreated until the second growth.

In this chapter, a technique is presented for growing epilayers by MBE on wafers which have previously formed structures, such as grown layers or etched patterns. Such structures would be entirely removed by the cleaning procedures applied to flat substrates, and it was necessary to find modified procedures which would still provide a surface suitable for MBE, yet not etch away structures which are on the surface of the wafer. Fabrication of new types of devices by this method was proposed by Tsang and Cho in 1977;[4] however, this is the first technique which allows the realization of device quality material over surfaces with pre-existing features less than a few microns in extent. The technique involves a cleaning process for GaAs surfaces in preparation for MBE growth which removes comparatively little material while still providing a surface which allows epitaxial growths with oval defect densities of $\sim 10^4$ cm^{-2}(still about an order of magnitude higher than the density achievable on flat substrates). The ability to grow high quality epilayers by MBE on previously processed wafers allows greater flexibility in design of devices, for example the integration of devices which require different layer structures,[5] and devices such as the laser structure which will be discussed in a later chapter which use the three dimensional features of the growth for their basic structure. In addition, the growth over a patterned surface yields new information about the MBE growth process itself, such as surface diffusion lengths and propagation of steps.

3.3 Cleaning technique

The cleaning procedure is a modification of the procedure described in chapter 2, with the etching times reduced, the bromine/methanol polishing solution at a much lower dilution, and the use of a specially designed pyrex apparatus which continuously distilled trichloroethylene (TCE) and flowed it over the surface of the wafer. The distillation apparatus, which is shown in figure 3.1, consists of a flask in which TCE is evaporated, and an attached reflux coil which condenses the TCE vapor and supplies it to a drip tube over the wafer, which is held in a pyrex basket. The TCE is then returned to the flask for reuse. The flow of water through the condensing coil is controlled so that the temperature of the liquid TCE coming off of the coil is still near its boiling point, but not so high that TCE vapor is lost from the apparatus. The purpose of the apparatus is to supply a continuous flow of hot, high purity TCE over the wafer for an extended period, which dissolves, dilutes, and washes away contaminant organics, without etching the surface.

Wafers are patterned using conventional photolithography, and etched in 1:8:8 (H_2SO_4, H_2O_2, H_2O) to form features of the desired depth. The patterned wafers are then boiled in TCE for ten minutes and placed in the TCE distillation apparatus for 24 hours. The wafers are then lightly hand polished on a lens tissue soaked in a very dilute bromine/methanol solution ($\sim$ 1 minute in a .005% solution by volume), and dip etched in 4:1:1 (H_2SO_4, H_2O_2, H_2O) for 10 seconds (which removes $\sim$ .4 microns of material). Next, they are soaked in HCl for two minutes, and a protective oxide is grown in a deionized water bath.[7] Finally, the samples are then rinsed and mounted for growth.

Figure 3.2 shows the surface of a [100] wafer which was etched with a pattern of 9 micron squares. Note the absence of growth-enhanced oval defects. Flat areas of the substrates have approximately 10^4 oval defects per square centimeter, which is in the upper range for growths reported on flat substrates in the literature. By comparison, flat substrates prepared by the technique presented in the previous chapter have typical oval defect densities of $10^3 \mathrm{cm}^{-2}$ after growth. The lack of oval defects in the patterned region may be accounted for by the fact that the macroscopic raised feature of the defects is due to an enhanced growth rate for facets which are nucleated by the defect, but in the concave sections of a patterned wafer, the facets which form are the fast growth facets, and outpaces growth in the region around the defects. Another possibility is that the etching of the material in the patterned areas provides additional cleaning which improves the oval defect density in those areas.

In summary, a technique for growing high quality GaAs/GaAlAs epilayers over patterned substrates has been presented. Substrates with surface features as small as two microns have been prepared by this technique, as will be discussed in chapter 6. This technique should prove useful in the fabrication of a wide variety of devices which take advantage of the geometric features of regrowth by MBE, several of which are discussed in later chapters, and for fabrication of integrated structures.

References:

[1] A. Y. Cho, M. B. Panish, J. Appl. Phys. 43, 5118 (1972).

[2] S. P. Kowalczyk, D. L. Miller, J. K. Waldrop, P. G. Newman, and R. W. Grant, J. Vac. Sci. Technol. Vol. 19, pp. 255 (1981).

[3] Y. J. Chang and Herbert Kroemer, J. Vac. Sci. Technol. B Vol. 3, No. 2, pp. 518, (1985).

[4] W. T. Tsang and A. Y. Cho, *J. Appl. Phys.*, **30**, 293 (1977).

[5] T. Sanada, S. Yamakoshi, O. Wada, T. Fujii, T. Sakurai, M. Sasaki, *Appl. Phys. Lett.* **44**, 325, (1984).

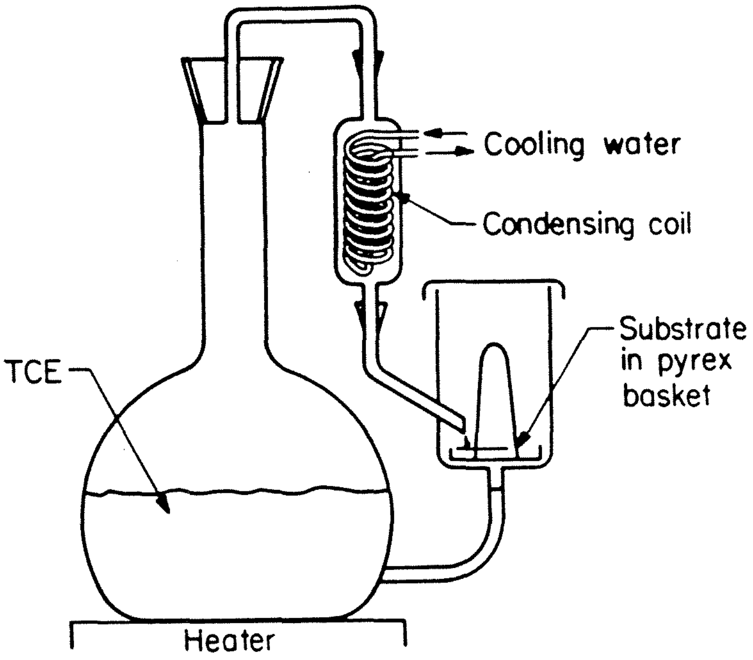


Figure 3.1) Trichloroethylene distillation system for decontaminating wafers.

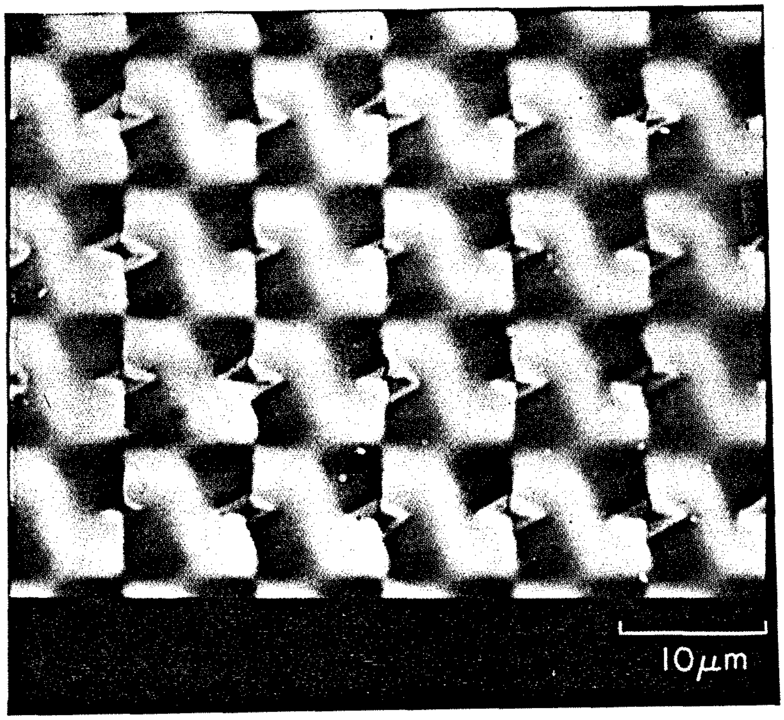


Figure 3.2) Growth of 3 μm of $Ga_{.7}Al_{.3}As$ over a (100) substrate etched with a pattern of 9 μm squares

Chapter IV

MBE facet formation and growth properties

4.1 Introduction

One of the long standing limitations of MBE has been its lack of lateral definition. Despite the ability to deposit layers which in some cases can be as thin as monolayers of atoms in the vertical direction,[1] the lateral definition of the devices formed can only come in later steps. Growth over patterned substrates has been proposed as a useful technique to increase the flexibility of the molecular beam epitaxy (MBE) process,[2] allowing fabrication of devices which have three dimensional structure[3] or the integration of devices which have different epitaxial layer requirements.[4] However, the low tolerance of MBE to contamination of the substrate has limited these growths to structures with surface features on the order of ten microns or more, which can be cleaned by techniques which etch the original surface extensively. Other methods which allow the use of MBE to grow epilayers with lateral definition include mechanical masks[5], masking with dielectric layers[2], and growth over etched[3,6] or previously grown layers.[4] However, except for regrowths over coarsely featured substrates,[4,7] few electrical or optical properties, or devices which use these techniques have been presented in the literature, presumably because of difficulty with the quality of the epitaxial layers.

This chapter discusses a series of experiments that were done to investigate temperature and flux ratio dependence of MBE growth and faceting for temperatures ranging from 580° C to 700° C and for As/Ga flux ratios from 1.4:1 to 4:1, for the purposes of finding growth geometries which might be useful for devices, and as a probe into the MBE growth mechanism itself.

Despite its outward similarity to vacuum deposition of metals and dielectric layers, MBE growth is not the simple uniform deposition of material which might at first be supposed. It was first noted by Tsang et al.[8] that MBE growths over grooves etched in a GaAs substrate revealed faceting at the edges of grooves, and a tapered edge growing under shadowed areas. This occurs because, despite the quite nonequilibrium way the constituent species are delivered to the surface of the growing wafer, the incorporation of the atoms into the lattice is actually at near equilibrium with a layer of molecules which are mobile on the surface. Significant departures from equilibrium in the growth process itself would result in a poor quality semiconductor, in contrast to the very high quality material which is produced by MBE. The presence of this mobile population means that materials which are supplied to the surface can move over the surface before they are incorporated, allowing the formation of facets.

4.2 Experimental results

The growths to be discussed in this chapter were produced by the following procedure. Grooves were etched with a nine micron period in the $[0\bar{1}\bar{1}]$ and $[0\bar{1}1]$ directions, respectively, on (100) substrates. After cleaning by the procedure discussed in chapter 2, one of two growth schedules was used. In the first, alternating 1000Å layers of GaAs and $Ga_{.2}Al_{.8}As$ were grown to allow accurate determination of the surface shape as a function of time. In the second growth schedule, alternating 4500Å layers of GaAs and 500Å marker layers of $Ga_{.8}Al_{.2}As$ were grown, minimizing the effect of aluminum on the surface, and allowing the growth rates for pure GaAs to be measured.

The growths were carried out with continuous rotation of the wafer at 12 rpm. This rotation prevents the asymmetrical growth noted in earlier papers[3,6,7] due to the angle of incidence of the molecular beams. As shown in appendix C, the flux on a facet with a tilt of θ is, for a rotating substrate, proportional to $\cos\theta$ for all of the Knudsen cells, provided the angles of the facets and of the effusion cells are not large enough to cause shadowing during some part of the rotation.

Figures 4.1 and 4.2 show the development of facets on ridges in the $[0\bar{1}\bar{1}]$ and $[0\bar{1}1]$ directions, respectively. The differences in the substrate groove shapes are due to the anisotropy of the etch used to form them. The etching during the cleaning steps is sufficient to remove the dovetail overhang occurring on ridges in the $[0\bar{1}\bar{1}]$ direction. If the overhang is not completely removed, leaving an oblique edge, gaps form in the epitaxial layers as shown in figure 4.3, and the layers are broken and of poor quality.

4.3 Crystal growth

On a concave surface, facets which grow more slowly than their neighboring facets get smaller as the growth progresses, and eventually disappear, as shown in figure 4.4. On convex facets, slow growing facets get larger, as shown in figure 4.5. Thus, observation of facet formation on crystal surfaces yields upper or lower bounds for the growth rates of the facets at intermediate angles to the observed facets. The relative growth rates of adjacent facets can be accurately measured from the propagation angle of their vertex. If the tilt angles of two adjacent facets are taken as α_1 and α_2 and θ is the angle of propagation of their intersection during the growth,

$$\frac{x_1}{x_2} = \frac{\cos\phi_1}{\cos\phi_2} = \frac{\cos(\alpha_1 - \theta)}{\cos(\alpha_2 + \theta)} \tag{4.1}$$

where x_1/x_2 is equal to the ratio of their growth rates, as shown in figure 4.6. Using this result, the ratio of relative growth rates for adjacent facets can be measured. Inverting equation 1 to find θ as a function of the growth rates,

$$\tan\theta = \frac{\cos\alpha_1 - x_1/x_2\cos\alpha_2}{\sin\alpha_1 - x_1/x_2\sin\alpha_2} \tag{4.2}$$

The restriction on the growth rate of the intermediate facet on a concave surface such that it will get smaller as the growth progresses is then found by requiring that $\theta_1 < \theta_2$, where θ_1 is the facet propagation angle for the left vertex and θ_2 is the facet propagation angle for the right vertex, and we have

$$x_2 < \frac{x_1 \sin(\alpha_2 - \alpha_3) + x_3 \sin(\alpha_1 - \alpha_2)}{\sin(\alpha_1 - \alpha_2)} \tag{4.3}$$

where x_1, x_2 and x_3 are the growth rates of the left, center and right facets, respectively, and α_1, α_2, and α_3 are the facet tilt angles. Similarly, for a convex surface,

$$x_2 > \frac{x_1 \sin(\alpha_2 - \alpha_3) + x_3 \sin(\alpha_1 - \alpha_2)}{\sin(\alpha_1 - \alpha_2)} \tag{4.4}$$

This implies, for example, that the $(8\bar{1}\bar{1})$ and (811) facets are preferred over any facet on the (100) faces tilted ±10° in the $[0\bar{1}\bar{1}]$ direction, and the $(4\bar{1}\bar{1})$ and (411) facets are the least preferred of facets on the (100) faces tilted ±19.5° in the $[0\bar{1}\bar{1}]$ direction. "Preferred" means that of two adjacent facets, the more preferred one will grow at the expense of the other facet. In the perpendicular direction, the (100) facet is the slowest growing facet for tilt angles ±19.5° in the [011] direction.

The wafer in figure 4.7 was produced by etching a pattern of squares with a 9 μm period in both the $[0\bar{1}\bar{1}]$ direction and the $[0\bar{1}1]$ direction, followed by a growth of 3 μm of $Ga_{.7}Al_{.3}As$. Note the development of the (110), (101), $(1\bar{1}0)$ and $(10\bar{1})$ facets during the growth over the peak of each of the raised square sections. Facets in the range of angles bracketed by the peak of the pyramid include all of the possible facet orientations on GaAs, taking into account the symmetries, and thus the (110), (101), $(1\bar{1}0)$ and $(10\bar{1})$ facets are the least preferred facets.

From figures 1 and 2 we can infer the growth rate of some of the low index facets normalized for the incident flux. The facets developed during the growth shown in figure 3 indicate that the growth rate is slowest for the (110), (101), $(1\bar{1}0)$ and $(10\bar{1})$ surfaces.

$$\left\{\begin{matrix}(110)\\(010)\\(1\bar{1}0)\\(10\bar{1})\end{matrix}\right\} < \left\{\begin{matrix}(111)\\(1\bar{1}\bar{1})\end{matrix}\right\} < \left\{\begin{matrix}(411)\\(4\bar{1}\bar{1})\end{matrix}\right\} < \{(100)\} < \left\{\begin{matrix}(811)\\(8\bar{1}\bar{1})\end{matrix}\right\} \tag{4.5}$$

and

$$\{(100)\} < \left\{\begin{matrix}(4\bar{1}1)\\(41\bar{1})\end{matrix}\right\} \tag{4.6}$$

To determine the temperature and As/Ga flux ratio dependence, a number of substrates were prepared with ridges in the [011] and $[0\bar{1}1]$ direction, and a series of growths were carried out at substrate temperatures ranging from 580° C to 700° C, and for flux ratios from 1.4 to 4.0. Figures 4.8-4.15 and 4.24-4.27 are scanning electron microscope photographs of growths which were stained with a pH 7.02 H_2O_2 solution to highlight the thin GaAlAs marker layers. Table 4.1 gives the growth rate ratios for the substrate growth temperatures 580° C, 620° C, 660° C, and 700° C, and they are plotted in figures 4.16, 4.18, 4.20, and 4.22. Table 4.2 gives the growth rate ratios for a substrate temperature of 660° C, and for As/Ga flux ratios of 1.4, 2.0, and 4.0, and the growth rate ratios are plotted in figures 4.17, 4.19, 4.21, and 4.23. Higher temperatures increase the differential growth rates, while a high As flux decreases them to some extent, but the relative ordering of the facets remains the same.

Since the long axis of the oval defects is oriented along the $[0\bar{1}1]$ direction, it is likely that this observed tendency of facets to propagate away from raised disturbances in the $[0\bar{1}1]$ direction and to propagate nearly vertically in the [011] direction is the cause of the shape of the oval defects, independent of the origin of the defects.

Diffusion lengths for group V elements have been reported to be quite small (< 200 Å) for low temperature growths ~ 550° C.[8] The diffusion lengths were estimated by growing an epilayer which was partially shadowed, and measuring the distance over which the grown edge tapers down. One difficulty with this method is, as shown in the experimental results shown here, that the properties of the tilted surfaces are quite different from those of the (100) surface, and the surface diffusion length measured in this manner may not be related to the diffusion length on the (100) facet. Additional complications arise because the arsenic flux is also shadowed, and the diffusion constants for the group three elements is possibly a function of the excess population of the group five elements. Variations in layer thickness next to edges on the (100) facets in this study for the first layers grown show diffusion lengths of $\sim$ .3 microns at 580° C, rising to somewhat in excess of 1 micron at 700° C for both the $[0\bar{1}\bar{1}]$ and $[0\bar{1}1]$ directions, and they do not appear to be a strong function of the As/Ga ratio over the range tested.

4.4 Macroscopic modeling

The notion of surface diffusion lengths relies on the following simple model. If the sticking coefficients of gallium and aluminum on the surface as a whole remain near unity, facets grow at rates different from the local incident flux by surface diffusion. Taking the case of the growth of GaAs, assuming an excess of arsenic at all points on the surface, we can divide the gallium atoms into two groups: atoms which are adsorbed on the surface and relatively free to move, and those which have been incorporated into the lattice and which are relatively immobile. Using mass balance for the mobile population,

$$\frac{\partial P}{\partial t} = F - L_i - L_d - (\mathrm{Diff}) \tag{4.7}$$

where P is the local population of adsorbed gallium atoms, F is the local flux incident on the surface, L_i is the losses due to incorporation into the lattice, L_d

is the losses due to desorption, and (Diff) is the net diffusion of atoms to other parts of the surface. Modeling desorption and lattice incorporation as simple lifetime processes, taking diffusion to be independent of facet angle, and taking the derivative with respect to time to be zero under steady state conditions,

$$0 = F(\vec{r}) - \frac{P(\vec{r},t)}{\tau_i(\vec{r})} - \frac{P(\vec{r},t)}{\tau_d(\vec{r})} + D\nabla^2 P(\vec{r},t) \tag{4.8}$$

where $\tau_i(\vec{r})$ is the characteristic time for incorporation of atoms into the lattice, $\tau_d(\vec{r})$ is the characteristic time for desorption of atoms from the surface, $P(\vec{r},t)$ is the mobile population, $F(\vec{r})$ is the flux at each point on the surface, D is the diffusion constant for atoms adsorbed on the surface, and the gradient is taken on the surface. If we reduce the problem to the case where there is only variation along one dimension, and introduce a variable x which is measured along the surface, we have

$$D\frac{\partial^2 P(x)}{\partial x^2} - P(x)\left(\frac{1}{\tau_d(x)} + \frac{1}{\tau_i(x)}\right) = -F(x) \tag{4.9}$$

and the growth rate of the lattice at position x is given by

$$R(x) = \frac{P(x)}{\tau_i(x)} \tag{4.10}$$

If the desorption rate is taken to be negligible, and we consider areas of the surface where the flux and incorporation times are relatively constant, the solutions are of the form

$$R = F_f + Ae^{x/l} + Be^{-x/l} \tag{4.11}$$

where A and B are constants which are determined by the boundary conditions at the edge of the facet, $F_f = F_0 \cos\theta$, where F_0 is the flux on the (100) facet, and $l = \left(\sqrt{D\tau_i}\right)^{-1}$ is the diffusion length for atoms on the facet.

One difficulty with this model is that nonuniform growth on the surface changes the facets on the surface. In addition, small tilts away from low index surfaces may change the growth rate substantially, because of the steps which are introduced on the surface. Despite these problems, this simple model makes it clear that a diagram of growth rate vs. facet angle cannot be constructed because of the large effect of the neighboring surface on the growth rate at a particular point of the surface. Unfortunately, the facet sizes of interest for device applications are comparable to the diffusion length, and neither of two simplifying approximations can be made, that surface diffusion is negligible and geometric flux ratios determine the growth rate, or that the surface is in equilibrium over a suitably large region and the growth rate over a facet can be taken as independent of its immediate neighborhood.

4.5 Microscopic structure

The vacuum environment of growth by MBE allows the use of high energy electron diffraction during the growth, which yields details about the growing interface which are not available for any other growth technique. Cho[9] discovered that an initially rough substrate surface, yielding a HEED pattern which is a field of dots due to transmission through asperities (bumps), quickly smooths out when growth is started, yielding a HEED pattern which is a set of parallel lines. He also noted the appearance of suborder diffraction lines which varied depending on the ratio of incident arsenic and gallium fluxes and the substrate temperature, with the easily reproduced arsenic stabilized surface yielding either a centered (2×8) or (2×4) structure (these two structures cannot be distinguished by first order HEED features). These features in the HEED patterns are caused by surface features which have periodicities of 2 times and 4 or 8 times that of the surface, respectively. He also observed several other surface structures for reduced arsenic to gallium ratios, including (3×1), (1×6), (2×4), and centered (2×8) for a gallium stabilized surface. Figure 4.28 is the "MBE phase diagram" which summarizes his findings. Cho later showed[10] that the two-fold periodicity corresponds to the direction of the dangling bonds of the uppermost layer, and suggested that it was caused by bond-pairing, alternate rows of atoms pulling together because of the interaction of their dangling bonds. The longer periodicity in the perpendicular direction is thought to be due to a long range order in the partially completed upper layer. Recent experiments [11] on the time dependence of HEED patterns upon initiation of growth suggest that growth on the (100) surface progresses through the development of domains formed at random positions, generating steps which propagate rapidly in the $[0\bar{1}1]$ direction and less rapidly in the $[0\bar{1}\bar{1}]$ direction. These results are consistent with the greatly enhanced growth for facets tilted slightly towards the $[0\bar{1}1]$ direction, with the maximum for the $(4\bar{1}1)$ plane. The maximum growth rate for the $(8\bar{1}\bar{1})$ plane, along with the minimum growth rate plane $(4\bar{1}\bar{1})$ is interestingly correlated with the (2×4) or $C(2\times8)$ HEED diffraction patterns seen for the growing (100) surface. In seems likely that the same atomic configurations account for both effects.

The simplest $(8\bar{1}\bar{1})$ surface terminated on an arsenic surface consists of a series of steps two atoms high and four atoms wide, as shown in figure 4.29. The growth rate for this facet being a local maximum as a function of tilt angle in the $[0\bar{1}]$ direction implies that interactions at a distance of at least four lattice constants are important.

Because of the complexity of the many body problem of the growing interface, modeling has consisted of taking energies for small configurations of a few atoms either determined analytically or as parameters, and using them in large Monte Carlo simulations of the growth kinetics.[12] The results presented here suggest that longer range effects are important, and probably will need to be taken into account.

4.6 Conclusion

The quality of the material grown over the facets tilted in the $[0\bar{1}1]$ direction is comparable to that of material grown over the (100) facet, as can be seen by the uniform electroluminescence seen in figure 4.30, and by the laser array work which will be discussed in a later chapter. This technique should prove useful in the fabrication of devices which take advantage of unique features formed during regrowth by MBE. The enhanced growth rate over the $[4\bar{1}1]$ surface has been used to form waveguide geometries for lasers and laser arrays, as discussed later in this thesis. Theoretical treatment of growth mechanisms is still weak for most surfaces because of the complexity of the many body problem, but these experimental results should prove useful in testing methods such as the Monte-Carlo simulations recently presented.[12]

References for chapter 4:

[2] A. Y. Cho and W. C. Ballamy, *J. Appl. Phys.* **46**, 783, (1975).

[3] W. T. Tsang and A. Y. Cho, *J. Appl. Phys.*, **30**, 293 (1977).

[4] T. Sanada, S. Yamakoshi, O. Wada, T. Fujii, T. Sakurai, M. Sasaki, *Appl. Phys. Lett.* **44**, 325, (1984).

[5] A. Y. Cho and F. K. Reinhart, *Appl. Phys. Lett.* **21**, 355, (1972).

[6] S. Nagata, T. Tanaka, and M. Fukai, *Appl. Phys. Lett.* **30**, 505 (1977).

[7] Y. H. Wu, M. Werner, K. L. Chen, and S. Wang, *Appl. Phys. Lett.* **44**, 834, (1984).

[8] Seiichi Nagata and Tsuneo Tanaka, J. Appl. Phys., Vol. 48, No. 3, (1977).

[9] A. Y. Cho, J. Appl. Phys., 42, 2074, (1971).

[10] A. Y. Cho, J. Appl. Phys., 47, 2841, (1976).

[11] J. H. Neave, B. A. Joyce, P. J. Dobson, and N. Norton, Appl. Phys. A 31, 1-8 (1983).

[12] S. V. Ghaisas and A. Madhukar, J. Vac. Sci. Technol. B, 3, 2, (1985).

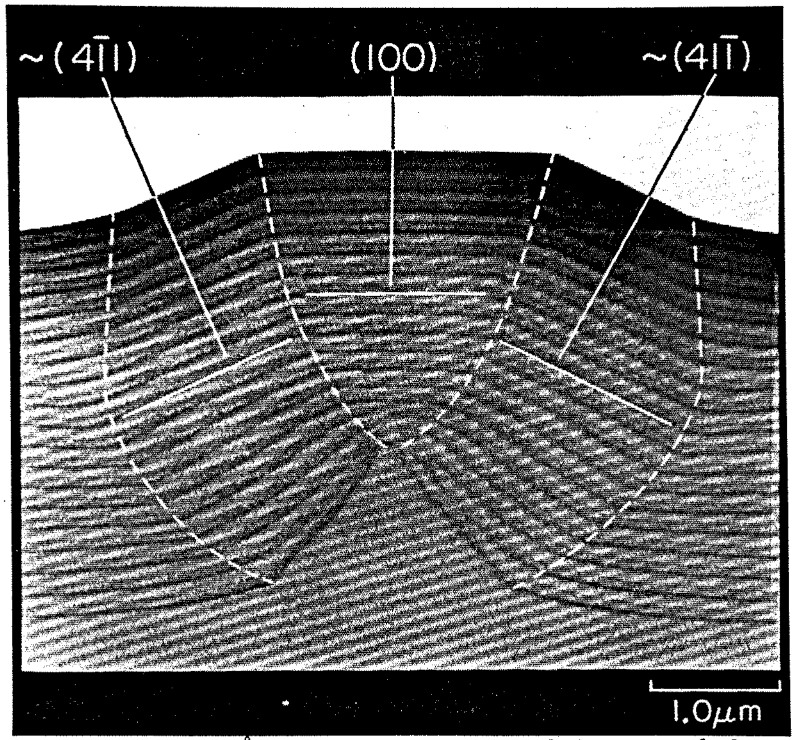


Figure 4.1) SEM of 1000Å alternating layers of GaAs and Ga$_{.3}$Al$_{.7}$As grown over a (100) substrate ridged in the (a) [0$\bar{1}\bar{1}$] direction at a substrate temperature of 680°C and a As/Ga flux ratio of 2, with facet indices marked.

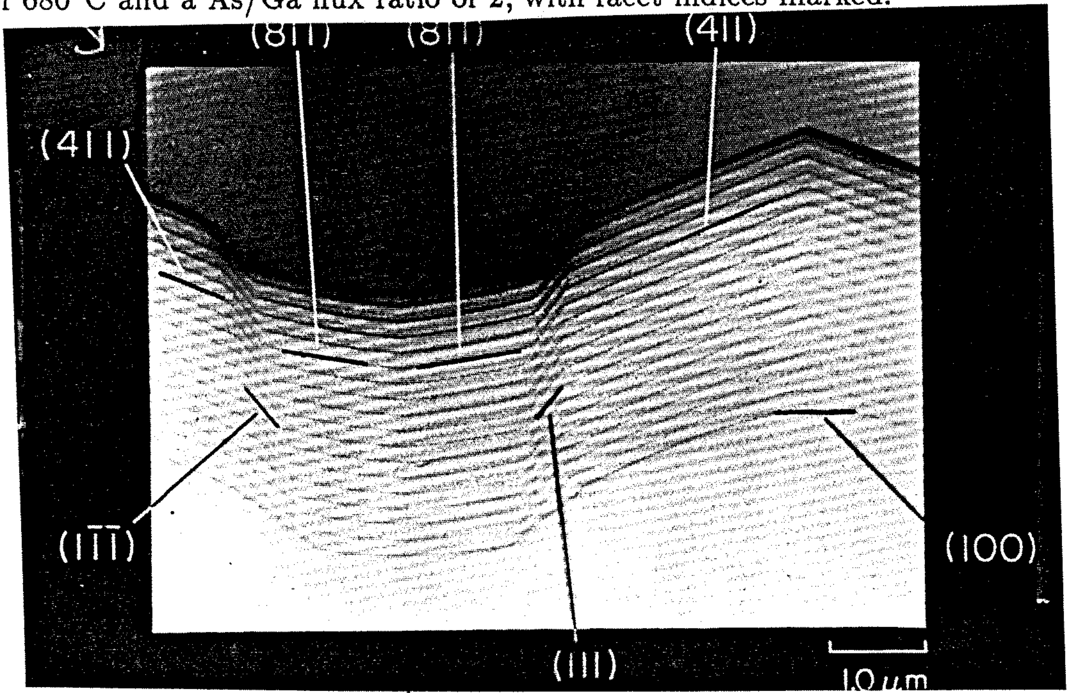


Figure 4.2) SEM of 1000Å alternating layers of GaAs and Ga$_{.3}$Al$_{.7}$As grown over a (100) substrate ridged in the (a) [0$\bar{1}$1] direction at a substrate temperature of 680°C and a As/Ga flux ratio of 2, with facet indices marked.

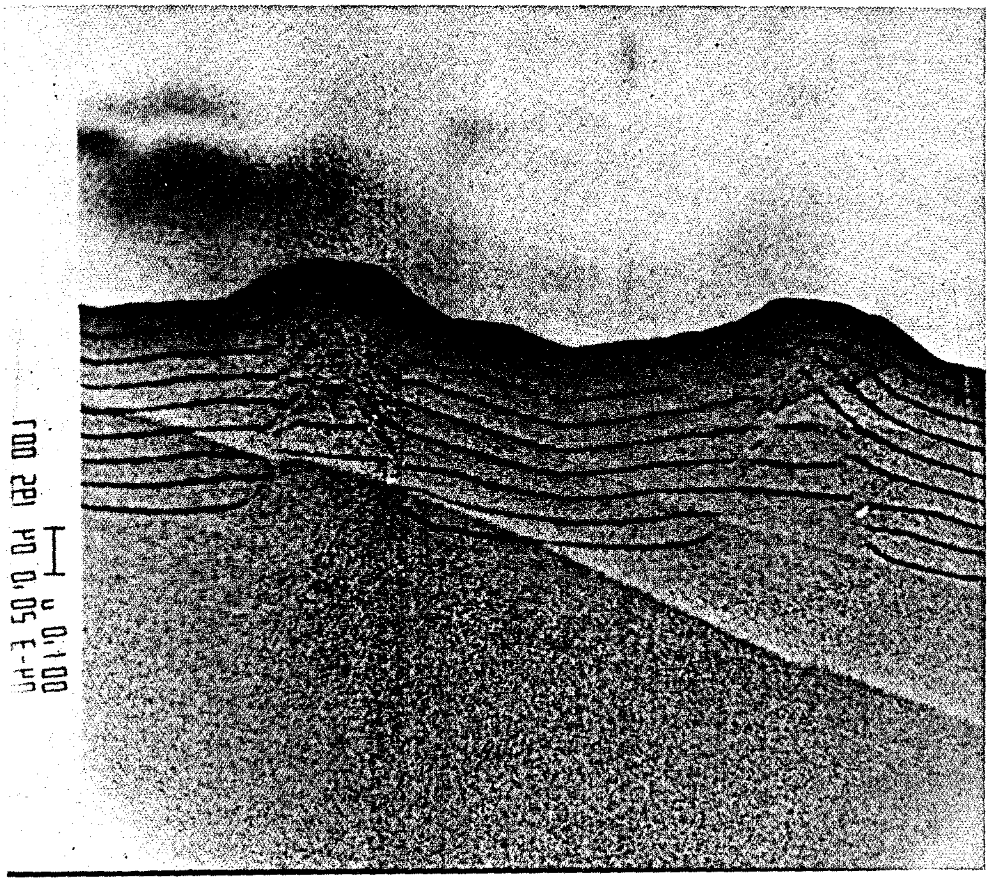


Figure 4.3) a) and b) poor growth morphology resulting from insufficient etching of the overhang on a $[0\bar{1}1]$ ridge.

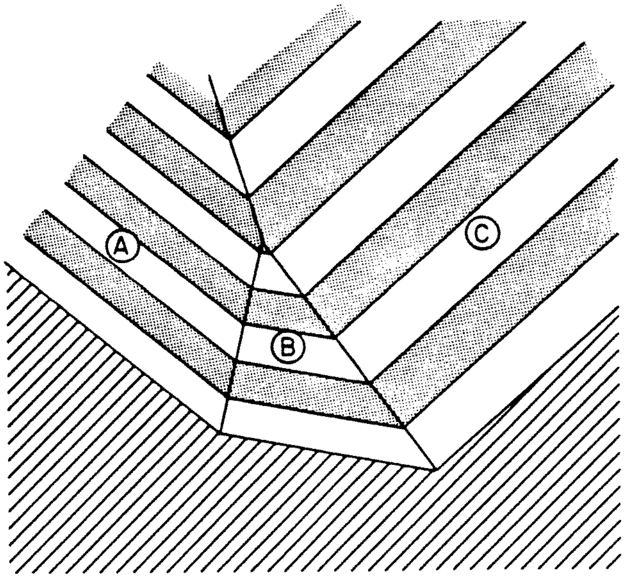


Figure 4.4) diagram illustrating disappearance of slow growing facets on concave surfaces.

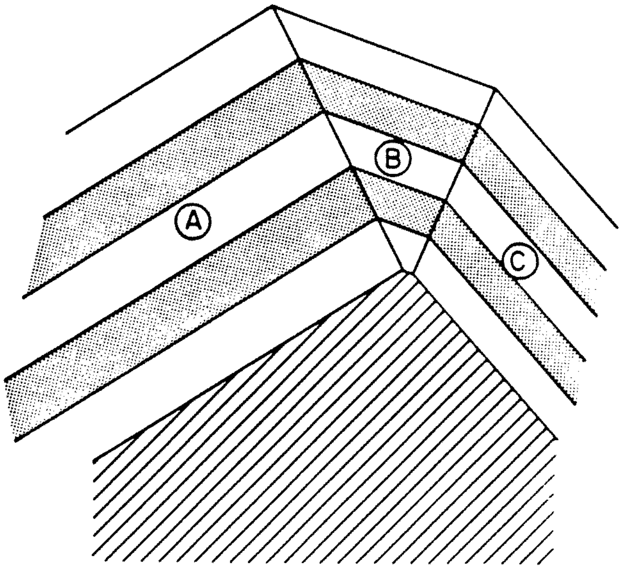


Figure 4.5) diagram illustrating expansion of slow growing facets on convex surfaces.

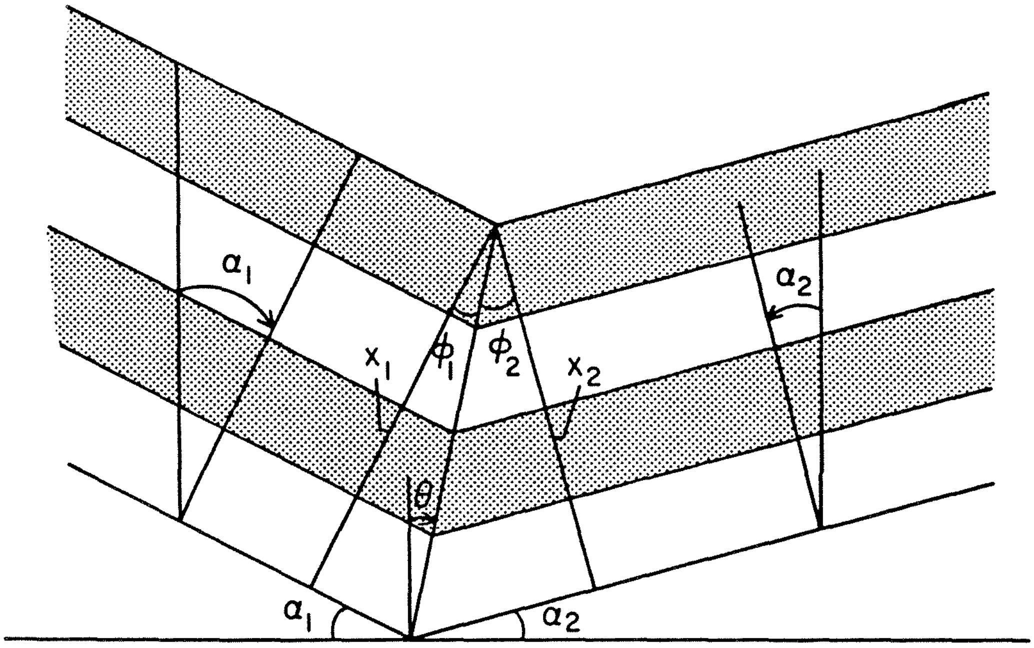


Figure 4.6) Diagram relating growth rates of two adjacent facets to the propagation of their vertex.

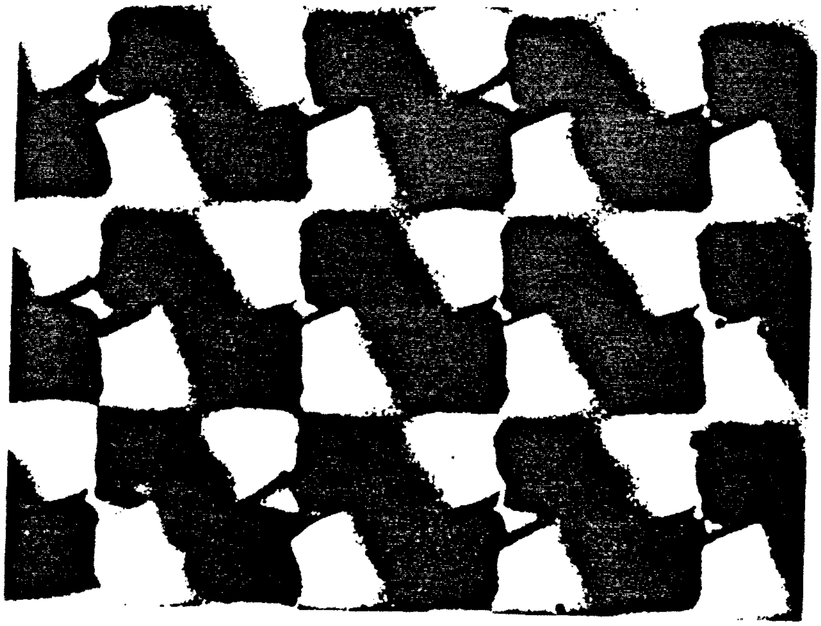


Figure 4.7) Growth of 3 μm of $Ga_{.8}Al_{.2}As$ over a (100) substrate etched with a pattern of 9 μm squares showing development of the (110), (101), $(1\bar{1}0)$ and $(10\bar{1})$ facets.

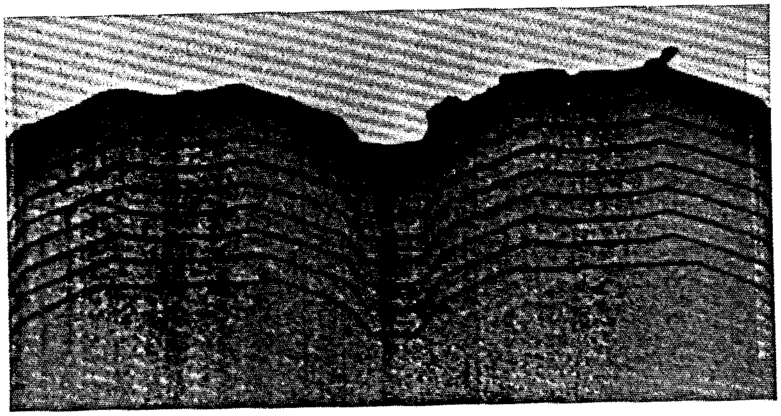


Figure 4.8) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}1]$ direction. The substrate temperature was 580° C, and the As/Ga flux ratio was 2.

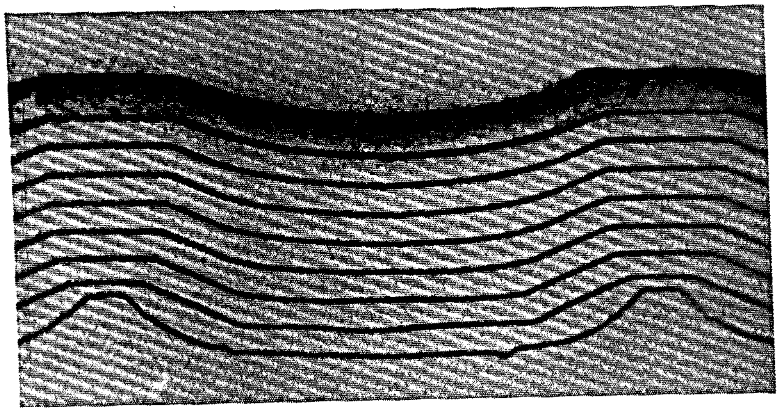


Figure 4.9) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}\bar{1}]$ direction. The substrate temperature was 580° C, and the As/Ga flux ratio was 2.

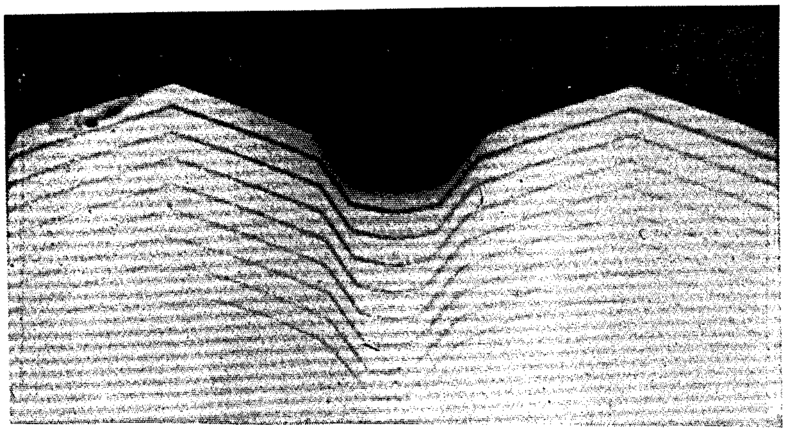


Figure 4.10) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}1]$ direction. The substrate temperature was 620° C, and the As/Ga flux ratio was 2.

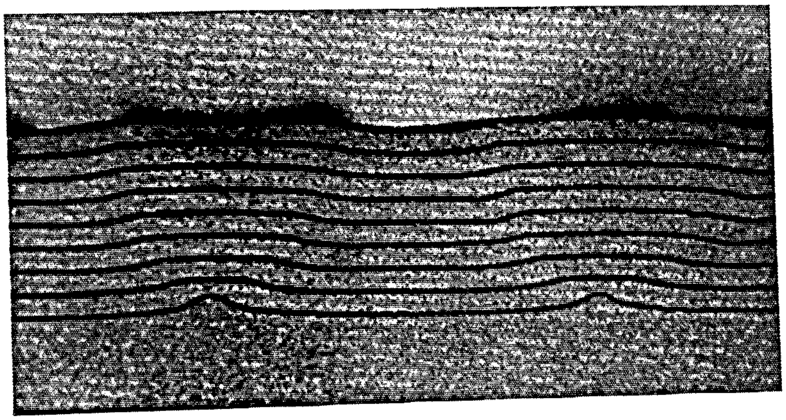


Figure 4.11) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}\bar{1}]$ direction. The substrate temperature was 620° C, and the As/Ga flux ratio was 2.

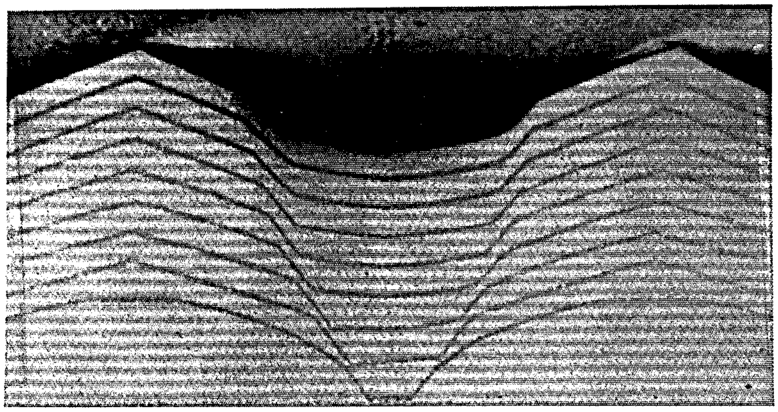


Figure 4.12) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}1]$ direction. The substrate temperature was 660° C, and the As/Ga flux ratio was 2.

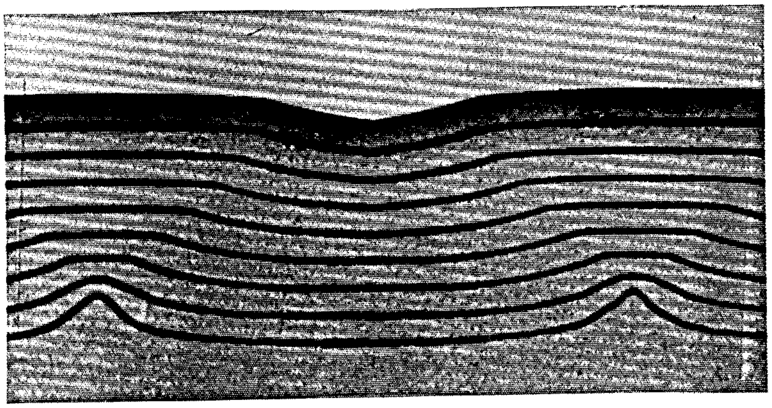


Figure 4.13) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}\bar{1}]$ direction. The substrate temperature was 660° C, and the As/Ga flux ratio was 2.

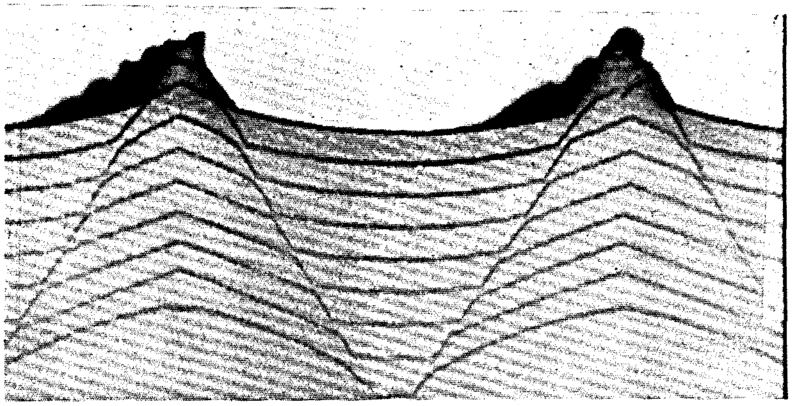


Figure 4.14) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}1]$ direction. The substrate temperature was 700° C, and the As/Ga flux ratio was 2.

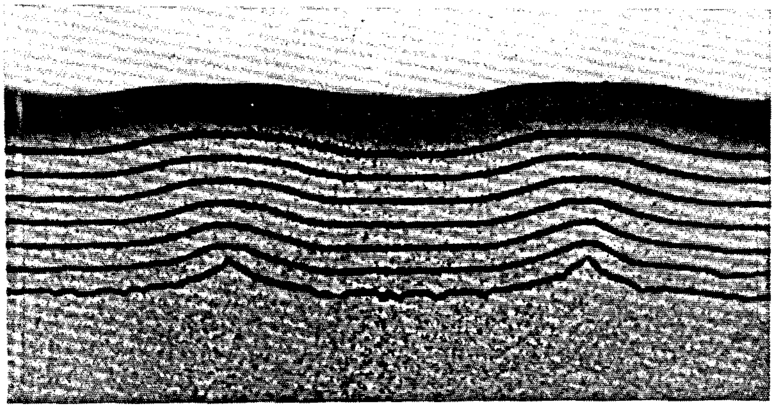


Figure 4.15) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}\bar{1}]$ direction. The substrate temperature was 700° C, and the As/Ga flux ratio was 2.

	$4\bar{1}1$	$4\bar{1}\bar{1}$	$1\bar{1}\bar{1}$	$8\bar{1}\bar{1}$
Temp.	vs. 100	vs. 100	vs. $4\bar{1}\bar{1}$	vs. $1\bar{1}\bar{1}$
580	1.01 ± 0.01	0.90 ± 0.03	0.59 ± 0.02	1.79 ± 0.03
	1.07 ± 0.01	0.97 ± 0.03	0.95 ± 0.02	1.07 ± 0.03
620	1.20 ± 0.08	0.78 ± 0.02	0.55 ± 0.02	2.0 ± 0.1
	1.27 ± 0.08	0.83 ± 0.02	0.89 ± 0.04	1.2 ± 0.1
660	1.24 ± 0.04	$< .5$	0.41 ± 0.06	2.9 ± 0.5
	1.31 ± 0.04	$< .5$	0.67 ± 0.10	1.7 ± 0.3
700	1.11 ± 0.05	$< .5$	0.21 ± 0.04	5.3 ± 0.8
	1.18 ± 0.05	$< .5$	0.35 ± 0.05	3.3 ± 0.5

Table 4.1) Table of relative growth rates for various facets formed by MBE growth as a function of temperature, at a As/Ga flux ratio of 2.0. The range indicated allows for errors in measurements of the photographs. The bottom pair of figures in each case are normalized for the cosine θ dependence of the flux.

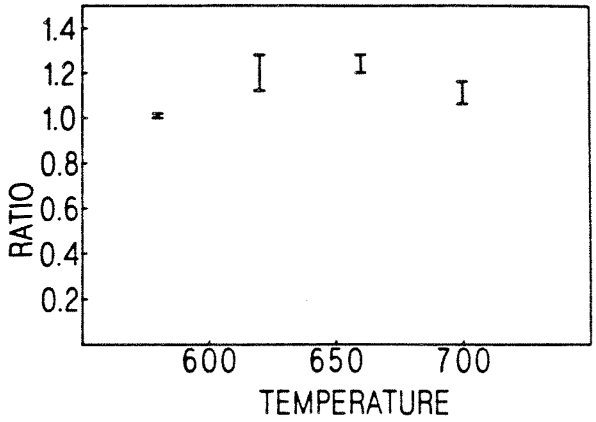


Figure 4.16 growth rate for the $(4\bar{1}1)$ facet vs. (100) facet as a function of temperature.

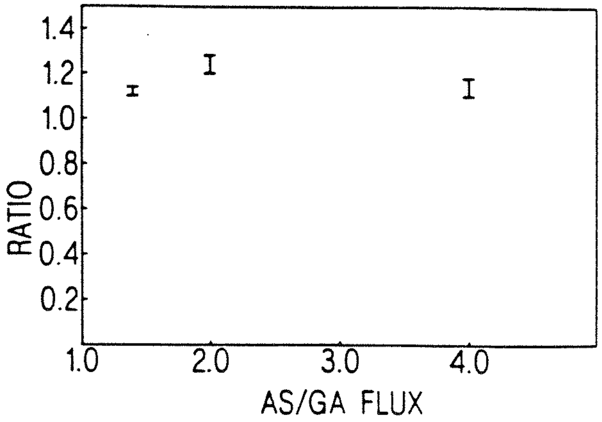


Figure 4.17 growth rate for the $(4\bar{1}1)$ facet vs. (100) facet as a function of As/Ga flux ratio.

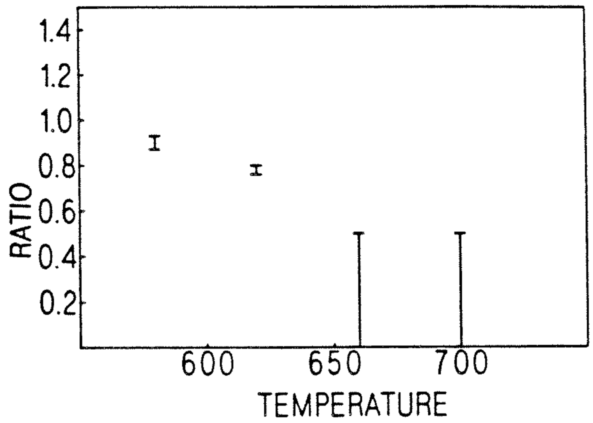


Figure 4.18 growth rate for the $(4\bar{1}\bar{1})$ facet vs. (100) facet as a function of temperature.

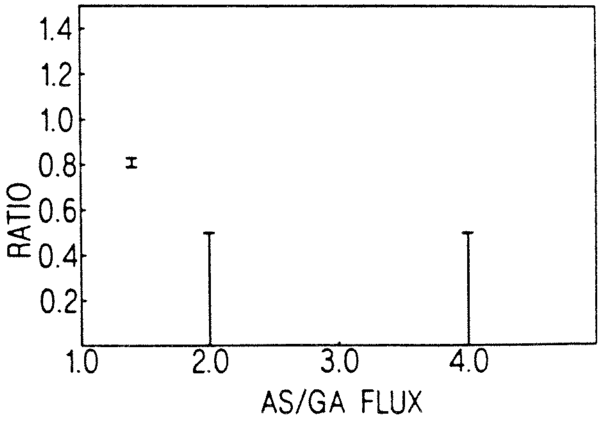


Figure 4.19 growth rate for the $(4\bar{1}\bar{1})$ facet vs. (100) facet as a function of As/Ga flux ratio.

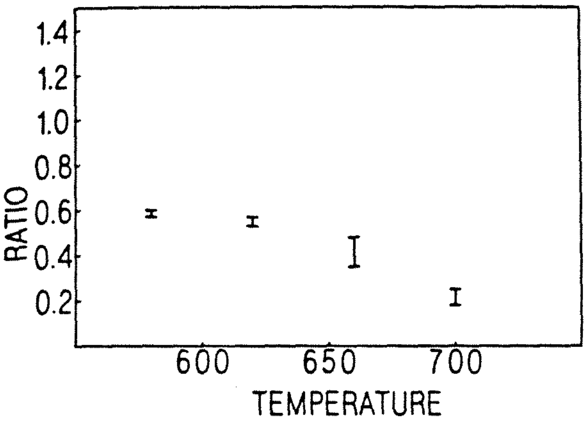


Figure 4.22 growth rate for the $(1\bar{1}\bar{1})$ facet vs. $(4\bar{1}\bar{1})$ facet as a function of temperature.

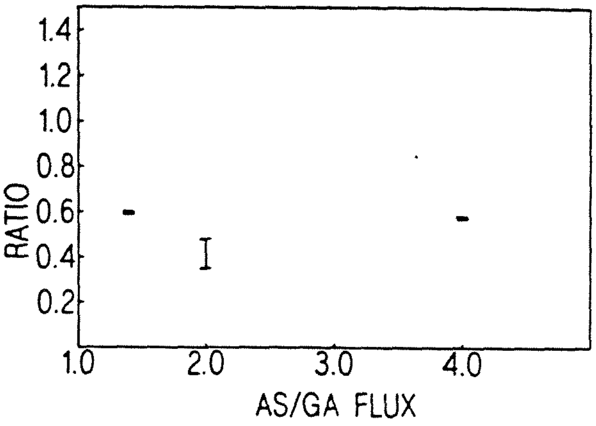


Figure 4.23 growth rate for the $(1\bar{1}\bar{1})$ facet vs. $(4\bar{1}\bar{1})$ facet as a function of As/Ga flux ratio.

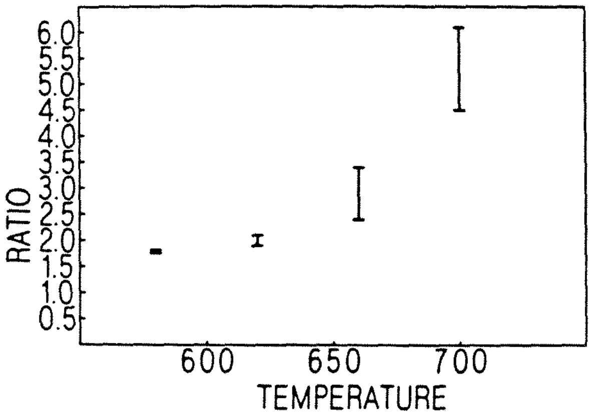


Figure 4.20 growth rate for the $(8\bar{1}\bar{1})$ facet vs. $(1\bar{1}\bar{1})$ facet as a function of temperature.

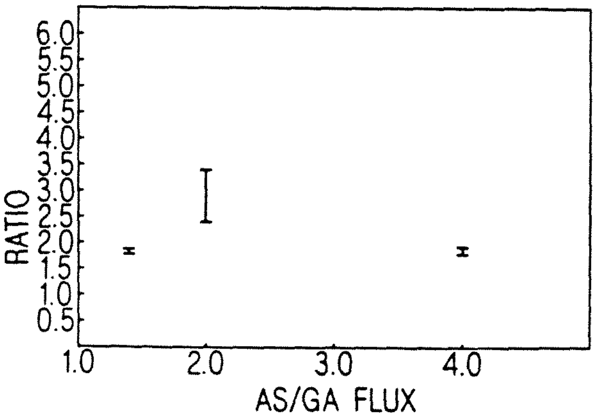


Figure 4.21 growth rate for the $(8\bar{1}\bar{1})$ facet vs. $(1\bar{1}\bar{1})$ facet as a function of As/Ga flux ratio.

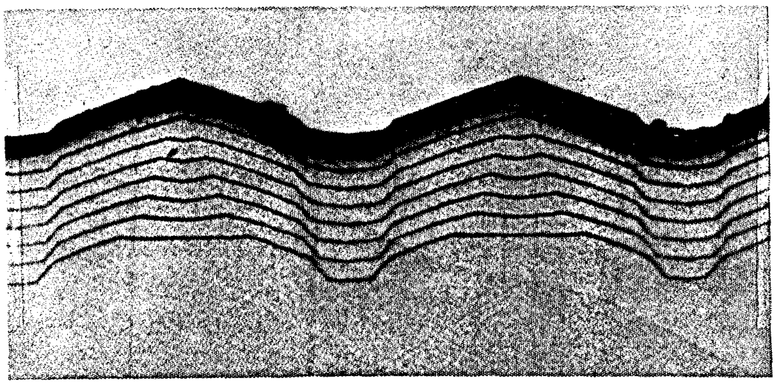


Figure 4.24) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}1]$ direction. The substrate temperature was 660° C, and the As/Ga flux ratio was 1.4.

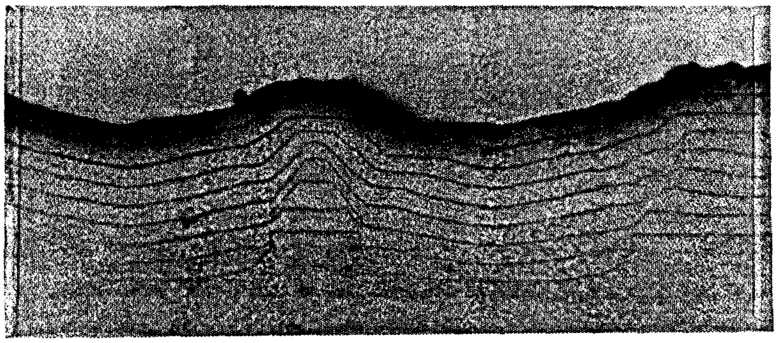


Figure 4.25) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}\bar{1}]$ direction. The substrate temperature was 660° C, and the As/Ga flux ratio was 1.4.

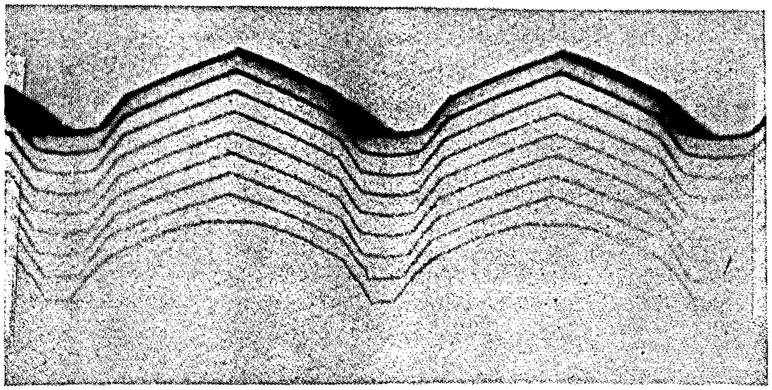


Figure 4.26) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}1]$ direction. The substrate temperature was 660° C, and the As/Ga flux ratio was 4.

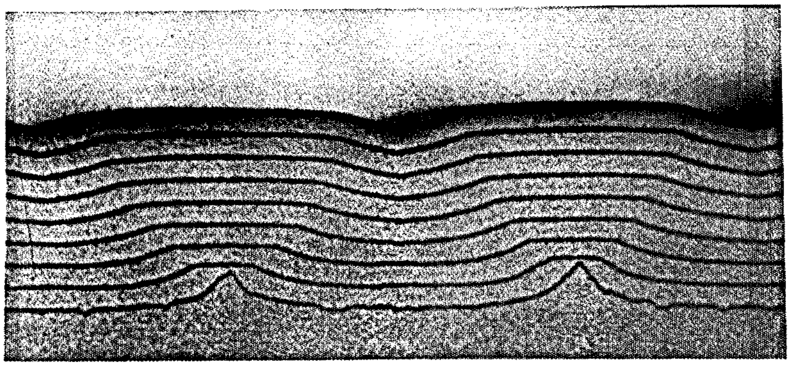


Figure 4.27) alternating 4500Å GaAs layers and 500Å marker layers of $Ga_{.8}Al_{.2}$ As grown over substrates ridged with a nine micron period in the $[0\bar{1}\bar{1}]$ direction. The substrate temperature was 660° C, and the As/Ga flux ratio was 4.

As/Ga ratio	$4\bar{1}1$ vs. 100	$4\bar{1}\bar{1}$ vs. 100	$1\bar{1}\bar{1}$ vs. $4\bar{1}\bar{1}$	$8\bar{1}\bar{1}$ vs. $1\bar{1}\bar{1}$
1.4	1.12 ± 0.02	0.81 ± 0.02	0.59 ± 0.01	1.83 ± 0.05
	1.18 ± 0.02	0.86 ± 0.02	0.95 ± 0.01	1.10 ± 0.03
2	1.24 ± 0.04	$< .5$	0.41 ± 0.06	2.9 ± 0.5
	1.31 ± 0.04	$< .5$	0.67 ± 0.10	1.7 ± 0.3
4	1.14 ± 0.04	$< .5$	0.57 ± 0.01	1.86 ± 0.07
	1.21 ± 0.04	$< .5$	0.92 ± 0.01	1.11 ± 0.04

Table 4.2) Table of relative growth rates for various facets formed by MBE growth for various As/Ga flux ratios, at a substrate temperature of 660° C. The range indicated allows for errors in measurements of the photographs. The bottom pair of figures in each case are normalized for the cosine θ dependence of the flux.

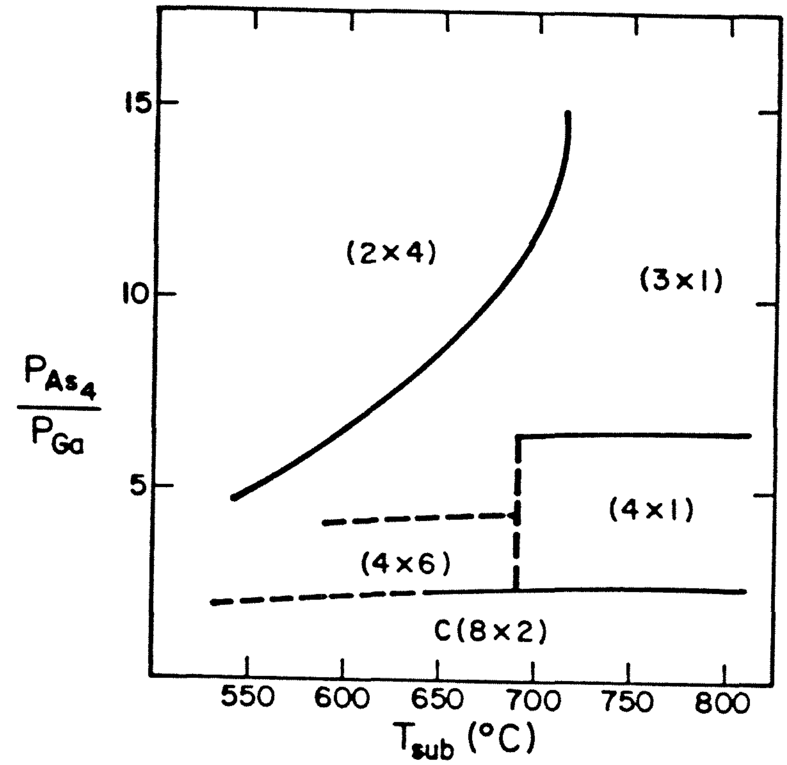


Figure 4.28) "MBE phase diagram" showing approximate zones of substrate temperature and pressure for the various surface reconstructions.

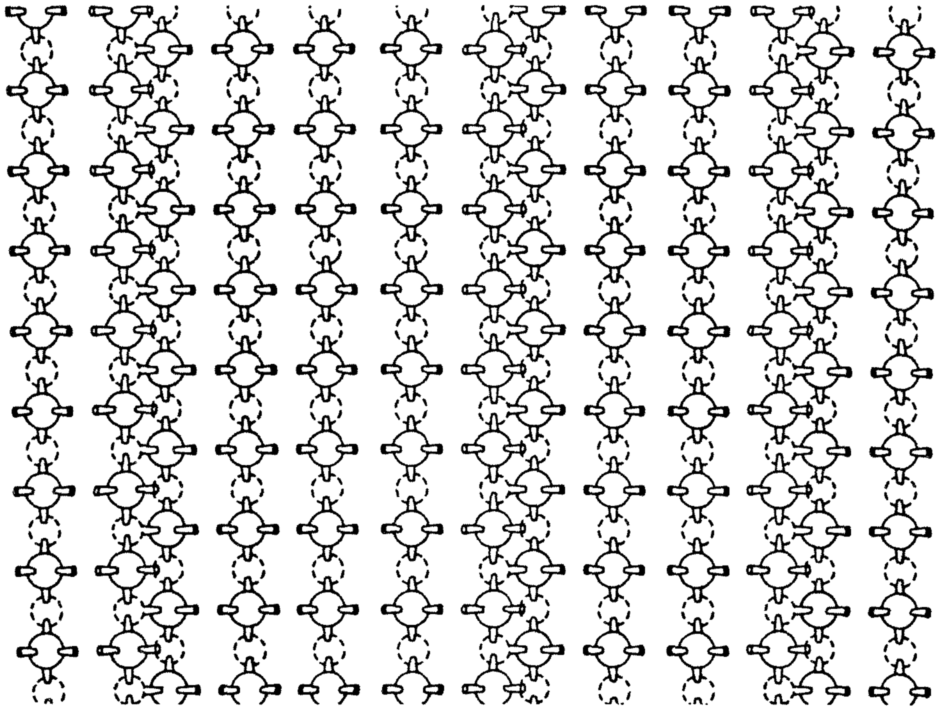


Figure 4.29) Diagram of simplest arsenic terminated $(8\bar{1}\bar{1})$ facet.

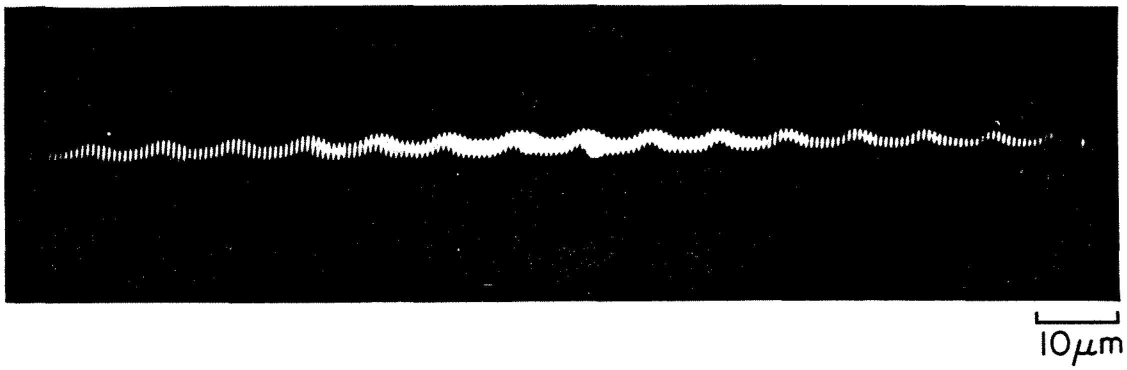


Figure 4.30) Infrared photograph of electroluminescence from a double heterostructure grown over a substrate ridged in the [$0\bar{1}\bar{1}$] direction.

Appendix C

Angle dependence of incident fluxes for a rotating substrate MBE

In an MBE chamber, there are as many as eight separate cells supplying molecular beams to the growing surface, which are arrayed in a cluster pointing at the substrate. The range of angles of the incoming beams creates nonuniform fluxes across the wafer which are partially compensated by rotating the substrate about its own axis. In the case of growths over patterned substrates, the rotation also eliminates the asymmetry due to some facets being tilted away from and others tilted towards the molecular beam.

The flux on a facet of the wafer from a given cell is proportional to the cosine of the angle between the normal of the facet and the direction of incidence of the flux. In order to find the average flux on a facet while the wafer is rotating, the cosine of this angle is averaged over one rotation. If we take a unit vector normal to the facet to be $\vec{F}$, and the unit vector pointing in the direction of a given Knudsen cell to be $\vec{K}$,

$$\cos\psi = \vec{F}\cdot\vec{K} \tag{C.1}$$

where ψ is the angle between the normal to the facet and the direction of the Knudsen cell. Defining a cartesian coordinate system by taking the z axis to be the normal of the wafer and the y axis to be along the direction of the projection of $\vec{F}$ in the plane of the wafer, we have

$$\vec{F} = (\sin\theta, 0, \cos\theta) \tag{C.2}$$

and

$$\vec{K} = (\sin\alpha\cos\phi, \sin\alpha\sin\phi, \cos\alpha) \tag{C.3}$$

where θ is the angle between $\vec{F}$ and the normal of the wafer, α is the angle between $\vec{K}$ and the normal of the wafer, and ϕ is the angle between them in the plane of the wafer. Taking the dot product, we obtain

$$\cos\psi = \sin\theta\sin\alpha\cos\phi + \cos\theta\cos\alpha \tag{C.4}$$

Where $\cos\psi$ is positive, it is proportional to the flux per unit area on the facet, provided the facet is not shadowed by an overhanging part of the surface. When ψ is greater than 90°, $\cos\psi$ is negative, the facet is entirely shadowed, and the flux is zero. If ψ is not greater than 90° during any part of the rotation, and there is no overhanging facet, the average flux on a given facet can be calculated by integrating over ϕ.

$$\begin{aligned}\overline{\text{Flux}} &\propto \frac{1}{2\pi}\int_0^{2\pi}(\sin\theta\sin\alpha\cos\phi + \cos\theta\cos\alpha)\,d\phi \\ &\propto \cos\theta\cos\alpha\end{aligned} \tag{C.5}$$

Thus for a rotating substrate, the flux on a facet at angle θ is proportional to $\cos\theta$ for all of the Knudsen cells, provided the angles of the facets and of the effusion cells are not large enough to cause shadowing during some part of the rotation. The result is that the fluxes are the same as they would be if all the materials were coming in normal to the wafer. Note that in the case of ψ being greater than 90° during some part of the rotation, the average flux is larger than this value. This can easily be seen in the case of a $\theta = 90°$ facet which has a nonzero flux for an off-axis oven because it is tilted toward the oven during part of the rotation.

If $\theta + \alpha > 90°$, then the flux is zero during part of the rotation, and the average flux is given by:

$$\begin{aligned}\overline{\text{Flux}} &\propto \frac{1}{2\pi}\int_0^{2\pi}\left(\max\left(0, \sin\theta\sin\alpha\cos\phi + \cos\theta\cos\alpha\right)\right) d\phi \\ &\propto \frac{1}{2\pi}\int_0^{\pi-\phi_0}\left(\sin\theta\sin\alpha\cos\phi + \cos\theta\cos\alpha\right) d\phi \\ &\quad + \frac{1}{2\pi}\int_{\pi+\phi_0}^{2\pi}\left(\sin\theta\sin\alpha\cos\phi + \cos\theta\cos\alpha\right) d\phi\end{aligned} \tag{C.6}$$

where $\phi_0 = \cos^{-1}(\cot\theta\cot\alpha)$, $\pi - \phi_0$ and $\pi + \phi_0$ being the roots of the integrand. Integrating, we have

$$\begin{aligned}\overline{\text{Flux}} &\propto \frac{1}{2\pi}\left(\sin\theta\sin\alpha\sin\phi + \phi\cos\theta\cos\alpha\right)\Big|_0^{\pi-\phi_0} \\ &\quad + \frac{1}{2\pi}\left(\sin\theta\sin\alpha\sin\phi + \phi\cos\theta\cos\alpha\right)\Big|_{\pi+\phi_0}^{2\pi} \\ &\propto \frac{1}{2\pi}\left(2\sin\theta\sin\alpha\sin\phi_0 + (2\pi - 2\phi_0)\cos\theta\cos\alpha\right)\end{aligned} \tag{C.7}$$

Substituting for ϕ_0, we have

$$\begin{aligned}\overline{\text{Flux}} \propto \frac{1}{2\pi}\Big\{&2\sin\theta\sin\alpha\sqrt{1 - (\cot\theta\cot\alpha)^2} \\ &+ \left(2\pi - 2\cos^{-1}(\cot\theta\cot\alpha)\right)\cos\theta\cos\alpha\Big\}\end{aligned} \tag{C.8}$$

Figure C.1 shows the flux for facets at various angles for effusion cells at angles from 0° to 90°. Figure C.2 shows the ratio between the flux on the facet at the given angle and that on the 0° facet. MBE effusion cells are placed at a range of angles from 10° to 25° and the fluxes on the facets discussed in chapter 3 are all independent of the cell angles within this range, except for the vertical [010] and [001] facets.

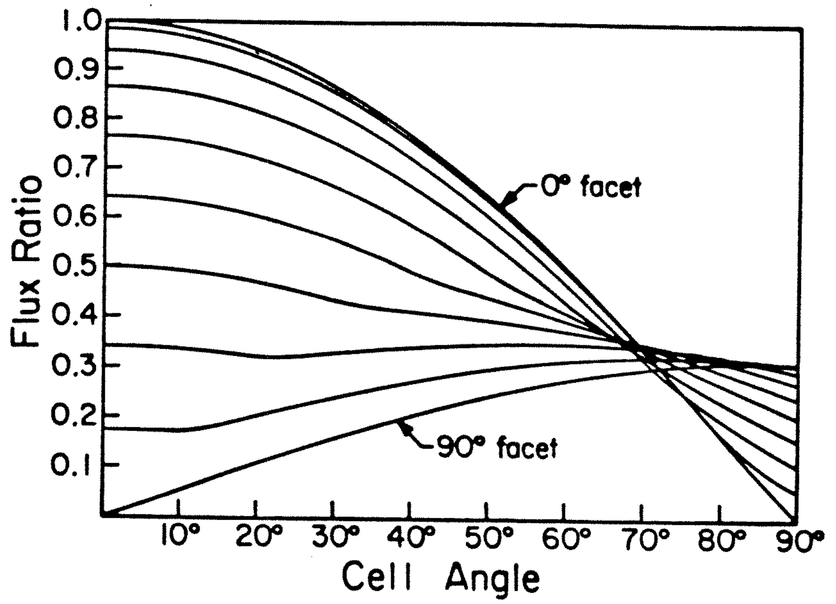


Figure C.1) Plot of molecular flux incident on various facets as a function of effusion cell angle.

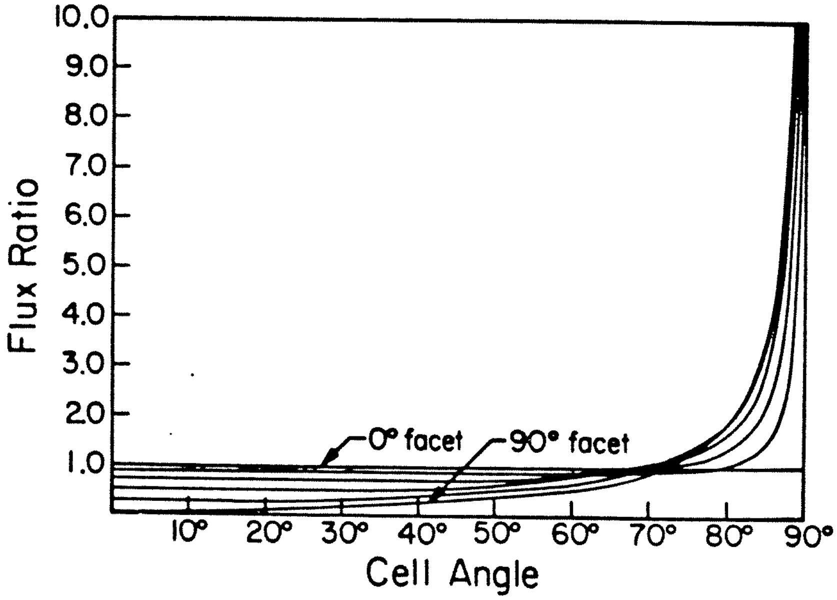


Figure C.1) Plot of molecular flux incident on various facets as a function of effusion cell angle, normalized for the flux on the 0° facet.

Chapter V

The 411 laser structure

5.1 Introduction

Semiconductor laser diodes are compact, highly efficient, and are pumped with an electrical current. They can also be directly modulated through their pump current. These characteristics of laser diodes make them the most desirable laser devices for use in many applications, such as communications and information storage. Research on semiconductor laser diodes has been directed in several main directions including lowering the current threshold for lasing action, improving the lifetime of the laser diodes, increasing the modulation response, and increasing the total output power. Threshold current densities have been reduced from in excess of $50,000 \mathrm{A/cm}^2$ twenty years ago to as low as $160 \mathrm{A/cm}^2$.[1] Devices with expected lifetimes in excess of 10^6 hours have been fabricated,[2] and modulation bandwidths have been extended as high as 20 gigahertz.[3] One thing that becomes apparent upon a review of the laser structures that advance beyond the experimental stage is that they are all very simple structures in comparison with many of the exotic designs which have been proposed in the literature. This is presumably because complex structures are difficult to fabricate, and yields are low.

Optoelectronic devices are intrinsically more complex than most electronic devices because they require a three dimensional structure, to form a waveguide to control propagation of light, as well for control of electrical pump currents. The first semiconductor lasers were homojunction lasers, simple p-n junctions diffused into a GaAs wafers. It was only later realized that they depended on a fortuitous waveguide mechanism due to the high doping away from the compensated region around the junction.[4] Kroemer[5] proposed in 1963 that semiconductor lasers could be made by sandwiching a low bandgap material with a higher bandgap material which would have improved characteristics because of the waveguide formed by the lower index of the higher bandgap material. The higher bandgap material also provides a confining structure for the injected carriers, keeping them from diffusing away from the optical channel. This structure, known as a double heterostructure because of the two interfaces, is now common to all low threshold semiconductor lasers.

In addition to the double heterostructure geometry, which allowed threshold current densities to come down to a few kiloamps per square centimeter by providing waveguiding and current confinement in the vertical direction, to achieve low total thresholds it is also necessary to limit the extent of the cavity in the horizontal direction. Lateral confinement can be divided into two types, index-guiding and "gain-guiding." In a gain guided structure, pumping is restricted to a narrow stripe by creation of high resistivity material using ion implantation or by a dielectric layer, but there is no confining structure provided for the light, and light which diffracts from the cavity is simply absorbed in the adjacent unpumped region. These lasers are simple and inexpensive to make, but thresholds are higher than those of index-guided lasers, and they operate in multiple longitudinal modes and exhibit "kinks" in their light-current curve due to shifts in their mode distribution.[6]

The lateral optical guiding of index-guided lasers is provided by one of several mechanisms. Buried lasers[7] use a two growth process which provides a slab of high index material completely surrounded by a lower index cladding layer, forming the optical guide, as shown in figure 5.1. A second class of lasers is that of the channeled substrate planer (CSP)[8] which uses an LPE growth over a channeled substrate as shown in figure 5.2. The section of the guide layer over the channel grows thicker, forming an effective index guide in the lateral direction. A third method of providing a lateral index guide structure is that of the transverse junction stripe laser (TJS).[9] In the TJS laser, a double heterostructure is grown which is heavily doped n type, and an acceptor is diffused in to provide a transverse junction. The lateral guiding mechanism is the same as the vertical guiding mechanism of a homostructure laser.

The first step in fabrication of buried heterostructure lasers such as that shown in figure 5.1, is growth of a double heterostructure, in this case GaAlAs cladding layers surrounding a (GaAs) active layer. This double heterostructure is then etched away except for a narrow stripe (1-2 microns wide), the resulting structure placed in an LPE furnace, and GaAlAs is grown surrounding the original ridge. This technique relies on several properties of LPE: its relative insensitivity to surface contamination, and the fact that growth will not nucleate over GaAlAs with aluminum content over 10 percent, but it will grow up against a ridge once it has started on a GaAs surface. This results in a structure which is near the ideal, with a waveguide consisting of a high dielectric constant material entirely embedded in a lower index material, and good current confinement due to a blocking layer included in the regrowth step. The main drawback to this geometry is the requirement for a complex LPE regrowth step, reducing the yield of the process and making these lasers comparatively expensive.

LPE growth rates are strongly dependent on the surface shape, grooves are filled as the growth proceeds, and projections or ridges may even be etched away by the meltback process, both effects tending to smooth the surface. Fabrication of CSP lasers use this property of LPE growths to form a waveguide by starting the growth on a substrate which has had shallow grooves etched in it. A double heterostructure is then grown over the wafer, and the part of the double heterostructure which grows over the grooves grows faster than that on the surrounding area, as shown in figure 5.2. This thicker part of the active layer forms the waveguide.

The composition control capability of MBE has lead to the development of laser structures in which the simple double heterostructure is replaced with a more complex set of layers. In these structures, which include the large optical cavity laser, quantum well lasers, and graded index separate confinement heterostructure (GRIN-SCH) lasers, separate layers are used for confinement of the carriers and the optical mode. With this separation, the optical cavity and the active layers can be individually optimized, the optical cavity for large mode volume and high power handling capability, and the active layers for low threshold. Electrons confined to layers less than $\sim$ 150 Å in extent show quantum size effects, one of which is an increase in the optical interaction coefficient, resulting in the gain being higher for the same inversion density. Threshold current densities as low as 160 A/cm^{-2} have been achieved with the GRIN-SCH laser.[1] One drawback to these very low threshold density MBE-grown lasers is that the lateral confinement must be provided by either the simple but undesirable gain guided stripe geometry, or by complex additional processing into buried lasers with an additional growth process in an LPE furnace. The zinc diffusion process used to form a TJS structure enhances diffusion of the constituent elements, and destroys the quantum well structure.[10]

5.2 The 411 laser: structure

From tables 4.1 and 4.2, it can be seen that the $(4\bar{1}1)$ facet grows faster than the adjacent (100) surfaces under a wide range of conditions, even when unnormalized for the lower flux density incident on it. The higher growth rate on this facet allows the growth of structures in a single MBE process which are analogous to the LPE-grown CSP laser, one of the commercially most important index guided laser structures. This laser structure, referred to as the 411 structure because of its dependence on the enhanced growth rate over the $(4\bar{1}1)$ facet, makes possible the fabrication of index guided MBE lasers, with their potentially very low thresholds, in a very straightforward process involving only one growth step.

A scanning electron microscope photograph of a cross section of the laser is shown in figure 5.3, and a schematic diagram is shown in figure 5.4. This structure was fabricated by the following steps. A series of stripes was photolithographically defined in the $[1\bar{1}1]$ direction on a [100] GaAs n^+ wafer. The wafer was then etched in 1:8:8 (H_2SO_4, H_2O_2, H_2O) for 20 seconds, leaving a series of ridges on the surface. Next, the wafer was cleaned by the technique discussed in chapter 2. The polishing step and the etch in 4:1:1 (H_2SO_4, H_2O_2, H_2O) remove the overhang which is left by the masked 1:8:8 etch. The wafer was then mounted and loaded into the MBE system for the growth. The growth consisted of a 0.1 micron GaAs n^+ buffer layer, a 1.3 micron $Ga_{.7}Al_{.3}As$ n^+ lower cladding layer, a 1500Å GaAs active layer, a 1.3 micron $Ga_{.7}Al_{.3}As$ p^+ upper cladding layer, and a .1 micron p^+ GaAs contact layer. A 2000 Å layer of SiO_2 was deposited by CVD, and stripe openings were etched through it over the waveguide formed in the MBE growth. A Cr/Au contact was applied to the upper surface, and a AuGe contact was alloyed to the substrate. The wafer was then cleaved into bars 350 micron long.

Typical threshold current levels for this simple structure are approximately 50 milliamps per 350 μm cavity, and they typically lase in a single longitudinal mode. Several things can be done to improve this basic structure, including growth of GRIN-SCH layers instead of the simple double heterostructure used here, and use of a diffused zinc contact to provide improved current confinement, similar to current confinement structure used in the CSP structure.

5.3 Proposed variation

As discussed in appendix B, if the sum of the facet angle and the angle to an effusion cell is greater than 90°, then the flux on the facet is no longer proportional to the cosine of its tilt angle. For sufficiently large effusion cell angles, the incident flux rate on tilted facets can be higher than the flux on the flat part of the wafer. The ratio of the flux on the $(4\bar{1}1)$ facet to the (100) surface is shown in figure 5.5 for a range of effusion cell angles. It is apparent that with an MBE with special cells placed at high angles, it would be possible to preferentially deposit materials on the $(4\bar{1}1)$ facet or any other facet angled with respect to the $[100]$ surface.

Figure 5.6 shows the calculated structure for an example of a growth using a high angle gallium cell and the following procedure: With an extra gallium cell at an grazing angle of 5°, the wafer is loaded, with a ridged surface as described in the previous section. The growth proceeds through a conventional startup and the lower cladding of $Ga_{.7}Al_{.3}As$ with the standard cells. At the start of the active layer, rather than the aluminum cell being shuttered off, the high angle gallium cell is operated at a flux calibrated to 0.8 microns per hour for a flat substrate. This results in an aluminum mole fraction of 0.1 in the flux on the region over the $[4\bar{1}1]$ facet, while that on the $[100]$ surfaces only drops to 0.19. After the active layer has reached the desired thickness, the high angle gallium cell is shuttered off, and the growth continues as before. This sequence provides an active layer which is completely embedded in a higher bandgap material, similar to a buried heterostructure laser, but in a single step. If a similar cell were used for a dopant, it would be possible to embed a conductive channel which follows the propagation of the $[4\bar{1}1]$ facet, providing current confinement as well.

5.4 Conclusion

Lasers using MBE growth over ridged substrates have the potential of combining the advantages of the MBE process, excellent control over the growing layers, with the structural advantages of the CSP laser grown by LPE. Ease of fabrication, low thresholds and good yields even at this early stage in their development, along with its desirable single longitudinal mode oscillation, may make this laser structure into more than just a laboratory curiosity.

References for chapter 5

[1] W. T. Tsang, "Extremely low threshold (AlGa)As graded-index waveguide separate-confinement heterostructure lasers grown by molecular beam epitaxy," Appl. Phys. Lett., vol. 40, no. 3, pp 217-219, (1982).

[2] W. T. Tsang, R. L. Hartman, H. E. Elder, and W. R. Holbrook, Appl. Phys. Lett. 37, 141, (1980).

[3] J. E. Bowers, B. R. Hemenway, A. H. Gnauck, T. J. Bridges, E. G. Burkhardt, D. P. Wilt and S. Maynard, Appl. Phys. Lett., Vol 47, No. 2, pp. 79, (1985).

[4] W. L. Bond, B. G. Cohen, R. C. C. Leite, and A. Yariv, "Observation of the dielectric waveguide mode of light propagation in p-n junctions," Appl. Phys. Lett., vol. 2, no. 3, pp 57-59, (1963).

[5] H. Kroemer, proc. IEEE 51, 1782, (1963).

[6] G. H. B. Thompson, Physics of Semiconductor Laser Devices, John Wiley & Sons, New York, pp. 374, (1980).

[7] T. Tsukada, J. Appl. Phys. 45, 4899, (1974).

[8] K. Aiki, M. Nakamura, T. Kuroda, and J. Umeda, Appl. Phys. Lett. 30, 649 (1977).

[9] H. Namizaki, H. Kan, M. Ishii, and A. Ito, J. Appl. Phys., 45, 2785, (1974).

[10] T. Fukuzawa, S. Semura, T. Ohtay, Y. Uchida, T. Narusawa, K. L. I. Kobayashi, and H. Nakashima, Appl. Phys. Lett. 45, 1, (1984).

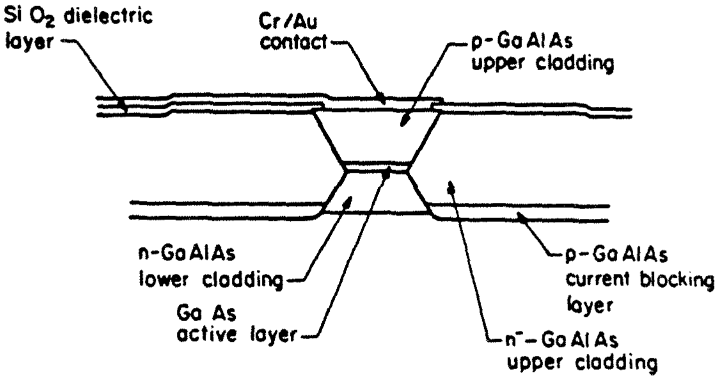


Figure 5.1) Diagram of a buried laser.

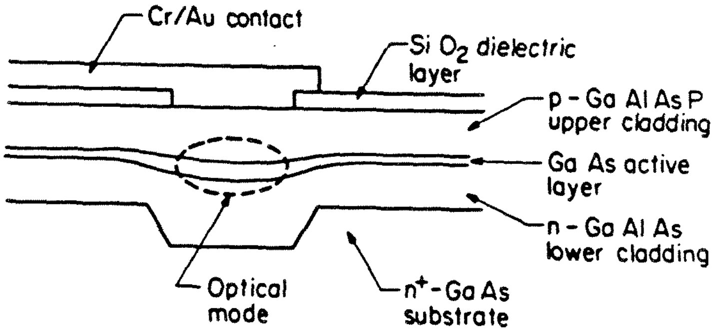


Figure 5.2) Diagram of the channeled substrate planner (CSP) laser structure.

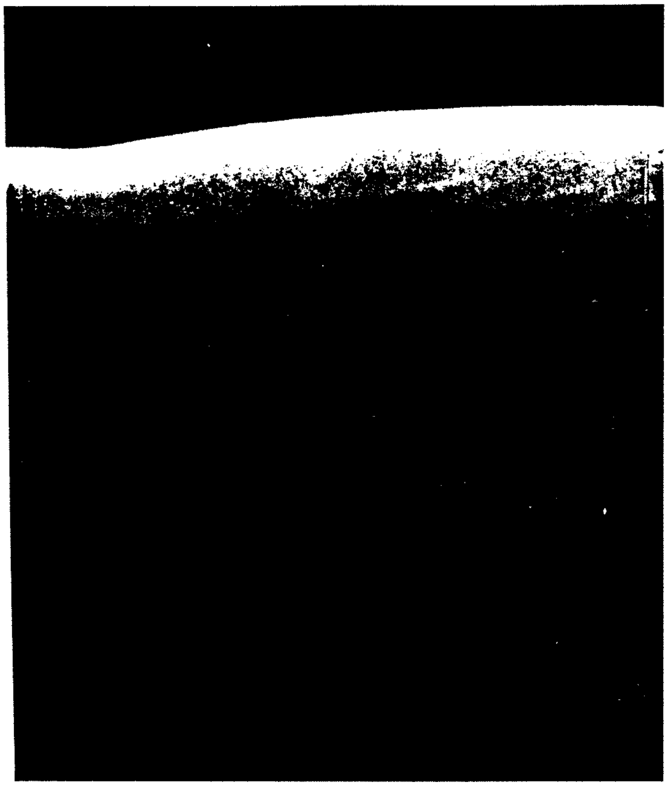


Figure 5.3) Scanning electron micrograph of a cross section of the 411 laser.

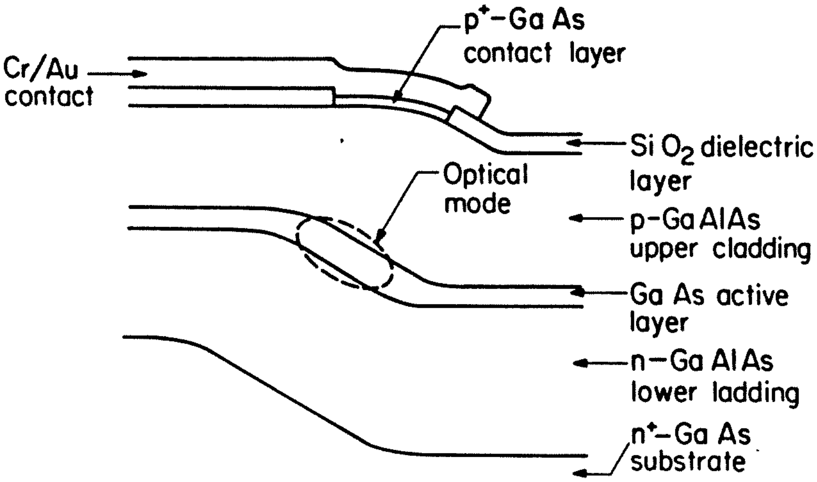


Figure 5.4) Diagram of the 411 laser structure.

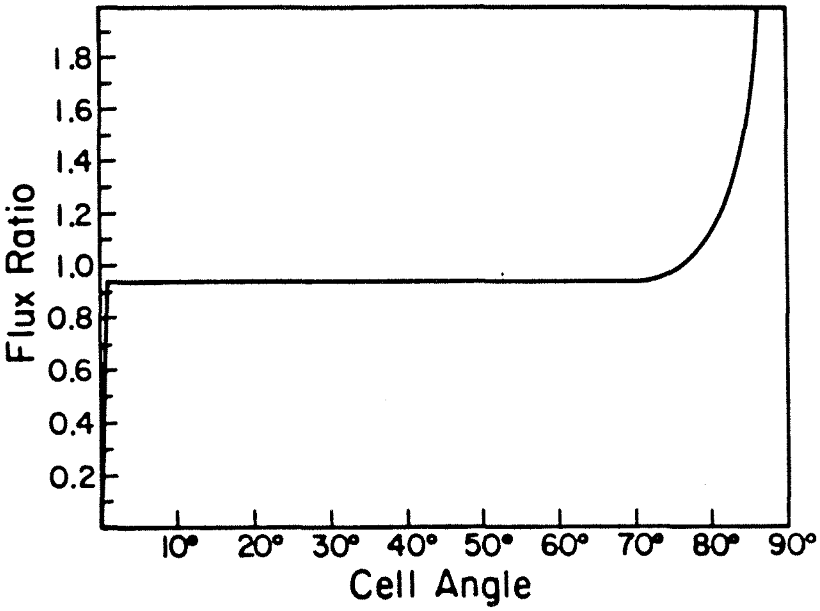


Figure 5.5) Ratio of the flux on the $(4\bar{1}1)$ facet to that on the (100) facet as a function of effusion cell angle.

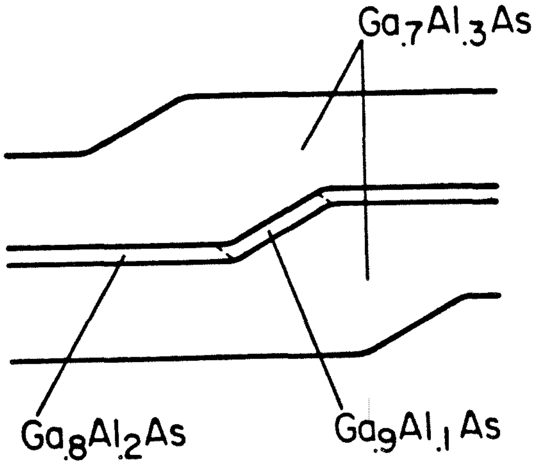


Figure 5.6) Diagram of the proposed enhancement of the 411 laser geometry.

Chapter VI

Index guided arrays of 411 lasers

6.1 Introduction

Efforts to raise the total output power of semiconductor laser diodes have been less successful than the impressive advances in improving the other characteristics such as threshold and reliability. The highest power levels for coherent single mode laser diode devices is still less than 500 milliwatts. It is useful to compare semiconductor lasers with other laser systems which have been successfully scaled to high output powers. A typical neodymium YAG laser has a threshold power density of $\sim$ 1watt/cm^3. A semiconductor laser, by comparison, requires a power density of 7×10^7watts/cm^3 to bring the material to transparency. This difference comes from the fact that the YAG laser is a four level laser and its lower laser level is effectively empty when unpumped, whereas the population of the semiconductor laser starts in the lower level and must be pumped with enough power to keep the population in the higher energy state. Since this threshold power is lost to spontaneous emission, to achieve high efficiency levels it is necessary that the pump power levels be at least twice this value.

With this level of pumping, the bulk semiconductor material can easily produce gains of 10-100 cm^{-1}. That semiconductor lasers can operate in a continuous wave mode at all is attributable to the efficiency of the direct current pumping and this extremely high gain which, along with the wave guide geometry, allows very small laser cavities to oscillate.

The strong interaction between the media and the light in the cavity causes large variations not only in the gain or loss of the material, but in the real index as well, through the Kramers-Kronig relationship,[1] and through a free-carrier-plasma interaction.[2] Where the local inversion is decreased towards transparency, the real index is lowered. This can cause a feedback effect, where a channel forms in which the inversion is depleted, creating a real index guide for light propagating down the channel, which in turn increases the depletion due to stimulated emission. Laser cavities in which this effect is not controlled break down into many independently oscillating filaments, and the coherence of the output is lost. The main methods of controlling filamentation in semiconductor lasers are to provide a built-in real index guide which is smaller than the intrinsic size of filaments, or to limit the pumped cavity to a stripe which is not wide enough to allow filaments to form.[3]

The high power density necessary to bring the semiconductor material to transparency, together with the self-focusing phenomenon which causes filamentation, is the major obstacle to fabrication of large, high power semiconductor lasers. The high power density requires the laser structure to be thin and in close proximity to a heat sink for cooling, and filamentation limits the transverse extent of the cavity. The total power from a solid state laser is then limited by the product of the small cross section and the power density at which material damage occurs. The material damage power level is on the order of 10 megawatts per square centimeter,[4] and may be reduced by an order of magnitude or more where non-radiative recombination is enhanced, such as at the facets of a GaAs laser. These limits give an upper bound of approximately 0.5 watt for conventional geometry coherent semiconductor laser emitters.

One approach to achieve substantially higher powers is to coherently combine the power of several laser diodes.[5,6] This requires that the laser diodes be phase locked. In order to avoid many optical elements with critical alignments, it is also necessary that the laser elements be tightly packed together so that they form an emitter array.[7] The elements of the array must also have 0° phase shift between them, and be separated by no more than the width of the individual elements for the output to be a single beam.[8] The elements of the array may be phase locked either by the coupling of the optical field within the array structure itself, or through feedback from an external mirror or a separate coupling structure integrated on the wafer. The simplest way to form laser diode arrays is to confine the current to multiple adjacent stripes, with intervening unpumped regions, and to place them closely enough so that they are coupled within the array structure. The current confinement can be provided by oxide stripes, proton implantation, or any of several other methods. However, the intrinsic feature of arrays of this type, the lossy zone between the elements of the array, causes the optical mode which has a null between each of the elements to have a higher gain than the mode which has no nulls. The mode with nulls between the laser elments has a 180° phase shift between the elements of the array and emits a double-lobed far field.

Arrays of index guided lasers do not have this built-in tendency to lase with a double lobed far field if losses between the guides can be avoided, but because of the tight optical confinement of index guided lasers, they must be placed very close together in order to achieve coupling within the array structure. The fabrication procedure for most types of index guided structures does not allow them to be fabricated as close together as necessary for coupling. In the case of buried lasers, the etches used to define the mesas in the first growth undercut the mask by approximately the same amount as the depth of the etch. This leads to a separation of at least 5 microns between adjacent lasers, which is too far for them to couple and phase lock. Similar problems occur for CSP lasers, where the initial channel is etched eight microns or so wide in order to end up with with a guide four or five microns wide, and the TJS structure, where the isotropic diffusion requires a separation of at least twice the thickness of the upper cladding layer.

6.2 Arrays of 411 lasers

The laser structure described in the previous chapter lends itself to index guided structures because of the nearly vertical propagation of the features formed on the substrate into the epitaxial layers. This makes it possible to make index guided structures which are very closely spaced, allowing the possibility of fabrication of tightly coupled arrays.

Using this technique, arrays of double heterostructure lasers were grown over a substrate with ridges etched in the $[1\bar{1}1]$ direction with a 9 micron period. The growth consisted of a .1 μm GaAs n^+ buffer layer, a 1.3 μm $Ga_{.7}Al_{.3}As$ n^+ lower cladding layer, a 1500Å GaAs active layer, a 1.3 μm $Ga_{.7}Al_{.3}As$ p^+ upper cladding layer, and a .1 μm p^+ GaAs contact layer. A 150 μm wide Cr/Au contact was applied, and the wafer was cleaved into bars 350 μm long. The laser cross section is shown in figure 6.1a, and the spontaneous emission nearfield is shown in figure 6.1b. The 1.5× threshold nearfield is shown in figure 6.1c with 15 adjacent pairs of devices of the single contact array lasing, each pair formed on either side of a ridge. The typical threshold current for these arrays was 50-60 milliamps per element. The elements of the arrays are index guided due to the greater thickness of the active layer over the $(4\bar{1}1)$ facets as shown in figure 3. In the 9 micron ridged array, individual pairs of lasers were phase locked and typically operated in a single longitudinal mode, but the individual pairs oscillated independently. Yields of individual lasers exceeded 90 percent, and thresholds were uniform to 10 percent. A typical threshold curve for a 150 micron wide device is shown in figure 6.2.

With the objective of increasing the coupling and producing arrays which are fully phase locked, similar structures were grown on substrates with periodicities of 5 microns and 4 microns. The cross sections of these laser arrays are shown in figures 6.4 and 6.5. As can be seen in the cross sections, the growth ridges start to overlap at these spacings, and the laser geometry changes. The spectrally resolved nearfield of the 4 micron arrays, shown in figure 6.6, shows that they are mode locked across the array, but the far field pattern is double lobed, showing that they are phase locking in the 180° out of phase mode. In addition, the array oscillates in several longitudinal modes as seen in the spectrally resolved nearfield, in contrast to the predominantly single longitudinal mode behavior of the 9 micron period array.

6.3 Conclusion

The present results show that is is possible to produce tightly packed arrays of index guided lasers using the 411 structure, but the objective of a single mode, single lobed farfield laser array has not been achieved. It is possible that additional measures to force the oscillation of the fundamental optical mode of the array, such as pumping the sections in between the array elements, will meet with success. External optical feedback from a mirror would probably result in fundamental mode oscillation, as it has in other marginal cases, but such apparatus defeats some of the main advantages of the laser diode array, namely its compact size and integrated construction.

references for chapter 6

[1] G. H. B. Thompson, Opto-electronics, 4, 257, (1972).

[2] A. K. Jonscher and M. H. Boyle, Proceedings of IPPS Symposium on GaAs, Reading, pp.78, (1966).

[3] H. C. Casey, Jr., and M. B. Panish, Heterostructure lasers, Academic Press, New York, pp 207, (1978).

[4] T. Kamejima and H. Yonezu, "Catastrophic Optical Damage Generation Mechanism in (AlGa)As DH Lasers", Japan. J. Appl. Phys. supplement 19-1, pp 425-429 (1979).

[5] D. R. Scifres, W. Streifer, and R. D. Burnham, Appl. Phys. Lett. 34(4), 259, (1979).

[6] D. E. Ackley, Appl. Phys. Lett. 42(2), 152, (1983).

[7] J. K. Butler, D. E. Ackley, and D. Botez, Appl. Phys. Lett. 44(3), 293, (1984).

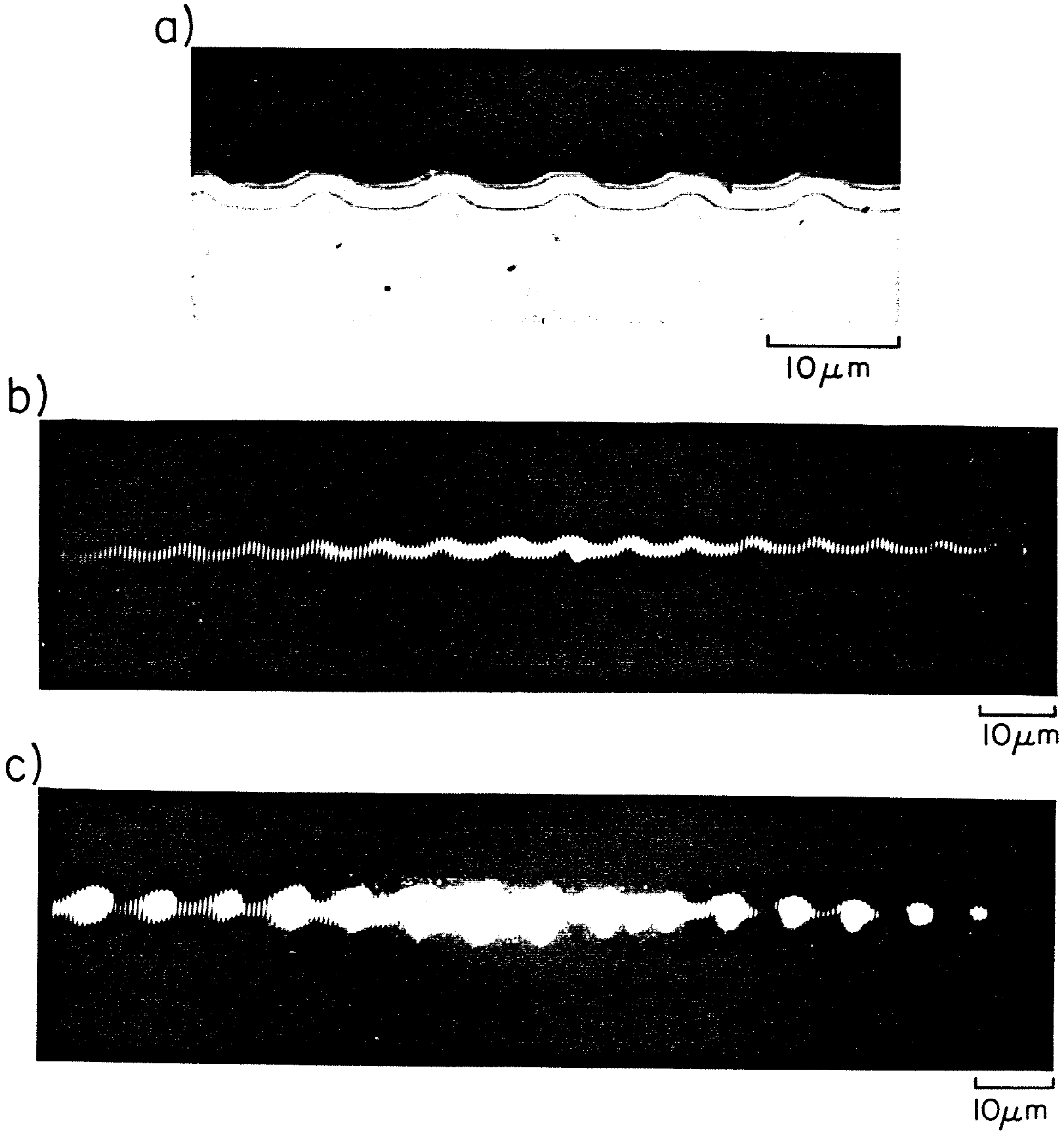


Figure 6.1) (a) Cross section of laser array. (b) Nearfield of laser array below threshold. (c) Nearfield of laser array at 1.5× threshold.

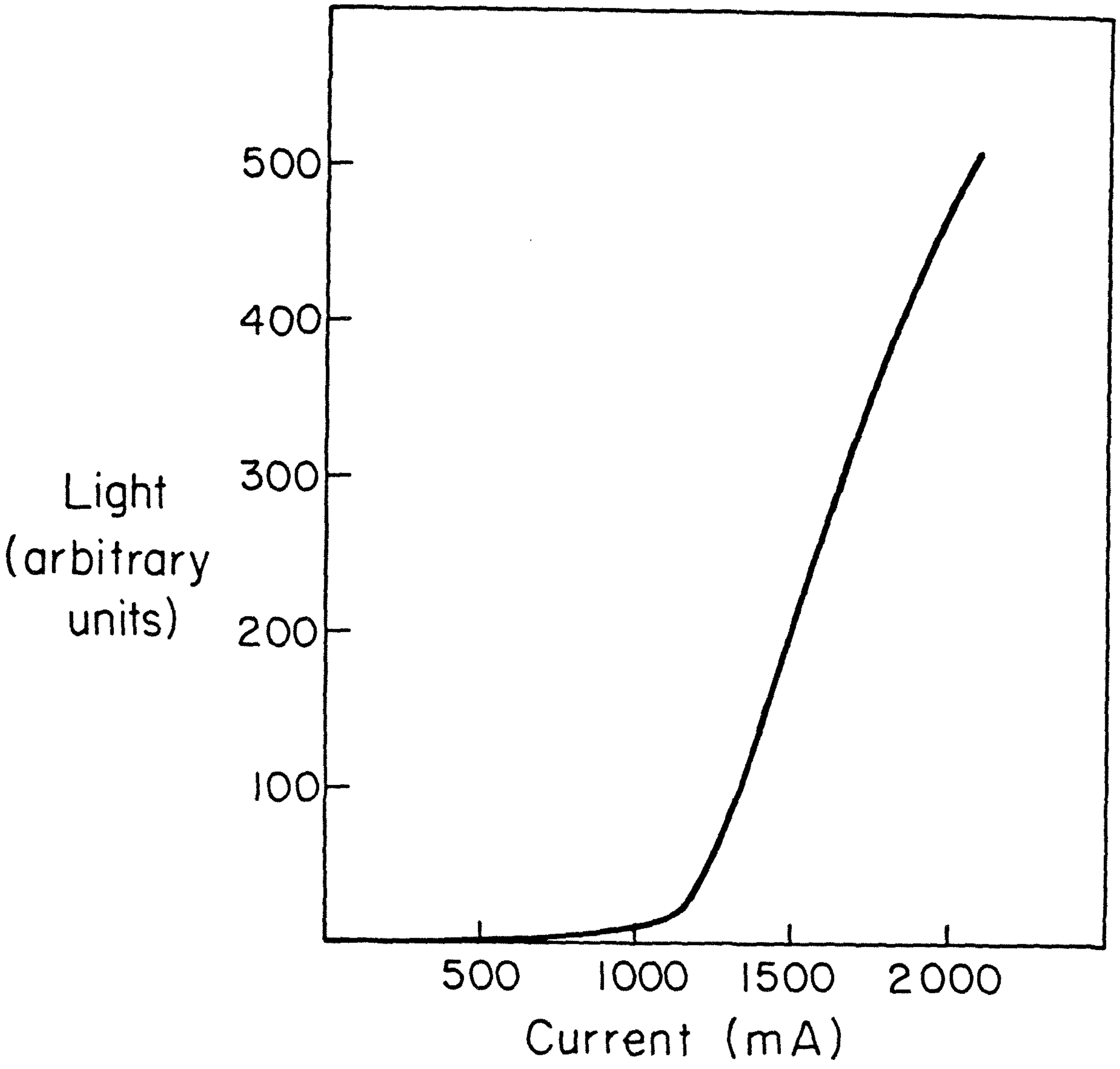


Figure 6.2) A typical threshold curve for a 150 micron wide array of 9 micron period 411 lasers

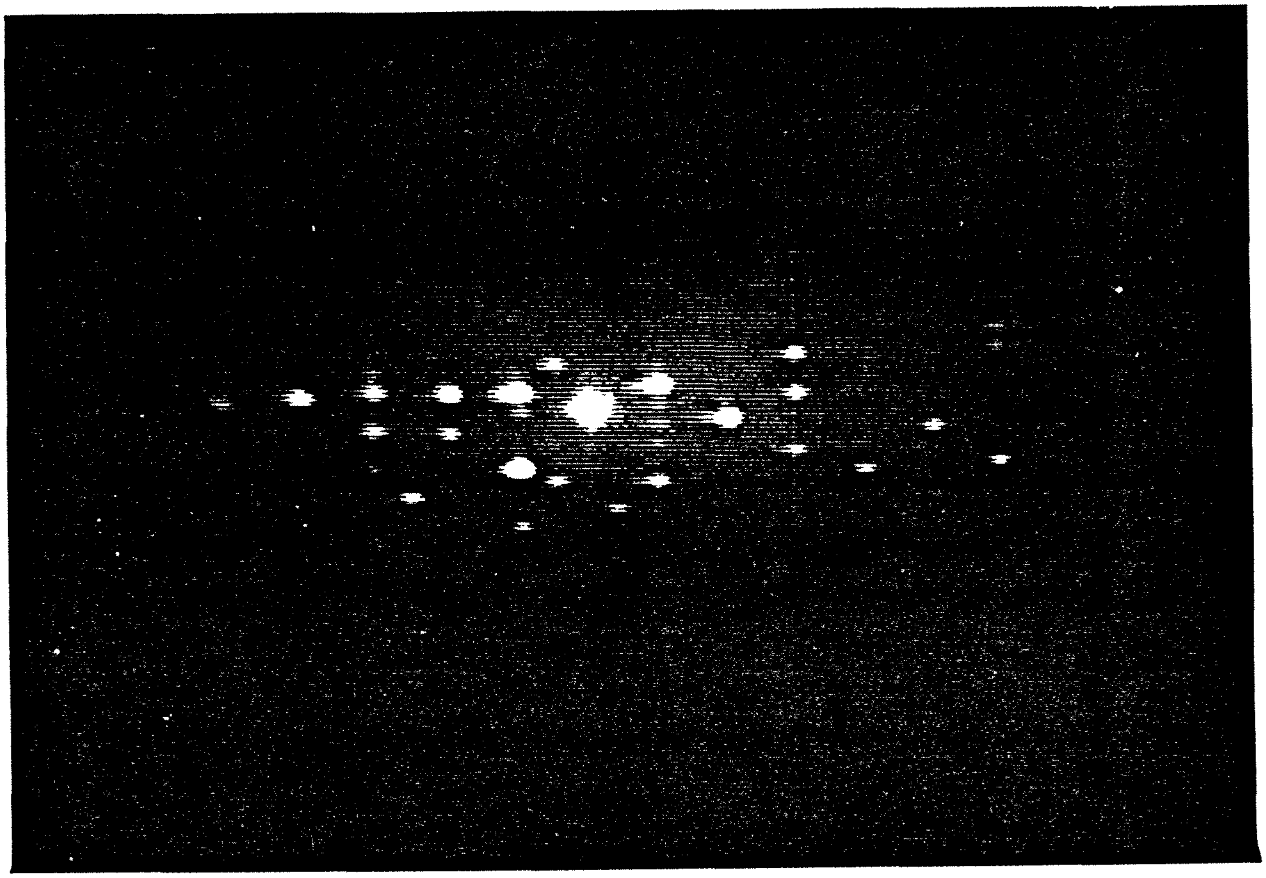


Figure 6.3) Nearfield of nine micron laser array, viewed through a spectrometer to produce an image which yields the emission wavelength across the lateral dimension of the array. This is refered to as a spectrally resolved nearfield.

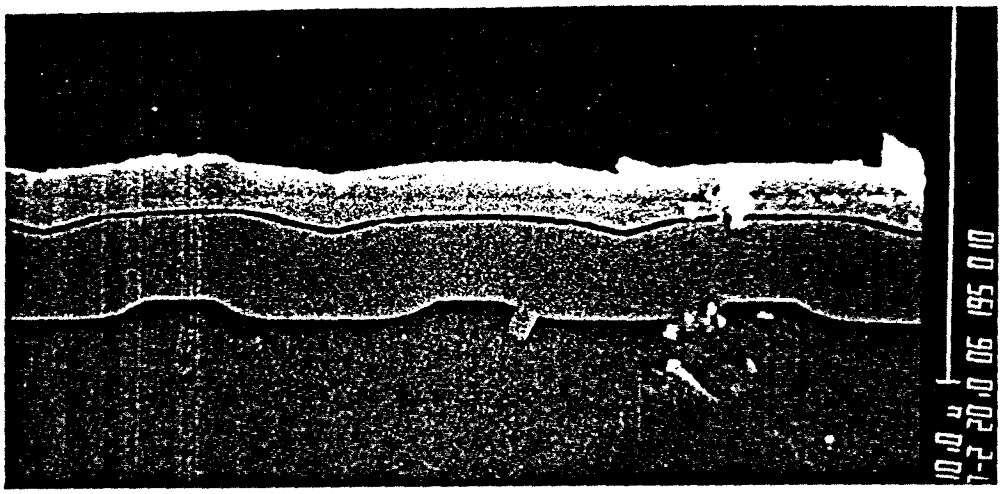


Figure 6.4) Cross section of five micron period laser array

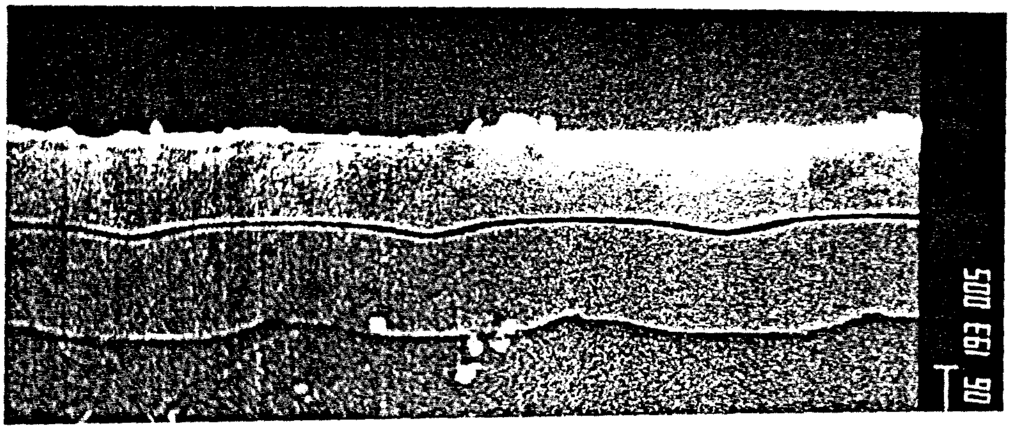


Figure 6.5) Cross section of four micron period laser array

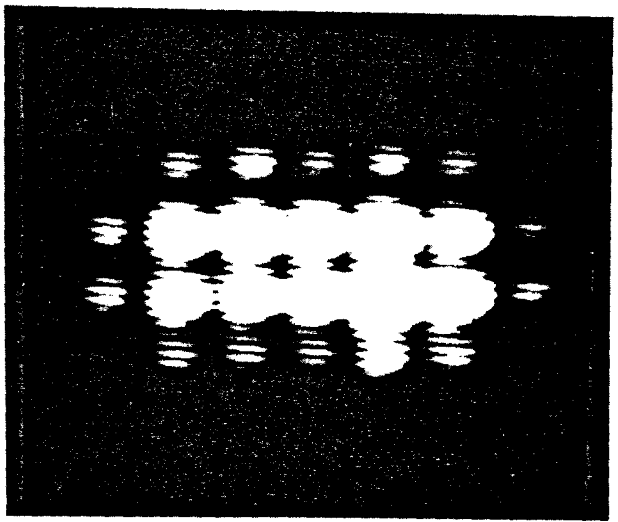


Figure 6.6) Spectrally resolved nearfield of four micron laser array.

Chapter VII

Other applications for MBE

Regrown structures

7.1 Introduction

This chapter discusses the application of the techniques discussed in previous chapters to the fabrication of several novel structures, a vertical channel field effect transistor (VFET) with a high electron mobility channel, a technique for making "quantum wire" structures, and a structure which allows contact to be made to individual quantum well layers in a multiple quantum well structure. This technique should also prove useful in the fabrication of other devices which take advantage of the geometric features of regrowth by MBE, and for fabrication of integrated structures.

7.2 Regrown channel vertical field effect transistor

This section discusses a series of experiments directed towards extending the performance of high speed transistors, using a vertical field effect transistor (FET) geometry to achieve very short channel lengths in combination with a modulation doped channel structure for high carrier mobilities and peak velocities. This geometry avoids the contact problems that the HEMT structure usually entails.

A rough estimate of the speed of a semiconductor switch is $\tau = Q/I$, where τ is the switch delay time, I is the current passed by the device, and Q is the charge necessary at the input to get the switch from the off state to the on state. The current drive of a given device is $I \sim qvn/d$ where q is the unit charge, v is the average velocity of the carriers through the structure, n is the number of moving carriers, and d is the distance the carriers must move. Since Q is also proportional to n, improvements in the speed of electronic devices come through either decreasing the scale of the devices, which reduces the distance the carriers must travel, or increasing the velocity of the carriers.

FET structures fabricated to the same scale as BJTs posess the potential of operating much faster than the latter because they are unipolar devices, and the mobility of electrons is much higher in most semiconductor materials than the mobility of holes.[1] However, it is more difficult to fabricate FETs than BJTs with a given critical dimension because the critical dimension of a FET, the gate length, is in the horizontal direction whereas the critical dimension of a BJT, the base thickness, is in the vertical direction, which is much easier to control as discussed before.

Many structures have been designed in which the gate structure of a FET is defined on the side of a channel etched in the substrate, with the length of the gate defined by the thickness of a grown or diffused layer.[2,3] These structures are collectively known as vertical FETs or VFETs. Not only do these structures gain the advantages of shorter channel lengths but also higher packing densities and higher voltage handling capability.[4] For high speed applications, it is necessary to use special techniques such as angle evaporation, to avoid a large overlap of the gate on the source and drain, to avoid excess parasitic capacitance. Combination of the short channel length available with the VFET geometry with the higher mobilities available in GaAs or other III-IV compounds should result in an increase in the state of the art performance for high speed transistors.

The dopant atoms which supply the mobile carriers in semiconductors also cause a reduction in the mobility of the carriers due to coulomb scattering. Esaki and Tsu[5] proposed in 1969 that a heterostructure junction could be used to separate the dopant atoms from the carriers as shown in figure 7.1, reducing the coulomb interaction between the carriers and their parent impurities and raising mobilities to near their values in the undoped material. After experimental demonstrations of the high electron mobility in such structures, FETs were demonstrated which used modulation doped channel layers.[6] High electron mobility transistors (HEMT) have been shown to have superior high frequency performance to MESFETs in the same materials.[7]

The structure of a typical HEMT, shown in figure 7.2, consists of a thick undoped layer of GaAs, followed by a thin (30 – 100Å) spacer layer of GaAlAs, followed by a highly doped layer of GaAlAs. A Schottky barrier gate is deposited on the grown layers, and the drain and source contacts are made by alloying to the GaAlAs layers on the sides. One of the remaining drawbacks to the HEMT structure is the difficulty of contacting the channel through the GaAlAs layer which is a poor conductor. If the layers could be inverted, with the GaAs layer on top, it would be easier to contact the channel, but heterointerfaces grown in the reverse order, GaAs on top of GaAlAs, have been found to have poor interface smoothness, resulting in low mobilities.

The device geometry studied here is shown in figure 7.3. NPN structures were grown on N^+ 100 substrates, and vee-grooves were etched in the $[0\bar{1}1]$ direction with 1:8:8 (H_2SO_4, H_2O_2, H_2O). The channels were formed by either implantation of Si^+ ions into the surface of the vee-groove, or by MBE regrowth using the techniques presented in earlier chapters. Aluminum gates were deposited by vacuum evaporation, and upper and lower contacts of Au/Ge were alloyed to the source and to the drain (substrate) as shown in the figure. The main advantages of this type of structure are that the gate length is determined by the thickness of a MBE grown layer and can therefore be very short and accurate, then the "vertical", the direction of accurate dimensional control, is changed by the etching of the v-groove so that the thickness of the channel is also controlled very accurately. The source and drain contacts are supplied by thick, heavily doped layers immediately adjacent to the channel, reducing the on resistance of the device.

7.3 Implanted channel device

The device whose I-V characteristic curves are shown in figure 7.4 was fabricated by the following procedure.[8] An MBE epitaxial growth was performed on a n^+ [100] substrate consisting of 3 microns of GaAs doped n type to $2 \times 10^{17}\text{cm}^{-3}$ with silicon followed by .15 microns of GaAs doped p type to $1 \times 10^{17}\text{cm}^{-3}$ with beryllium and a final layer 1.5 microns thick of GaAs doped n type to $2 \times 10^{17}\text{cm}^{-3}$. A layer of photoresist was spun on and 10 micron stripes were opened in the $[0\bar{1}1]$ direction. The wafer was then etched in 1:8:8 (H_2SO_4, H_2O, H_2O_2) for twenty five seconds which forms vee-grooves to a depth of 2.5 microns, just penetrating the p-type layer. The wafer was then implanted with Si^+ ions at a dose of $3 \times 10^{12}\text{cm}^{-2}$ at an energy of 120 Kev. A Si_3N_4 cap layer was then deposited by CVD,and the ion damage was then annealed out of the implanted layer at 850° C. AuGe contacts were then applied to the top layer and the substrate and separation grooves were etched through the epilayers to separate the devices. A Schottky gate Al layer was then applied to the top surface of the wafer by vacuum evaporation, and then patterned by photolithography and etching to remove all but the contact pad and the tab over the channel section as shown, which acts as the gate. The gate width defined was 120 microns. The typical transconductance of the device was 250 mS/mm.

7.4 Regrown channel device

The ion implanted structure discussed in the previous section has several features that are less than ideal. The channel region is formed by implanting n type dopant silicon atoms into a p type layer, which is then annealed. This results in a channel which is compensated due to the presence of the p type as well as the n type dopants, and ionized impurity scattering is increased. The ion damage may also not be completely annealed out, and the conduction channel is near the surface with its trap states. All of these reduce the electron mobility in the channel, and the performance of the device depends on the extremely short channel length.

Replacement of the ion implanted channel with a separately grown channel would eliminate the compensation problem, and the quality of the grown material may be higher than that of annealed ion implanted material. In addition, the regrown structure can incorporate a modulation doped structure, resulting in enhanced electron mobilities in the channel. A schematic diagram of the device is show in figure 7.5.

A regrown vertical channel of the type discussed above was formed by the following procedure. Starting with an n^+ substrate prepared by the procedures described in chapter 1, a 3 layer GaAs structure was grown, the first layer 2 microns thick doped with silicon to $2 \times 10^{17}\mathrm{cm}^{-2}$, followed by a beryllium doped layer 0.4 microns thick, and finally a 2 micron layer doped with silicon to $2 \times 10^{17}\mathrm{cm}^{-2}$. Vee-grooves were etched as described in section 7.3. The substrate was then cleaned following the procedure of chapter 3, and reloaded into the MBE. Layers with thicknesses calibrated for the (100) surface were grown as follows. 0.4 microns doped to $5 \times 10^{16}\mathrm{cm}^{-2}$ with tin, followed by a 0.2 micron layer grown with the tin shutter closed, but doped due to the surface segregation of the tin, decreasing to near the background level by the next interface. A 30 Å layer of $\mathrm{Ga_{.7}Al_{.3}As}$ spacer layer was grown, followed by a 0.15 micron layer of $\mathrm{Ga_{.7}Al_{.3}As}$ doped to $2 \times 10^{18}\mathrm{cm}^{-2}$ with silicon. Gold/germanium contacts were then applied and annealed. The I-V curve for a 100 micron wide channel is shown in figure 7.6. Figure 7.7 shows the I-V curve for the transistor structure formed by depositing an aluminum Schottky gate over the channel, and operating it as a FET. The performance of these devices has not been high thus far, with transconductances of approximately 20 mS/mm, and large leakage currents. Possible reasons include defects at the regrown interface, which is near the channel layer, and possible difficulties with the smoothness of the HEMT channel. It would be preferable to grow the HEMT channel over the $(4\bar{1}1)$ facet because of the proven high quality of the material, but the cleaning procedure is a selective etch for p-type GaAs, leaving an undercut as shown in figure 7.8.

7.5 Quantum wire structure and quantum well contact structures

The following describes a proposed procedure for forming "quantum wires," structures in which electrons are confined to channels which have two dimensions small enough to produce quantum size effects. In addition to the applications for quantum well structures, quantum well lasers and high electron mobility devices, these structures are interesting from a pure research viewpoint due to the effectively reduced dimensionality of the electrons which are constrained to move in only one dimension.[9]

One method of obtaining such structures is to use the intersection of two MBE epilayers such as that used in the VFET described above. With the first growth a fully depleted doping superlattice rather than the npn structure shown above, a set of quantum wires would be formed as shown in figure 7.9.

One limit on the use of quantum well structures is the difficulty of making electrical contact to the quantum wells. Etching, alloying and implantation procedures do not have the resolution necessary to single out a quantum well. An interesting solution to this problem would be to use the fact that growth over patterned substrates can result in quantum well layers which widen into layers as much as 10 times thicker on an adjacent facet, such as that shown in figure 7.10.

This would allow the use of conventional selective etching and contact deposition to apply contacts to the individual layers.

references for chapter 7

[1] P. M. Solomon, Proc. IEEE, vol. 70, no. 5, (1982).

[2] T. M. S. Heng, J. G. Oakes, and R. A. Wickstrom, Proc. Int. Electron Devices Meet., Suppl., p. 7, (1976).

[3] J. I. Nishizawa, T. Terasaki, and J. Shibata, IEEE Trans. Electron Devices, vol. ED-22, p. 185, (1975).

[4] K. P. Lisiak and J. Berger, IEEE Trans. Electron Devices, vol. ED-25, no. 10, p. 1229, (1978).

[5] L. Esaki and R. Tsu, "Superlattice and negative conductivity in semiconductors," IBM RES., Internal Rep. RC 2418, Mar. 26, (1969).

[6] P. Delescluse et al., "Transport properties in GaAs-AlGaAs heterostructures and MESFET application," Electron. Lett., Vol. 17, pp. 342–344, (1981).

[7] D. Delagebeaudeuf and N. T. Linh, IEEE trans. electron devices, vol ED-29, no. 6, (1982).

[8] Z. Rav-Noy, U. Schreter, S. Mukai, E. Kapon, J. S. Smith, L. C. Chiu, S. Margalit, and A. Yariv, IEEE Elect. Dev. Lett., vol EDL-5, no. 7, (1984).

[9] J. Hautman and L. M. Sander, Superlattices and Microstructures, vol. 1, no. 1, p. 39, (1985).

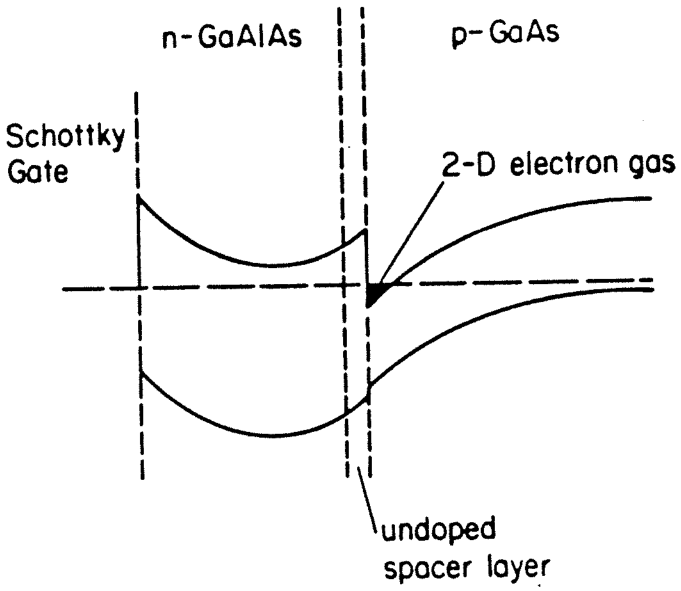


figure 7.1) Diagram of modulation doping structure.

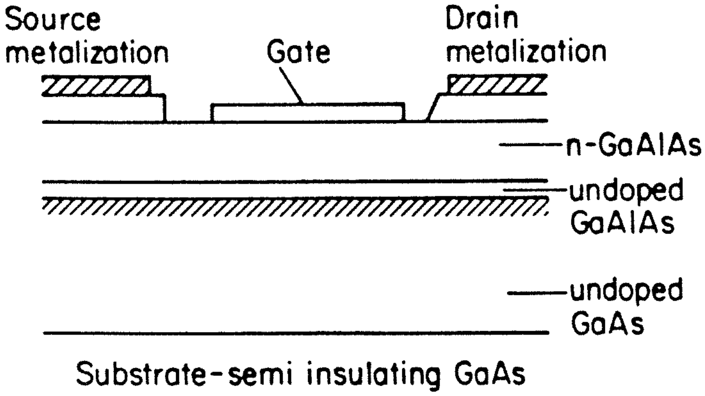


figure 7.2) Diagram of high electron mobility transistor.

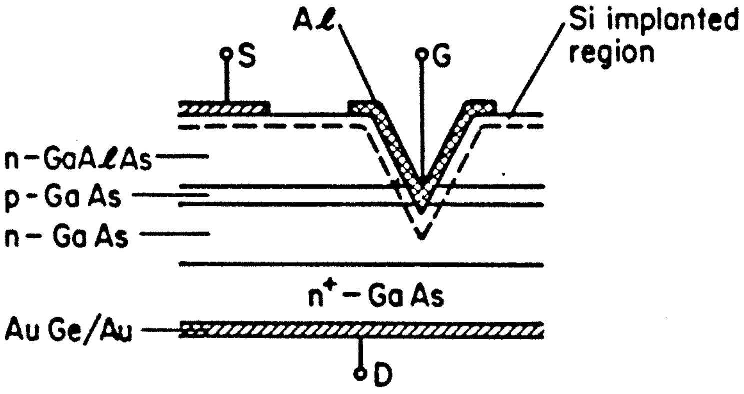


figure 7.3) Diagram of implanted channel VFET on GaAs.

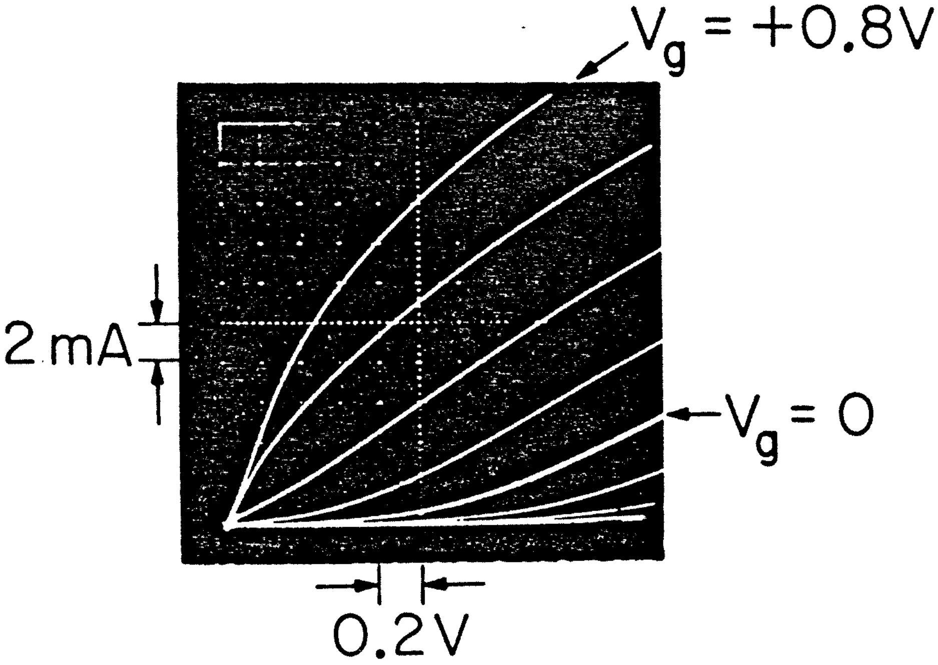


figure 7.4) I-V curve of implanted channel GaAs VFET.

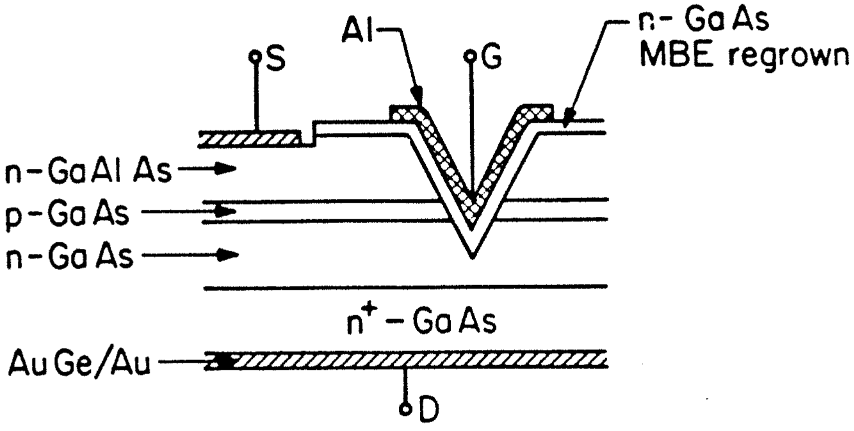


figure 7.5) Schematic diagram of HEMT VFET structure.

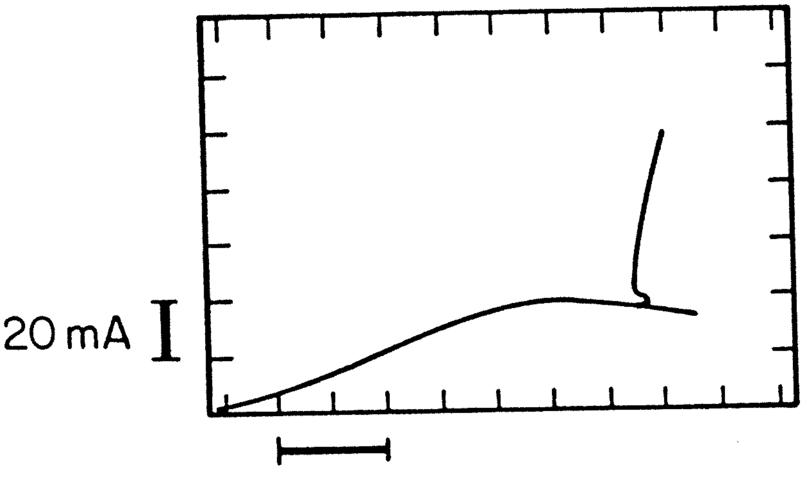


figure 7.6) I-V curve for regrown HEMT vertical short channel.

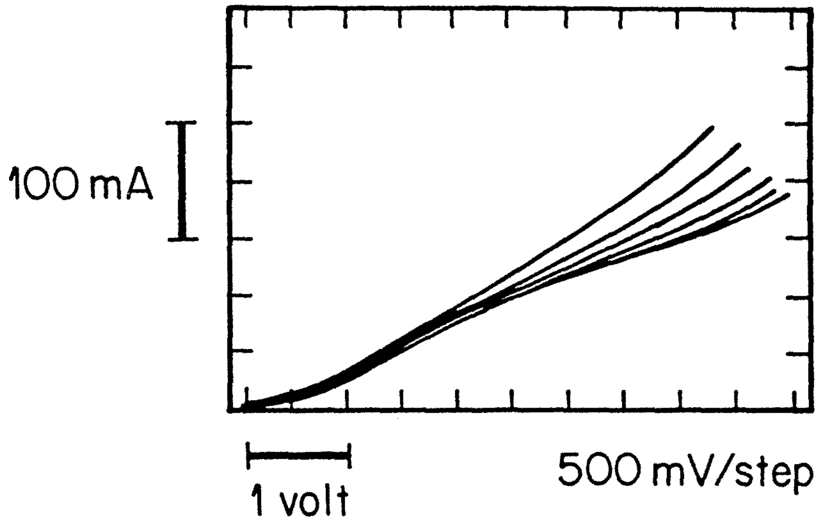


figure 7.7) I-V curve of HEMT VFET.

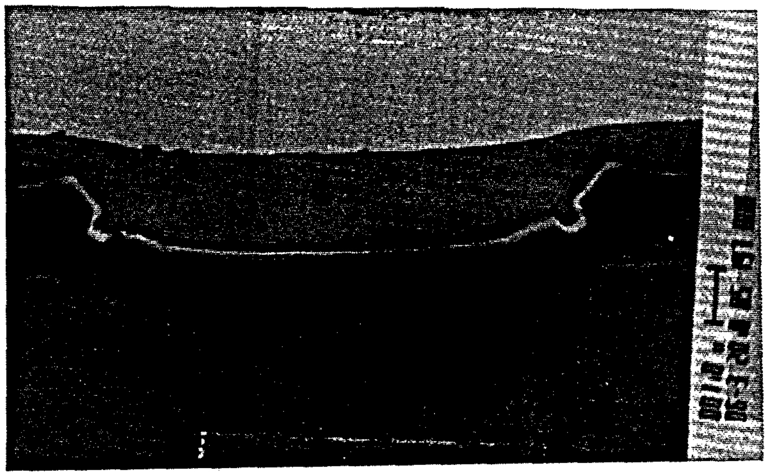


figure 7.8) Undercut of npn layers by cleaning procedure on $0\bar{1}\bar{1}$ ridge.

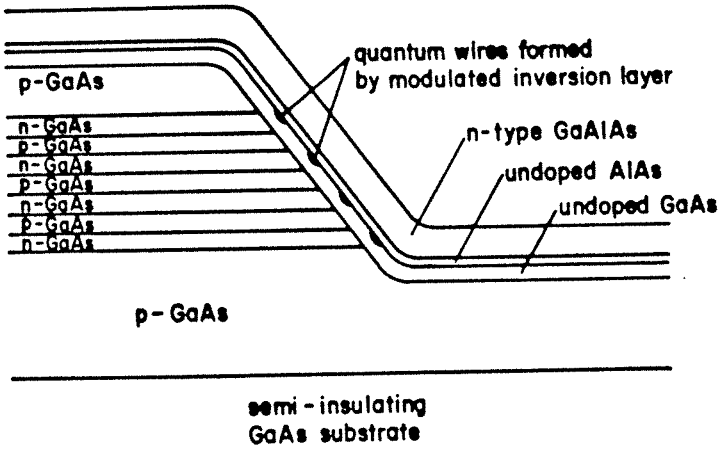


figure 7.9) Proposed quantum wire structure.

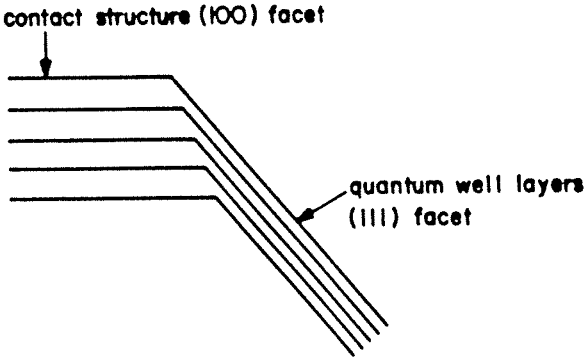


figure 7.10) Proposed geometry for contacting quantum well layers.

Chapter VIII

Ballistic transport over heavily doped quantum wells

8.1 Introduction

The purpose of this chapter is to discuss the transport of electrons transversely across heavily doped quantum wells, and its application to several devices, including a mid-infrared detector and a proposed new type of high speed transistor. Quantum well devices are interesting for several reasons, one being the production of two dimensional electronic states, and others which are not directly related to the quantum nature of the confinement, but simply due to the small dimension of the structure. An example of the second category is the high electron mobility transistor where the primary benefit of the quantum well structure is the spatial separation of the electrons from their parent donor atoms, which greatly increases the mobility of the electrons due to the reduction of ionized donor scattering. The devices discussed in this chapter fall into this latter category, using the small size of the quantum well to allow ballistic, nonequilibrium transport of hot electrons over a quantum well which confines a degenerate electron gas which is at a low temperature, in equilibrium with the crystal lattice.

The geometry which applies to the case to be considered in this chapter is that of a thin layer of a low bandgap material, such as GaAs, embedded in a crystal of a larger bandgap material, such as GaAlAs, with the larger bandgap material undoped and the smaller bandgap material heavily doped with a donor such as silicon, forming a quantum well with a dense electron population. In the cases where an array of these quantum wells is considered, adjacent wells are separated by a large enough distance so that tunneling between adjacent quantum wells can be neglected, and the quantum wells can be analyzed as if they were in isolation.

The devices are operated at temperatures sufficiently low that in equilibrium, all of the electrons are in a degenerate fermi gas confined to the quantum well, and the occupancy of states in the conduction band of the GaAlAs is negligible. An electric field is then applied to the structure transverse to the quantum wells, and electrons are introduced into the barrier regions by injection from the contacts, absorption of photons, or by other means. The injected electrons gain energy from the electric field as they accelerate toward the quantum well layer. If the quantum well layer is sufficiently thin, and the electron has sufficient energy, it may traverse the well before it loses enough energy to be captured by it. This not only includes true ballistic electrons which undergo no scattering events as they cross the well, but also electrons which do undergo scattering events, but lose insufficient energy to become captured. Under some circumstances, this transmission probability can be high, and can be put to use in the design of several interesting devices. If the energies of the injected electrons are sufficiently high, the probability that the incoming electron is captured is less than the probability that an additional electron is scattered out of the quantum well, resulting in avalanche gain. Avalanche gain is desirable in cases where the primary source of noise is the detection amplifier rather than the device itself, in which case the excess noise of the avalanche mechanism itself is less important than the gain it provides.

8.2 Solid State Photomultiplier

The first device to be considered is an electron multiplier which is a solid state analog of a vacuum electron multiplier. The band structure under an applied field is shown in figure 1. The principle of operation of the device is that electrons corresponding to a signal are injected into the device and are accelerated by the applied electric field. While the electrons are in the GaAlAs, they gain energy from the field until their energy losses due to inelastic scattering phenomena, predominantly phonon emission, balance the energy input from the field. Additionally, the electrons undergo nearly elastic collisions with acoustic phonons and ionized impurities which also randomize their momentum. While an electron is in the vicinity of a quantum well, it can also undergo Coulombic interactions with the trapped electrons and transfer energy to them. There are three possible outcomes for each electron injected into the well:

(1) the electron loses sufficient energy to become trapped in the quantum well without exciting any other electrons out of the quantum well;

(2) one or more e-e collisions occur resulting in more than one electron having sufficient energy to escape the quantum well;

(3) either no interactions occur and the electron escapes the well alone, or a series of processes result in only one electron escaping the quantum well.

Averaging over all of these possible outcomes, an average of n electrons are ejected from the quantum well for each electron which goes in. If n is greater than one, the possibility exists of using a multiple layer structure as an electron multiplier. Note that only one carrier type is involved in this avalanche process, in contrast to the P-I-N avalanche photodiode in which impact ionization generates both carrier types. This type of avalanche avoids the excess noise caused by the statistical variations in the multiplication in the P-I-N diode when the impact ionization rates for the two carrier types is similar.[1] It is necessary to supply an external bias structure to the device to replenish the electrons, otherwise, the structure would remain positively charged from the removal of the electrons, and current would flow continuously, independent of the injected current.

The device is thus a close analog of the vacuum electron multiplier, with the dynodes of the vacuum device corresponding to the low bandgap layers of the solid state device. Potential advantages of this device over a vacuum photomultiplier are its small size and the possibility of fabricating integrated arrays of such multipliers.

The calculation of the numerical electron gain per layer is broken down into two parts. First, the probability distribution $p(\vec{k})$ for electrons entering the quantum well is calculated, then a calculation is made of $n(\vec{k})$, the expected number of particles exiting the well given a particular incident $\vec{k}$ vector. The statistical average of n, is then given by

$$\langle n\rangle = \int p(\vec{k})n(\vec{k})d^3k \tag{1}$$

which is the expected number of electrons which leave the well for each electron entering.

This is a useful separation if: (1) the barrier layers are sufficiently wide so that the distribution function $p(\vec{k})$ is at a quasi-steady state value, that is, the distribution in k space is that of a bulk material with the same electric field; and (2) the quantum wells are thin enough so that few interactions occur during the transit time of an electron across the quantum well.

8.3 Electron distribution under a moderate electric field

The band structure for GaAs is shown in figure 2. We are primarily interested in the conduction band, as these are unipolar devices. The lowest energy in the conduction band in GaAs is at the bottom of the single gamma (Γ) valley. The valley is spherically symmetric in the regions of interest, and the electron effective mass is $.06m_o$, where m_o is the free electron mass. The next higher valleys are the four equivalent L valleys in the [111] symmetry directions, which are .29 eV above the gamma valley. The effective mass in the L valleys is much higher than that in the gamma valley, and they are elliptical. The averaged effective electron mass in the L valley is $.222m_o$. In $\mathrm{Ga_{1-x}Al_xAs}$, the electron effective mass is[2]

$m_\Gamma = (0.067 + 0.083)\mathrm{x}m_o$, and the separation of the gamma valley and the L valley becomes $\Delta E_{\Gamma L} = 0.284 - 0.605\mathrm{x}(eV)$.

The phonon spectrum of GaAs is shown in figure 3. GaAs, having two atoms per unit cell, has six phonon branches, three acoustic and three optical. The energy of phonons in the optical branches is on the order of .038 eV. At low fields, where the electrons have insufficient energy to emit optical phonons, the dominant processes for momentum relaxation are acoustic phonon scattering and ionized impurity scattering, and energy relaxation is dominated by the acoustic phonon scattering. In moderate fields in GaAs, the energy of some electrons exceed the optical phonon energy, and the dominant scattering process for both energy relaxation and momentum relaxation is L-O (longitudinal optical) phonon emission.[3] At higher field strengths, electrons gain sufficient energy to transfer into the L valleys. The dominant scattering process at energies above the minimum of the L valleys is inter-valley phonon scattering.[4] Phonons with wavevectors large enough for conservation of crystal momentum in an intervalley scattering process have energies comparable to the optical phonons, regardless of their type. These scattering processes are very fast because of the much higher density of states compared to the gamma valley alone, due to the high electron effective mass and the presence of four valleys, and the electrons do not heat up further until the electric field is much higher.

The "moderate electric fields" referred to at the beginning of this section are fields in which an electron propagating through the crystal has a relatively high probability of having an energy in excess of the optical phonon energy, but less than that which would cause large excursions above the bottom of the L-valley where the interaction rates are higher. This range is approximately 3 to 10 kv/cm. This is the energy range that we are interested in because the electrons which approach the well must have a kinetic energy at least as large as the binding energy of the electrons in the quantum well if they are to be able to scatter a cold electron out of the quantum well, but the electric field should not be so high that some electrons impact ionize, creating electron-hole pairs.

In a homogeneous material, we can define a distribution $f(\vec{k},t)$ as the probability for an electron to possess a wavevector $\vec{k}$ at time t. In the case of relatively infrequent collisions and short collision times, the Boltzmann transport equation can be used to derive $f(\vec{k},t)$ in the presence of the accelerating electric field and scattering mechanisms. The Boltzmann equation for a homogeneous material is[5]

$$\frac{\partial f(\vec{k},t)}{\partial t} + \vec{\nabla}_k f(\vec{k},t) \cdot \frac{d\vec{k}}{dt} + \sum_j I^j_{\vec{k}} f(\vec{k},t) = 0 \tag{2}$$

where $I^j_{\vec{k}}$ is the scattering rate operator for the j^{th} scattering mechanism, $\vec{k}$ is the wavevector which acts as the index for the one-electron states $|\Psi(\vec{k})\rangle$, and $f(\vec{k},t)$ is normalized so that $\int \left(\frac{2V_o}{(2\pi)^3}\right) f(\vec{k},t)d^3k = 1$, where V_o is the integration volume, and the distribution function is taken to be independent of the electron spin. The equation of motion for an electron in an applied electric field is

$$\frac{d\vec{k}}{dt} = -\left(\frac{q}{\hbar}\right)\vec{\mathcal{E}} \tag{3}$$

and the scattering rate operator applied to $f(\vec{k},t)$ is[6]

$$I^j_{\vec{k}} f(\vec{k},t) = \frac{V_o}{(2\pi)^3}\int W(\vec{k},\vec{k}')f(\vec{k}',t)d\vec{k}' - \frac{V_o f(\vec{k},t)}{(2\pi)^3}\int W(\vec{k}',\vec{k})d\vec{k}' \tag{4}$$

where $W(\vec{k},\vec{k}')$ is the transition rate from state $|\Psi(\vec{k}')\rangle$ to the state $|\Psi(\vec{k})\rangle$. The first term corresponds to the scattering rate for electrons into the state $\vec{k}$ from all other states $|\Psi(\vec{k}')\rangle$ and the second term corresponds to the scattering rate out of state $\vec{k}$ to all other states $|\Psi(\vec{k})\rangle$. Setting the derivative with respect to time to zero to arrive at the steady state equation,

$$\frac{q}{\hbar}\vec{\mathcal{E}} \cdot \vec{\nabla}_k f(\vec{k}) = \frac{V_o}{(2\pi)^3}\int \sum_j W_j(\vec{k},\vec{k}')f(\vec{k}')d\vec{k}' - \frac{V_o f(\vec{k})}{(2\pi)^3}\int \sum_j W_j(\vec{k}',\vec{k})d\vec{k}' \tag{5}$$

At this point, the main approximation which has been made is that the scattering mechanisms are a small perturbation to the energy eigenstates. Adding the assumption that the time between collisions is large compared to the duration of an individual collision, and that the interaction potentials are small, we can use Fermi's golden rule to find the transition rate due to the scattering mechanisms. $W(\vec{k},\vec{k}') = \sum_j W_j(\vec{k},\vec{k}')$ where W_j is the scattering rate due to each of the scattering mechanisms, and

$$W_j(\vec{k},\vec{k}') = \frac{2\pi}{\hbar}|\langle\psi_f \mid H'_j \mid \psi_i\rangle|^2\delta(E_f - E_i) \tag{6}$$

where H'_j is the interaction Hamiltonian, ψ_i and ψ_f are the initial and final states, and E_i and E_f are the total initial and final energies.

In the energy spectrum above the threshold for L-O phonon emission but below the minimum of the L-valley, L-O phonon emission dominates other scattering processes in $\mathrm{Ga_{1-x}Al_xAs}$. The matrix element for L-O phonon emission is[7]

$$|\tilde{H}'|^2 = \xi(\vec{k},\vec{k}')\frac{C}{|\vec{k}-\vec{k}'|^2} \tag{7}$$

where $\xi(\vec{k},\vec{k}') = |\frac{1}{V_o}\int_r u_{1'}^*(\vec{r})u_1(\vec{r})d^3r|^2$ is the overlap integral for the periodic part of the electron states, absorption and stimulated emission have been omitted because of the low occupancy factor for the optical phonons at these lattice temperatures,

$$C = \alpha_p\frac{\hbar 2\pi\sqrt{2}(\hbar\omega_{\mathrm{op}})^{3/2}}{m_e^{1/2}V_o} \tag{8}$$

and

$$\alpha_p \equiv \frac{q^2}{\hbar\epsilon}\left(\frac{m_e}{2\hbar\omega_{\mathrm{op}}}\right)^{1/2}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) \tag{9}$$

is the dimensionless Fröhlich constant, ϵ_∞ is the relative dielectric constant at high frequencies, ϵ_0 is the static relative dielectric constant, $\hbar\omega_{\mathrm{op}}$ is the energy of the optical phonon, and V_o is the volume of the crystal.

The transition rate is

$$W(\vec{k}_i, \vec{k}_f) = \left(\frac{2\pi C/\hbar}{|\vec{k}_i - \vec{k}_f|^2} \right) \xi(\vec{k}, \vec{k}')\delta\left(\frac{\hbar^2}{2m_e}(|\vec{k}_f|^2 - |\vec{k}_i|^2) - \hbar\omega_{\mathrm{op}} \right) \tag{10}$$

Inserting the transition rate into equation 8.5,

$$\begin{aligned} \frac{q}{\hbar}\vec{\mathcal{E}} \cdot \vec{\nabla} f(\vec{k}) &= \frac{C}{(2\pi)^2\hbar}\xi(\vec{k}, \vec{k}') \int \frac{f(\vec{k}')}{|\vec{k}' - \vec{k}|^2}\delta\left(\frac{\hbar^2}{2m_e}(|\vec{k}'|^2 - |\vec{k}|^2) - \hbar\omega_{\mathrm{op}} \right) d\vec{k}' \\ &- \frac{C}{(2\pi)^2\hbar}\xi(\vec{k}, \vec{k}') f(\vec{k}) \int \frac{1}{|\vec{k}' - \vec{k}|^2}\delta\left(\frac{\hbar^2}{2m_e}(|\vec{k}'|^2 - |\vec{k}|^2) - \hbar\omega_{\mathrm{op}} \right) d\vec{k}' \end{aligned} \tag{11}$$

carrying out the integration in the second term,

$$\begin{aligned} \frac{q}{\hbar}\vec{\mathcal{E}} \cdot \vec{\nabla} f(\vec{k}) &= \frac{C}{(2\pi)^2\hbar}\xi(\vec{k}, \vec{k}') \int \frac{f(\vec{k}')}{|\vec{k}' - \vec{k}|^2}\delta\left(\frac{\hbar^2}{2m_e}(|\vec{k}_f|^2 - |\vec{k}_f|^2) - \hbar\omega_{\mathrm{op}} \right) d\vec{k}' \\ &- \frac{C}{(2\pi)^2\hbar}\xi(\vec{k}, \vec{k}') f(\vec{k}) \frac{2m_e\pi}{\hbar^2|\vec{k}|} \ln\left(\frac{\left(\frac{\hbar^2}{2m_e}\right)^{1/2} |\vec{k}| + \sqrt{\frac{\hbar^2 k^2}{2m_e} - \hbar\omega_{\mathrm{op}}}}{\left(\frac{\hbar^2}{2m_e}\right)^{1/2} |\vec{k}| - \sqrt{\frac{\hbar^2 k^2}{2m_e} - \hbar\omega_{\mathrm{op}}}} \right) \end{aligned} \tag{12}$$

At this point, it is possible to generate numerical solutions to this equation using a Monte-Carlo calculation, or an analytical result can be found by assuming a drifted thermal distribution for the electrons, however, the result would not be physically meaningful for the following reasons. The last term in equation 12 is the total scattering rate out of the state $\vec{k}$. Figure 4 shows the variation of this scattering rate as a function of the magnitude of the electron wavevector. The scattering rate for gamma point electrons is $10^{13}\ \mathrm{s}^{-1}$, to a good approximation, for energies from slightly above the optical phonon energy to .29 eV, where scattering into the L valley commences. This gives a probability distribution for the time between collisions, when the electron's energy is within this range, $P(t) = \frac{1}{\tau}e^{-t/\tau}dt$, the probability of scattering between t and $t + dt$. If we assume that an optical phonon emission completely randomizes the momentum, the momentum after a scattering event has magnitude k_o and its direction probability distribution is uniform. Taking the $\hat{z}$ direction in the direction of the electric field, the probability distribution for k_z is $P(k_z) = (\pi/2)\left(1 - k_z^2/k_o^2\right)^{1/2} dk_z \quad \{|k_z| < k_o\}$. Using the equation of motion for the electrons, the change in energy over a collision cycle is $\Delta E = \frac{\hbar^2}{2m_e}[-q/\hbar\mathcal{E}tk_{zo} + (q/\hbar\mathcal{E}t)^2] - \hbar\omega_{\text{op}}$, where the $\hbar\omega_{\text{op}}$ term accounts for the loss of the energy of one optical phonon, and k_{zo} is the initial value of the component of the wavevector in the direction of the electric field. The probability distribution for the energy loss ΔE is

$$P(\Delta E) = \frac{d}{dE}\int \int_{E'(k_{zo},t)<E} P(k_{zo})P(t)dk_{xo}dt \tag{13}$$

which unfortunately does not analytically reduce to a usefully simple expression, however, several conclusions can be drawn by finding the mean value of the energy change and an approximation for its variance. Starting again with the equation of motion for the electron,

$$\frac{d\vec{k}}{dt} = (\frac{q}{\hbar})\vec{\mathcal{E}}$$

$$\langle\vec{k}\rangle = \int_0^{\tau_{\text{op}}} \langle\dot{\vec{k}}\rangle dt = \int_0^{\tau_{\text{op}}} (\frac{q}{\hbar})\vec{\mathcal{E}}\,dt = \tau_{\text{op}}(\frac{q}{\hbar})\vec{\mathcal{E}} \tag{14}$$

In the effective mass approximation, $\vec{v}_g = (\hbar/m)\vec{k}$ and the average rate power gained by the electron from the field is

$$P = q\vec{E}\cdot\langle\vec{v}\rangle = \tau\frac{q^2}{m}|\vec{\mathcal{E}}|^2 \tag{15}$$

Each optical phonon emitted carries $\hbar\omega_{\text{op}}$ of energy with it, so the average power loss to the phonon field is $P_l = \hbar\omega/\tau_{op}$. Note that this power loss is constant once the electron exceeds the optical photon energy, while the energy input from the field continues to rise. The point at which the energy lost to the optical photons does not balance the energy input from the electric field is

$$|\vec{\mathcal{E}}|^2 = \frac{\hbar\omega_{op}m_e}{\tau_{op}^2 q^2} \tag{16}$$

For GaAs, this occurs at approximately 4000 V/cm. Below this field strength, energy balance is maintained by the changing ratio of electron population above and below the optical phonon energy. At higher fields, the optical phonon interaction is insufficient to maintain a steady state, and other processes which occur at higher energies become important. The process which limits the further heating of the electron gas is intervalley phonon scattering. Not only does the net mean energy loss over a collision cycle reduce to zero and eventually become negative, the variance of the probability distribution grows, initially at zero (probability of energy lost equal one for $\Delta E = \hbar\omega_{op}$, and zero otherwise), it becomes $\sim \frac{\hbar q \pi}{8m} \mathcal{E} \tau k_o$, (where the variance due to the phonon emission time has been neglected), which becomes comparable to the total energy of the electron. This large variance comes from the fact that after a collision the electron's direction is randomized. If it propagates with the field, it gains energy rapidly, but if it is against the field, it loses energy rapidly. The distribution function of the electrons flattens out, the effective temperature of the electrons goes to infinity, and the distribution can even become inverted, as shown in the Monte-Carlo simulation in reference ([8]). Note that the calculation in that reference was undertaken at a time when the energy ordering of the conduction valleys in GaAs was thought to be $\Gamma - X - L$, rather than the actual $\Gamma - L - X$, but good agreement with experimental results was achieved anyway. No comparable Monte-Carlo calculations seem to have been published since then with the proper ordering which include the electron distribution, but the essential features of the transport seem to be relatively independent of details such as the identity of the higher energy, low mobility valley, and of the exact scattering rates in the higher valley. Over a range of electric fields, the distribution can be approximated as uniform, independent of energy in the gamma valley. The portion of electrons which remain in the gamma valley can be estimated by noticing that the mobility of electrons in the gamma valley remains constant at the value which it has in lower fields, or even increases slightly, where the electrons in the L-valley have a greatly reduced mobility. Since average electron velocity saturates at $\sim 2 \times 10^7$cm/s, the proportion of electrons which remain in the gamma valley in moderate electric fields is approximately $n_\Gamma/n = \frac{v_{\rm sat}}{\mu_\Gamma \mathcal{E}}$.

Electrons which are in the L-valley as they approach the quantum well remain in the L-valley within the quantum well.[9] Once they are in the quantum well layer, they quickly lose energy to intervalley phonon emission until they are at the minimum energy of the L-valley, which is a state which is confined to the quantum well. However, to cool further, they must make the transition to the gamma valley. At this point, they have an energy above the conduction band edge of the GaAs of $E_{\Gamma L}$. If the depth of the quantum well is less than this, the electron must lose the difference before diffusing out of the quantum well if it is to be permanently trapped in the quantum well. Thus, for the purpose of calculation of capture probabilities, L-valley electrons which are injected into the quantum well are considered equivalent to electrons injected in the Γ valley at an energy of $E_{\Gamma L}$.

The above approximation yields the probability distribution in $\vec{k}$ space for an arbitrary electron. The distribution of carriers *entering* the quantum well is weighted for the average velocity perpendicular to the quantum well. The average drift velocity electrons in the gamma valley is high and relatively independent of energy, and the drift velocity of electrons in the L-valley is much lower. The mobility in the L-valley is estimated to be ~ 900 cm^2/V-s,[10] compared to ~ 7000 for electrons in the gamma valley.

The probability distribution in k space for electrons passing through a plane of the crystal, for example, the x-y plane, can be found by weighting the distribution by the average velocity in the $\hat{z}$ direction and renormalizing.

$$p(\vec{k}) = \frac{v_z f(\vec{k})}{\int\limits_{v_z>0} v_z f(\vec{k}) d^3k} \tag{17}$$

Since the drift velocity of the electron depends only on which valley it is in, in this approximation, the integral collapses to

$$\int\limits_{v_z>0} v_z f(\vec{k}) d^3k = n_\Gamma v_{z\Gamma} + n_l v_{zL} \tag{18}$$

The ratio of electrons injected into the quantum well which are in the gamma valley to the total electrons injected is then

$$r_\Gamma = \frac{v_\Gamma n_\Gamma}{v_\Gamma n_\Gamma + v_L(n - n_\Gamma)} \tag{19}$$

and the probability distribution in k space is

$$p(\vec{k}) = \begin{cases} \dfrac{r_\Gamma}{\frac{4\pi}{3}\left(\frac{2m_e}{\hbar^2}E_{\Gamma L}\right)^{3/2}} & \left\{ |k| < \sqrt{\dfrac{2m_e}{\hbar^2}E_{\Gamma L}} \right\} \\ 0 & \left\{ |k| > \sqrt{\dfrac{2m_e}{\hbar^2}E_{\Gamma L}} \right\} \end{cases} \tag{20}$$

with effectively a delta function at $E = E_{\Gamma L}$ containing $(1 - r_{\Gamma L})$ of the electrons. If momentum-randomizing collisions occur often enough, the electron will undergo several collisions as it passes over the quantum well, and the direction of its initial wavevector is unimportant. In this case, the probability distribution as a function of energy (only) is used as the initial condition.

$$p(E) = \int\limits_{E(\vec{k})=E} p(\vec{k})|\vec{k}|^2 \sin(\theta) d\phi d\theta \left(\frac{dE}{d|k|}\right)^{-1} \tag{21}$$

$$p(E) = \begin{cases} \dfrac{3}{2} r_\Gamma \dfrac{E^{1/2}}{E_{\Gamma L}^{3/2}} & E < E_{\Gamma L} \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

where there is again a delta function at $E_{\Gamma L}$ corresponding to the electrons which were in the L-valley before injection. The diffusion and drift of the hot carrier are then taken to be the mean values calculated for an electron at that energy. Only the energy losses are taken to be statistical. This approximation is justified provided that there are statistically large number of momentum randomizing collisions over the quantum well.

8.4 Capture probability due to L-O phonon emission

Now that we have an approximate probability distribution for the electrons entering the quantum well, we need to calculate the probability with which an electron at a particular wave vector will either be captured by the quantum well or will interact with the electrons in such a way that a total of two or more escape the quantum well.

The mobility over the quantum well layers is lowered by ionized donor scattering in addition to the scattering mechanisms which are important in the undoped GaAlAs layer. The mobility in GaAs doped in the $10^{17}-10^{18}$ cm^{-3}range is ~2000 cm^2/v-s.[11] The electric field in the quantum well layer depends on the distribution of the doping in the layer, which can be controlled, and the distribution of the electrons in the layer, which depends on the electric field. If the quantum well layer is uniformly doped, the field in the quantum well layer is lower than that in the barrier layers due to the displacement of the confined electrons in the electric field. Assuming that the doping distribution is tailored to maintain the same field over the quantum well (for a specific value of the electric field), the mean time the hot electron spends in the vicinity of a quantum well is

$$t_q = \frac{w}{\mu_q \mathcal{E}} \tag{23}$$

where w is the width of the quantum well and μ_q is the mobility of the hot electron in the quantum well layer. For a typical quantum well width of 80 Å and an electric field of 10000 v/cm, this time is approximately 5×10^{-14} seconds.

Since the probability of emitting a L-O phonon is approximately independent of the electron's energy if it is above the optical phonon energy, the probability of emitting N L-O phonons during a time t_q is given by the Poisson distribution.

$$P(N, \tau_{\mathrm{op}}/t_q) = \frac{\left(\tau_{\mathrm{op}}/t_q\right)^N}{N!} e^{-(\tau_{\mathrm{op}}/t_q)} \tag{24}$$

The energy loss necessary for the electron to become captured in the quantum well is approximately the kinetic energy which it had when injected into the quantum well. The probability that the electron will be captured due to optical phonon emission is

$$P_{\mathrm{capture}}(E) \approx 1 - \sum_{x=0}^{L(E)} \frac{\left(\tau_{\mathrm{op}}/t_q\right)^x}{x!} e^{-(\tau_{\mathrm{op}}/t_q)} \tag{25}$$

where $L(E) = \mathrm{int}(E/\hbar\omega_{\mathrm{op}})$ is the number of optical phonons which must be emitted for the electron to become captured. Figure 5 shows the probability of capture by optical phonon emission as a function of the energy at injection and (τ_{op}/t_q), the ratio of the time over the quantum well and the optical phonon emission lifetime. Integrating over energy yields the probability that an individual electron will become captured due to phonon emission as it propagates across a quantum well.

$$P_{\mathrm{capture}} \approx \int_0^{E_{\Gamma L}} 3/2 \frac{E^{1/2}}{E_{\Gamma L}^{3/2}} \left[1 - \sum_{x=0}^{\mathrm{int}(E/\hbar\omega_{\mathrm{op}})} \frac{\left(\tau_{\mathrm{op}}/t_q\right)^x}{x!} e^{-(\tau_{\mathrm{op}}/t_q)}\right] dE \tag{26}$$

It can be seen that for the range of parameters which we are interested, where the electron energies are high and the transit time is comparable to the photon emission time, the probability of electron capture solely due to optical phonon emission is extremely small.

8.5 Calculation of perturbation matrix elements in the vicinity of a quantum well

In the following sections, in calculating the scattering interactions of the electrons, we will need to integrate over functions which have factors which vary rapidly over a unit cell, and other factors which vary slowly over a few unit cells. If we impose periodic boundary conditions over an integration volume V_o, and we have an integral

$$I = \int_{V_o} \mathcal{A}(\vec{r})\mathcal{B}(\vec{r})d^3r \tag{27}$$

where $\mathcal{A}(\vec{r})$ varies rapidly and $\mathcal{B}(\vec{r})$ varies slowly over a unit cell, we can decompose $\mathcal{A}(\vec{r})$ and $\mathcal{B}(\vec{r})$ into Fourier series $\mathcal{A}(\vec{r}) = \sum_{\vec{k}} A_{\vec{k}} e^{i\vec{k}\cdot\vec{r}}$ and $\mathcal{B}(\vec{r}) = \sum_{\vec{k}} \mathcal{B}_{\vec{k}} e^{i\vec{k}\cdot\vec{r}}$. The integral I is then given by

$$I = \sum_{\vec{k}_1}\sum_{\vec{k}_2} \mathcal{A}_{\vec{k}}\mathcal{B}_{\vec{k}} \int e^{(\vec{k}_1+\vec{k}_2)\cdot\vec{r}} d^3r \tag{28}$$

If the functions do not overlap in $\vec{k}$ space, there are no terms where $\vec{k}_1 = -\vec{k}_2$ where either $\mathcal{A}_{\vec{k}_1}$ or $\mathcal{B}_{\vec{k}_2}$ is not zero, and only the $\vec{k} = 0$ terms contribute to I. The factors which vary rapidly over a unit cell should not have nonzero components corresponding to wavevectors smaller than the reciprocal of the cell dimensions, and the slowly varying factors should not have nonzero components larger wavevectors. The Fourier components $\mathcal{A}_o$ and $\mathcal{B}_o$ are $\mathcal{A}_o = \frac{1}{V_o}\int_{V_o} \mathcal{A}(\vec{r})d^3r$ and $\mathcal{B}_o = \frac{1}{V_o}\int_{V_o} \mathcal{B}(\vec{r})d^3r$, and we have

$$I = \frac{1}{V_o}\int_{V_o} \mathcal{A}(\vec{r})d^3r \int_{V_o} \mathcal{B}(\vec{r})d^3r \tag{29}$$

which we will use in the later evaluations.

The electronic states in the vicinity of a quantum well can be approximated by a wavefunction which is separated into two parts in the effective mass approximation, a periodic part which varies rapidly over a unit cell, and one which varies slowly over a single cell.

$$\Psi(\vec{r}) = u(\vec{r})\psi(\vec{r}) \tag{30}$$

Where $u(\vec{r})$ is periodic over a unit cell, and $\psi(\vec{r})$ varies slowly from one unit cell to the next. The orthonormalization of these wavefunctions is given by

$$\int_{r} \Psi_n^*(\vec{r})\Psi_m(\vec{r})d^3r = \delta(n,m) \tag{31}$$

For low energy electrons near the bottom of the gamma valley in GaAs or GaAlAs, the rapidly varying part of the wavefunction is approximately the same as that of the Bloch wavefunction $\langle\vec{r}|\vec{k}\rangle = u_{\vec{k}}(\vec{r})e^{i\vec{k}\cdot\vec{r}}$ which is independent of $\vec{k}$ for $\vec{k}$ small.

When the material composition of a semiconductor alloy is varied, the first order effect on the electronic wavefunctions is to establish an effective potential which is equal to the portion of the bandgap change which appears as an offset in the conduction band. In $\mathrm{Ga_{1-x}Al_xAs}$, the bandgap is approximated by $E_g^\Gamma(eV) = 1.424+1.247\mathrm{x}$ for $\mathrm{x} <\sim .35$,[12] and approximately 85% of the change shows up in the variation of the gamma point. The effective potential for electrons in $\mathrm{Ga_{1-x}Al_xAs}$ is then given by $V(\vec{r}) = -.85 \cdot 1.247 \cdot \mathrm{x}(\vec{r})$, where $\mathrm{x}(\vec{r})$ is the composition of the material as a function of position in the crystal, and we have chosen the zero potential to be that in the GaAs.

The effective potential in the vicinity of a quantum well is given by

$$V(\vec{z}) = \begin{cases} 0 & |z| < \dfrac{w}{2} \\ \delta V_c & |z| > \dfrac{w}{2} \end{cases} \tag{32}$$

where $\delta V_c(\mathrm{eV}) = 1.247 \times \mathrm{x}$ is the difference in the effective potential between the GaAs and the GaAlAs. With this potential, the equation can be separated into one dimensional equations for spatial variables perpendicular to and parallel to the quantum well. The constants of separation are taken to be E_x, E_y and E_z, and $E_x + E_y + E_z$ is the energy of the wavefunction, with the zero taken to be the GaAs conduction band edge. The wavefunctions in this case are given by

$$\Psi(\vec{r}) = Au(\vec{r})e^{i\vec{k}_\rho \cdot \vec{\rho}}\phi(z) \tag{33}$$

where $\vec{\rho} = x\hat{x} + y\hat{y}$ is the position vector in the plane of the quantum well, $\vec{k}_\rho = k_x\hat{x} + k_y\hat{y}$ is the wavevector in the plane, and $\phi(z)$ is the solution for the simple particle over a finite potential square well with an energy (separation constant) of E_z, normalized so that $\int_{-\ell/2}^{\ell/2} \phi^*(z)\phi(z)dz = 1$. The wavefunctions are normalized such that $\int_{V_o} \Psi^*(\vec{r})\Psi(\vec{r})dz = 1$, and the normalization constant is $A = \mathcal{A}^{-1/2}$, where $\mathcal{A}$ is the area of the quantum well structure, $E_x = \frac{\hbar^2 k_x^2}{2m_e}$, and $E_y = \frac{\hbar^2 k_y^2}{2m_e}$.

In the case of a state which is not bound to the quantum well layer, with a separation constant E_z in the $\hat{z}$ direction corresponding to an energy greater than the depth of the quantum well, the solutions are taken to be

$$\phi(z) = \begin{cases} \phi_I(z) = A_I e^{ik_I z} + A_I' e^{-ik_I z} & z < -\dfrac{w}{2} \\ \phi_{II}(z) = A_{II}\left(e^{ik_{II}z} \pm e^{-ik_{II}z}\right) & |z| < \dfrac{w}{2} \\ \phi_{III}(z) = \pm A_I' e^{ik_I z} \pm A_I e^{-ik_I z} & z > \dfrac{w}{2} \end{cases} \tag{34}$$

where k_I is the magnitude of the $\hat{z}$ wavevector in the $\mathrm{Ga_{1-x}Al_xAs}$ and $k_{II} = \left(\left(\frac{2m_e V_c}{\hbar^2}\right) + k_I^2\right)^{1/2}$ is the magnitude of the $\hat{z}$ wavevector in the GaAs. The z origin is taken to be at the center of the quantum well, and we have chosen the symmetric and antisymmetric solutions, the upper sign corresponding to the symmetric solutions, and the lower sign corresponding to the antisymmetric solutions. This choice results in solutions which are similar to the wavefunctions of a homogeneous crystal far from the quantum well, but rather than having independent solutions for $\pm k$, there are an odd set of solutions and an even set of solutions, which must be explicitly summed over. The boundary conditions at $\pm\frac{w}{2}$ for the function and its derivative, and the normalization

$$\int_{-\ell/2}^{+\ell/2} \phi_n^*(z)\phi_m(z)dz = \delta(m,n) \tag{35}$$

where ℓ is the dimension of the integration volume perpendicular to the quantum well, serve to fix the values of the constants A_I, A_I', and A_{II} to within a phase.

The wavefunctions corresponding to a separation constant E_z less than V, are

$$\phi(z) = \begin{cases} \phi_I(z) = \pm A_I e^{-\kappa|z|} & |z| > \frac{w}{2} \\ \phi_{II}(z) = A_{II}\left(e^{ik_{II}z} \pm e^{-ik_{II}z}\right) & |z| < \frac{w}{2} \end{cases} \tag{36}$$

Applying continuity of the function and its derivative at the interfaces,

$$\begin{cases} k_{II}w\tan(k_{II}\frac{w}{2}) = \kappa w & \text{even solutions} \\ k_{II}w\tan(k_{II}\frac{w}{2}) = \kappa w & \text{odd solutions} \end{cases} \tag{37}$$

and $\kappa^2 + k_{II}^2 = \frac{2mV_c}{\hbar^2}$. There are n solutions to these equations, where

$$n = \text{int}\sqrt{\frac{2m_e V w^2}{\hbar^2\pi^2}} \tag{38}$$

For typical values of $w \sim 100$ Å and $V \sim .2$eV, only one bound state exists. In the calculations which follow, we will assume that the quantum well is thin enough so that only one bound state exists.

If the bound eigenstate is tightly confined to the low bandgap layer, the wavefunctions can for some purposes be considered to be approximately those of an infinite quantum well with the same width. The energies of the eigenstates of an infinite potential well are

$$E_n = n^2\frac{\pi^2\hbar^2}{2m_e w^2} \tag{39}$$

for $w = 100$ Å in GaAs, $E_1 \sim .06eV$, reasonably tightly bound for a well depth of $\sim .2$ eV. The exponential length in the barrier regions for the finite potential is ~ 67 Å. The amplitude of the wavefunction is considerably smaller in the barrier region as well, leading to the conclusion that the evanescent tails can be neglected for quantum well widths ~ 100 Å.

In the following sections, we will need to calculate matrix elements of the form $\langle \psi_1 | \tilde{H} | \psi_2 \rangle$, where one of the eigenstates is that of a wavefunction confined to the quantum well. The values of the coefficients A_I,A_I' of the unbound eigenstate are not needed for the calculation of the interaction matrix elements which involve an integration over both a bound and an unbound eigenstate because the integral is zero outside of the quantum well. However, the magnitude of A_{II} is needed. The equation from continuity at $\frac{w}{2}$ is

$$\pm A_I e^{-ik_I \frac{w}{2}} \pm A_I' e^{+ik_I \frac{w}{2}} = A_{II} e^{-ik_{II} \frac{w}{2}} \pm A_{II} e^{+ik_{II} \frac{w}{2}} \tag{40}$$

and the equation for continuity of the derivative at $\frac{w}{2}$ is

$$\pm A_I e^{-ik_I \frac{w}{2}} \mp A_I' e^{+ik_I \frac{w}{2}} = A_{II} (k_{II}/k_I) e^{-ik_{II} \frac{w}{2}} \mp A_{II} (k_{II}/k_I) e^{+ik_{II} \frac{w}{2}} \tag{41}$$

The boundary conditions at $-\frac{w}{2}$ are handled automatically by the symmetry (or antisymmetry) which we have imposed. Resolving these two equations yields

$$\frac{|A_{II}|^2}{|A_I|^2 + |A_I'|^2} = \left\{ 1 + \left(\frac{k_I}{k_{II}} \right)^2 \pm \left[1 - \left(\frac{k_I}{k_{II}} \right)^2 \right] \cos(k_{II} w) \right\} \tag{42}$$

Noting that the majority of the contribution to the normalization integral comes far from the quantum well, $\{|A_I|^2 + |A_I'|^2\} \cdot \ell = 1$.

$$|A_{II}|^2 = 1/\ell \left\{ 1 + \left(\frac{k_I}{k_{II}} \right)^2 \pm \left[1 - \left(\frac{k_I}{k_{II}} \right)^2 \right] \cos(k_{II} w) \right\} \tag{43}$$

and since we will want to index the functions by k_I, we substitute in $k_{II}^2 = \left(\left(\frac{2m_e V_c}{\hbar^2} \right) + k_I^2 \right)$.

$$|A_{II}|^2 = 1/\ell \Bigg\{ 1 + \left(\frac{k_I}{\left(\left(\frac{2m_e V_c}{\hbar^2} \right) + k_I^2 \right)} \right) \pm \left[1 - \left(\frac{k_I}{\left(\left(\frac{2m_e V_c}{\hbar^2} \right) + k_I^2 \right)} \right) \right]$$

$$\cos \left(\left(\left(\frac{2m_e V_c}{\hbar^2} \right) + k_I^2 \right)^{1/2} w \right) \Bigg\} \tag{44}$$

The bound wavefunctions are also normalized so that $\int_{-\ell/2}^{\ell/2} \phi_b^*(z)\phi_b(z)dz = 1$, and they are taken to be those of the infinite potential well,

$$\phi_b(z) = \begin{cases} 0 & |z| > \frac{w}{2} \\ A_b cos(z\pi/w) & |z| < \frac{w}{2} \end{cases} \tag{45}$$

and $A_b = (w/2)^{-1/2}$.

8.6 Comment on wavefunction approximations

The analysis of the devices in this chapter use the approximation that the quantum wells are isolated and that effects due to the presence of multiple quantum wells add incoherently. This approximation is justified in the case that the individual quantum wells are separated by a sufficiently large distance.

The optical phonon interaction reduces the lifetime of the momentum state $|\vec{k}\rangle$ of an electron with an energy in excess of the optical phonon energy to approximately 10^{-13} seconds. The finite lifetime of the state causes broadening of the energy of the state by $\Delta E \sim \hbar/\Delta t$, approximately 6 meV. For an electron propagating in a homogeneous crystal, the energy uncertainty can be related to an uncertainty in the wavevector, $\Delta|\vec{k}| = \Delta E \left(\frac{m_e}{2\hbar^2 E}\right)^{1/2}$. If a packet state is constructed with a position uncertainty limited to Δx, the uncertainty in the wavevector can be held to $\Delta|\vec{k}| = 1/\Delta x$. The spatial extent of packets which fall within the uncertainty caused by the optical phonon interaction is given by $\Delta x = \left(\frac{2E}{m_e}\right)^{1/2} \Delta t$. Since the electron group velocity is $v_g = \frac{\hbar|\vec{k}|}{m_e}\Delta t$, we have $\Delta x = v_g \Delta t$, which has a simple intuitive interpretation: on the average, when the electron has gone a distance Δx it undergoes a collision, which destroys its coherence. This coherence length varies from 140Å for gamma point electrons at .29 eV(=$E_{\Gamma L}$), down to 52Å for gamma point electrons just above the optical phonon energy. The local electronic wavefunctions are insensitive to conditions outside this range, and the hot electrons behave as they would over an isolated quantum well.

An additional consideration that must be made is that of the current which is due to tunneling between the adjacent quantum wells. This leakage current can be estimated by using the Kronig-Penny model for the electronic states within the superlattice. Taking the $\hat{z}$ separation constant to be E_z, the wavefunctions within the lower bandgap layers are $\psi(z) = Ae^{iKz} + Be^{-iKz}$, where $E_z = \frac{\hbar^2 K^2}{2m}$. Within the barrier regions, the wavefunctions are $\psi(z) = Ce^{i\kappa z} + Be^{-i\kappa z}$, where $\delta V_c - E_z = \frac{\hbar^2\kappa^2}{2m}$ and δV_c is the conduction band edge offset. Satisfying the boundary conditions at each of the interfaces, and using the Bloch theorem, we find the relation[13]

$$\left[\frac{(\kappa^2 - K^2)}{2\kappa K}\right] \sinh(\kappa b)\sin(Kw) + \cosh(\kappa b)\cos(Kw) = \cos[k(w+b)] \qquad (46)$$

where b is the width of the barrier between the quantum wells, and kz is the phase change of the Bloch wavefunction, where $\psi_k(z) = U(z)e^{ikz}$, and $U(\vec{r})$ is periodic over one period of the superlattice. These form an implicit relation for $E(k)$ in the subbands formed by the superlattice.

Since κb is large, $\sinh(\kappa b)$ and $\cosh(\kappa b)$ can can be simplified to $\frac{1}{2}e^{(\kappa b)}$.

$$\frac{1}{2}e^{(\kappa b)}\left\{\left[\frac{(\kappa^2 - K^2)}{2\kappa K}\right]\sin(Kw) + \cos(Kw)\right\} = \cos[k(w+b)] \qquad (47)$$

states with real k wavevectors only exist when the left hand side has magnitude smaller than one, so the term $\left\{\left[\frac{(\kappa^2-K^2)}{2\kappa K}\right]\sin(Kw) + \cos(Kw)\right\}$ can be expanded about $E = E_o$, where the expression is zero at E_o, which is the energy of a state in an isolated quantum well.

$$E \sim E_o + E_o^{\frac{1}{2}} e^{(-\kappa_o b)} w^{-1} \left(\frac{\hbar^2}{2m_e}\right)^{\frac{1}{2}} cos[k(w+b)] \qquad (48)$$

taking the second derivitive with respect to k,

$$\frac{d^2E}{dk^2} = -E_o^{\frac{1}{2}} e^{(-\kappa_o b)} w^{-1} \left(\frac{\hbar^2}{2m_e}\right)^{\frac{1}{2}} (w+b)^2 \cos[k(w+b)] \qquad (49)$$

The effective mass in the band is $m^* = \hbar^2 \left[\frac{d^2E}{dk^2}\right]^{-1}$, the mobility is given by $\mu = \frac{q}{m^*}\langle\tau_m\rangle$, where $\langle\tau_m\rangle$ is the momentum relaxation lifetime, and the current density j is $j = nq\mu\mathcal{E}$, so we have

$$j = nq^2\langle\tau_m\rangle\mathcal{E}\,2E_o^{\frac{1}{2}}e^{(-\kappa_o b)}\frac{(w+b)^2}{\hbar sqrt2m_e w} \tag{50}$$

where n is the overall doping density of the superlattice. Evaluating this expression for a device with a barrier region of 300 Å, a quantum well width of 80 Å, doping density of 1×10^{18}, and a confinement energy difference of .2 eV, the leakage current comes to $\sim$ 18 A/cm^2, or 180 μA for a typical device area of 1000 μm^2. The leakage current falls exponentially with the barrier width, and it is down several orders of magnitude for $b = 500$ Å.

8.7 Charge shielding due to a degenerate fermi gas of electrons in a quantum well

The field from a point charge in an electron gas is partially shielded by movements of the charge carriers around it, which effectively changes the interaction potential of an electron-electron collision. The induced charge density in the surrounding area is a function of the potential, $\rho = q[-n(\phi) + N_d]$ where $n(\phi)$ is the density of carriers as a function of the potential and N_d is the donor density. Following (^{14}), the potential is then determined by the Poisson equation

$$\nabla^2\phi(r) = \left(\frac{1}{r}\right)\frac{d^2\left(r\phi(r)\right)}{dr^2} = -\frac{\rho}{\epsilon} \tag{51}$$

where spherical symmetry about the charge has been assumed. Expanding about $\phi = 0$,

$$n(\phi) = n(\phi = 0) + \frac{\partial n(\phi = 0)}{\partial\phi}\Big|_{\phi=0} \times \delta\phi \tag{52}$$

Assuming the quasi-neutrality, $n(\phi = 0) = N_d$ and

$$\rho(r) = -q\phi\left[\frac{\partial n(\phi)}{\partial\phi}\Big|_{\phi=0}\right] \tag{53}$$

With the boundary condition $\phi \to 0$ as $\vec{r} \to \infty$, we have from equation 51:

$$r\phi = Ce^{-r\beta} \tag{54}$$

where

$$\beta^2 \equiv \frac{q}{\epsilon}\left[\frac{\partial n(\phi)}{\partial \phi}\Big|_{\phi=0}\right] \tag{55}$$

For $r \to 0$, $\phi(r) \to \frac{q}{4\pi\epsilon r}$ so $C = \frac{q}{4\pi\epsilon}$ and

$$\phi(r) = \frac{q}{4\pi\epsilon r}e^{-r\beta} \tag{56}$$

yielding a scattering potential of the Yukawa form, with a screening length of $1/\beta$.

To find the screening length β, we need the charge density as a function of the local potential. For the bulk material, $n(\vec{r})$ is found by summing the occupancy of all the states in the conduction band.

$$n(\phi) = \frac{2}{\pi^2}\int_0^\infty \frac{k^2}{e^{(E(\vec{k})-\mu-q\phi)/K_bT}+1}dk \tag{57}$$

where μ is the chemical potential, which can be found by numerically inverting this expression for n. Taking the derivative with respect to ϕ,

$$\beta^2 = \frac{q}{\epsilon}\frac{dn(\phi)}{d\phi}\Big|_{\phi=0} = \frac{q^2}{\epsilon K_bT\pi^2}\int \frac{k^2 e^{(E(\vec{k})-\mu)/K_bT}}{(e^{(E-\mu)/K_bT)}+1)^2}dk \tag{58}$$

Within the quantum well layer, the eigenstates can be approximated by the states of an infinite barrier well which are given by

$$\langle\vec{r}|\Psi\rangle = (\frac{2}{wA})^{1/2}u(\vec{r})e^{i\vec{k}_\rho\cdot\vec{\rho}}\sin(\frac{j\pi z}{w}) \tag{59}$$

where the z origin has been taken at the edge of the quantum well, A is the area of the quantum well, $\rho = x\hat{x} + y\hat{y}$ is the position vector in the plane of the quantum well and j is an integer. The energies of the eigenstates are

$$E_j = \frac{\hbar^2}{2m_e}[|\vec{k}_\rho|^2 + (\frac{j\pi}{w})^2] \tag{60}$$

n is again found by summing over the occupancy of the states. The sums in the $\hat{x}$ and $\hat{y}$ directions are converted into integrals, but the sum corresponding to the $\hat{z}$ direction remains discrete.

$$n = \frac{2}{\pi w} \int_{k_\rho} \sum_j \frac{k_\rho}{e^{\left(\frac{\hbar^2}{2m}[k_\rho^2 + (\frac{j\pi}{w})^2] - \mu - q\phi\right)/K_b T} + 1} dk_\rho \tag{61}$$

Taking the derivative with respect to ϕ at $\phi = 0$,

$$\beta^2 = \frac{2q^2}{\pi w \epsilon} \int_0^\infty \sum_j \frac{k_\rho \frac{2}{K_b T} e^{\left[\frac{\hbar^2}{2m}[k_\rho^2 + (\frac{j\pi}{w})^2] - \mu\right]/K_b T}}{\left\{e^{\left[\frac{\hbar^2}{2m}[k_\rho^2 + (\frac{j\pi}{w})^2] - \mu\right]/K_b T} + 1\right\}^2} dk_\rho \tag{62}$$

If the electron gas is very degenerate, this expression for β reduces to

$$\beta^2 = \frac{q^2}{\epsilon \pi w} j_{\text{filled}} \frac{2m_e}{\hbar^2} \tag{63}$$

where j_{filled} is the quantum number of the highest transverse state which has started to fill. For the electron densities and quantum well dimensions that we are interested in here, the screening length is approximately 50Å.

8.8 Electron-electron scattering cross section for a hot electron over a degenerate electron gas confined to a quantum well

We are now ready to calculate the interaction rate for electron-electron collisions over the quantum well. The transition rate from state $|\Psi_1\rangle|\Psi_2\rangle$ to state $|\Psi_1'\rangle|\Psi_2'\rangle$ can be approximated using Fermi's golden rule

$$W_{i \to f} = \frac{2\pi}{\hbar} |\tilde{H}_{if}'|^2 \delta(E_f - E_i) \tag{64}$$

The perturbation Hamiltonian is that due to the screened Coulomb potential

$$V(r) = \frac{q e^{-\beta|\vec{r}_1 - \vec{r}_2|}}{4\pi\epsilon|\vec{r}_1 - \vec{r}_2|} \tag{65}$$

where $1/\beta$ is the screening length calculated in section 8.7. The matrix elements of the perturbation Hamiltonian are given by

$$\begin{aligned}
&\langle\Psi_{2'}|\langle\Psi_{1'}|\tilde{H}'_{if}|\Psi_1\rangle|\Psi_2\rangle = \\
&\int_{r_1}\int_{r_2} A^*_{2'}u^*_{2'}(\vec{r}_2)e^{-i\vec{k}_{\rho 2'}\cdot\vec{\rho}_2}\phi^*_{2'}(z_2)A^*_{1'}u^*_{1'}(\vec{r}_1)e^{-i\vec{k}_{\rho 1'}\cdot\vec{\rho}_1}\phi^*_{1'}(z_1)\frac{q^2 e^{-\beta|\vec{r}_1-\vec{r}_2|}}{4\pi\epsilon|\vec{r}_1-\vec{r}_2|}\times \\
&A_1 u_1(\vec{r}_1)e^{i\vec{k}_{\rho 1}\cdot\vec{\rho}_1}\phi_1(z_1)A_2 u_2(\vec{r}_2)e^{i\vec{k}_{\rho 2}\cdot\vec{\rho}_2}\phi_2(z_2)d^3r_1 d^3r_2
\end{aligned} \tag{66}$$

where ρ_1 and ρ_2 are the components of r_1 and r_2 in the plane of the quantum well, and z_1 and z_2 are the components perpendicular to the quantum well. From section 8.5, we have $|A^*_{1'}A^*_{2'}A_1A_2| = \frac{1}{\mathcal{A}^2}$. The rapidly varying parts of the wavefunctions $u(\vec{r})$ are approximately periodic over a unit cell, and the other terms of the integrand vary slowly over a unit cell. Separating along the lines discussed earlier,

$$\begin{aligned}
&\langle\Psi_{2'}|\langle\Psi_{1'}|\tilde{H}'_{if}|\Psi_1\rangle|\Psi_2\rangle = \\
&\left\{\int_r \frac{1}{V_o}u^*_{1'}(\vec{r})u_1(\vec{r})d^3r\right\}\left\{\int_r \frac{1}{V_o}u^*_{2'}(\vec{r})u_2(\vec{r})d^3r\right\}\left\{\frac{q^2}{4\pi\epsilon}\right\}\times \\
&\frac{1}{\mathcal{A}^2}\int_{r_1}\int_{r_2} e^{-i\vec{k}_{\rho 2'}\cdot\vec{\rho}_2}\phi^*_{2'}(z_2)e^{-i\vec{k}_{\rho 1'}\cdot\vec{\rho}_1}\phi^*_{1'}(z_1)\times \\
&\frac{e^{-\beta|\vec{r}_1-\vec{r}_2|}}{|\vec{r}_1-\vec{r}_2|}e^{i\vec{k}_{\rho 1}\cdot\vec{\rho}_1}\phi_1(z_1)e^{i\vec{k}_{\rho 2}\cdot\vec{\rho}_2}\phi_2(z_2)d^3r_1 d^3r_2
\end{aligned} \tag{67}$$

Making the substitution $\vec{r}' = \vec{r}_1 - \vec{r}_2$ in the r_1 integration,

$$\begin{aligned}
&\langle\Psi_{2'}|\langle\Psi_{1'}|\tilde{H}'_{if}|\Psi_1\rangle|\Psi_2\rangle = \\
&\left\{\int_r \frac{1}{V_o}u^*_{1'}(\vec{r})u_1(\vec{r})d^3r\right\}\left\{\int_r \frac{1}{V_o}u^*_{2'}(\vec{r})u_2(\vec{r})d^3r\right\}\left\{\frac{q^2}{4\pi\epsilon}\right\}\times \\
&\frac{1}{\mathcal{A}^2}\int_{r'}\int_{r_2} e^{-i\vec{k}_{\rho 2'}\cdot\vec{\rho}_2}\phi^*_{2'}(z_2)e^{-i\vec{k}_{\rho 1'}\cdot(\vec{\rho}'+\rho_2)}\phi^*_{1'}(z'+z_2)\frac{e^{-\beta|\vec{r}'|}}{|\vec{r}'|}\times \\
&e^{i\vec{k}_{\rho 1}\cdot(\vec{\rho}'+\rho_2)}\phi_1(z'+z_2)e^{i\vec{k}_{\rho 2}\cdot\vec{\rho}_2}\phi_2(z_2)d^3r' d^3r_2
\end{aligned} \tag{68}$$

$$\begin{gathered}
\langle\Psi_{2'}|\langle\Psi_{1'}|\tilde{H}'_{if}|\Psi_1\rangle|\Psi_2\rangle = \\
\left\{\int\limits_r \frac{1}{V_o} u^*_{1'}(\vec{r})u_1(\vec{r})d^3r\right\}\left\{\int\limits_r \frac{1}{V_o} u^*_{2'}(\vec{r})u_2(\vec{r})d^3r\right\}\left\{\frac{q^2}{4\pi\epsilon}\right\}\times \\
\frac{1}{\mathcal{A}^2}\int\limits_{\rho_2} e^{i\left[-\vec{k}_{\rho 2'}-\vec{k}_{\rho 1'}+\vec{k}_{\rho 1}+\vec{k}_{\rho 2}\right]\cdot\vec{\rho}_2} d^2\rho_2 \times \\
\int\limits_{\rho'}\int\limits_{z'}\int\limits_{z_2} e^{-i\left[\vec{k}_{\rho 1'}+\vec{k}_{\rho 1}\right]\cdot(\vec{\rho}')}\phi^*_{2'}(z_2)\phi^*_{1'}(z'+z_2) \\
\frac{e^{-\beta|\vec{r}'|}}{|\vec{r}'|}\phi_1(z'+z_2)\phi_2(z_2)d^2\rho' dz' dz_2
\end{gathered} \tag{69}$$

The integral over the exponential term yields a pair of Kronecker delta functions $\delta(k_{x2'}-k_{x1'}, k_{x1}+k_{x2})$ and $\delta(k_{y2'}-k_{y1'}, k_{y1}+k_{y2})$. The last integral can be simplified in the case that the quantum well is heavily doped so that the screening length given by equation 8.62 is small compared to the width of the quantum well. Since the integral over z_2 is small outside of $|z_2| < \frac{w}{2}$ due to the presence of the bound state ψ_1, the wavefunctions in the integrand can be taken to possess the same functional form that they have inside the quantum well. The extent of the error which is introduced is the extent of the protrusion of the interaction potential of a bound electron beyond the edge of the quantum well into the region where the wavefunctions used do not have the proper form.

$$\begin{gathered}
\langle\Psi_{2'}|\langle\Psi_{1'}|\tilde{H}'_{if}|\Psi_1\rangle|\Psi_2\rangle = \\
\left\{\int\limits_r \frac{1}{V_o} u^*_{1'}(\vec{r})u_1(\vec{r})d^3r\right\}\left\{\int\limits_r \frac{1}{V_o} u^*_{2'}(\vec{r})u_2(\vec{r})d^3r\right\}\left\{\frac{q^2}{4\pi\epsilon}\right\} \\
\mathcal{A}^{-1}\delta(\vec{k}_{\rho 2'}+\vec{k}_{\rho 1'}=\vec{k}_{\rho 1}+\vec{k}_{\rho 2})\int\limits_{\rho'}\int\limits_{z'}\int\limits_{-\frac{w}{2}}^{\frac{w}{2}}\left\{e^{-i\left[\vec{k}_{\rho 1'}\pm\vec{k}_{\rho 1}\right]\cdot(\vec{\rho}')}\times\right. \\
\left\{A_{\mathrm{II}}(1)\left(e^{-ik_{1\mathrm{II}}(z'+z_2)}\pm e^{ik_{1\mathrm{II}}(z'+z_2)}\right)\right\} \\
\left\{A_{\mathrm{II}}(2)\left(e^{-ik_{2\mathrm{II}}z_2}+e^{ik_{2\mathrm{II}}z_2}\right)\right\}\times
\end{gathered} \tag{70}$$

$$\frac{e^{-\beta|\vec{r}'|}}{|\vec{r}'|}\left\{A^*_{\rm II}(1')\left(e^{-ik_{1'{\rm II}}(z'+z_2)} \pm e^{ik_{1'{\rm II}}(z'+z_2)}\right)\right\} \times$$
$$\left\{A^*_{\rm II}(2')\left(e^{-ik_{2'{\rm II}}z_2} \pm e^{ik_{2'{\rm II}}z_2}\right)\right\}\Bigg\}dz_2d^2\rho'dz'$$

where $\mathcal{A}$ is the area of the quantum well. The ± signs come from the symmetry of each of the wavefunctions. Note that the state ψ_2 is the symmetrical confined state. The integral over $\vec{r}_2$ can now be carried out. The integral

$$\int e^{i\vec{k}\cdot\vec{r}}\frac{e^{-\beta|r|}}{|r|}d^3r = \frac{4\pi}{|\beta|^2+|\vec{k}|^2} \tag{71}$$

is the Fourier transform of the Yukawa potential, which is evaluated, for example, in reference ([15]). Introducing a special notation to help keep track of the many terms in the equation, let the variables ν_1, $\nu_{1'}$, and $\nu_{2'}$ take on the values $(0,1)$ according the values of the sign for the symmetry of each of the wavefunctions ψ_1, $\psi_{1'}$, and $\psi_{2'}$, 0 for even symmetry and 1 for odd symmetry. Additionally, four dummy variables ν'_1, $\nu'_{1'}$, ν'_2, and $\nu'_{2'}$ take on the values $(0,1)$, converting the ennumerated terms into a compact notation. Defining $K_{\rm II} = (-1)^{\nu'_1}k_{1{\rm II}} + (-1)^{\nu'_2}k_{2{\rm II}} + (-1)^{\nu'_{1'}}k_{1'{\rm II}} + (-1)^{\nu'_{2'}}k_{2'{\rm II}}$,

$$\begin{gathered}\langle\Psi_{2'}|\langle\Psi_{1'}|\tilde{H}'_{if}|\Psi_1\rangle|\Psi_2\rangle = \\ \left\{\int_{r_1}\frac{1}{V_o}u^*_{1'}(\vec{r})u_1(\vec{r})d^3r\right\}\left\{\int_{r}\frac{1}{V_o}u^*_{2'}(\vec{r})u_2(\vec{r})d^3r\right\}\times \\ \left\{\frac{q^2}{4\pi\epsilon}\right\}\mathcal{A}^{-1}\delta(\vec{k}_{\rho2'}+\vec{k}_{\rho1'}=\vec{k}_{\rho1}+\vec{k}_{\rho2})\times \\ \left[A^*_{\rm II}(2')A^*_{\rm II}(1')A_{\rm II}(2)A_{\rm II}(1)\right]\frac{4\pi}{\beta^2+|\vec{k}_1-\vec{k}_{1'}|^2}\times \\ \sum_{\substack{\nu'_1=0,1\\ \nu'_{1'}=0,1\\ \nu'_{2'}=0,1\\ \nu'_2=0,1}}(-1)^{[\nu_1\nu'_1+\nu_{1'}\nu'_{1'}+\nu_{2'}\nu'_{2'}]}\frac{\sin(\frac{w}{2}K_{\rm II})}{K_{\rm II}}\end{gathered} \tag{72}$$

So the transition rate from state $|\Psi_1\Psi_2\rangle$ to state $|\Psi_{1'}\Psi_{2'}\rangle$ is

$$\begin{gathered}
R_{1,2\rightarrow 1',2'} = \xi_{1,1'}\xi_{2,2'}\frac{2\pi}{\hbar}\delta(E_f - E_i)\times \\
\left\{\frac{q^2}{4\pi\epsilon}\right\}^2 A^{-2}\delta(\vec{k}_{\rho 2'} + \vec{k}_{\rho 1'} = \vec{k}_{\rho 1} + \vec{k}_{\rho 2}) \times \\
\left|\left[A^*_{\mathrm{II}}(2')A^*_{\mathrm{II}}(1')A_{\mathrm{II}}(2)A_{\mathrm{II}}(1)\right]\right|^2 \left|\frac{4\pi}{\beta^2 + |\vec{k}_1 - \vec{k}_{1'}|^2}\right|^2 \times \\
\left|\sum_{\substack{\nu'_1=0,1\\ \nu'_{1'}=0,1\\ \nu'_{2'}=0,1\\ \nu'_2=0,1}} (-1)^{[\nu_1\nu'_1+\nu_{1'}\nu'_{1'}+\nu_{2'}\nu'_{2'}]}\frac{\sin(\frac{w}{2}K_{\mathrm{II}})}{K_{\mathrm{II}}}\right|^2
\end{gathered} \tag{73}$$

where $\xi_{1,1'}$ and $\xi_{2,2'}$ are the overlap integrals for electron 1 and 2, respectively.

To find the total scattering rate for the initial electron from state $|\Psi_1\rangle$ to state $|\Psi_{1'}\rangle$, we sum over the possible initial and final states of the secondary electron.

$$R_{1\rightarrow 1'} = \sum_{2(\text{occupied})}\ \sum_{2'(\text{empty})} R_{1,2\rightarrow 1',2'} \tag{74}$$

In the case we wish to consider, the secondary electrons are in a degenerate Fermi gas, confined to the quantum well. The occupancy of the states above the quantum well is negligible, and the sum over the initial, occupied states is taken over the two dimensional states. The sum over the final states is separated into two parts, corresponding to final states which are energetically confined to the quantum well, and to those which correspond to unconfined electrons. Note that there is a distinction here between electrons which are energetically confined to the quantum well layer and electrons which are in states which are confined to the quantum well but whose energy is greater than the discontinuity in the conduction band, and can therefore scatter again and escape the quantum well layer. All of the electrons which end up in 3-D states are energetically untrapped, but electrons in 2-D states can be either trapped or not depending on their wavevector. In addition to the spatial states, we must also sum over the spin states of the scattered electron. The initial spin gives a factor of 2, but there is also another potential factor of 2, due to a possible spin exchange. The Kronecker delta function $\delta(\vec{k}_{\rho 2'} + \vec{k}_{\rho 1'} = \vec{k}_{\rho 1} + \vec{k}_{\rho 2})$ makes all of the terms in the sum over the initial states zero except for the term where $\vec{k}_{\rho 2} = \vec{k}_{\rho 2'} + \vec{k}_{\rho 1'} - \vec{k}_{\rho 1}$. Converting the sum over the final states into integrals and substituting in $R_{1,2\rightarrow 1',2'}$, the scattering rate into states where the secondary electron is unconfined is given by:

$$R_{1\rightarrow 1'[\text{confined}]} = \left\{ \int_{E(k_{\rho 2'} < \delta V_c} \int \frac{2\mathcal{A}}{(2\pi)^2} \xi_{1,1'} \xi_{2,2'} \frac{2\pi}{\hbar} \delta(E_f - E_i) \left\{ \frac{q^2}{4\pi\epsilon} \right\}^2 \times \right.$$

$$\mathcal{A}^{-2} \left| \left[A^*_{\text{II}}(2') A^*_{\text{II}}(1') A_{\text{II}}(2) A_{\text{II}}(1) \right] \right|^2 \left| \frac{4\pi}{\beta^2 + |\vec{k}_1 - \vec{k}_{1'}|^2} \right|^2 \times \tag{75}$$

$$f(\vec{k}_{\rho 2'} + \vec{k}_{\rho 1'} - \vec{k}_{\rho 1}) \left[1 - f(k_{\rho 2'}) \right]$$

$$\left. \left| \sum_{\substack{\nu'_1 = 0.1 \\ \nu'_{1'} = 0,1 \\ \nu'_{2'} = 0,1 \\ \nu'_2 = 0,1}} (-1)^{[\nu_1 \nu'_1 + \nu_{1'} \nu'_{1'} + \nu_{2'} \nu'_{2'}]} \frac{\sin(\frac{w}{2} K_{\text{II}})}{K_{\text{II}}} \right|^2 dk_{2'x} dk_{2'y} \right\}$$

And the scattering rate into states where the secondary electron is energetically confined to the quantum well layer is given by

$$R_{1\rightarrow 1'[\text{unconfined}]} = \left\{ \int_{E(k_{\rho 2'} > \delta V_c} \int \frac{2\mathcal{A}}{(2\pi)^2} \xi_{1,1'} \xi_{2,2'} \frac{2\pi}{\hbar} \delta(E_f - E_i) \left\{ \frac{q^2}{4\pi\epsilon} \right\}^2 \times \right.$$

$$\mathcal{A}^{-2} \left| \left[A^*_{\text{II}}(2') A^*_{\text{II}}(1') A_{\text{II}}(2) A_{\text{II}}(1) \right] \right|^2 \left| \frac{4\pi}{\beta^2 + |\vec{k}_1 - \vec{k}_{1'}|^2} \right|^2 \times$$

$$f(\vec{k}_{\rho 2'} + \vec{k}_{\rho 1'} - \vec{k}_{\rho 1}) \left[1 - f(k_{\rho 2'}) \right]$$

$$\left|\sum_{\substack{\nu'_1=0,1\\ \nu'_{1'}=0,1\\ \nu'_{2'}=0,1\\ \nu'_2=0,1}} (-1)^{[\nu_1\nu'_1+\nu_{1'}\nu'_{1'}+\nu_{2'}\nu'_{2'}]}\frac{\sin(\frac{w}{2}K_{\mathrm{II}})}{K_{\mathrm{II}}}\right|^2 dk_{2'x}dk_{2'y}\Bigg\}+$$

$$\left\{\sum_{\nu_{2'}=0,1}\int_{k_{2'x}}\int_{k_{2'y}}\int_{0\atop k_{2'z}}^{\infty} \frac{2V_o}{(2\pi)^3}\xi_{1,1'}\xi_{2,2'}\frac{2\pi}{\hbar}\delta(E_f-E_i)\left\{\frac{q^2}{4\pi\epsilon}\right\}^2 A^{-2}\times\right. \tag{76}$$

$$\left|\left[A^*_{\mathrm{II}}(2')A^*_{\mathrm{II}}(1')A_{\mathrm{II}}(2)A_{\mathrm{II}}(1)\right]\right|^2\left|\frac{4\pi}{\beta^2+|\vec{k}_1-\vec{k}_{1'}|^2}\right|^2\times$$

$$\left.\left|\sum_{\substack{\nu'_1=0,1\\ \nu'_{1'}=0,1\\ \nu'_{2'}=0,1\\ \nu'_2=0,1}} (-1)^{[\nu_1\nu'_1+\nu_{1'}\nu'_{1'}+\nu_{2'}\nu'_{2'}]}\frac{\sin(\frac{w}{2}K_{\mathrm{II}})}{K_{\mathrm{II}}}\right|^2 f(\vec{k}_{\rho 2'}+\vec{k}_{\rho 1'}-\vec{k}_{\rho 1})dk_{2'x}dk_{2'y}dk_{2'z}\right\}$$

Defining:

$$\mathrm{I}_1 \equiv \frac{\hbar^2}{2\mathrm{m_e}}\int\int_{\mathrm{E}(\mathrm{k}_{\rho 2'}<\delta \mathrm{V_c}} \delta(\mathrm{E_f}-\mathrm{E_i})\mathrm{f}(\vec{\mathrm{k}}_{\rho 2'}+\vec{\mathrm{k}}_{\rho 1'}-\vec{\mathrm{k}}_{\rho 1})\left[1-\mathrm{f}(\mathrm{k}_{\rho 2'})\right]\mathrm{dk_{2'x}dk_{2'y}} \tag{77}$$

The change in energy $E_f - E_i$ is

$$E_f - E_i = \frac{\hbar^2}{2m_e}\left\{k^2_{2'}+k^2_{1'}-k^2_2-k^2_1\right\} \tag{78}$$

and since $\vec{k}_{\rho 2}=\vec{k}_{\rho 1'}+\vec{k}_{\rho 2'}-\vec{k}_{\rho 1}$ from the delta function in $R_{1,2\rightarrow 1',2'}$,

$$E_f - E_i = \frac{\hbar^2}{2m_e}\{-2(\vec{k}_{\rho 1})^2-2\vec{k}_{\rho 1'}\cdot\vec{k}_{\rho 2'}$$

$$+2\vec{k}_{\rho 2'}\cdot\vec{k}_{\rho 1}+2\vec{k}_{\rho 1'}\cdot\vec{k}_{\rho 1}+k^2_{1'\mathrm{II}}-k^2_{1\mathrm{II}}\} \tag{79}$$

defining

$$K_1 \equiv \frac{\left[2\vec{k}_{\rho 1}\cdot(\vec{k}_{\rho 1}-\vec{k}_{\rho 1'})+k^2_{1\mathrm{II}}-k^2_{1'\mathrm{II}}\right]}{2|\vec{k}_{\rho 1}-\vec{k}_{\rho 1'}|} \tag{80}$$

and defining k_b, C_1, C_2, and C_3,

$$k_b = \left[\frac{2m_e\delta V_c}{\hbar^2} - \left(\frac{\pi}{w}\right)^2\right]^{\frac{1}{2}} \tag{81}$$

$$C_1 = \left[k_f^2 - \mathcal{K}_1^2\right] \tag{82}$$

$$C_2 = \left[k_f^2 - \left(\mathcal{K}_1 - |\vec{k}_{\rho 1} - \vec{k}_{\rho 1'}|\right)^2\right] \tag{83}$$

$$C_3 = \left[k_b^2 - \mathcal{K}_1^2\right] \tag{84}$$

and performing the integrals over $k_{2'x}$ and $k_{2'y}$, we have

$$I_1 = \frac{\max\left(0, \sqrt{\max\left(0, \left(\min\left(C_2, C_3\right)\right)\right)} - \sqrt{\max\left(0, C_1\right)}\right)}{|\vec{k}_{\rho 1} - \vec{k}_{\rho 1'}|} \tag{85}$$

and defining

$$\mathrm{I_2} \equiv \frac{\hbar^2}{2\mathrm{m_e}} \int_{\mathrm{E(k_{\rho 2'}} > \delta \mathrm{V_c}} \int \delta(\mathrm{E_f} - \mathrm{E_i})\mathrm{f}(\vec{\mathrm{k}}_{\rho 2'} + \vec{\mathrm{k}}_{\rho 1'} - \vec{\mathrm{k}}_{\rho 1}) \left[1 - \mathrm{f}(\mathrm{k}_{\rho 2'})\right] \mathrm{dk_{2'x} dk_{2'y}} \tag{86}$$

we have

$$I_2 = \frac{\max\left[0, \left(\max\left[0, C_2\right]\right)^{\frac{1}{2}} - \left(\max\left[0, C_3\right]\right)^{\frac{1}{2}}\right]}{|\vec{k}_{\rho 1} - \vec{k}_{\rho 1'}|} \tag{87}$$

similarly for the second term, we define

$$\mathrm{I_3} \equiv \frac{\hbar^2}{2\mathrm{m_e}} \int_{\mathrm{k_{2'x}}} \int_{\mathrm{k_{2'y}}} \delta(\mathrm{E_f} - \mathrm{E_i})\mathrm{f}(\vec{\mathrm{k}}_{\rho 2'} + \vec{\mathrm{k}}_{\rho 1'} - \vec{\mathrm{k}}_{\rho 1}) \mathrm{dk_{2'x} dk_{2'y}} \tag{88}$$

and the difference in energy between the initial and final states is

$$E_f - E_i = \frac{\hbar^2}{2m_e}\{-2|\vec{k}_{\rho 1}|^2 - 2\vec{k}_{\rho 1'} \cdot \vec{k}_{\rho 2'} + 2\vec{k}_{\rho 2'} \cdot \vec{k}_{\rho 1}$$

$$+2\vec{k}_{\rho 1'} \cdot \vec{k}_{\rho 1} + k_{1'\mathrm{II}}^2 - k_{1\mathrm{II}}^2 + k_{2'\mathrm{II}}^2 - (\pi/w)^2\} \tag{89}$$

and defining

$$\mathcal{K}_2 \equiv \frac{\left[2\vec{k}_{\rho 1} \cdot (\vec{k}_{\rho 1} - \vec{k}_{\rho 1'}) + k_{1\mathrm{II}}^2 - k_{1'\mathrm{II}}^2 + k_{2'\mathrm{II}}^2 - (\pi/w)^2\right]}{2|\vec{k}_{\rho 1} - \vec{k}_{\rho 1'}|} \tag{90}$$

and integrating over $k_{2'x}$ and $k_{2'y}$,

$$\mathrm{I}_3 = \begin{cases} 0 & \text{if } |K_2| > k_f \\ \dfrac{\left(k_f^2 - K_2^2\right)^{1/2}}{\left[|\vec{k}_{\rho 1'} - \vec{k}_{\rho 1}|\right]} & \text{if } |K_2| < k_f \end{cases} \tag{91}$$

Rewriting the equation for $R_{1\rightarrow 1'[\text{confined}]}$,

$$R_{1\rightarrow 1'[\text{confined}]} = \frac{m_e \pi q^4}{\hbar^3 \mathcal{A}^2 2\pi^2 \epsilon^2} \left| \frac{4\pi}{\beta^2 + |\vec{k}_1 - \vec{k}_{1'}|^2} \right|^2 \times$$

$$\left[\left\{ \mathrm{I}_1 \frac{\mathcal{A}}{(2\pi)^2} \left| \left[\mathrm{A}^*_{\mathrm{II}}(2') \mathrm{A}^*_{\mathrm{II}}(1') \mathrm{A}_{\mathrm{II}}(2) \mathrm{A}_{\mathrm{II}}(1) \right] \right|^2 \times \right.\right. \tag{92}$$

$$\left. \left| \sum_{\substack{\nu_1'=0,1 \\ \nu_{1'}'=0.1 \\ \nu_{2'}'=0,1 \\ \nu_2'=0,1}} (-1)^{[\nu_1 \nu_1' + \nu_{1'} \nu_{1'}' + \nu_{2'} \nu_{2'}']} \frac{\sin(\frac{w}{2} K_{\mathrm{II}})}{K_{\mathrm{II}}} \right|^2 \right\} +$$

and rewriting the equation for $R_{1\rightarrow 1'[\text{unconfined}]}$,

$$R_{1\rightarrow 1'[\text{unconfined}]} = \frac{m_e \pi q^4}{\hbar^3 \mathcal{A}^2 2\pi^2 \epsilon^2} \left| \frac{4\pi}{\beta^2 + |\vec{k}_1 - \vec{k}_{1'}|^2} \right|^2 \times$$

$$\left[\left\{ \mathrm{I}_2 \frac{\mathcal{A}}{(2\pi)^2} \left| \left[\mathrm{A}^*_{\mathrm{II}}(2') \mathrm{A}^*_{\mathrm{II}}(1') \mathrm{A}_{\mathrm{II}}(2) \mathrm{A}_{\mathrm{II}}(1) \right] \right|^2 \times \right.\right.$$

$$\left. \left| \sum_{\substack{\nu_1'=0.1 \\ \nu_{1'}'=0,1 \\ \nu_{2'}'=0,1 \\ \nu_2'=0.1}} (-1)^{[\nu_1 \nu_1' + \nu_{1'} \nu_{1'}' + \nu_{2'} \nu_{2'}']} \frac{\sin(\frac{w}{2} K_{\mathrm{II}})}{K_{\mathrm{II}}} \right|^2 \right\} + \tag{93}$$

$$\left\{ \sum_{\nu_{2'}=0,1} \int\limits_{\substack{0 \\ k_{2'Iz}}}^{\infty} \mathrm{I}_3 \frac{\mathrm{V}_\circ}{(2\pi)^3} \left| \left[\mathrm{A}^*_{\mathrm{II}}(2') \mathrm{A}^*_{\mathrm{II}}(1') \mathrm{A}_{\mathrm{II}}(2) \mathrm{A}_{\mathrm{II}}(1) \right] \right|^2 \right.$$

$$\left.\left. \left| \sum_{\substack{\nu_1'=0,1 \\ \nu_{1'}'=0,1 \\ \nu_{2'}'=0,1 \\ \nu_2'=0.1}} (-1)^{[\nu_1 \nu_1' + \nu_{1'} \nu_{1'}' + \nu_{2'} \nu_{2'}']} \frac{\sin(\frac{w}{2} K_{\mathrm{II}})}{K_{\mathrm{II}}} \right|^2 dk_{2'z} \right\} \right]$$

Now that we have expressions for the scattering rates from states $\vec{k}_1$ to states $\vec{k}_{1'}$ for the incoming electron, we will sum over the final states $\vec{k}_{1'}$ to arrive at a scattering rate distribution for an electron which is initially in state $\vec{k}_1$ as a function of the energy lost, for each of the two cases: secondary electron confined, or unconfined. The scattering rate from state $\vec{k}_1$ to all states which have an energy less than E is given by

$$R(\vec{k}_1, < E) = \sum_{E(\vec{k}_{1'})<E} R_{1\rightarrow 1'}$$

Converting the sum into a pair of integrals over the 2–D and 3–D states,

$$R(\vec{k}_1, < E) = \int\limits_{E(k_\rho)=E_f}^{E(k_\rho)=E} \int\limits_0^{2\pi} \frac{2A}{(2\pi)^2} R_{1\rightarrow 1'} k_\rho dk_\phi dk_\rho +$$

$$\sum_{\nu_1'=0,1} \int\limits_{k_r=0}^{E(k_r)=E} \int\limits_0^{\frac{\pi}{2}} \int\limits_0^{2\pi} \frac{2V_o}{(2\pi)^3} R_{1\rightarrow 1'} k_r^2 \sin(k_\theta) dk_\phi dk_\theta dk_r \tag{94}$$

where the second term is zero if $E_u(0) > E$, (meaning that the energy E is below that required to produce an 3–D electron). E as a function of $|k|$ is substituted into the integrals, and the expression is then differentiated, and we arrive at the distribution function for the scattering rate out of state $|\Psi_1\rangle$ as a function of energy after the scattering event.

$$R(\vec{k}_1, E)dE = \begin{cases} \int\limits_0^{2\pi} \frac{2A}{(2\pi)^2} R_{1\rightarrow 1'} dk_\phi \frac{m_e}{\hbar^2} dE & \text{if } E < \delta V_c \\ \int\limits_0^{2\pi} \frac{2A}{(2\pi)^2} R_{1\rightarrow 1'} dk_\phi \frac{m_e}{\hbar^2} dE & \\ \quad + \sum_{\nu_1'=0,1} \int\limits_0^{2\pi} \int\limits_0^{\frac{\pi}{2}} \frac{2V_o}{(2\pi)^3} R_{1\rightarrow 1'} & \text{if } E > \delta V_c \\ \quad \times \sin(k_\theta) \left(2m_e(E - \delta V_c)\right)^{\frac{1}{2}} \frac{m_e}{\hbar^3} dk_\phi dk_\theta dE & \end{cases} \tag{95}$$

where $E(k_\rho) = \frac{\hbar^2}{2m_e}\left[k_\rho^2 + (\frac{\pi}{w})^2\right]$ and $E(k_r) = \left[k_r^2\right] + \delta V_c$ where the first term accounts for transitions where the scattered electron is in a 2-D state, and the second accounts for transitions where it is in a 3-D state. Notice that the integral over θ is restricted to $0 - \pi/2$, and the even and odd symmetry states are summed over, as before.

Equations 8.95, along with equations 8.92 and 8.93 and the earlier definitions constitute a multiple integral relation for the the scattering rates as a function of the incident electron wavevector $\vec{k}$ for the three cases: a) both electrons are energetically trapped in the quantum well, b) only one electron escapes the quantum well, and c) both electrons escape the quantum well. Figures 6 and 7 show the dependence of these scattering rates as a function of the magnitude and angle of the incident k vector. (They are independent of the azimuthal angle due to symmetry). The parameters of the multiple quantum well structure were taken to be: well width 80Å, doping density, $5 \times 10^{17}/\mathrm{cm}^{-}3$, and an aluminum mole fraction of 0.15. The scattering rates peak at approximately 2×10^{14}, for an incident k vector of $5 - 6 \times 10^{8}$. The scattering rates for incident electrons with lower k vectors are smaller due to the degeneracy of the electron gas, only electrons near the fermi level can participate in the scattering events. At higher energies, the scattering rates start to drop again due to the $|\vec{k}_i - \vec{k}_f|$ factor in the denominator for coulomb scattering. Combining the results which are displayed in figure 8.6 with the transit times of equation 8.23, it is clear that electrons with k vectors above $3 - 4 \times 10^{8}\mathrm{m}^{-1}$ are inevitably scattered before they can cross the quantum well, but once they have lost sufficient energy so that their wavevector is below $\approx 2 \times 10^{8}\mathrm{m}^{-1}$ the scattering rate is relatively small and they can escape the quantum well. For this range of parameters, for electrons with an incoming distribution given by equation 8.22, the probability of exciting a secondary electron out of the quantum well in this initial collision is close to 50%, and the probability of undergoing collisions which result in capture of the electron are roughly 10%, counting electron-electron scattering processes only. Other multiple scattering processes which include emission of optical phonons will decrease the total number of electrons escaping the quantum well, but the scattering rates are low enough so that the total multiplication rate will remain on the order of 1.3-1.4 for a single quantum well. This total multiplication rate for a multiple quantum well structure containing as few as 40 wells can therefore be as high as $\sim 10^6$.

8.9 Quantum well transistor

One potential application of the low capture probability for hot electrons propagating over a quantum well is a transistor structure similar to a npn transistor with the base replaced by a quantum well. The advantages of a highly conductive, very thin base transistor are well known, and many devices have been proposed which use a highly conductive metal layer in various geometries. Several difficulties have been encountered with these types of structures, related to the difficulty of providing a high quality semiconductor layer on top of the metal base layer. In a bipolar transistor, the doping level of the base region is limited due to the need to maintain a high emitter efficiency. The emitter efficiency, the ratio of the injected electrons to the total current, is given by [16].

$$\gamma = \left[1 + \frac{n_e}{p_b}\frac{D_e}{D_b}\frac{L_b}{L_e}\tanh(W/L_b)\right]^{-1} \tag{96}$$

where n_e and p_b are the dopant densities in the emitter and the base, D_e and D_b are the diffusion constants for the minority carriers in the base and the emitter, and L_e and L_b are the diffusion lengths for electrons in the emitter and holes in the base, respectively. The current gain is

$$\beta_o = \frac{\gamma\alpha_o}{1 - \gamma\alpha_o} \tag{97}$$

where α_o is the the base transport factor, which is close to one for a narrow base conventional transistor. The emitter efficiency can be improved by using a heterostructure at the emitter-base junction, but the electron densities are still limited. Conductivity of p GaAs at these doping densities is an order of magnitude less than that of n GaAs at the same doping densities,[17] and a quantum well device with a doping density of $\sim 5 \times 10^{18}\text{cm}^{-3}$ and a thickness of 100 Å should have a conductivity several times that of the base of a npn GaAs transistor. Difficulties which need to be addressed before experimental realization of these devices is attempted include that of forming a control contact to the quantum well itself. If the transistor were operated in the regime treated in the previous section, the current from the emitter to the collector would scatter electrons from the base, decreasing the number of electrons in the base region. The result would be a negative input impedance to the base, and the transistor would have a negative beta (current gain). The electrical model of such a device would include a large positive feedback term, which would probably be undesirable for most applications. However, these results imply that there is a regime in which the current gain is very high, since it goes to positive infinity before changing sign and becoming negative. This is in contrast with the metal base types of structures which usually have current gains below one and must therefore be limited to power amplification applications.

8.10 Multiple quantum well infrared photoconductor

The next device to be discussed is a new type of infrared photodetector which uses the high degree of control available with MBE to enable the tailoring of such characteristics as the cutoff wavelength and the tradeoff between the photoconductive gain and response time.[18]

The body of the device consists of a heavily doped multiple quantum well structure formed by alternating layers of semiconductors which form heterojunctions with a conduction-band discontinuity of the desired magnitude, in this case GaAs and GaAlAs. With a typical band diagram as shown in Fig. 8.1. The device is operated at a sufficiently low temperature such that nearly all of the electrons are trapped in the quantum wells. When a bias voltage is applied across the structure, little current flows in the absence of light due to the lack of carriers outside the quantum wells. Absorption of infrared radiation excites electrons out of the quantum wells, allowing current to flow in the external circuit. The device acts as a photoconductor, similar in some respects to extrinsic silicon and germanium devices,[19] with the impurity traps of those devices corresponding to the quantum wells of the multiple quantum well device. An advantage of the quantum well device relative to the extrinsic photoconductive devices is that the concentration of states participating in the absorption interaction can be much larger. The allowable density of absorbing impurities in extrinsic photoconductive materials is limited by the formation of impurity bands which cause the leakage current to increase. This problem is avoided in the quantum well device because of the separation of absorption sites into separate layers. The structure becomes very conductive in the direction parallel to the quantum wells, but the dark current perpendicular to the quantum wells can remain low, provided the quantum wells are separated by a sufficient distance to avoid electron tunneling through the barrier layers.

The cutoff wavelength of this device is determined by the depth of the conduction band quantum wells. The dark current of the device is due to the thermal excitation of electrons over the same barriers. If the device is fabricated with a lower barrier height, the cutoff wavelength will be increased, and the dark current at a given temperature will also increase. Direct control of this tradeoff between cutoff wavelength, device sensitivity and cooling requirements of the device by merely varying the aluminum mole fraction of the barrier regions may make possible the fabrication of particular devices optimized for several bands of interest, including the 3 to 5 micron band and the 8 to 10 micron band.

The mechanism of this device is similar to that of the Schottky-barrier detector,[20] with the metal replaced with a heavily doped semiconductor. The low density of electrons in the semiconductor compared to that in a metal is partially made up for by the large number of active interfaces in this structure, however, the total number of electrons participating in the absorption is still small compared to that in a metal.

In order to estimate the absorption rate for infrared photons in the multiple quantum well structure, we will assume that the quantum wells are separated by a distance large enough that the probability of tunneling between the quantum wells is negligible. The absorption rate for a single quantum well in a crystal volume will be found, and the absorption rate of the multiple quantum well structure will be taken to be the sum of the absorption rates of the individual quantum wells. To find the total number of photons absorbed in a unit volume, we need to sum the transition rate for all initial states $|\Psi_i\rangle$ to all states $|\Psi_f\rangle$ due to the perturbing Hamiltonian of the electromagnetic field. Applying Fermi's golden rule,

$$R_{\text{total}} = \sum_i \sum_f \frac{2\pi}{\hbar} |H'_{if}|^2 f_i (1 - f_f) \delta(E_f - E_i) \tag{98}$$

where f_i and f_f are the occupation probabilities of the initial and final states, respectively.

The perturbation Hamiltonian due to an electromagnetic plane wave, in the dipole approximation, neglecting the term quadratic in the vector potential and the magnetic interaction with the spin, is given by,[21]

$$\tilde{H} = -\frac{q}{m_e} \vec{P} \cdot \vec{A} \tag{99}$$

where $\vec{P}$ is the momentum operator and $\vec{A}$ is the operator of the vector potential.

$$\vec{A} = A_o \hat{a} \cos(\vec{q} \cdot \vec{r} - \omega t) = \frac{A_o}{2} \hat{a} [e^{(iwt - i\vec{q} \cdot \cdot r)} + e^{-(iwt - i\vec{q} \cdot \cdot r)}]$$

Converting the momentum operator to the $\vec{r}$ representation,

$$\tilde{H} = \frac{iq\hbar A_o}{2m_e}\hat{a}\cdot\nabla \tag{100}$$

where A_o is the magnitude of the vector potential and $\hat{a}$ is its direction. The matrix element is given by

$$\langle\Psi_f|\tilde{H}|\Psi_i\rangle = \frac{iq\hbar A_o}{2m_e}\int \Psi_f^*(\vec{r})(\hat{a}\cdot\nabla)\Psi_i(\vec{r})d^3r \tag{101}$$

$$\langle\Psi_f|\tilde{H}|\Psi_i\rangle = \frac{iq\hbar A_o}{2m_e}\int A_f^* u_f^*(\vec{r})e^{-i\vec{k}_{\rho f}\cdot\vec{\rho}}\phi_f^*(z)(\hat{a}\cdot\nabla)A_i u_i(\vec{r})e^{i\vec{k}_{\rho i}\cdot\vec{\rho}}\phi_i(z)d^2\rho dz \tag{102}$$

Most of the integrals over the terms which result from application of the gradient to the wavefunction are zero for various reasons. In the terms which result from application of the gradient to the rapidly varying periodic part $u_{i,f}(\vec{r})$, the orthogonality of the slowly varying parts of the wavefunction causes the integral to be zero for a nonzero energy change. Similarly, terms which result from application of the gradient to the $e^{ik\rho}$ terms are zero due to a condition which is identical to conservation of crystal momentum in a bulk material. In other words, the crystal momentum of the electron in the direction parallel to the quantum well is conserved, the photon supplies negligible momentum, and transitions to states which have different transverse energies are disallowed. The only integral which is not zero is the one in which the gradient is applied to the $\hat{z}$ component of the wavefunction.

$$\begin{gathered}\langle\Psi_f|\tilde{H}|\Psi_i\rangle = \frac{iq\hbar A_o}{2m_e}A_f^* A_i\frac{1}{V_o}\int u_f^*(\vec{r})u_i(\vec{r})d^3r\times \\ \int_\rho e^{-i\vec{k}_{\rho f}\cdot\vec{\rho}}e^{i\vec{k}_{\rho i}\cdot\vec{\rho}}d^2\rho\int_z \phi_f^*(z)(\hat{a}\cdot\nabla)\phi_i(z)dz\end{gathered} \tag{103}$$

integrating,

$$\begin{gathered}\langle\Psi_f|\tilde{H}|\Psi_i\rangle = \frac{iq\hbar A_o}{2m_e}A_f^* A_i\frac{1}{V_o}\int u_f^*(\vec{r})u_i(\vec{r})d^3r\times \\ \mathcal{A}\delta(k_{xf},k_{xi})\delta(k_{yf},k_{yi})\hat{a}\cdot\hat{z}\int_z \phi_f^*(z)(\frac{d}{dz})\phi_i(z)dz\end{gathered} \tag{104}$$

Noting that all of the occupied initial electronic states are bound states, it is also apparent that there are no absorption transitions where the final state is also bound in the $\hat{z}$ direction due to the delta functions over the crystal momentum in the $\hat{x}$ and $\hat{y}$ directions and conservation of energy, which requires that the separation constant in the $\hat{z}$ direction change to that of an unbound state. The summation over the final state in the absorption equation can now be converted into an integral, indexing them by their wavevector far from the quantum well. Indexing the states in this way results in the density of states being the same as that of a homogeneous crystal. In other words, for a crystal width large compared to the width of the quantum well, there is a new state for every change in k_z such that $\ell \cdot \Delta k_z = 2\pi$. The sum over the initial states collapses due to the Kroniker delta functions, leaving only a factor of two from the two possible spin orientations, together with the factor of two due to the possibility of spin flip. Converting the sum over the final state into integrals,

$$R_{\text{total}} = \sum_{\pm} \int\limits_{k_{xf}} \int\limits_{k_{yf}} \int\limits_{k_{zf}=0}^{\infty} \frac{V_o}{(2\pi)^3}$$

$$4[N_{\mathrm{QW}}\ell]\left(\frac{2\pi}{\hbar}\right)|H'_{if}(k_{\rho i}=k_{\rho f})|^2 f_i(k_{\rho i}=k_{\rho f})(1-f_f)\delta(E_f-E_i)d^3k_f \qquad (105)$$

where N_{QW} is the number of quantum wells per unit length, A is the area of the quantum wells, ℓ is the thickness of the multiple quantum well structure. The $\sum_{\pm}$ indicates that the odd and even symmetry final states are summed over. Since the change in energy does not depend on the transverse momentum of the electron (since it remains the same), the integral over the transverse k vectors can be evaluated separately:

$$\int\limits_{k_{xf}} \int\limits_{k_{yf}} \frac{2\mathcal{A}}{(2\pi)^2} f_i(\vec{k}_{\rho i}=\vec{k}_{\rho f})dk_{ix}dk_{iy} = \mathcal{A}n \times (w+s) \qquad (106)$$

where n is the average doping density of the structure, and s is the separation between the quantum wells. Converting the delta function over energy into delta functions over k_z, neglecting the occupancy factor of the final state, and inserting the matrix element,

$$R_{\text{total}} = \int\limits_{k_{zf}=0}^{\infty} \sum_{\pm} \frac{V_o}{(2\pi)} n \cdot (w+s) 2[N_{\text{QW}}\ell] \left(\frac{2\pi}{\hbar}\right) \left| \frac{-iq\hbar A_o}{2m_e V} A_f^* A_i \right|^2 \xi_{i,f} \mathcal{A}^2 \times$$

$$[\hat{a} \cdot \hat{z}]^2 \left| \int\limits_{z} \phi_f^*(z) \frac{d}{dz} \phi_i(z) dz \right|^2 \frac{1}{|\frac{\hbar^2 k'}{m_e}|} (\delta(k_{zf} - k') + \delta(k_{zf} + k')) dk_{xf} \tag{107}$$

where $k' = \left[\frac{(\hbar\omega - V)2m_e}{\hbar^2} + \left(\frac{\pi}{w}\right)^2\right]^{1/2}$, s is the separation between the quantum wells, and $n \cdot (w+s)$ is the density of the electrons per unit area of a quantum well. Note that the term coresponding to the odd symmetry final state is zero. The integral over z is negligible outside the quantum well region. Inside the quantum well region, $\phi_i(z) = A_b \cos(\pi z/w)$ and $\phi_f^*(z) = A_{\text{II}}(e^{ik_{\text{II}2}z} \pm e^{-ik_2 z})$ where $k_{II2} = \left[\left(\frac{2m\hbar\omega}{\hbar^2}\right) + \left(\frac{\pi}{w}\right)^2\right]^{1/2}$ is the wavevector of the final state over the quantum well. Evaluating the integral,

$$R_{\text{total}} = \frac{V_o}{(2\pi)} n \cdot (w+s) 2[N_{\text{QW}}\ell] \left(\frac{2\pi}{\hbar}\right) |\frac{q\hbar A_o}{2m_e} A_f^* A_i|^2 \xi_{i,f} \mathcal{A}^2 [\hat{a} \cdot \hat{z}]^2$$

$$\frac{4|A_{\text{II}}(k')A_b|^2}{\frac{\hbar^2 k'}{m_e}} \left(\frac{\pi}{w}\right)^2 \left\{ \frac{\sin[(k_{\text{II}2} - \pi/w)\frac{w}{2}]}{k_{\text{II}2} - \pi/w} + \frac{\sin[(k_{\text{II}2} + \pi/w)\frac{w}{2}]}{k_{\text{II}2} + \pi/w} \right\}^2 \tag{108}$$

The absorption constant α can be found from the absorption per unit volume for a given vector potential by finding the energy flux corresponding to the value of A_o versus the energy loss rate due to the absorption of photons, and is given by[22]

$$\alpha = \frac{\mu_o c 2\hbar^2}{n_r \hbar\omega} \left(\frac{R}{A_o^2 V_o}\right) \tag{109}$$

where n_r is the index of refraction of the material. Combining,

$$\alpha = \frac{\mu_o c 2\hbar^2}{n_r \hbar\omega} \frac{1}{(2\pi)} n \cdot (w+s)[N_{\text{QW}}\ell] \left(\frac{2\pi}{\hbar}\right) |\frac{q\hbar}{2m_e} A_f^* A_i|^2 \xi_{i,f} \mathcal{A}^2 [\hat{a} \cdot \hat{z}]^2$$

$$\frac{4|A_{\text{II}}(k')A_b|^2}{|\frac{\hbar^2 k'}{m_e}|} \left(\frac{\pi}{w}\right)^2 \left\{ \frac{\sin[(k_{\text{II}2} - \pi/w)\frac{w}{2}]}{k_{\text{II}2} - \pi/w} + \frac{\sin[(k_{\text{II}2} + \pi/w)\frac{w}{2}]}{k_{\text{II}2} + \pi/w} \right\}^2 \tag{110}$$

$$\alpha = \frac{\mu_o c q^2}{n_r \omega |k'| 2 m_e} n[w+s][N_{\mathrm{QW}}\ell]\xi_{i,f} \mathcal{A}^2 |A_f^* A_i|^2 [\hat{a}\cdot\hat{z}]^2$$
$$|A_{\mathrm{II}}(k')A_b|^2 \left(\frac{\pi}{w}\right)^2 \left\{ \frac{\sin[(k_{\mathrm{II}2}-\pi/w)\frac{w}{2}]}{k_{\mathrm{II}2}-\pi/w} + \frac{\sin[(k_{\mathrm{II}2}+\pi/w)\frac{w}{2}]}{k_{\mathrm{II}2}+\pi/w} \right\}^2 \tag{111}$$

$N_{\mathrm{QW}} = \frac{1}{(w+s)}$, $|A_i|^2 = |A_f|^2 = \mathcal{A}^{-1}$ and $|A_b|^2 = (2/w)$ and we can approximate $\xi_{i,f} = 1$, remembering that $k_{\mathrm{II}2}$ is a function of k'

$$\alpha = \frac{\mu_o c q^2}{n_r \omega |k'| 2 m_e} n\ell |A_{\mathrm{II}}(k')|^2 2/w \left(\frac{\pi}{w}\right)^2 \times$$
$$\left\{ \frac{\sin[(k_{\mathrm{II}2}-\pi/w)\frac{w}{2}]}{k_{\mathrm{II}2}-\pi/w} + \frac{\sin[(k_{\mathrm{II}2}+\pi/w)\frac{w}{2}]}{k_{\mathrm{II}2}+\pi/w} \right\}^2 \tag{112}$$

Figure 8.8 shows the absorption constant α as a function of wavelength for a range of doping densities. These results for direct absorption are of the same magnitude as those estimated for the second order interaction involving emission of a phonon during the absorption process.[23] In either case, absorption lengths are on the order of ~ 1000μm, and surface detecting devices are limited to efficiencies of ~ 1% or less.

8.11 Experimental results on infrared photoconductor

An infrared detector of the above type was fabricated and tested for dark current, photoresponse, and response time.[24] To increase the sensitivity, the device demonstrated used a waveguide geometry allowing more of the light to be absorbed due to the increased interaction length, as shown in figure 8.9.

The device consisted of 20 80Å quantum wells of GaAs and 300Å barrier layers of $\mathrm{Ga_{.7}Al_{.3}As}$. The structure was doped with a uniform flux of Sn, providing a donor density of approximately $1 \times 10^{19}\mathrm{cm}^{-3}$, however, freeze-out of the carriers limits the free carrier density to approximately 2×10^{18}. Due to the aluminum effusion cell being operated at a constant current rather than a constant temperature, the aluminum layers were graded slightly. This can be expected to enhance electron transport in the direction of the grading.

Upper and lower claddings 1.5 microns thick of $Ga_{.6}Al_{.4}As$ were grown to provide an optical waveguide. The device was grown on a N^+ GaAs wafer, and a 2000Å N^+ GaAs contact layer was used to provide Ohmic contact at the low operating temperatures. Following the growth, the wafer was removed from the mounting block and the back was lapped to remove the indium solder and allow photolithography. Transverse isolation was provided by etching away the contact layer and part of the upper cladding outside of a 15-micron stripe. Silicon nitride was then deposited by chemical vapor deposition, openings were etched over the previously etched mesas, and AuGe/Au contacts were alloyed to the top and bottom. The wafer was cleaved into bars to provide a geometry similar to that of a stripe contact double heterostructure laser, as shown in Fig. 8.9, to be used as an edge detector. The completed device had a receiving aperture of 0.5 by 15 microns, and an interaction length of 150 microns.

The device was mounted on a copper block and submerged in a bath of liquid nitrogen and was illuminated with approximately 10^{-8} W of broadband near-infrared radiation (a 2700°C source with a GaAs filter was used). The dc current-voltage curve is shown in Fig. 8.10.

The device displayed a very slow response time, with the current continuing to decrease for more than 1 s after the light was turned off. The response of the device when stimulated by a very short pulsed excitation from a YAG mode-locked laser is shown in figure 8.11, showing the non-exponential behavior with a long slow tail. This slow response is the result of a very low electron capture probability for the structure at the applied bias, and the long dielectric relaxation time of the structure, similar to the effects observed in the extrinsic photoconductive devices. When a photon is absorbed, exciting an electron out of a quantum well, the latter is accelerated by a field of approximately 10^4 V/cm and has a low capture probability in succeeding wells. The structure is then left with a unit positive charge. The electron current rises above its dark level until an electron is captured, neutralizing the excess charge. Between the initial absorption and the eventual electron capture, a large number of electrons can pass through the structure. Thus the photoconductive gain is large, but the device response time is long, approximately 1 s, this being the average time between a photon absorption and the final electron capture. The non-exponential fall time is a result of the reduction of the dielectric relaxation time due to the presence of the photo-induced carriers.

Potential advantages of this device structure include the increased efficiency of the device over the Schottky-barrier device due to elimination of absorption far from the interface which does not contribute to the detected current, and the built-in electric field on each of the quantum wells due to the separation of the electrons from their parent donors in the barrier region that increases the likelihood that excited carriers will escape the well region, which also increases the efficiency. The small transit time for such a device structure implies that the potential speed is very high, since the energetic carrier lifetime can be made short by increasing the width of the quantum wells, although this would occur at the expense of the photoconductive gain. The probability of capture of a high-energy electron by a quantum well is one of the important parameters of the device, determining the photoconductive gain and the response time. This probability can be tailored during the growth by varying the width of the quantum wells, with a wider well allowing electrons to thermalize and be captured by the well before making it across.

The high degree of transparency of the structure for electrons allows the fabrication of devices with a thick quantum well structure, which would be needed in the surface detecting geometry due to the long absorption length. However, even with epitaxial layer thicknesses of ten microns, the absorption efficiency is still rather small, on the order of 0.5 to 1.0 percent. This, together with the GaAs/GaAlAs construction, which would allow integration with other electronic devices such as amplifiers and multiplexers, is necessary for the construction of large two-dimensional arrays. A schematic diagram of one such structure is shown in figure 8.12. The bias contact is shown as an implanted and reannealed structure, allowing the voltage to be increased into the avalanche gain region.

8.12 Conclusions

In this chapter, the nonequilibrium transport of hot electrons over degenerately doped quantum wells has been discussed. The collisions which tend to bring the hot electrons into equilibrium with the cold electrons in the quantum well have been analyzed, and it has been shown that under certain circumstances transmission probabilities are high, and regions exist where electron multiplication can occur due to scattering of secondary electrons out of the quantum well. Application to several types of devices has been discussed, including an infrared detector, and a proposed quantum well base transistor.

references for chapter 8

[1] R. J. McIntyre, IEEE Trans. Electron. Devices ED-13, 164 (1966).

[2] H. C. Casey and M. B. Panish, "Heterostructure lasers," Academic Press, New York, (1978).

[3] E. M. Conwell, "High Field Transport in semiconductors," pp. 156, Academic Press, New York, (1967).

[4] B. R. Nag, "Electron transport in compound Semiconductors," pp. 98, Springer-Verlag, Berlin, (1982).

[5] L. Reggiani, "Hot-Electron Transport in Semiconductors," p. 10, Springer-Verlag, New York, (1985).

[6] ibid, p. 12.

[7] ibid, p. 66.

[8] W. Fawcett, A. D. Boardman, and S. Swain, *J. Phys. Chem. Solids,* Vol. 31, pp. 1963-1990, (1970).

[9] C. Mailhiot, thesis, California Institute of Technology, Pasadena, CA, (1983).

[10] K. Seeger, "Semiconductor Physics," p. 234, Springer-Verlag, Berlin, (1982).

[11] S. M. Sze, "Physics of Semiconductor Devices," p. 29, J. Wiley and Sons, New York, (1981).

[12] H. C. Casey and M. B. Panish, "Heterostructure lasers," vol. B, p. 196, Academic Press, New York, (1978).

[13] A. Yariv, "Theory and applications of Quantum Mechanics," p. 192, John Wiley and Sons, New York, (1982).

[14] B. R. Nag, "Electron transport in compound Semiconductors," p. 105, Springer-Verlag, Berlin, (1982).

[15] C. Cohen-Tannoudji, B. Diu, and F. Laloe, "Quantum Mechanics," p. 958, vol. 1, John Wiley and Sons, New York, (1977).

[16] S. M. Sze, "Physics of Semiconductor Devices," p. 141, J. Wiley and Sons, New York, (1981).

[17] ibid, p. 33.

[18] J. S. Smith, L. C. Chiu, S. Margalit, A. Yariv and A. Y. Cho, paper 9.7 International Conference on Metastable and Modulated Semiconductor Structures, Pasadena, California, (1982).

[19] P. R. Bratt, in "Semiconductors and Semimetals," edited by R. K. Willardson and A. C. Beer, Academic Press, New York,Vol. XII, Chap. 2, (1977).

[20] See, for example, S. M. Sze, "Physics of Semiconductor Devices" Wiley & Sons, New York, Chap. 13, p. 760, (1981).

[21] C. Cohen-Tannoudji, B. Diu, and F. Laloe, "Quantum Mechanics," p. 1307, vol. 1, John Wiley and Sons, New York, (1977).

[22] K. Seeger, "Semiconductor Physics," p. 310, Springer-Verlag, Berlin, (1982).

[23] L. C. Chiu, J. S. Smith, S. Margalit and A. Yariv, Infrared Phys. vol. 23, no. 2, pp. 93-97, (1983).

[24] J. S. Smith, L. C. Chiu, S. Margalit, A. Yariv and A. Y. Cho, J. Vac. Sci. Technol. B. Vol. 1, no. 2, (1983).

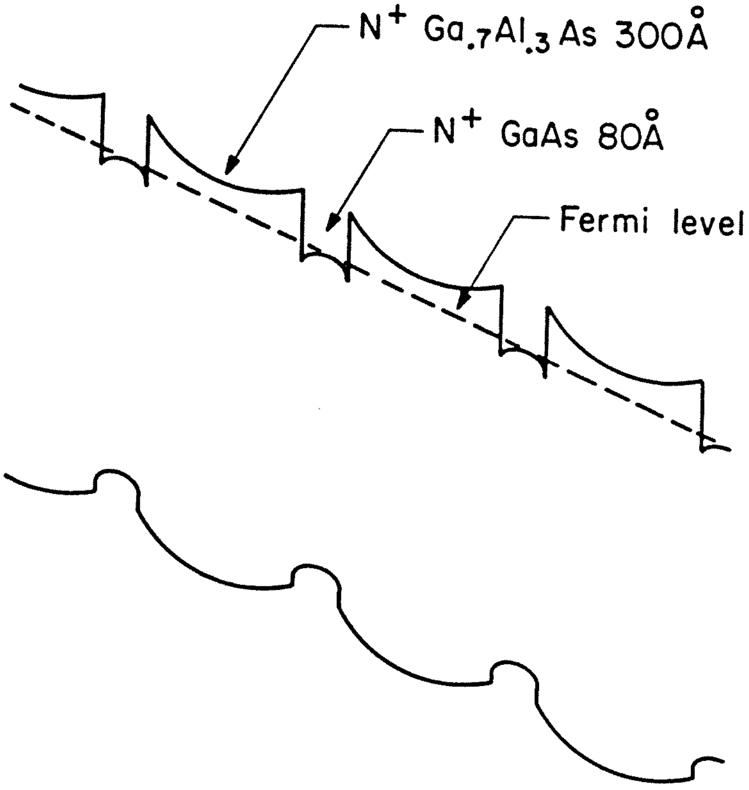


Figure 8.1) The band structure of the multiple quantum well device under an applied field.

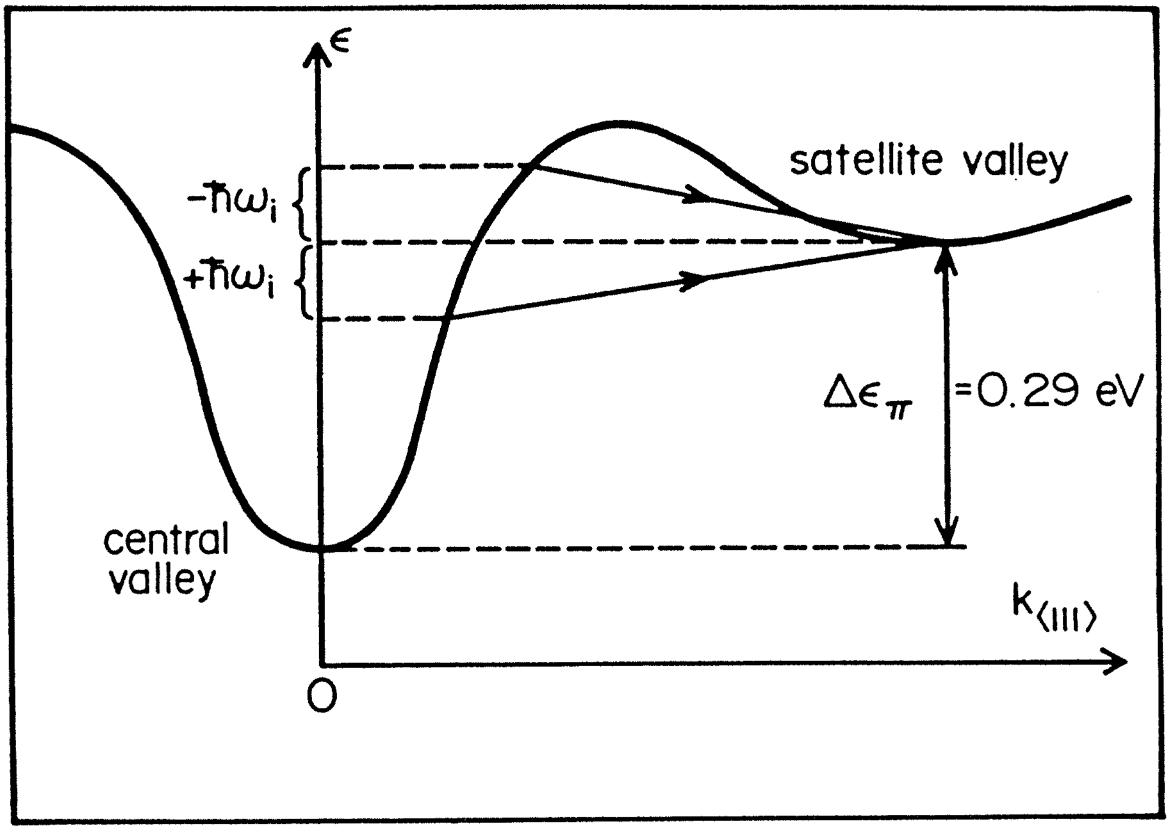


Figure 8.2) The band structure of GaAs.

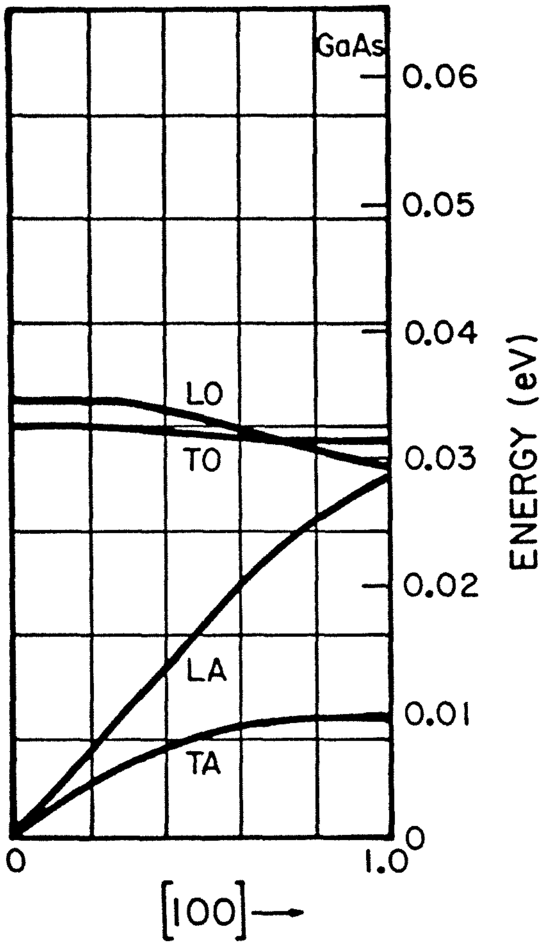


Figure 8.3) The phonon spectrum of GaAs.

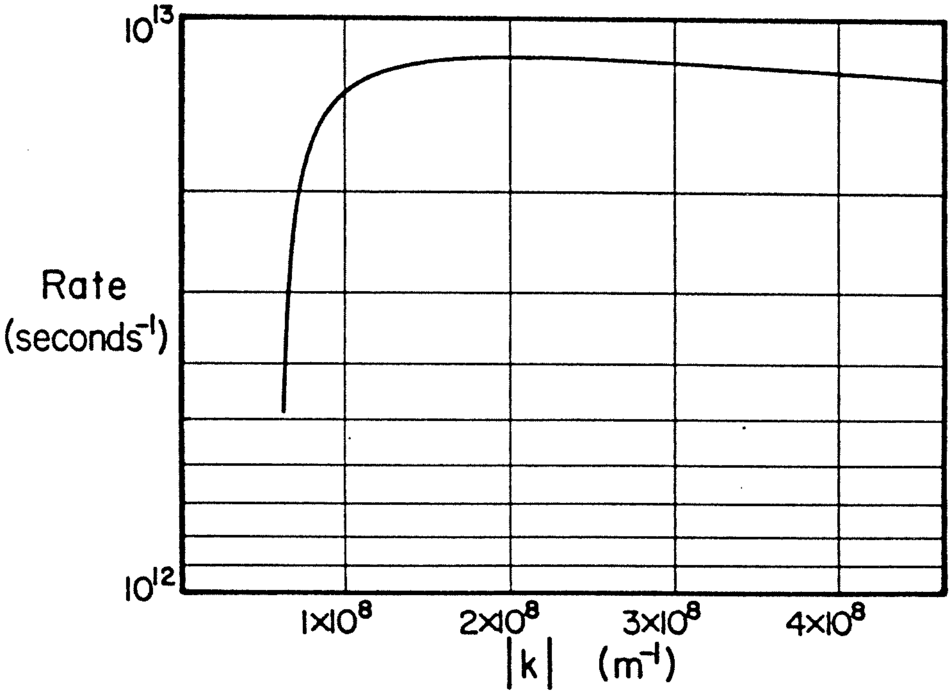


Figure 8.4) Total scattering rate out of the state $\vec{k}$ due to optical phonon emission, as a function of electron energy.

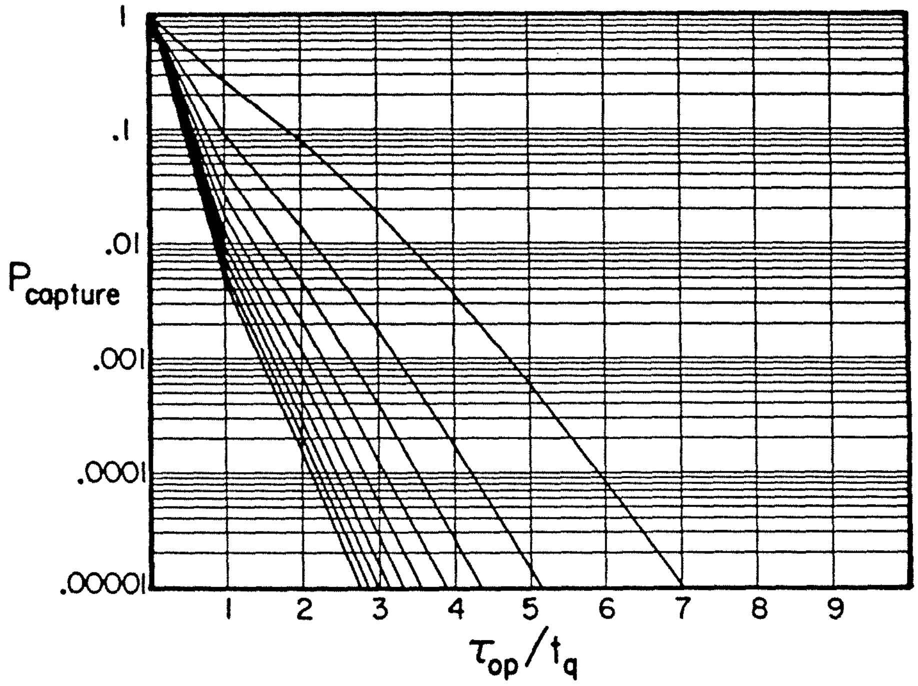


Figure 8.5) probability of capture by optical phonon emission as a function of the energy at injection and (τ_{op}/t_q)

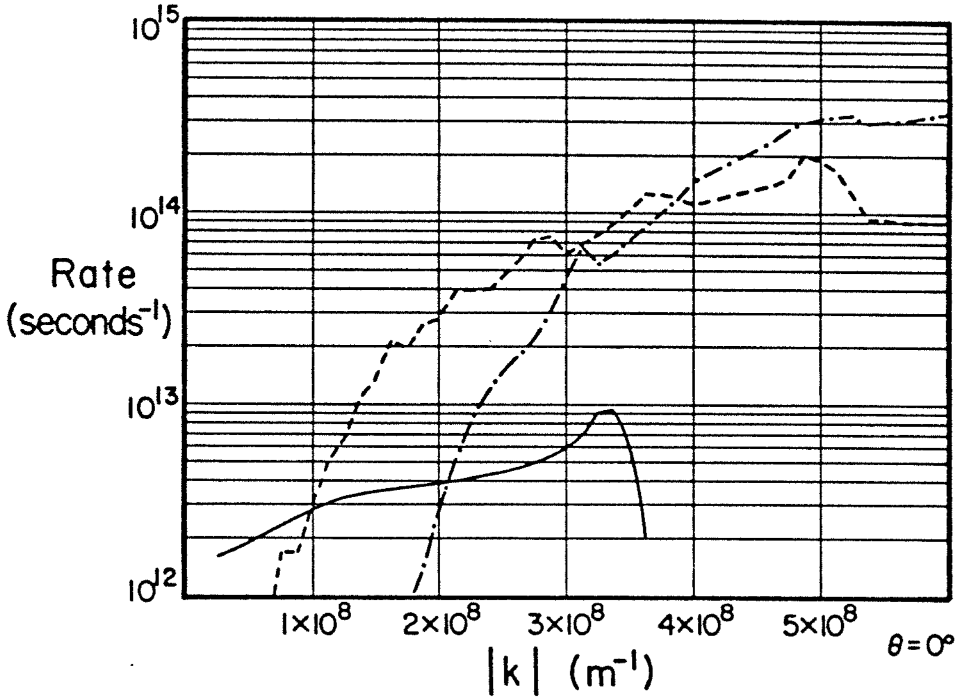


Figure 8.6) scattering rates as a function of the magnitude of the incident k vector. The solid line is the scattering rate for events in which neither electron escapes the quantum well, the dashed line for those where one electron escapes, and the dot-dashed line shows the rate for events in which both electrons are energetically unconfined

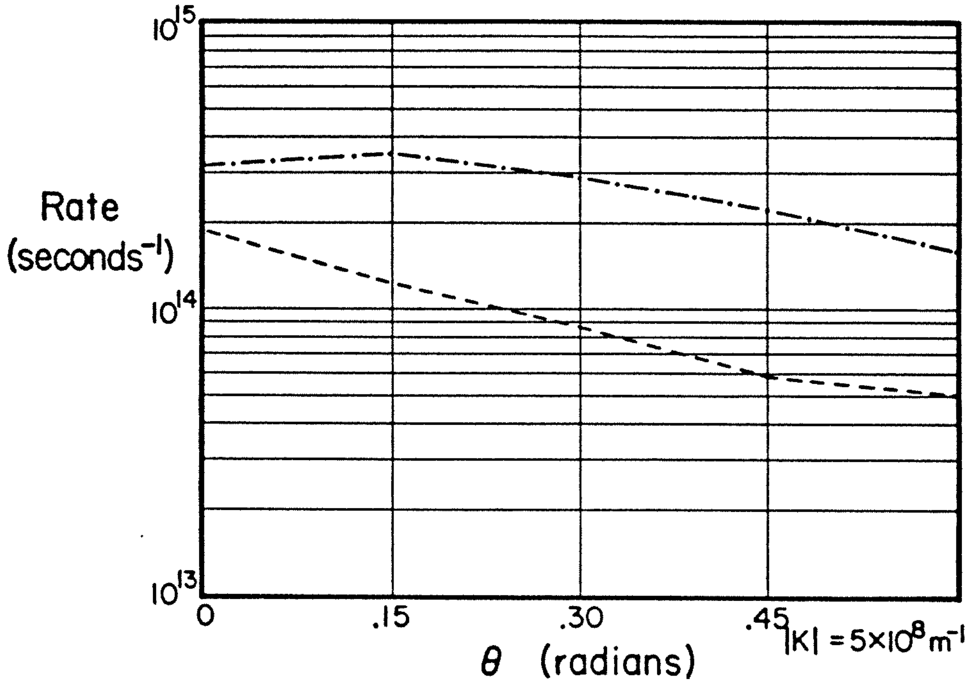


Figure 8.7) scattering rates as a function of the angle of the incident k vector. The solid line is the scattering rate for events in whch neither electron escapes the quantum well, the dashed line for those where one electron escapes, and the dot-dashed line shows the rate for events in which both electrons are energetically unconfined

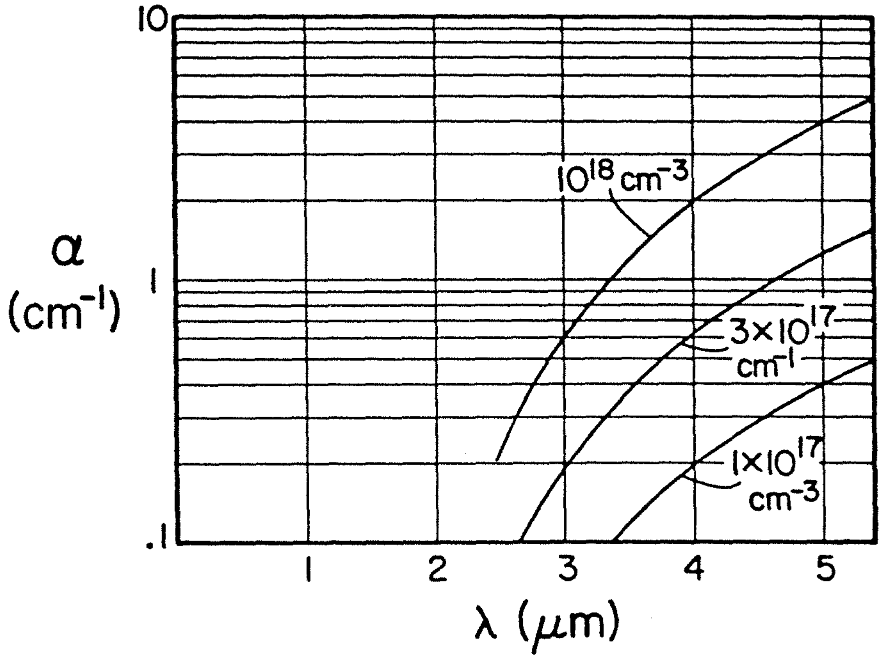


Figure 8.8) absorption constant α as a function of wavelength for a range of doping densities.

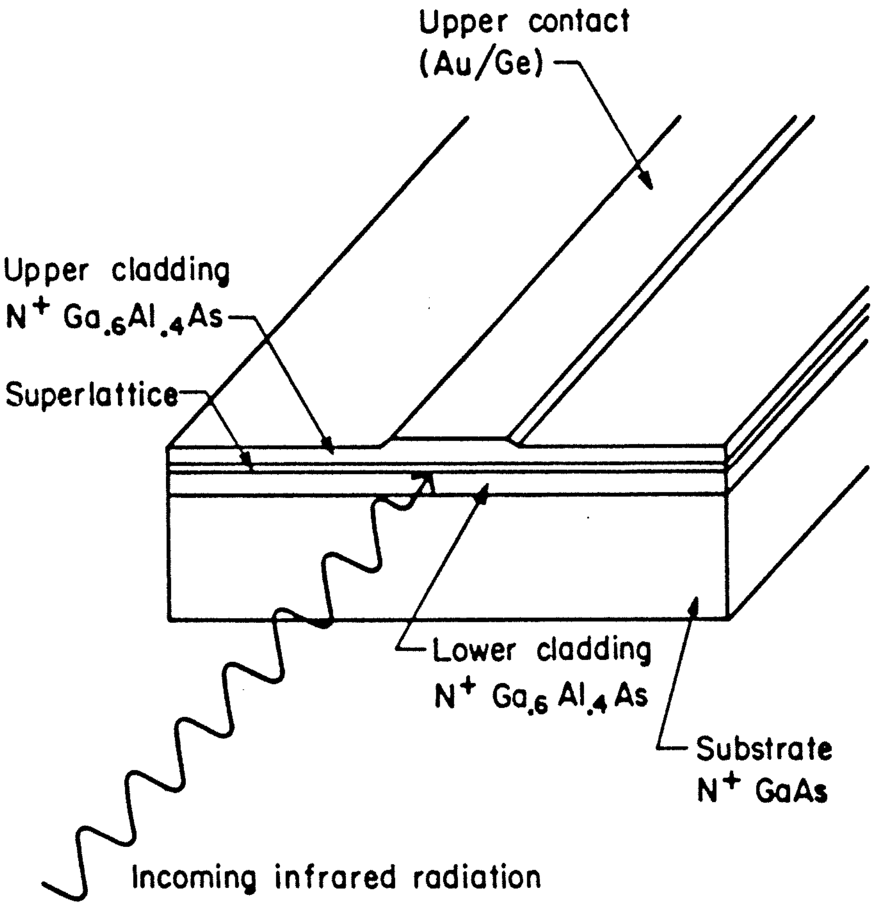


Figure 8.9) Diagram of the edge-detecting structure.

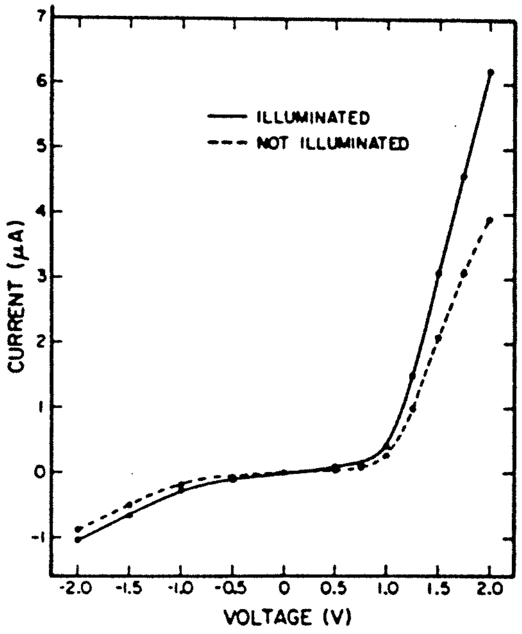


Figure 8.10) Current-voltage characteristic at 77° K with illumination of 10^{-8} W of near-infrared radiation.

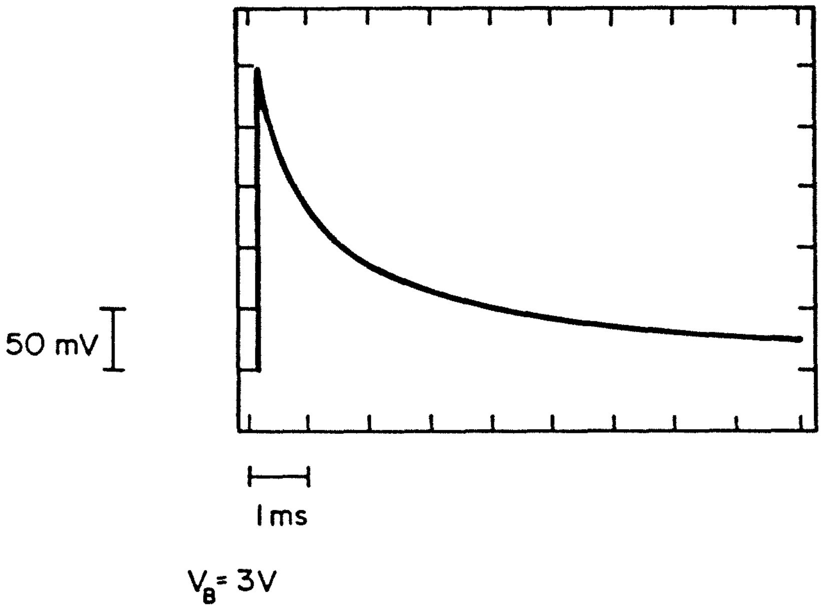


Figure 8.11) Time response of infrared detector to pulsed excitation from a YAG mode-locked laser. Note the non-exponential behavior with a long tail.

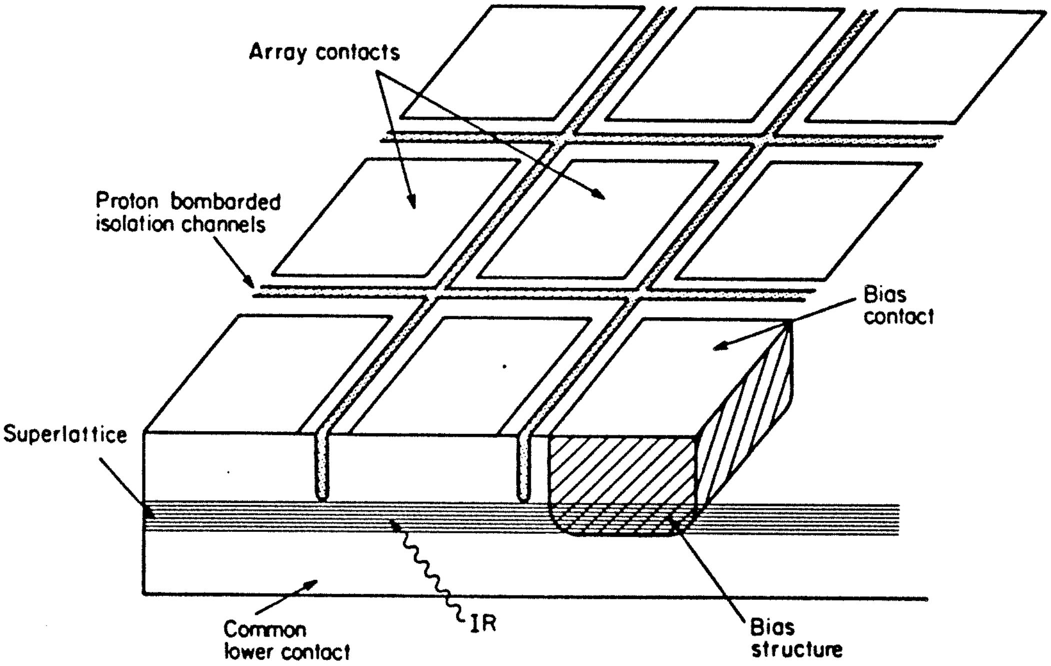


Figure 8.12) Schematic diagram of a proposed array of detectors.

4 14~~13~~ 12