Investigation of Novel Semiconductor Heterostructure Systems:

I: Cerium Oxide/Silicon Heterostructures

II: 6.1 Å Semiconductor-Based Avalanche Photodiodes

Thesis by

Edward J. Preisler

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2003

(Submitted March 25, 2003)

To Cori

Acknowledgements

The largest share of my gratitude will always be with my wife. Being a spouse of a scientist is just as difficult as being a scientist yourself and Cori managed to perfect this art form flawlessly. If it hadn't been for her, the Caltech experience would have made me a stranger and more maladjusted person (and I guess that's saying something).

When I sit and think about the effect my advisor, Tom McGill, has had on me, I come up with two things. First, the best and worst thing he has ever done for me as a student is convincing me that I am doing good work and that I am a legitimate researcher. I don't know if it's true, but having Tom's endorsement is the next best thing to believing it myself. Second, Tom runs a group in which we have to learn to do *everything*. It seems like a pain, but after five years, you find you've received an extremely broad education in performing research.

I benefited in very real ways from interactions with all of my contemporaries in the group. Firstly I need to thank Tim Harris for insulating all of us from the financial side of science and letting us just worry about research. Bob Beach is the only guy in the world who, if you told him you wanted to build a synchrotron out of spare parts in the lab, not only would tell you it was a good idea, but could also probably find most of the parts. Cory Hill taught me everything I ever wanted to know but was afraid to ask about MBE, plus he forced me to keep my short game healthy. Xavier Cartoixà always had reasonable (and correct) explanations of every solid-state physics question I ever had. Plus he calculated the band structure of every possible permutation of binary 6.1 Å semiconductor superlattices for me in about an hour (literally). Justin Brooke gave all of my projects a real shot in the arm, plus, while he was here I didn't have to be the only muck-raking liberal around. And he makes a mean mojitos (not that there's anything wrong with that). I want to thank Neal Oldham for his courage and enthusiasm when it comes to all things UHV. What would a vent be without

Neal around? I seem to have a knack for finding the "don't cares" in our computers' truth tables, so I have to thank Steve Ichiriu for cleaning up those messes. He'll have a lot more time on his hands to do research when I'm gone. Lastly, I've got to thank Rob Strittmatter for being the most dependable source of knowledge and insight I've come across (most people probably wonder if this is the same Rob). I guess I always got more of his yin than his yang or vice versa, but there are several parts of this thesis I couldn't have done without his help.

I have also had the help of several group associates in my research. I had nothing but positive experiences with the late Jim McCaldin who was one of the few great academics who didn't talk down to you even though it was probably his right to. Gerry Picus forced me to always make sure I knew what I was talking about and when I did (which was rare) he'd trust me. Ogden Marsh has an idea for how to measure anything and everything. He is by far the best person to talk to when things seem completely hopeless.

Finally I'd like to thank my family. In addition to the usual love and support stuff, my mom made certain that I never smoothed any data, my dad is the most rapt and competent audience in the world for talking about science, and my sister, going through an undergrad physics program at the same time as I went through Caltech, constantly exposed how little I actually understood about physics and forced me to relearn things.

Thanks to all these people, I think I finally know enough to take some measurements. Oh well.

Abstract

The work presented in this thesis concerns the development of two different semiconductor heterostructure technologies.

Part I describes research in the CeO_2/silicon heterostructure system. Details are presented concerning the growth of CeO_2 on silicon and the reactions that take place at the CeO_2/silicon interface. The evolution of this interface as a function of annealing temperature and annealing ambient are studied via *in situ* x-ray photoelectron spectroscopy (XPS). Studies of metal-CeO_2-silicon capacitors are also presented which help to determine the usefulness of this oxide as an alternative gate dielectric for silicon-based device applications.

Part II involves research into the fabrication of avalanche photodiodes (APD's) utilizing the 6.1 Å semiconductor system. Certain alloys of $Al_xGa_{1-x}Sb$ are shown to greatly favor hole multiplication which is beneficial for both noise characteristics and gain-bandwidth product. Further, details are presented on the current investigation into using 6.1 Å superlattices to acheive even more desirable detector performance.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Thesis Overview

This thesis is a report on two different semiconductor heterostructure systems. Part I is a general study of aspects of the cerium oxide/silicon heterostructure system. Cerium oxide has become the subject of considerable attention in the field of silicon electronics because of its excellent lattice match with silicon, its insulating properties and its chemical stability. However, certain aspects of this materials system are as of yet poorly understood and need extensive further examination to bring the system to the level of maturity needed for commercial devices.

In particular, the following areas are of critical importance. First, the behavior of the semiconductor-insulator interface formed between silicon and cerium oxide needs to be quantified. Among the most important pieces of information are a precise knowledge of the species formed at the interface, the interface trap state density and the conduction band offset with silicon. These are all critical parameters in determining cerium oxide's usefulness for metal-oxide-semiconductor (MOS) applications. Second, more exact knowledge of the nature of cerium oxide in thin film form will need to be ascertained in order to predict its applicability as a tunnel barrier in a silicon-based resonant tunneling device.

In order to clarify the issues addressed above, the work presented in Part I details a host of different types of analysis. These include, but are not limited to, x-ray photoelectron spectroscopy (XPS) to determine interface properties and the nature of chemical bonding in the oxide thin films, and capacitance and conductance profiling of MOS capacitor devices to examine the electrical behavior of films.

In the end, it is the aim of this project to expand the level of understanding of the cerium oxide/silicon heterojunction system from its current state of perhaps only scientific interest to the point where it is known whether or not it has the properties

necessary to be used in practical devices. Along the way it becomes clear that much about the nature of oxide thin films and semiconductor-insulator junctions in general is exemplified by this system.

Part II concerns research on the 6.1 Å family of semiconductors. The three semiconductors, InAs, AlSb and GaSb, known as the 6.1 Å family because they share roughly that lattice constant, are gaining importance in the fields of optoelectronics and high speed electronic devices. They have bandgaps that span from the visible to the near infrared and exhibit type I, type II-staggered and type II-broken gap band offsets with respect to one another. These properties, combined with the ability to grow all three materials epitaxially on one another, enable the creation of an incredibly rich and diverse range of superlattices that can extend the usefulness of the system.

The work presented in Part II is a study of the application of this system to a specific kind of optoelectronic device known as an avalanche photodiode (APD). The work concerns adapting an alloy of $Al_xGa_{1-x}Sb$ for use as a multiplication region in an APD as well as progress towards the use of various 6.1 Å superlattices in APD's. Aspects of growth, materials characterization, device fabrication and finally device characterization are discussed.

The purpose of this project is to advance the technology of 6.1 Å APD's to the point where they have properties that are comparable or superior to competitive infrared APD material systems such as $Hg_xCd_{1-x}Te$ and $In_xGa_{1-x}As$.

1.2 Motivation

1.2.1 CeO_2/Silicon Heterostructures

The Future of Silicon Electronics

As semiconductor devices continue to scale down in size and up in speed, several fundamental physical limits begin to loom. For metal-oxide-silicon field effect transistor (MOSFET) technology, for instance, these limits include the minimum gate oxide

thickness that can still prevent gate leakage and the maximum saturation velocity of carriers in silicon that ultimately determines the operation frequency. Although alternative semiconductors provide some answers (such as III-V's for high speed devices) for a path to circumvent these problems, there also may be some ways to extend the life of silicon itself. Two such paths are indicated below: alternative gate dielectrics and silicon-based heterostructures, both of which can be facilitated by the introduction of cerium oxide materials into traditional silicon-based electronic device structures.

Alternative Gate Dielectrics

The ability to "scale" down the size of MOSFET's while maintaining roughly the same electrical behavior is the reason that microprocessor technology has been able to increase in speed and exponentially increase in device density over time (Moore's law).

The basic MOSFET scaling argument is as follows (see, for instance, reference [1]). The main parameter that shrinks with each successive generation of transistor is the gate width. Smaller gate widths allow more transistors to be packed into a smaller area and reduce the transit times for carriers going from source to drain. However, in order to keep the transistor normally off, as is required for complementary MOS (CMOS) logic, the semiconductor region under the gate cannot be fully depleted at zero bias. This leads to the need for higher doping in the gate region so that the source and drain depletion widths will not extend to the point where they meet each other (punch-through). The higher doping requirement in turn leads to higher charge density at the semiconductor-insulator interface. Thus, in order to invert the channel and thus turn on the device, a higher gate voltage will be necessary. Finally, in order to keep the same threshold voltages to turn on the device, a higher gate capacitance is needed. A higher gate capacitance also improves the subthreshold voltage current swing and the on-state drive current. Traditionally the gate capacitance is increased by simply thinning the gate oxide.

Theoretically the gate capacitance diverges as the oxide thickness goes to zero but

of course in reality, other effects take over as the thickness approaches the "quantum" limit (roughly the de Broglie wavelength of an electron). Most importantly quantum tunneling from the gate electrode to the semiconductor becomes significant, especially under an applied bias. This translates to a "leakage" current which consumes excess power and reduces transistor gain and noise margins. Also, the uniformity of devices across a large wafer becomes more and more sensitive to fluctuations in the actual oxide thickness. For instance, for a 20Å thick gate oxide, a deviation of even one atomic layer of silicon atoms (about 2Å) would create a 10% fluctuation in thickness and hence about a 10% fluctuation in threshold voltage which would greatly affect the consistency of device performance across the chip. In fact, Intel researchers came to the conclusion that for their purposes the most pressing physical limit standing in the way of future device scaling is the gate oxide thickness [2, 3].

Alternative gate dielectrics have been vigorously sought after as a way to avoid the gate oxide thickness problem. That is, gate insulators made out of materials with a higher dielectric constant than that of SiO_2 can be made thicker while maintaining the same device performance (same gate capacitance per unit area) in order to prevent gate leakage.

There are several candidates for possible materials to replace SiO_2. Buchanan [4] gives an excellent overview of the use of silicon oxynitrides as gate dielectrics and recently Yeo et al. [5] have successfully made silicon nitride MOS devices with promising characteristics. However, it is not clear how much can be gained from using nitrided oxides or pure silicon nitride because their bandgap and band offsets with silicon are always going to be less than those of SiO_2 without gaining much in dielectric constant (κ=7.5 for Si_3N_4 as compared to κ=3.9 for SiO_2). Also, while there is some knowledge about the interface state density [6], little is known about how the stoichiometry of silicon nitride and deposition conditions can affect the device.

Many other metal oxides have also been proposed. Most notable are zirconium oxide [7] or yttria stabilized zirconium oxide (YSZ) [8], tantalum oxide [9], hafnium oxide [10], and titanium oxide [11]. A comparison of the pertinent parameters for gate dielectrics of these and various other dielectrics along with those of cerium oxide

are shown in Table 1.1.

As one can see from the table, each dielectric has its own advantages and disadvantages. For instance, while $SrTiO_3$ has a huge dielectric constant and relatively good interface properties, it has almost no conduction band offset to silicon and grows in a perovskite structure that can be very difficult to control. Or for the case of Si_3N_4, while it has a large (2.4 eV) band offset to silicon, it has a relatively low dielectric constant and poor interface qualities.

Comparatively little research has been done on the application of the cerium oxides as gate dielectrics. The values for the bandgap and the interface state density vary widely in the literature [20, 21, 22]. Also, even though Hubbard and Schlom have calculated CeO_2 to be thermodynamically unstable on silicon at 1000K [14], it has been shown that it can be grown epitaxially on silicon anyway with no amorphous SiO_{2-x} layer at the interface through the use of pulsed laser deposition (PLD) at very low temperatures [23]. Even if the CeO_2 reacts with the silicon, it will reduce to Ce_2O_3 (see chapter 3), which is thermodynamically stable on silicon even at high temperatures and still acts as an insulator. In other words an ultrathin suboxide layer between the semiconductor and the CeO_2 will barely affect most device properties.

Another potentially critical advantage of CeO_2 over some of the other materials listed in Table 1.1 is the ability to grow CeO_2 epitaxially on silicon. The mobility of carriers in the channel of a MOSFET is almost exclusively governed by scattering at the silicon-insulator interface either via coulomb interactions or interactions with a mechanically rough interface. For a true epitaxially grown insulator, the carriers in the channel would have mobilities much closer to their bulk values since there are no dangling bonds or drastic breaks in symmetry at a heteroepitaxial interface. In this case the channel would be a true two-dimensional electron or hole gas as is used for the channel in a heterojunction field effect transistor (HFET).

All in all, the cerium oxides have properties that are comparable to all of the other materials listed and yet have been studied far less than all the others in terms of their use as a gate dielectric.

Material	Dielectric Constant	Band Gap (eV)	CB Offset w.r.t Si (eV)	Interface Trap State Density ($cm^{-2}eV^{-1}$)	Crystalline Growth on Si?	Thermo-dynamically Stable on Si?
SiO_2	3.9	9	3.5	$\leq 1\times10^{10}$	No	NA
Si_3N_4	7.5	5.3 [12]	2.4 [12]	5×10^{11} [6]	No	NA
SiO_xN_y	3.9-7.5	5.3-9	2.4-3.5	?	No	NA
YSZ	25-29.7 [8]	$\sim$5.8	$\sim$1.4	2×10^{11} [8]	Yes	Yes? [8]
ZrO_2	20-25 [13]	5.8 [12]	1.4 [12]	3×10^{11} [13]	No	Yes [14]
HfO_2	30 [10]	6 [12]	1.5 [12]	1×10^{11} [10]	No	Yes [14]
TiO_2	40-86 [11]	3-3.5 [11]	1 [11]	1×10^{11} [15]	No	No [14]
Ta_2O_5	25 [9]	4.4 [12]	0.3 [12] (theory) 0.77 [17] (exp)	2×10^{11} [16]	Yes	No [14]
$SrTiO_3$	150 [11]	3.3 [12]	-0.1 [12]	6.4×10^{10} [18]	Yes	No [14]
CeO_2	20-26	5.5 [19]	?	2×10^{11} (see note)	Yes	No [14]
Ce_2O_3	?	3 [19]	?	?	?	Yes [14]

Table 1.1: **Comparison of the important properties of various candidates for alternative gate dielectrics for silicon MOSFET's. The interface trap state density for CeO_2 is taken from the data in Chapter 4. Since all of the other numbers in this table were taken on Si (100) substrates, the number has been divided by two to reflect that it is taken for growth on a (111) substrate which has a higher dangling bond density than (100).**

Silicon-Based Heterostructures

Different device paradigms and changes in the traditional structure of silicon devices have also been proposed as ways to further the development of silicon-based electronics. Devices that use quantum confinement such as resonant tunneling diodes have been suggested as a way to enable higher speeds and multi-state logic [24]. Stacking of layers of nanoelectronics has also been proposed as a way to increase device density for a given area of wafer. Finally, use of silicon-on-insulator (SOI) layers instead of bulk silicon has been shown to improve both the speed and power consumption of CMOS circuits [25]. To enable any of these techniques one must search for a material that can be grown on silicon and then have single-crystal silicon grown back on it. As of yet, nobody has created a true silicon heterostructure akin to those created in III-V semiconductor systems. While SiGe alloys provide somewhat of a solution, the maximum conduction band offset achievable is very low ($\sim$0.1eV [26]) which makes using this heterojunction for quantum confinement almost impossible at room temperature. An epitaxial insulator with a wide bandgap would effectively enable this technology as well as provide a true insulating substrate for growing single-crystal silicon for SOI. Table 1.2 gives a summary of some potential silicon heteromaterials and their properties.

Again it is clear that CeO_2 has properties that are comparable to those of the other materials in the table in terms of qualities for forming a silicon heterostructure. Ce_2O_3 is also shown to emphasize that it too can be made to be commensurate with the silicon lattice because of its convenient match to the hexagonal symmetry (111) face of Si. The experiments carried out by Kirk et al. on ZnS [33] and $BeSe_xTe_{1-x}$ [34] have shown that fairly good epitaxial silicon can be grown on these materials. They also have shown the $BeSe_xTe_{1-x}$ system to have reasonable band offsets and electrical qualities. The example of Pr_2O_3 is shown above as an example of a materials system very similar to that of the cerium oxides, in which relatively precise interfaces have been created and silicon overgrowth achieved [30]. However, Pr_2O_3 has a hexagonal crystal structure and thus only can be grown epitaxially on Si (111). $(La_xY_{1-x})_2O_3$

Material	Crystal Structure	Lattice Constant in Å (mismatch to Si)	Bandgap in eV	Conduction Band Offset (eV)
Silicon	Diamond	5.43 (0)	1.12	NA
Si_xGe_y	Diamond	5.43-5.65 (0-4%)	0.66-1.12	0-0.1 [26]
ZnS	Zinc Blende	5.42 (0.2%)	3.7	1.7 on (111) [27] 1.0 on (100) [28]
$BeSe_xTe_{1-x}$	Zinc Blende	5.63-5.15 (0-5.2%)	2.7-4.5	1.3 [29]
Pr_2O_3	Hexagonal [30]	(0.5%) (111) only [30]	2.5-3.9 [31]	0.5-1.5 [31]
$(La_xY_{1-x})_2O_3$	Variable [32]	(0-2.4%) [32]	?	?
CeO_2	Fluorite	5.41(0.4%)	5.5 [19]	?
Ce_2O_3	Hexagonal	a=3.89(1.2% on(111)) [19]	3 [19]	?

Table 1.2: **Comparison of the important properties of various candidates for silicon heterostructure materials.**

represents another oxide that has been used for successful regrowth of epitaxial silicon, also on (111), but may potentially be grown on (100) in the future because it can be made to form in a cubic crystal structure [32]. CeO_2 always grows in the cubic CaF_2 structure so it is theoretically possible to grow CeO_2 on any face of silicon.

All of these materials show promise as silicon heterojunction materials and it is probably only limited time and resources that have prevented these technologies from maturing further. The excellent lattice match and commensurate cubic crystal structure of CeO_2 make it an excellent candidate for a silicon heterostructure material. Combined with its favorable properties as an alternative gate dielectric, it is clear that there should be significant experimental effort put forth in determining the usefulness of applying CeO_2 to silicon-based electronics.

1.2.2 6.1 Å Avalanche Photodiodes

Overview of the 6.1 Å Semiconductor Materials

Over the last two decades or so, compound semiconductors such as GaAs have been studied nearly as much as silicon. However, the application of compound semiconductor devices lags far behind the now omnipresent silicon-based device industry. There are several reasons for this, most notably the failure to create a viable III-V MOS device. However, in the field of optoelectronics, which can loosely be defined as electronics applied to the creation of or detection of light, compound semiconductors dominate. This can be attributed mostly to two factors. Firstly, many compound semiconductors (most notably GaAs) have direct bandgaps that allow light to be emitted or absorbed efficiently. Secondly, the ability to "engineer" the bandgap of materials by creating alloys and heterostructures makes compound semiconductors far more flexible than silicon.

In Part II of this thesis, a relatively new compound semiconductor system is examined for its suitability in creating optoelectronic devices.

The three semiconductors InAs, AlSb and GaSb are known as the 6.1 Å family because they share roughly that lattice constant. They already have been successfully

employed in making infrared light detectors [35] and infrared lasers [36]. Although not discussed any further herein, they also have been applied to high-speed electronic devices [37, 38].

All three materials grow in the Zinc Blende crystal structure and the excellent lattice match between the three materials allows all three to be grown epitaxially on one another, as can many mixed III or mixed V alloys [39]. Their bandgaps span the range from the near infrared to the visible and between the three of them, they exhibit type I, type II-staggered gap and type II-broken gap band offsets [40] (see Fig. 1.1). InAs and GaSb are direct gap materials while AlSb is indirect.

Superlattices can be grown that take advantage of these properties and move the useful optical spectrum of the system out into the mid infrared [35]. In fact, because of the type II-broken gap band offset between InAs and GaSb, there is no theoretical lower limit to the effective bandgap that can be engineered within this system.

Even for use in optoelectronic devices, the electronic properties of the system will be important in creating practical devices. Table 1.3 shows the mobilities for these materials as compared to GaAs. InAs has an extremely high electron mobility and GaSb has a higher hole mobility than most of the other III-V's. Also, n-type InAs forms a negative Schottky barrier with most metals so ohmic contacts easily can be made to any n-type 6.1 Å material.

Uses of the Infrared Spectrum

The information contained in the infrared light spectrum is useful in several practical applications. One example is thermal imaging where one seeks to identify an object by detecting its blackbody radiation against a background of different temperature. Figure 1.2 (a) illustrates the fact that objects that are at practical temperatures on Earth emit most strongly in the mid to far infrared (greater than about 5μm). A second example is active imaging such as 3-D LIDAR where a reflected laser beam is used for detecting objects and determining their distance from the detector analogous to RADAR. In these applications it would be advantageous to have a high power laser that is still "eye-safe." Figure 1.2 (b) shows the maximum permissible energy dose

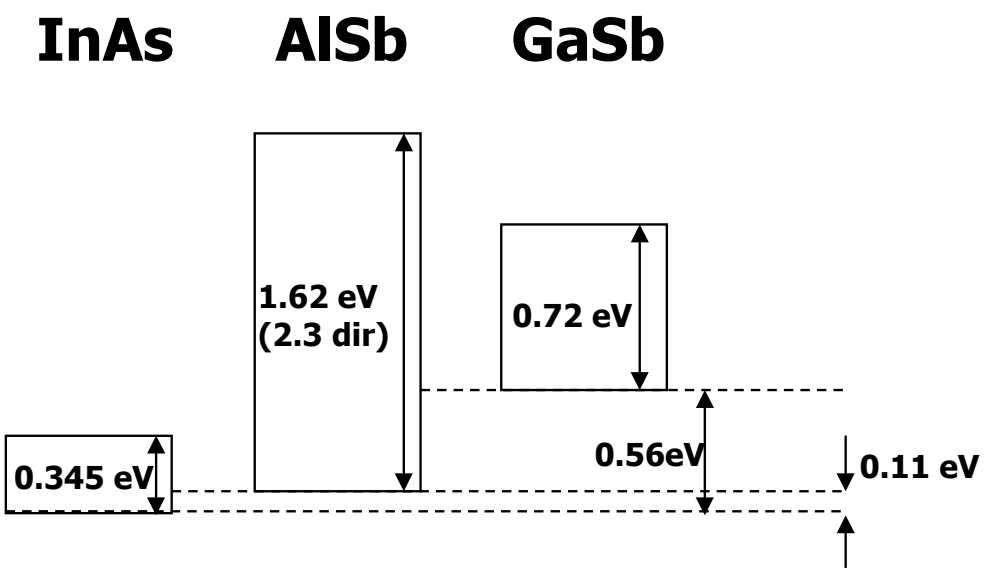


Figure 1.1: **Band alignments for the 6.1 Å semiconductor system.**

Material	Electron Mobility	Hole Mobility
	($cm^2V^{-1}s^{-1}$)	
GaAs	9200	400
InAs	33000	450
GaSb	3759	680
AlSb	200	400

Table 1.3: **Mobilities of 6.1 Å semiconductors compared with GaAs. Data taken from reference [41]**

(according to OSHA) absorbed by the human eye for lasers of various wavelengths. One can see that far more energy is permitted as the wavelength is increased, indicating that the infrared spectrum is most suitable for this application. A third example is fiber optic communications. Figure 1.2 (c) shows the well-established minimum at about 1.5 μm in the attenuation spectrum of a silica optical fiber. Detectors that operate at this wavelength are needed to process information that has been sent along the fiber. Finally, night vision is a field that benefits greatly from the use of the infrared spectrum. Table 1.4 summarizes the sources of light available for imaging at night. The bands are atmospheric transmission bands centered at the wavelength indicated in the table. On a night where moonlight is weak, the only light available comes from stars or the night airglow which originates from optical transitions of molecules in the atmosphere. One can see from the table that the strongest signal of these remaining light sources comes from the night airglow in the near infrared.

As mentioned above, the 6.1 Å semiconductors are poised to be of great benefit for infrared detection because of the ability to create materials with almost arbitrarily small bandgaps, and hence extremely long wavelength detection capabilities.

Application of 6.1 Å Materials to Infrared Avalanche Photodiodes

It is clear that for all of the applications of infrared light detection discussed above, desirable detector qualities will include sensitivity and high-speed performance. Avalanche photodiodes (APD's) can provide a substantial advantage with respect to both the signal-to-noise and gain-bandwidth product metrics of traditional semiconductor photodetectors [43]. While silicon APD's have excellent properties and are readily available commercially, they are limited in their detection wavelength by the bandgap of about 1.1 eV (about 1100 nm). Once again, the 6.1 Å materials can be made to absorb much further into the infrared than any of the other traditional semiconductor systems perhaps with the exception of $Hg_xCd_{1-x}Te$.

In most semiconductor photodetectors light is converted into an electrical signal by the absorption of photons with energies greater than the bandgap of the semiconductor material, leading to the creation of electron-hole pairs. The signal then

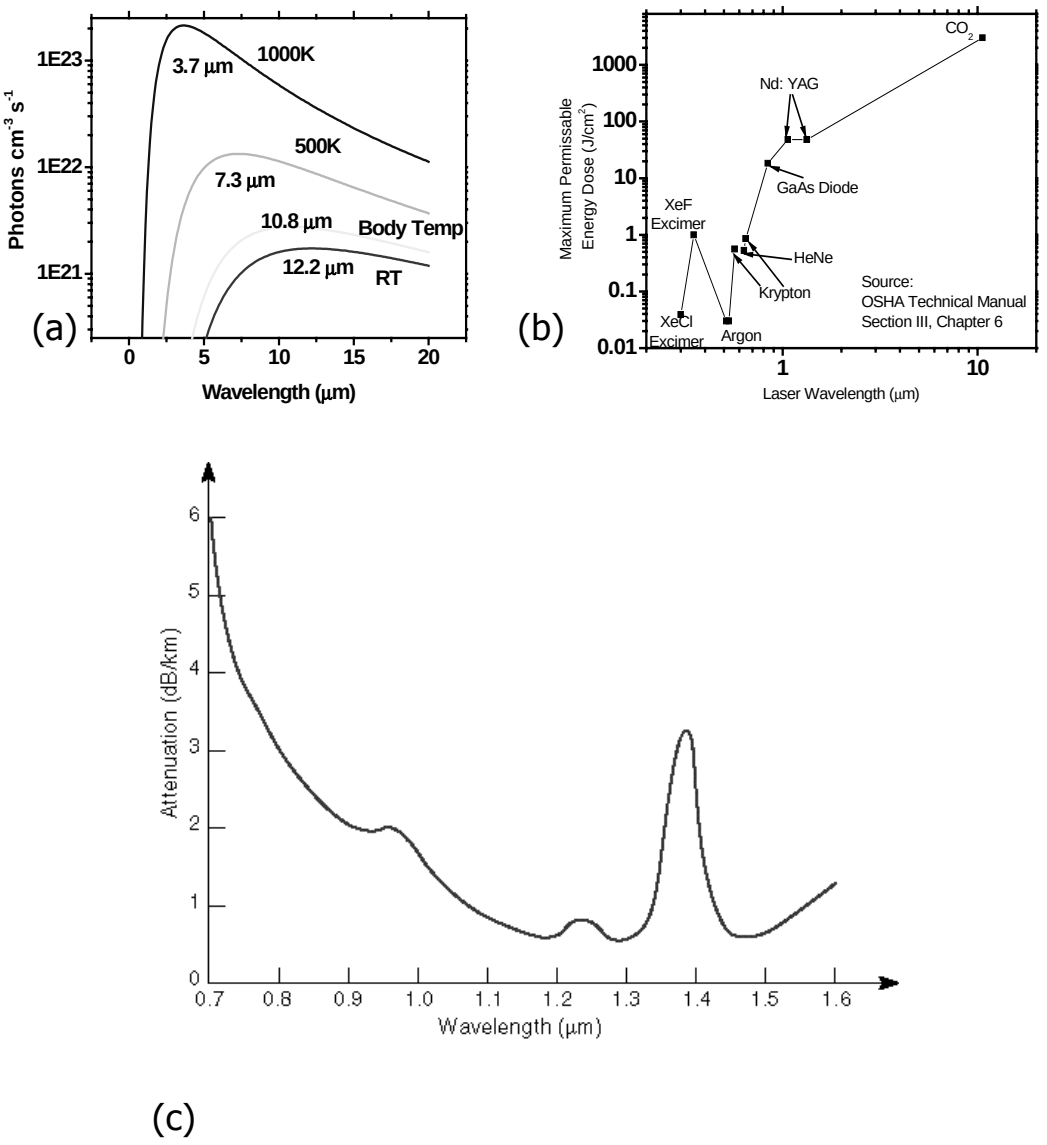


Figure 1.2: **Various applications of the infrared spectrum. Figure (a) shows the blackbody radiation spectrum at various temperatures, Fig. (b) shows the permissible energies for "eye-safe" lasers of different wavelengths and Fig. (c) shows the attenuation spectrum of a silica optical fiber taken from www.newport.com.**

Band	Wavelength	Moon	Stars	Night AirGlow
	μm	photons cm^{-2} s^{-1}		
1	0.58	5.83×10^{11}	1.21×10^{9}	4.2×10^{7}
2	1	1.83×10^{11}	1.24×10^{8}	
3	1.24	2.09×10^{11}	1.15×10^{8}	
4	1.59	1.99×10^{11}	8.67×10^{7}	4.5×10^{10}
5	2.1	1.37×10^{11}	6.02×10^{7}	2.2×10^{10}

Table 1.4: **Sources of light for night vision. Data taken from reference [42]. The data is listed in bands since these wavelengths represent minima in the absorption spectrum of our atmosphere.**

is measured by converting these newly created carriers into a current. This is done by using an electric field to accelerate the carriers and thus create a current. This electric field either can be applied externally or can be intrinsic to the device as in a p-n junction or Schottky barrier photodiode. In an ideal case, each photon creates one electron-hole pair and each electron and hole are collected at one terminal of the device or the other and measured as a current. Of course, in reality, this is never the case. Many photons are absorbed by other processes and many carriers are created by processes other than direct bandgap absorption. Thus detection of a small number of photons becomes difficult because of both a finite quantum efficiency (not all photons are converted to electron hole pairs) and a noise floor (some of the current comes from carriers not created by the light).

All photocurrent devices such as the ones discussed above must subsequently pass their signal into a transimpedance amplifier to convert the current signal to a voltage for measurement. At its most basic, a transimpedance amplifier is simply a resistor. The action of current passing through a resistor leads to both shot noise and thermal noise. So while the signal is being amplified by an amount proportional to the value of the resistor, noise is being added to the signal such that there is only an advantage in signal to noise ratio for a finite range of amplification.

It is this added noise coming from the subsequent amplification stage that avalanche photodiodes (APD's) seek to obviate. In an APD, there is already an intrinsic gain mechanism built into the photodetector itself, reducing or even eliminating the need for subsequent amplification. Basically, an APD is a photodiode that absorbs light by creating electron-hole pairs in the same way as the conventional detectors discussed above, but then the carriers enter a gain medium. In this gain medium the number of carriers is multiplied by creating several more electron-hole pairs for each initial carrier. Then the photocurrent that one would measure is the true photocurrent as in a conventional photodetector times this multiplication factor. This is the solid-state analog to a photomultiplier tube.

The intrinsic gain mechanism in APD's comes from the process of impact ionization where a carrier in an electric field in a semiconductor achieves sufficient energy

to excite or "ionize" a valence electron into the conduction band. This electric field usually is created by applying a large reverse bias to a p-n junction so that a large field exists in the depletion region. The result of the impact ionization event is two newly created carriers (the excited electron and the hole it left behind in the valence band) plus the initiating carrier, be it a hole or an electron. If this impact ionization event takes place in a region with high enough electric field, the three final carriers in turn can be accelerated and initiate new impact ionization events. Thus after two levels of events, there can be as many as nine carriers directly attributable to the one initial carrier. In certain situations this process can continue indefinitely, reaching a state known as "avalanche breakdown," hence the name of the device. When a device is operated in this regime, it is known as a Geiger-mode device. In carefully controlled situations, devices operated in this mode can detect single photons [44, 45, 46]. More typically, however, devices are operated in a regime where the gain is on the order of 10 to 1000.

However, this added gain does not come without a cost. In order for the APD structure to be advantageous compared to conventional photodetectors, certain criteria must be met by both the materials system and the device design. To illustrate this point, the signal-to-noise ratio of an APD will be examined. The mean squared signal current for a modulated input signal in an APD is given by,

$$< I_0^2 >= \frac{1}{2}\left(\frac{q\eta\lambda P_0 M}{hc}\right)^2, \tag{1.1}$$

where P_0 is the power of the optical signal, η is the quantum efficiency, λ is the wavelength of the optical signal and M is the multiplication of the APD. There are two noise sources in the APD and readout circuit: the shot noise coming from all the various currents in the circuit and the thermal noise from all the resistances in the circuit. Thus the mean squared noise current is given by

$$< I_{noise}^2 >= 2qFM^2B\left[I_0 + I_b + I_d\right] + 4kTB/R_{eq}, \tag{1.2}$$

where I_0 is the signal current, I_b is the photoelectric current from background light sources that are not part of the signal, I_d is the dark current, R_{eq} is the parallel sum of all the resistances in the circuit and F is called the excess noise factor which is an artifact of the avalanche multiplication process and will be discussed in detail in Chapter 6. Dividing Equation 1.1 by 1.2 gives the mean squared signal to noise ratio. An often cited figure of merit is the noise equivalent power (NEP) defined as the minimum signal power required to give a signal to noise ratio of one. Setting the signal to noise ratio equal to one and solving for P_0 gives

$$NEP = \frac{2hcBF}{\lambda\eta}\left[1+\sqrt{1+\frac{I_d+I_b+\frac{2kT}{qM^2R_{eq}}}{qBF^2}}\right]. \tag{1.3}$$

Figure 1.3(a) shows a plot of Equation 1.3 for various values of dark current. It is clear that it is vital to reduce the dark current (current not derived from photogenerated carriers) to improve the noise behavior of APD's.

The excess noise factor is an artifact of the avalanche multiplication process that does not exist in conventional photodetectors. Hence, it must be minimized in order to reap the benefits of the multiplication. As will be shown in Chapter 6, the excess noise factor is a function of multiplication with the worst-case scenario being F=M. It has been shown that APD's made from materials exhibiting a strong preference for impact ionization of one carrier type over the other (e.g., electrons over holes) have the lowest values of the excess noise factor and the highest gain-bandwidth product [47, 48] (this is discussed in much greater detail in Chapter 6). Figure 1.3 (b) shows the NEP for an APD with 100nA of dark current and varying values of "k," which is defined as the ratio of the hole impact ionization coefficient to the electron impact ionization coefficient. It is clear that the value of k also has a profound effect on the noise characteristics.

Gallium rich alloys of $Al_xGa_{1-x}Sb$ have been shown to greatly favor hole ionization over electron ionization because of a resonance in the energy spacings of the bandgap, E_g and the split-off band energy, Δ_{so} as demonstrated by Hildebrand et al. [49, 50]

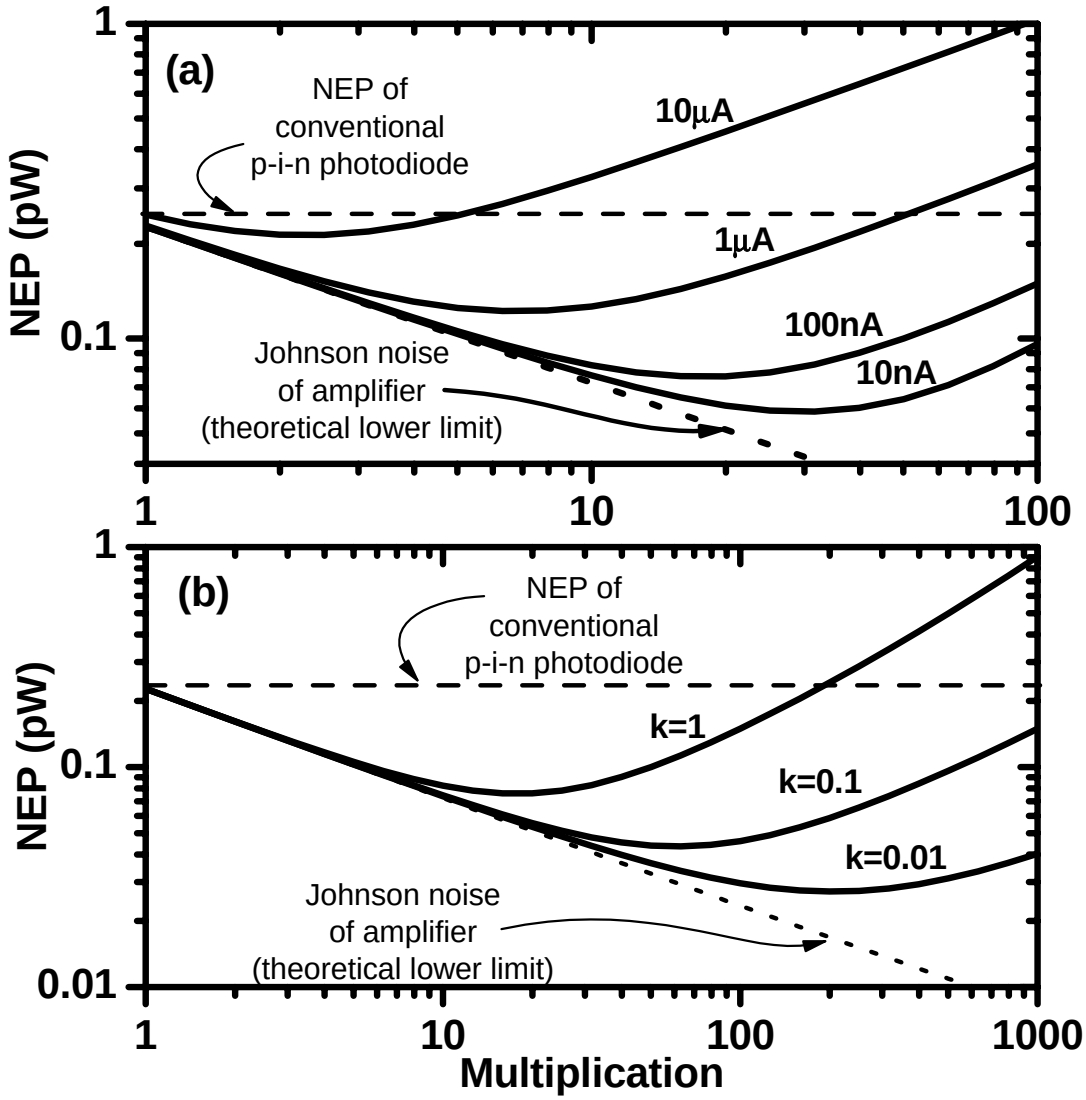


Figure 1.3: **Noise equivalent power at a bandwidth of 1 GHz for an APD with different dark currents and different values of k=β/α, where β is the hole impact ionization coefficient and α is the electron impact ionization coefficient. Figure (a) shows the NEP as a function of multiplication for an APD with various dark currents and k=1. Figure (b) shows the NEP as a function of multiplication for an APD with 100nA of dark current and various values of k. The dashed line is the NEP for a photodiode with no gain. The dotted line represents the noise floor of the external interaction circuit (last term in Equation 1.2).**

Thus there may be significant advantages in employing the 6.1 Å system to make APD's, not just in detection wavelength but in the actual performance of the devices as well.

Further, the ability to grow heterostructures of almost arbitrary bandgap and band offset discussed above leads to the possibility of many more improvements in APD design [51]. These improvements include the use of separate absorption and multiplication regions so that one layer of material can be designed so as to have optimum impact ionization characteristics while another layer can be tailored to absorb light of a precise wavelength. Also, by engineering the band offsets between absorption and multiplication regions, one can select either electrons or holes to carry the majority of the signal, enhancing device performance (see Chapter 6).

1.3 Summary of Results

1.3.1 CeO_2/Silicon Heterostructures

In Part I, the nature of thin films of CeO_2 on silicon is studied. First, the growth of CeO_2 on silicon is examined. It is found that single-crystal layers of CeO_2 can be deposited by electron beam evaporation, but only at elevated temperatures which has repercussions as discussed in the next paragraph. It is found that single-crystal layers of cerium oxide can be grown by pulsed laser ablation at much lower temperatures but require a significant additional background oxygen supply. These layers have extremely poor surface topology due to formation of particulates during the laser ablation process. The topology can be improved drastically by choosing a lower laser wavelength for growth, but only with a sacrifice in the crystal quality.

Next, the reactions that take place at the CeO_2/silicon interface are studied by XPS. It is shown that the interface is extremely reactive at elevated temperatures and that it may be difficult or even impossible to bring a stoichiometric layer of CeO_2 in contact with a silicon surface. It is found that when annealed in vacuum, the CeO_2 is reduced to Ce_2O_3 at the expense of a growing layer of SiO_{2-x} between the cerium

oxide and the silicon. It is found that when annealed in an oxygen-rich ambient, the CeO_2 remains stoichiometric but a thick layer of SiO_2 forms between the cerium oxide and the silicon.

In the next section, electrical characterization of metal-CeO_2-silicon capacitors is performed. The capacitance-voltage (C-V) characteristics elucidate the dielectric constant, trapped charge in the CeO_2 layer, and the interface trap state density at the CeO_2/silicon interface. Further, the effect of various annealing procedures on these parameters is discussed. It is shown that the effective dielectric constant of single-crystalline layers of stoichiometric CeO_2 is on the order of 10 but changes significantly with annealing. There is excessive trapped charge, but annealing seems to eliminate some of it including charges that lead to hysteresis in the C-V curves. The interface trap state density of the as-grown layers is on the order of 2×10^{12} $cm^{-2}eV^{-1}$ but can be reduced as low as 4×10^{11} $cm^{-2}eV^{-1}$ by annealing in hydrogen.

1.3.2 6.1 Å Avalanche Photodiodes

In Part II, the results of an investigation of avalanche photodiodes made from 6.1 Å semiconductor materials are discussed. P-i-n avalanche photodiodes using $Al_{0.05}Ga_{0.95}Sb$ as the intrinsic multiplication layer have been fabricated. Several unique techniques are used in order to create diodes with good electrical and photoresponse behavior. It is found that the treatment of the surfaces of mesa diodes is extremely important for obtaining good reverse bias leakage current behavior. It is also found that counterdoping of the intrinsic layers in p-i-n diodes is highly effective in enhancing both the electrical characteristics and the quality of the photoresponse of the diodes.

It is observed that these devices do indeed show a preference for hole multiplication as predicted. The ratio of the impact ionization coefficients, k=β/α is measured to be about five at high electric fields and much higher as the field is reduced. Avalanche gains of as high as 300 are achieved for devices operating at -20 volts at 150K. All of this data shows great promise for use of this material as a multiplication layer for infrared APD's.

Progress also has been made towards fabricating superlattice APD's with 6.1 Å materials. Devices have been fabricated using a superlattice designed specifically for enhanced hole impact ionization. However, it is found that the leakage current of these devices in unacceptably high. Thus the design of new device structures which should show much better leakage properties is discussed. These new structures not only should reduce the leakage current but also actively select the carrier type to be multiplied in the gain region, improving both noise and bandwidth characteristics.

Bibliography

[1] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H. S. P. Wong, Proc. IEEE **89**, 259 (2001).

[2] S. Thompson, P. Packan, and M. Bohr, Intel Technology Journal **Q3**, 1 (1998).

[3] B. Doyle, R. Arghavani, D. Barlage, S. Datta, M. Doczy, J. Kavalieros, A. Murthy, and R. Chau, Intel Technology Journal **6**, 42 (2002).

[4] D. A. Buchanan, IBM J. Res. Develop. **43**, 245 (1999).

[5] Y. C. Yeo, Q. Lu, P. Ranade, H. Takeuchi, K. J. Yang, I. Polishchuk, T. J. King, C. Hu, S. C. Song, H. F. Luan, and D. L. Kwong, IEEE Electron Device Lett. **22**, 227 (2001).

[6] J. Schmidt, F. M. Schuurmans, W. C. Sinke, S. W. Glunz, and A. G. Aberle, Appl. Phys. Lett. **71**, 252 (1997).

[7] M. Copel, M. Gribelyuk, and E. Gusev, Appl. Phys. Lett. **76**, 436 (2000).

[8] S. J. Wang, C. K. Ong, S. Y. Xu, P. Chen, W. C. Tjiu, A. Huan, W. J. Yoo, J. S. Lim, W. Feng, and W. K. Choi, Semicond. Sci. Tech. **16**, L13 (2001).

[9] D. Park, Y. King, Q. Lu, T. J. King, C. M. Hu, A. Kalnitsky, S. P. Tay, and C. C. Cheng, IEEE Electron Device Lett. **19**, 441 (1998).

[10] L. Kang, B. H. Lee, W. J. Qi, Y. Jeon, R. Nieh, S. Gopalan, K. Onishi, and J. C. Lee, IEEE Electron Device Lett. **21**, 181 (2000).

[11] S. A. Campbell, H. S. Kim, D. C. Gilmer, B. He, T. Ma, and W. L. Gladfelter, IBM J. Res. Dev. **43**, 383 (1999).

[12] J. Robertson, J. Vac. Sci. Technol. B **18**, 1785 (2000).

[13] C. M. Perkins, B. B. Triplett, P. C. McIntyre, K. C. Saraswat, S. Haukka, and M. Tuominen, Appl. Phys. Lett. **78**, 2357 (2001).

[14] K. J. Hubbard and D. G. Schlom, J. Mater. Res. **11**, 2757 (1996).

[15] H. Fukuda, S. Namioka, M. Miura, Y. Ishikawa, M. Yoshino, and S. Nomura, Jpn. J. Appl. Phys. **38**, 6034 (1999).

[16] B. Lai, N. H. Kung, and J. Lee, J. Appl. Phys. **85**, 4087 (1999).

[17] B. C. Lai, J. C. Yu, and J. Lee, IEEE Electron Device Lett. **22**, 221 (2001).

[18] R. Droopad, Z. Y. Yu, J. Ramdani, L. Hilt, J. Curless, C. Overgaard, J. L. Edwards, J. Finder, K. Eisenbeiser, J. Wang, V. Kaushik, B. Y. Ngyuen, and B. Ooms, J. Crystal Growth **227**, 936 (2001).

[19] N. V. Skorodumova, R. Ahuja, S. I. Simak, I. A. Abrikosov, B. Johansson, and B. I. Lundqvist, Phys. Rev. B **64**, 115108 (2001).

[20] L. Tye, N. A. Elmasry, T. Chikyow, P. Mclarty, and S. M. Bedair, Appl. Phys. Lett. **65**, 3081 (1994).

[21] T. Nakazawa, T. Inoue, M. Satoh, and Y. Yamamoto, Jpn. J. Appl. Phys. **34**, 548 (1995).

[22] Y. Nishikawa, T. Yamaguchi, M. Yoshiki, H. Satake, and N. Fukushima, Appl. Phys. Lett. **81**, 4386 (2002).

[23] M. Yoshimoto, K. Shimozono, T. Maeda, T. Ohnishi, M. Kumagai, T. Chikyow, O. Ishiyama, M. Shinohara, and H. Koinuma, Jpn. J. Appl. Phys. **34**, L688 (1995).

[24] J. P. Sun, G. I. Haddad, P. Mazumder, and J. N. Schulman, Proc. IEEE **86**, 641 (1998).

[25] G. G. Shahidi, IBM J. Res. Dev. **46**, 121 (2002).

[26] E. T. Yu, E. T. Croke, T. C. Mcgill, and R. H. Miles, Appl. Phys. Lett. **56**, 569 (1990).

[27] C. Maierhofer, S. Kulkarni, M. Alonso, T. Reich, and K. Horn, J. Vac. Sci. Technol. B **9**, 2238 (1991).

[28] B. Brar, R. Steinhoff, A. Seabaugh, X. Zhou, S. Jiang, and W. P. Kirk, Compound Semiconductors 1997 **156**, 167 (1998).

[29] W. P. Kirk, K. Clark, E. Maldonado, N. Basit, R. T. Bate, and G. F. Spencer, Superlattices And Microstructures **28**, 377 (2000).

[30] H. J. Osten, J. P. Liu, E. Bugiel, H. J. Mussig, and P. Zaumseil, Mater. Sci. Engineering B **87**, 297 (2001).

[31] H. J. Osten, J. P. Liu, and H. J. Mussig, Appl. Phys. Lett. **80**, 297 (2002).

[32] S. Guha, N. A. Bojarczuk, and V. Narayanan, Appl. Phys. Lett. **80**, 766 (2002).

[33] X. C. Zhou, S. Jiang, and W. P. Kirk, J. Appl. Phys. **82**, 2251 (1997).

[34] S. Jiang, P. Barrios, R. T. Bate, and W. P. Kirk, Nanotechnology **10**, 187 (1999).

[35] H. Xie, J. Piao, J. Katz, and W. I. Wang, J. Appl. Phys. **70**, 3152 (1991).

[36] T. C. Hasenberg, D. H. Chow, A. R. Kost, R. H. Miles, and L. West, Electron. Lett. **31**, 275 (1995).

[37] K. Segawa, H. Miki, M. Otsubo, K. Shirahata, and K. Fujibayashi, Jpn. J. Appl. Phys. **16**, 103 (1977).

[38] E. R. Brown, J. R. Soderstrom, C. D. Parker, L. J. Mahoney, K. M. Molvar, and T. C. Mcgill, Appl. Phys. Lett. **58**, 2291 (1991).

[39] T. H. Chiu and W. T. Tsang, J. Appl. Phys. **57**, 4572 (1985).

[40] E. T. Yu, J. O. Mccaldin, and T. C. Mcgill, Solid State Physics: Advances In Research And Applications, Vol 46 **46**, 1 (1992).

[41] *Semiconductors - Basic Data*, edited by O. Madelung (Springer, Berlin, Germany, 1994).

[42] X. C. Cheng, Ph.D. thesis, California Institute of Technology, 1999.

[43] G. E. Stillman and C. M. Wolfe, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beers (Academic Press, New York, New York, 1977), Vol. 12, pp. 291–392.

[44] H. Dautet, P. Deschamps, B. Dion, A. D. Macgregor, D. Macsween, R. J. Mcintyre, C. Trottier, and P. P. Webb, Appl. Opt. **32**, 3894 (1993).

[45] W. J. Kindt, N. H. Shahrjerdy, and H. W. vanZeijl, Sensors And Actuators A **60**, 98 (1997).

[46] W. J. Kindt and H. W. van Zeijl, IEEE Trans. Nucl. Sci. **45**, 715 (1998).

[47] R. J. Mcintyre, IEEE Trans. Electron Devices **Ed19**, 703 (1972).

[48] R. J. Mcintyre, IEEE Trans. Electron Devices **Ed13**, 164 (1966).

[49] O. Hildebrand, W. Kuebart, J. Lutz, and M. Pilkuhn, IEEE Trans. Electron Devices **27**, 2188 (1980).

[50] O. Hildebrand, W. Kuebart, K. W. Benz, and M. H. Pilkuhn, IEEE J. Quantum Electron. **17**, 284 (1981).

[51] K. F. Brennan and J. Haralson, Superlattices And Microstructures **28**, 77 (2000).

Part I

CeO_2/Silicon Heterostructures

Chapter 2 CeO_2 Growth

2.1 Overview

In this chapter, aspects of the epitaxial growth of CeO_2 on silicon are discussed. First, a brief comparison of the various growth techniques that have been shown to be successful in producing single-crystal layers of CeO_2 on silicon is given. Then a summary of the attempts to grow high quality CeO_2 layers in our research group is presented. Finally, an assessment of the optimum procedures for growing high-quality, defect-free thin films of CeO_2 on silicon with the best possible interface characteristics is given.

2.2 Background of CeO_2 Growth on Silicon

The original motivation for growing CeO_2 on silicon was to integrate the growth of high temperature superconducting materials with silicon. In particular, $YBa_2Cu_3O_{7-x}$ is well lattice-matched to CeO_2. Most of the early work was focused in this direction. It was shown that high-quality, insulating CeO_2 could be grown on Si (111) wafers [1]. Since then, most of the work has focused on improving the quality of this growth. Considerable strides have been made in this area including room temperature growth of CeO_2 on Si (111) [2] and growth on Si (100) substrates [3].

Some of the methods most commonly used for deposition of CeO_2 on silicon include e-beam evaporation [1], pulsed laser deposition (PLD) [4], sputtering [5] and dual ion beam deposition [6]. In the case of both e-beam deposition and sputtering, high substrate temperatures during growth or high temperature post-growth annealing are necessary in order to create highly crystalline layers on silicon. In the case of PLD, however, it has been shown that single-crystalline material can be grown even as cold as room temperature [2]. The role of temperature on the nature of the CeO_2/silicon

interface is discussed in great detail in Chapter 3. It is shown that a low-temperature growth method is preferred.

While growth of CeO_2 is theoretically possible on any silicon surface, it has been found that there are issues that complicate growth on all but the (111) surface. It has been shown that it is the oxygen atoms in CeO_2 interacting with dangling bonds on silicon atoms that tend to dictate the growth direction [7]. It turns out that even though CeO_2 forms a cubic crystal structure as does silicon, single-crystal growth is only intrinsically possible on (111) substrates because of this effect. On (001) substrates, two different orientations of {110} CeO_2 are energetically equivalent and favorable to any {001} orientation. Thus considerably more effort is required to grow single-crystal CeO_2 on Si (001) substrates. Nevertheless, Nagata et al. [3] showed that it is possible to achieve single-crystalline growth of CeO_2 on Si (001) substrates miscut slightly towards (110). However, the rest of the discussion in this chapter is focused on growth on Si (111).

2.3 Silicon Preparation

All of the growths used for this section of the thesis were done on 3" Si (111) wafers with various types of doping. The wafers first were degreased in acetone, isopropyl alcohol, and deionized water. The wafers then were etched with buffered oxide etchant (NH_4F_2 and HF in water) until the surface was completely hydrophobic, indicating that the native oxide was removed. This process left the surface hydrogen terminated after which the wafers were introduced quickly into vacuum to prevent reoxidation of the surface.

In order to create a repeatable starting surface for CeO_2 growth, a thin buffer layer of epitaxial silicon was grown. This was accomplished by first heating the substrates to 800°C in vacuum to desorb the hydrogen and residual oxygen contamination. Then silicon from an e-beam evaporation source was slowly deposited on the wafer until a strong (7×7) reconstruction could be observed by reflection high-energy electron diffraction (RHEED). The temperature then was lowered to about 600°C while the

rate of silicon deposition was slowly increased. A total of between 100Å and 500Å was grown and always resulted in a strong (7×7) surface reconstruction that was stable down to room temperature. This process created a very reproducible starting surface for CeO_2 growth.

2.4 E-Beam Deposition

The existing silicon molecular beam epitaxy (MBE) machine in our laboratory contains dual e-beam sources so very little effort was required to modify the chamber for CeO_2 deposition. By evaporating cerium oxide from a CeO_2 target and bleeding in molecular oxygen through a variable fine-leak valve, stoichiometric thin films of polycrystalline CeO_2 (as determined by *in situ* RHEED) routinely could be grown on Si (111) substrates. However, temperatures in excess of 700°C were required to obtain single-crystal films. At these elevated temperatures the growth morphology was extremely poor and relatively thick layers of both SiO_{2-x} and CeO_x were formed between the silicon and the crystalline CeO_2 as shown in Fig. 2.1. It is clear from these images that under the growth conditions in our system, device-quality single-crystalline CeO_2 thin films could not be grown on silicon via e-beam evaporation. Nevertheless, the lower temperature polycrystalline growths were useful for examining interface reactions (see Chapter 3).

2.5 Pulsed Laser Deposition

The failure to grow consistent high quality crystalline CeO_2 thin films via e-beam MBE led to an investigation of pulsed laser deposition of CeO_2. Figure 2.2 shows a schematic of the setup for using PLD to grow CeO_2 on silicon.

Pulsed laser deposition works by impinging high-intensity laser light on a material held in vacuum, which causes the material to ablate off and then redeposit elsewhere in the vacuum chamber. There are several mechanisms that cause the ablation which are outlined in great detail by Chrisey and Hubler [8]. The main advantage of PLD

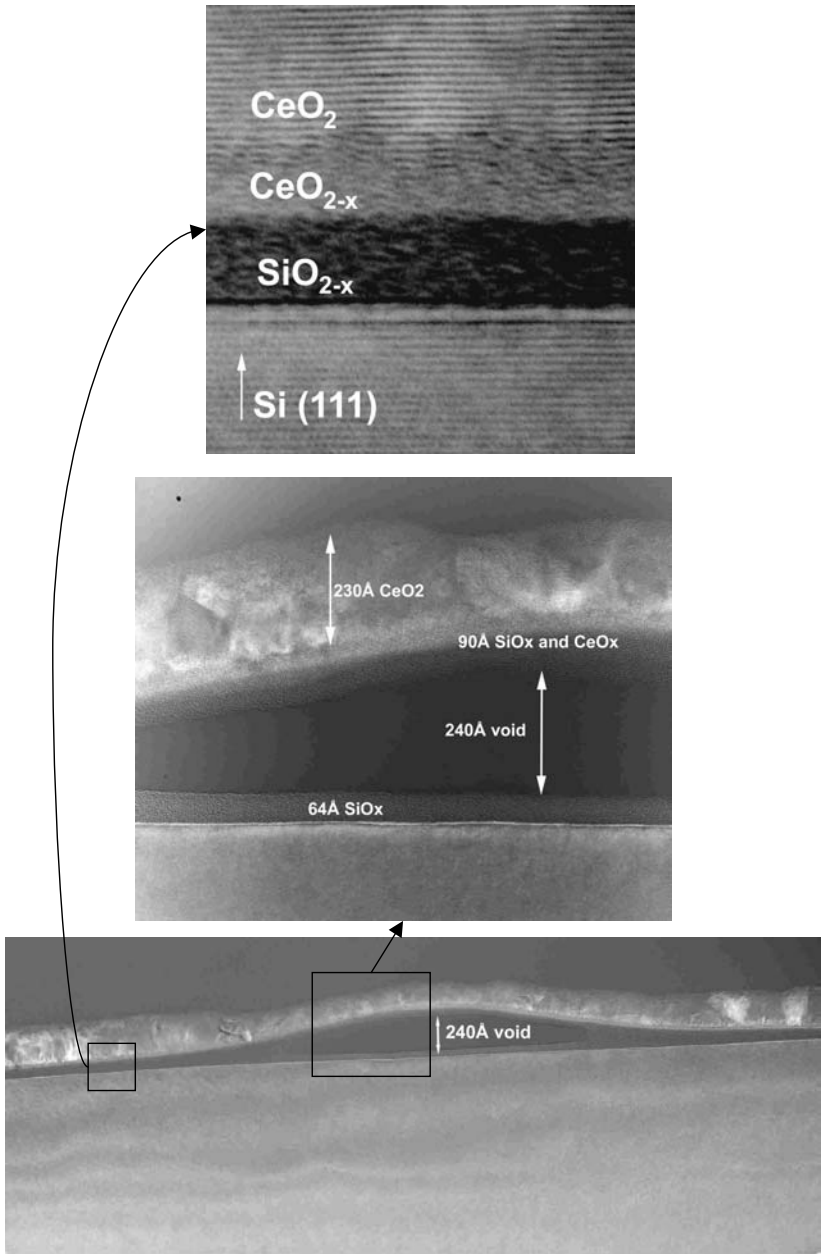


Figure 2.1: **Transmission Electron Microscopy (TEM) images of CeO_2 grown on silicon via e-beam deposition. The detailed sections show the various layers at the interface including what appear to be voids.**

over traditional thermal evaporation is the high energy of the expelled particles. Thermally evaporated particles will have energies only on the order of k_BT, or a few tenths of an eV, whereas laser ablated species can have energies as high as 100 eV [9], partially due to the interaction of the laser with the plume. For our purposes, the advantage of this extra energy is that the substrate will not need to be heated as much to obtain the amount of surface diffusion necessary for high quality epitaxial growth. Another advantage of PLD is that the growth can proceed just as easily in any ambient, so the introduction of molecular or atomic oxygen will not hamper the growth process. Almost all other ultra-high vacuum (UHV) deposition techniques involve filaments which easily can be damaged by operating in an excess of oxygen.

In the first experiments, the second harmonic of a Nd:YAG laser (532 nm light) was used to ablate cerium oxide from a stoichiometric CeO_2 target. As can be seen in Fig. 2.3 (a), this produced a streaky RHEED pattern when ablated onto Si (111). The spacing of the lines was commensurate with the spacing of the underlying silicon lines and the pattern repeated itself every 60 degrees of substrate rotation as required for the 6-fold symmetry of a {111} plane. Figure 2.3(c) shows a transmission electron microscopy (TEM) image of a CeO_2/silicon interface grown as described without any extra oxygen. Comparison with Fig. 2.1 shows that this technique is a considerable improvement over e-beam MBE as far as the formation of interfacial layers are concerned. However, by examining *in situ* x-ray photoelectron spectroscopy (XPS) spectra of the films ablated in vacuum (see Fig. 2.4) it was discovered that the layers were underoxidized, almost to the point where they were closer to Ce_2O_3 than CeO_2. While Ce_2O_3 has hexagonal symmetry instead of cubic like CeO_2 and silicon, a rough lattice match between Ce_2O_3 and silicon can be found on the hexagonal symmetry Si (111) face with about a 1% mismatch. There is also evidence that CeO_2 will remain in its cubic state even when a significant amount of oxygen vacancies are created [10]. Either of these scenarios would explain the fact that single-crystalline growth is possible for underoxidized cerium oxide species on silicon.

By bleeding in oxygen through a variable leak valve, the oxidation state easily could be brought back to that of CeO_2. Figure 2.4 shows that full CeO_2 stoichiometry

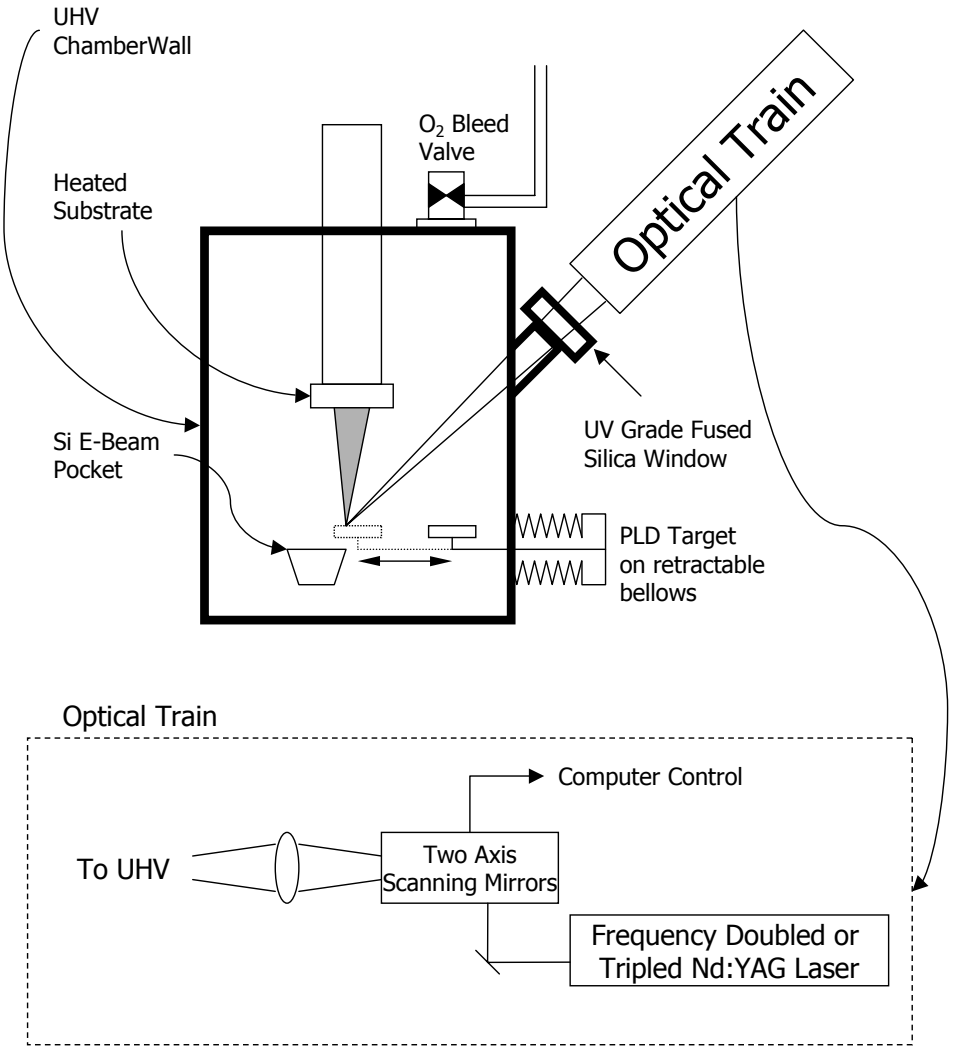


Figure 2.2: **Schematic of the setup used for PLD growth of CeO_2 on silicon.**

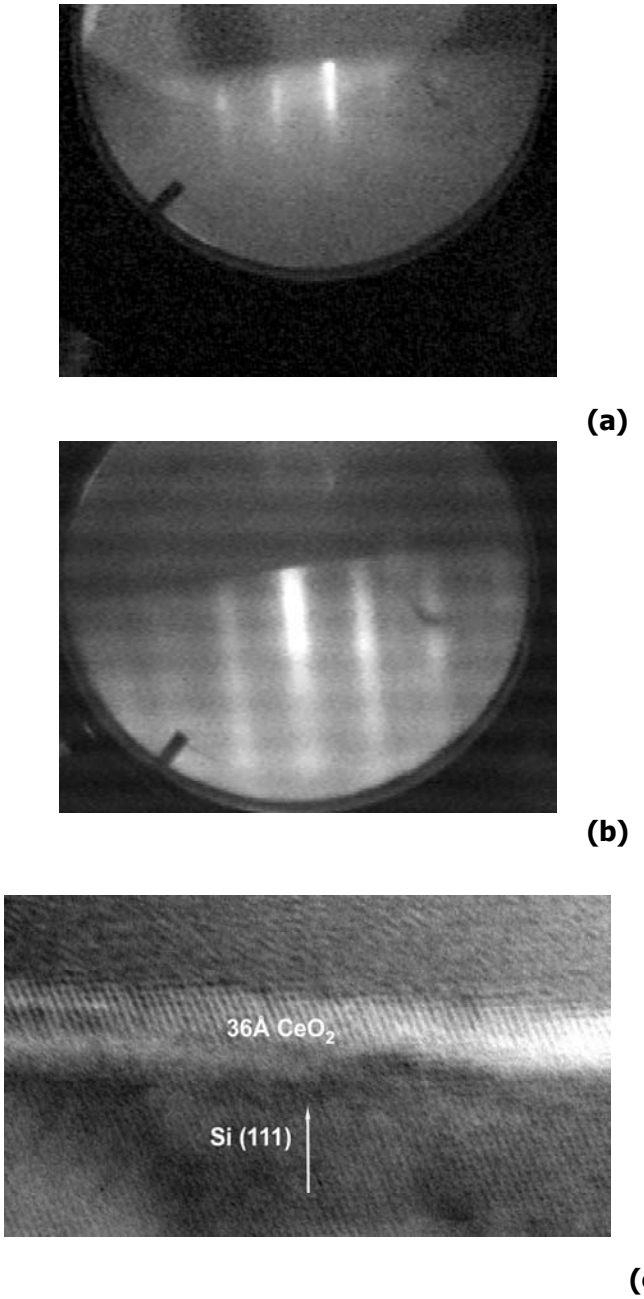


Figure 2.3: **Figure (a) shows the RHEED pattern of PLD-grown CeO_2 on silicon grown in vacuum. The pattern repeats itself every 60 degrees of substrate rotation indicative of 6-fold symmetry. Figure (b) shows the RHEED pattern of PLD-grown CeO_2 on silicon grown in 5×10^{-6} torr of O_2. The lines are much broader indicating that some charging is occuring and the lines do not change with substrate rotation. Figure (c) shows a TEM image of a thin film grown by PLD in vacuum.**

was achieved by bleeding in 5×10^{-6} torr of O_2 at a growth temperature of 450°C. The amount of oxygen required to achieve the full (IV) oxidation state changed somewhat with growth temperature.

Despite the fact that the oxidation state easily could be controlled, introducing O_2 changed the crystallinity of the growths. Instead of the streaky RHEED pattern shown in Fig. 2.3 (a), the pattern in Fig. 2.3 (b) was observed. This pattern did not change with substrate rotation indicating columnar growth where crystallites are all oriented along the [111] direction but oriented randomly in the plane of the substrate. This is actually to be expected in the case where the silicon surface sees oxygen before any cerium arrives. Yoshimoto et al. [7] observed the same result and explained it by the fact that an oxygen-saturated surface will preferentially attract the cerium atoms in CeO_2 instead of the oxygen atoms as is the case for a clean silicon surface (see Section 2.2). In CeO_2, {111} planes have the highest density of cerium atoms per unit area. Thus crystallites will tend to orient themselves along $\langle 111 \rangle$ axes perpendicular to the substrate surface. However, since there is no preference for orientation in the plane of the wafer after the surface has been oxidized, the crystallites will randomly orient themselves in this direction. It seems that exposure of even a monolayer or two of oxygen is enough to randomize the surface such that columnar growth is seen instead of single-crystalline growth.

In order to eliminate the columnar growth, the method of Yoshimoto et al. [2] was used whereby roughly 50Å of CeO_2 was first deposited in vacuum followed by bleeding in of O_2. This two step process enabled thick, single-crystalline layers of stoichiometric CeO_2 to be grown. The drawback to this method is of course that it is impossible to create ultra-thin layers of stoichiometric, single-crystalline CeO_2. Nevertheless, this growth technique proved ideal for creating the MOS structures studied in Chapter 4.

To verify the crystallinity, x-ray diffraction was performed on a relatively thick layer of CeO_2 grown in this manner. A detailed $\theta/2\theta$ scan of this sample is shown in Fig. 2.5. Besides the silicon substrate (111) peak and its multiples, there were no other peaks over the entire range of diffraction angles. This ensures that the film is

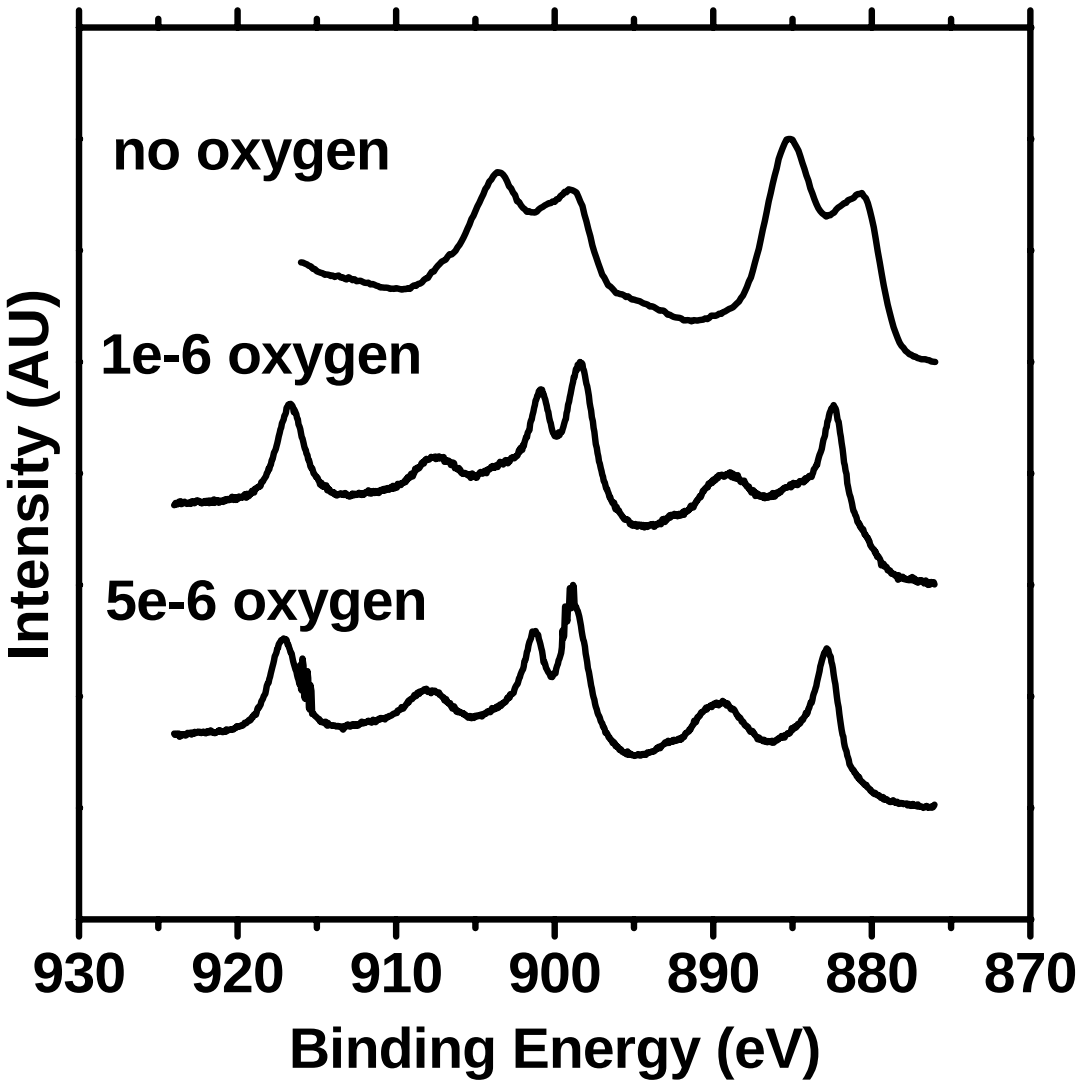


Figure 2.4: **Ce 3d XPS spectra of PLD-grown CeO_2 grown under various oxygen ambients. For a detailed interpretation of the spectra, refer to Chapter 3. The sample grown in vacuum mostly shows characteristics of the (III) oxidation state. The sample grown in 5×10^{-6} torr of O_2 only shows characteristics of the (IV) oxidation state. The sample grown in 1×10^{-6} torr of O_2 is somewhere in between but closer to the (IV) oxidation state. The samples shown here all were grown at 450°C.**

preferentially oriented in the growth direction. This combined with the fact that the RHEED repeated itself every 60 degrees of substrate rotation ensures that the film is epitaxial, oriented with the substrate [111] direction and also oriented in the plane of the wafer. The separation between the CeO_2 peak and the silicon peak suggests an out-of-plane lattice constant of 5.37 Å for the CeO_2, meaning that the layer is strained by 0.7% relative to its bulk value.

The morphology of PLD-grown materials is notoriously rough because of large (0.1–5 μm) particulates that are formed during the ablation process and redeposit onto the substrate [11]. Figure 2.6 (a) shows an atomic force microscopy (AFM) image of a 700Å CeO_2 layer grown by PLD. Kautek et al. [12] showed that the surface morphology of PLD-deposited oxides could be dramatically improved by decreasing the laser wavelength. This is probably due to a change in absorption depth of the laser light in the target. A shorter absorption depth will tend to concentrate more energy into a smaller volume at the surface of the target and thus reduce the probability of large chunks of material being ejected. In the case of CeO_2, several studies [13, 14, 15] including our own spectroscopic ellipsometry data have shown that the absorption coefficient begins to increase dramatically with decreasing wavelength right around 350 nm or so. Thus an attempt was made to perform PLD with the third Nd:YAG harmonic (355 nm). Figure 2.6 (b) shows an AFM image of a CeO_2 film grown at this wavelength. There is a dramatic difference. All of the smaller particulates are gone and there is a much lower density of the larger particulates. However, the drawback of growing at this wavelength was that it required a much higher temperature to overcome the columnar growth regime mentioned above. A temperature of 700°C was necessary for single-crystalline growth, making it hardly an improvement over the e-beam grown material. An explanation for this difference is that, in our system, there is about a factor of two difference in the power attainable between the two wavelengths. The lack of power in the 355 nm pulses may create a situation where, even with no background oxygen, the substrate surface sees much more oxygen before the first monolayer of cerium is formed simply because it takes longer to grow a monolayer. This would also explain the successes of other groups in depositing CeO_2

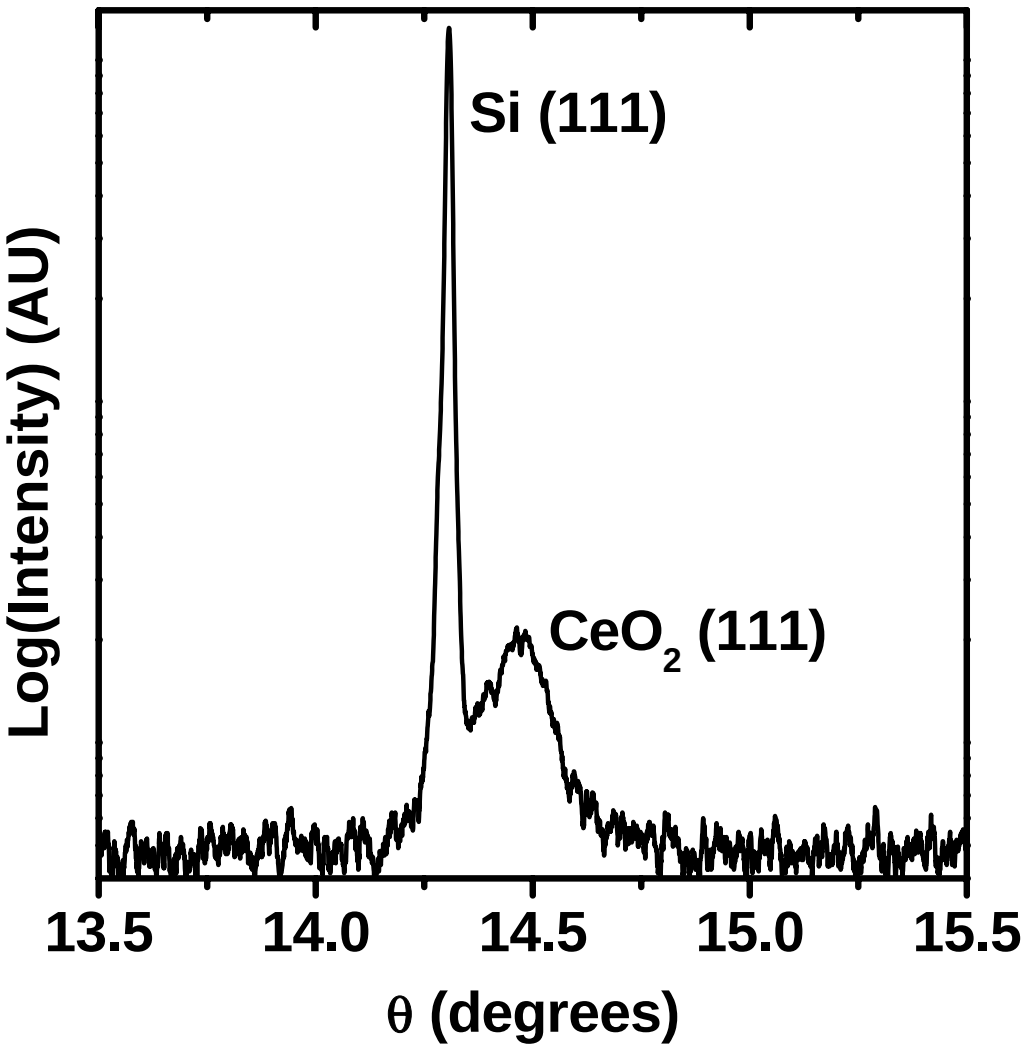


Figure 2.5: **XRD scan of a stoichiometric 700Å thin film of single-crystalline CeO_2 on silicon (111).**

with very high power excimer lasers operating even deeper in the ultraviolet [2].

2.5.1 PLD Growth Rates

Another difficulty with PLD is the lack of any dependable *in situ* growth rate monitor. The plume is so directional that a quartz crystal monitor would need to be extremely close to get a dependable reading and none of the techniques discussed above produced films smooth enough to make RHEED oscillations observable. However, if the growth rate is calibrated by other means, it will be very consistent since the laser pulses are very uniform in energy and shape over the course of hundreds of thousands of pulses. Thus two *ex situ* techniques were used to calibrate the growth rates. First, Fig. 2.7 shows x-ray reflectivity data for two samples grown with the 355 nm laser. The oscillations in the $\theta/2\theta$ scan for θ close to zero are interference fringes generated by the heterointerface and the sample surface. This method is very accurate because the x-rays hardly are absorbed or refracted at all over the thickness of the film. Thus, the thickness can be calculated from

$$t = \frac{n\lambda}{2(\sin\theta_1 - \sin\theta_2)} \sim \frac{n\lambda}{2\Delta\theta}. \tag{2.1}$$

The data in Fig. 2.7 suggests a growth rate of 0.58Å/s for growth with the 355 nm laser. A similar set of scans for samples grown with the 532 nm wavelength laser yielded a growth rate of 0.77Å/s. After these samples were measured via x-ray reflectivity, the known thickness value was used to help fit the data from spectroscopic ellipsometry. By taking several samples of known thicknesses, one is able to construct a spectrum of the optical constants for the CeO_2 films. Figure 2.8 shows the resulting data for PLD-grown CeO_2 films. Note that the sharp rise in the imaginary part of the refractive index is consistent with the theory proposed above for why PLD with 355 nm wavelength light produces much cleaner surfaces. This spectrum then could be used to fit data from films of unknown thickness. This technique is actually complementary to the x-ray reflectivity technique because films of much larger thickness can be analyzed with spectroscopic ellipsometry. Between the two, dependable thickness

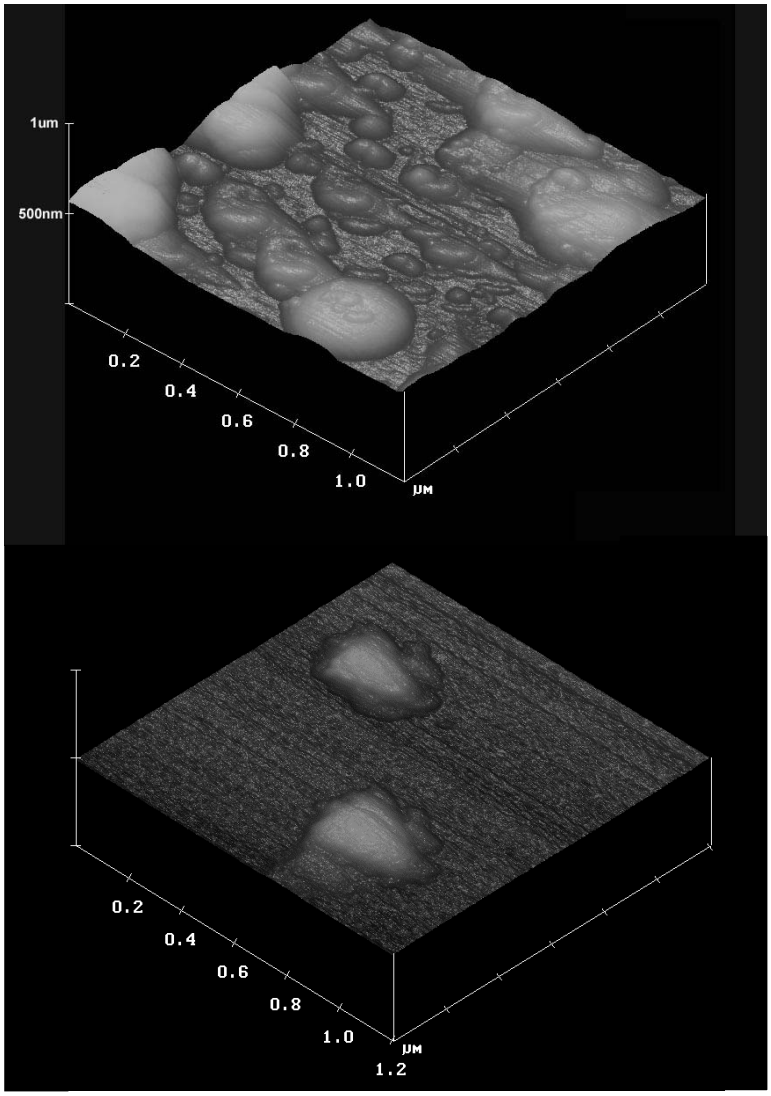


Figure 2.6: **AFM images of CeO_2 films grown by PLD at two different laser wavelengths. Figure (a) shows a film of CeO_2 grown with a laser wavelength of 532 nm. There is a high density of particulates ranging from tens of nanometers up to one micron. Figure (b) shows a film grown with a laser wavelength of 355 nm. All the small particulates have disappeared and the density of large particulates is much less.**

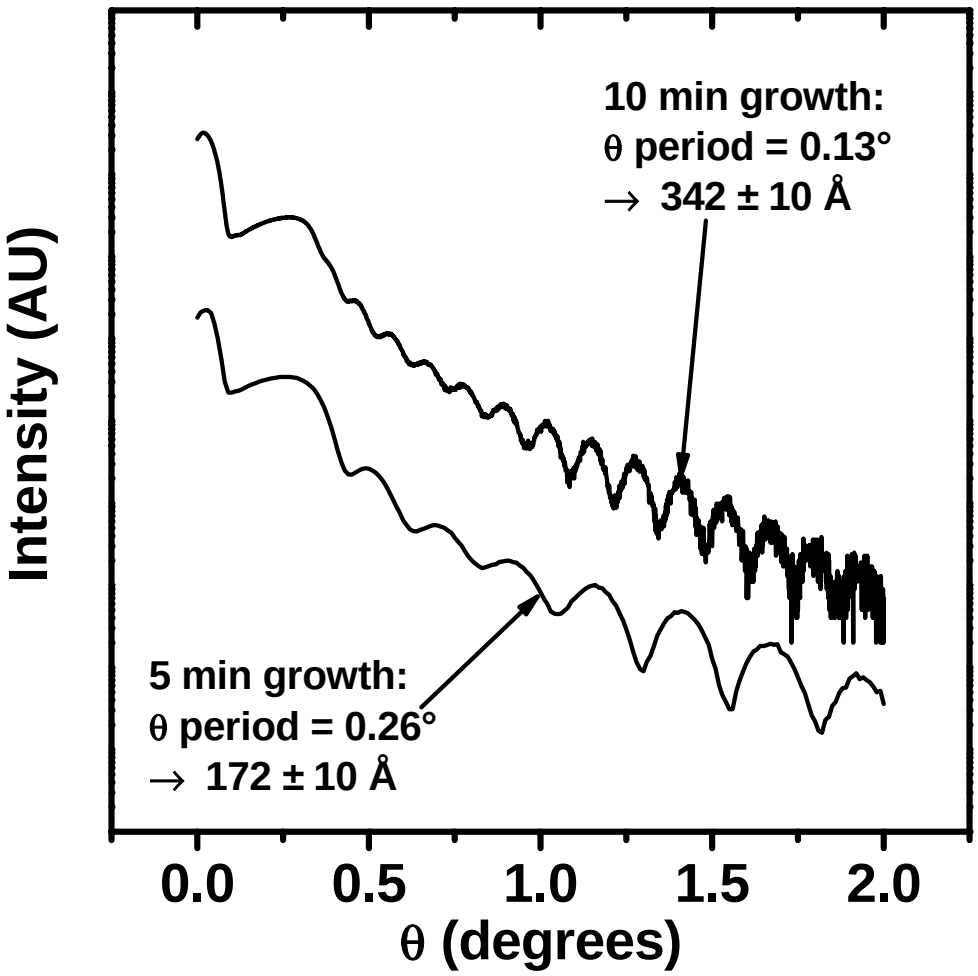


Figure 2.7: **X-Ray reflectivity for two CeO_2 samples of different thicknesses grown by PLD with the 355 nm wavelength laser.**

values could be found for all of our CeO_2 films.

2.6 Conclusion

To advance CeO_2 to the point where it can be used as a true heterojunction material for silicon systems, the issues discussed in this chapter will need to be addressed. Obtaining a true single-crystal thin film of stoichiometric CeO_2 on silicon proved to be a considerable challenge. While the final solution for growing epitaxial films of arbitrary thickness with good crystalline qualities and proper oxidation state never was obtained, adequate solutions were found for producing films that could be of use for various kinds of studies of the CeO_2/silicon system. Two examples of such studies are discussed in Chapters 3 and 4.

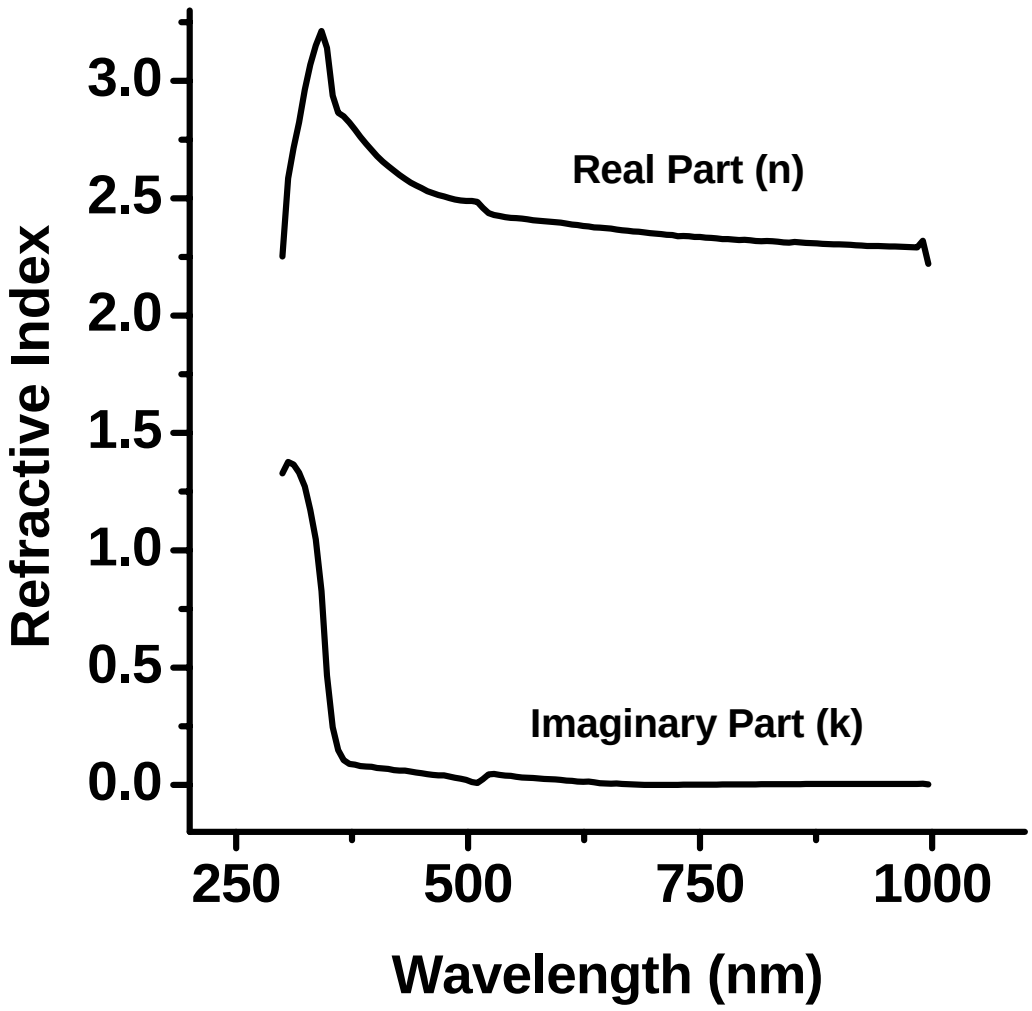


Figure 2.8: **Optical constants for PLD-grown CeO_2 films obtained from spectroscopic ellipsometry over the range 250 nm to 1000 nm. Note the steep rise in k below about 350 nm.**

Bibliography

[1] T. Inoue, Y. Yamamoto, S. Koyama, S. Suzuki, and Y. Ueda, Appl. Phys. Lett. **56**, 1332 (1990).

[2] M. Yoshimoto, K. Shimozono, T. Maeda, T. Ohnishi, M. Kumagai, T. Chikyow, O. Ishiyama, M. Shinohara, and H. Koinuma, Jpn. J. Appl. Phys. **34**, L688 (1995).

[3] H. Nagata, T. Tsukahara, S. Gonda, M. Yoshimoto, and H. Koinuma, Jpn. J. Appl. Phys. **30**, L1136 (1991).

[4] H. Koinuma, H. Nagata, T. Tsukahara, S. Gonda, and M. Yoshimoto, Appl. Phys. Lett. **58**, 2027 (1991).

[5] S. Yaegashi, T. Kurihara, H. Hoshi, and H. Segawa, Jpn. J. Appl. Phys. **33**, 270 (1994).

[6] D. D. Huang, F. G. Qin, Z. Y. Yao, Z. Z. Ren, L. Y. Lin, W. B. Gao, and Q. Y. Ren, Appl. Phys. Lett. **67**, 3724 (1995).

[7] M. Yoshimoto, H. Nagata, T. Tsukahara, and H. Koinuma, Jpn. J. Appl. Phys. **29**, L1199 (1990).

[8] R. Kelly and A. Miotello, in *Pulsed Laser Deposition of Thin Films*, edited by D. B. Chrisey and G. Hubler (John Wiley & Sons Inc., New York, NY, 1994).

[9] G. K. Hubler, in *Pulsed Laser Deposition of Thin Films*, edited by D. B. Chrisey and G. Hubler (John Wiley & Sons Inc., New York, NY, 1994).

[10] I. Riess, R. Koerner, M. Ricken, and J. Noelting, Solid State Ionics **28**, 539 (1988).

[11] L. C. Chen, in *Pulsed Laser Deposition of Thin Films*, edited by D. B. Chrisey and G. Hubler (John Wiley & Sons Inc., New York, NY, 1994).

[12] W. Kautek, B. Roas, and L. Schultz, Thin Solid Films **191**, 317 (1990).

[13] F. Marabelli and P. Wachter, Phys. Rev. B **36**, 1238 (1987).

[14] M. Niwano, S. Sato, T. Koide, T. Shidara, A. Fujimori, H. Fukutani, S. Shin, and M. Ishigame, J. Phys. Soc. Jpn. **57**, 1489 (1988).

[15] S. Guo, H. Arwin, S. N. Jacobsen, K. Jarrendahl, and U. Helmersson, J. Appl. Phys. **77**, 5369 (1995).

Chapter 3 XPS Studies of the CeO_2/Silicon Interface

3.1 Overview

The cerium oxide/silicon interface is studied using x-ray photoelectron spectroscopy (XPS). The oxidation and reduction of species at the interface are examined as a function of annealing temperature both in vacuum and oxygen ambient, in order to determine their relative stabilities. By depositing a very thin CeO_2 film ($\sim$30Å), the cerium and silicon core level peaks can be monitored simultaneously. The presence of characteristic chemical shifts of the Si 2p peak gives information about any SiO_{2-x} layer that may form at the interface. The oxidation state of the cerium can be probed from three different areas of the spectrum. From this information we can infer the oxidation state of both the silicon and the cerium.

3.2 Introduction

In all of the CeO_2 applications discussed in Section 1.2.1, a critical issue is the nature of the CeO_2/silicon interface. It has been shown that an SiO_{2-x} layer can form between the silicon and the CeO_2, and further that the CeO_2 sometimes can be reduced to CeO_{2-x} in the presence of silicon [1] (see Fig. 3.1). These effects are strongly dependent on temperature, oxygen partial pressure, and other growth conditions. The determination of the parameters governing the state of the layers in Fig. 3.1 are essential to delineating the range of application of this oxide.

Thermodynamics provides some guidance to the kinds of reactions that might take place. On the one hand, thermodynamics favors the formation of CeO_2 over the formation SiO_2. The Gibbs free energy of formation of CeO_2 from cerium metal

and O_2 gas is -21.08 eV/molecule (-2030 kJ/mol) at room temperature whereas for SiO_2 from silicon and O_2 it is -17.77 eV/molecule (-1712 kJ/mol) [2]. However, it is thermodynamically favored for silicon to reduce CeO_2 to its suboxide Ce_2O_3, while forming SiO_2 [3]. While these arguments may help indicate which reactions will or will not occur in bulk materials, an experimental study of what happens at a real interface is needed. Several studies, most recently Hirschauer et al. [4], have shed light on the subject by examining high resolution XPS of the silicon core levels and the cerium valence levels. In this chapter, information is obtained from almost every part of the XPS spectrum in order to produce a detailed picture of the interface.

It is confirmed that a silicon oxide layer, whose oxidation state depends on annealing conditions, forms underneath the cerium oxide overlayer (see Fig. 3.1). It is also determined that, when annealed in vacuum, the cerium oxide will reduce to Ce_2O_3 but no further, whereas it will fully oxidize to CeO_2 almost immediately when annealed in oxygen.

In Section 3.3, details are presented concerning the preparation of the samples and of the XPS experiment. In Section 3.4, the methods used in the analysis of the data are presented. These techniques include the deconvolution of the core spectra into to separate Gaussian peaks, the use of the attenuation of the spectrum by the inelastic collisions of the electrons to determine overlayer thicknesses, and an overall analysis of the valence band. Section 3.5 presents the results based on an interpretation of the information given by the core levels treated with the techniques mentioned above. The thermodynamic situation is clarified by comparing data from samples annealed at various temperatures and interpreting them in terms of the oxidation and reduction of the layers at the interface between the silicon and cerium oxide.

3.3 Experiment

Two samples were prepared by UHV (base pressure 3×10^{-10} torr) electron beam evaporation of a solid CeO_2 source onto polished three inch Si(111) wafers in a Perkin Elmer silicon MBE system. At the time of this study, the PLD apparatus was not

yet operational and hence the e-beam deposition method was used. While the growth method may affect the initial state of the sample, it should have little effect on the results of the subsequent annealing steps.

Wafers were prepared as described in Section 2.3 and thin layers of silicon were grown by MBE also as described in Section 2.3. The CeO_2 growth was performed at low temperature ($\sim$400°C) and without any background oxygen flow in order to minimize oxidation of the substrate prior to growth. However, the background pressure in the chamber due to oxygen outgassing from the source was close to 1×10^{-5} torr. Thickness was determined by a quartz crystal monitor to be approximately 30Å for both samples. Samples were also monitored by reflection high energy electron diffraction (RHEED) which showed a transformation from a typical Si (7×7) reconstruction to polycrystalline rings almost immediately after the source shutter was opened. The samples then were transferred in UHV to an analysis chamber for performing XPS and back to the growth chamber for subsequent annealing. The analysis chamber is equipped with a monochromated AlKα x-ray source and a Perkin Elmer model 10-360 hemispherical analyzer with a 16 channel electron multiplier detector with a minimum resolution of 0.6 eV. The samples were radiatively heated for 30 minutes at a time at temperatures from 300°C to 900°C. One was always annealed in vacuum and the other in 1×10^{-5} torr of O_2 to simulate growth conditions as described above. The temperatures were monitored by optical pyrometer in the range of 400°C to 800°C which is the pyrometer's range of validity and by thermocouple otherwise. It has been shown that exposure to AlKα radiation can itself reduce CeO_2 [5], so the Ce 3d scans of each data set were performed first. This being said, there was no noticeable change in the oxidation state before and after exposure.

3.4 Analysis Techniques and Fitting of Core Levels

Standard analysis techniques have been applied to the data for each of the core levels (Si 2p, Ce 3d, O 1s) and the valence band edge. In each case, this analysis was used to provide information on variations in the state of the sample in Fig. 3.1.

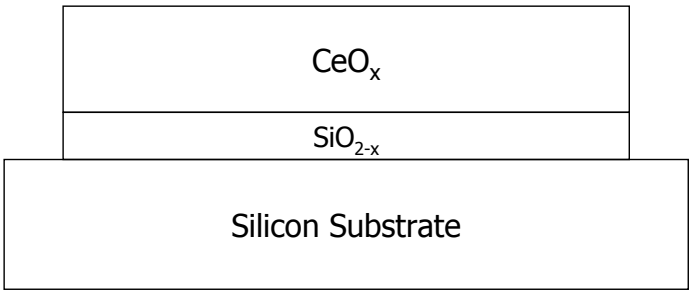


Figure 3.1: **A schematic of the samples used for the experiments in this chapter. Only CeO_2 is grown intentionally. The rest of the layers form as result of subsequent interfacial reactions which will be described throughout this chapter.**

Before examining the individual areas of the XPS spectrum, the methods used to obtain information from the raw data are described here. A Shirley background subtraction [6, 7] was performed on all the raw data. After this, peaks either were integrated to determine their spectral weight or fit with Gaussian peaks to deconvolve multiple peaks that are close together in binding energy (BE). The spectral weights of all the peaks then were used to determine either chemical information or the thickness of different layers of material.

To determine a thickness, the attenuation of the signal from electrons emitted from a buried layer by inelastic collisions with atoms in an overlayer is used. If the peak from electrons in the buried layer is chemically shifted to a different BE than the peak for electrons in the overlayer, the area of the two peaks can be compared in the same scan. The ratio of the area of the overlayer peak to the area of the buried layer peak should thus be related to the thickness of the overlayer. More precisely we use an exponential extinction law where the signal intensity from photoelectrons traveling through various materials decays as

$$I = I_0 \exp\left[-\frac{x}{\lambda}\right], \tag{3.1}$$

where I_0 is related to the cross section for the process that creates the electron, x is the path length the electron has to traverse to leave the sample and λ is the electron inelastic mean free path (IMFP).

As an example, the case of the Si 2p peaks is addressed. There are two peaks in question, one from the silicon in the substrate and one from the silicon in the SiO_{2-x} layer immediately on top of the substrate. The signal from the silicon in the substrate, after traveling through the SiO_{2-x} and CeO_x layers is given by

$$I_{Si} = \exp\left(-\left(\frac{t_{SiO_{2-x}}}{\lambda_{SiO_{2-x}}} + \frac{t_{CeO_x}}{\lambda_{CeO_x}}\right)\right) \times \int_0^{t_{Si}\to\infty} I_0 \exp\left(\frac{-x}{\lambda_{Si}}\right) dx, \tag{3.2}$$

where λ_{Si}, $\lambda_{SiO_{2-x}}$, and λ_{CeO_x} are the IMFP's of electrons in Si, SiO_{2-x} , and CeO_x respectively as described in Table 3.1, and the t's are the path lengths that the

photoelectrons must travel through each material. The strength of the signal from the silicon in the SiO_{2-x} layer will be

$$I_{SiO_{2-x}} = \exp\left(-\frac{t_{CeO_x}}{\lambda_{CeO_x}}\right) \times \int_0^{t_{SiO_{2-x}}} I_0 \exp\left(\frac{-x}{\lambda_{SiO_{2-x}}}\right) dx. \tag{3.3}$$

It is assumed that $\lambda_{SiO_{2-x}} \sim \lambda_{SiO_2}$ and furthermore that the cross sections for both processes (i.e., I_0) are the same since both peaks come from electrons emitted from silicon atoms. With these assumptions, and taking into account a 45° angle between the plane of the sample and the detector, the final expression for the thickness of the SiO_{2-x} layer becomes

$$t_{SiO_{2-x}} = \frac{\lambda_{SiO_{2-x}}}{\sqrt{2}} \times \ln\left(1 + \frac{\lambda_{Si}}{\mathbf{r} \times \lambda_{SiO_{2-x}}}\right), \tag{3.4}$$

where **r** is the ratio of the area of the SiO_{2-x} peak to the area of the silicon substrate peak. Throughout this chapter, values of λ for the cerium species were calculated from the expressions given by Tanuma et al. [8], and the values for silicon species were obtained directly from the literature [9]. All the values used in this study are summarized in Table 3.1.

3.4.1 Silicon 2p Core Level

An example of a Si 2p spectrum is shown in Fig. 3.2 and the peaks used for fitting are summarized in Table 3.2. The Si 2p spectrum was used to gain the following information. First, the position of the Si 2p substrate peak was measured in the sample after each annealing step and compared to data from a bare silicon (111) wafer to be used as a calibration point for the energy scale of all spectra. Secondly, the large chemical shifts to higher binding energy of silicon atoms in different oxidation states create an obvious satellite peak next to the substrate peak. The relative weights of these two peaks then can be used to monitor the growth of SiO_{2-x} species created at the interface. It is acknowledged here that what is being called SiO_{2-x} may actually have a finite amount of cerium in it [10, 11], but it should still be reasonable to assume

Overlayer Material	Photoelectron Source		
	Si 2p	Oxygen 1s	Ce 3d
Silicon	30	-	-
SiO_2	36	27	-
CeO_2	23.8	17.9	10.9
Ce_2O_3	27	20.3	12.3

Table 3.1: **Electron inelastic mean free paths, λ in Å. The photoelectron source indicates which type of atom created the photoelectron and the overlayer material indicates the type of material that the electrons will subsequently have to travel through on their way out of the sample. The data for silicon and SiO_2 overlayers are from reference [9], and the data for the cerium species are calculated from the expression given by Tanuma et al. [8]**

that the spectrum can be modeled by examining different oxidation states of silicon. Finally, by deconvolving the Si 2p spectrum, information can be obtained concerning the oxidation state of the silicon in the SiO_{2-x} .

Good fits (reduced $\chi^2 \leq 2$) were found for most spectra by varying only the height and full width at half maximum (FWHM) of these five Gaussian peaks, and letting the positions be determined by the known chemical shifts from the literature [12, 13]. Once the fits were performed, the area of oxidized silicon species (the sum of peaks C,D,and E in Fig. 3.2) was divided by the area of the non-oxidized species (sum of peaks A and B) to give the ratio **r** required in Eq. 3.4.

The relative weights of the four oxidized silicon peaks then were used to determine an average oxidation state for the SiO_{2-x} layer. Each of the oxidized peaks actually represents a doublet like the substrate peak. However, fitting each shifted component as a doublet introduces so many fitting parameters that the accuracy of the fit may be compromised. Hence, each shifted peak was modeled as one broadened Gaussian. With this assumption, the oxidation state is given by,

$$\text{Avg. Oxidation State of } SiO_{2-x} = \frac{2C + 3D + 4E}{C + D + E}. \tag{3.5}$$

3.4.2 Cerium 3d Core Level

Changes in the oxidation state of the cerium oxide were observed by analyzing the Ce 3d spectrum. The Ce 3d spectrum of CeO_2 is known for its complexity. Besides the spin-orbit splitting of the Ce $3d_{5/2}$ and Ce $3d_{3/2}$ peaks, there are several other splittings that are caused by a redistribution of the entire energy spectrum after a core hole is created. This phenomenon is discussed in detail by Fujimori [14, 15, 16] and it is not examined further here.

Table 3.3 summarizes the 10 peaks of the spectrum in the notation of Burroughs et al. [17] that are used for fitting, and Fig. 3.3 shows two sample spectra with the fitted peaks. There are four peaks (v_0, v', u_0, u') that are derived from cerium in the (III) oxidation state, and the other six are from the (IV) oxidation state [18]. By

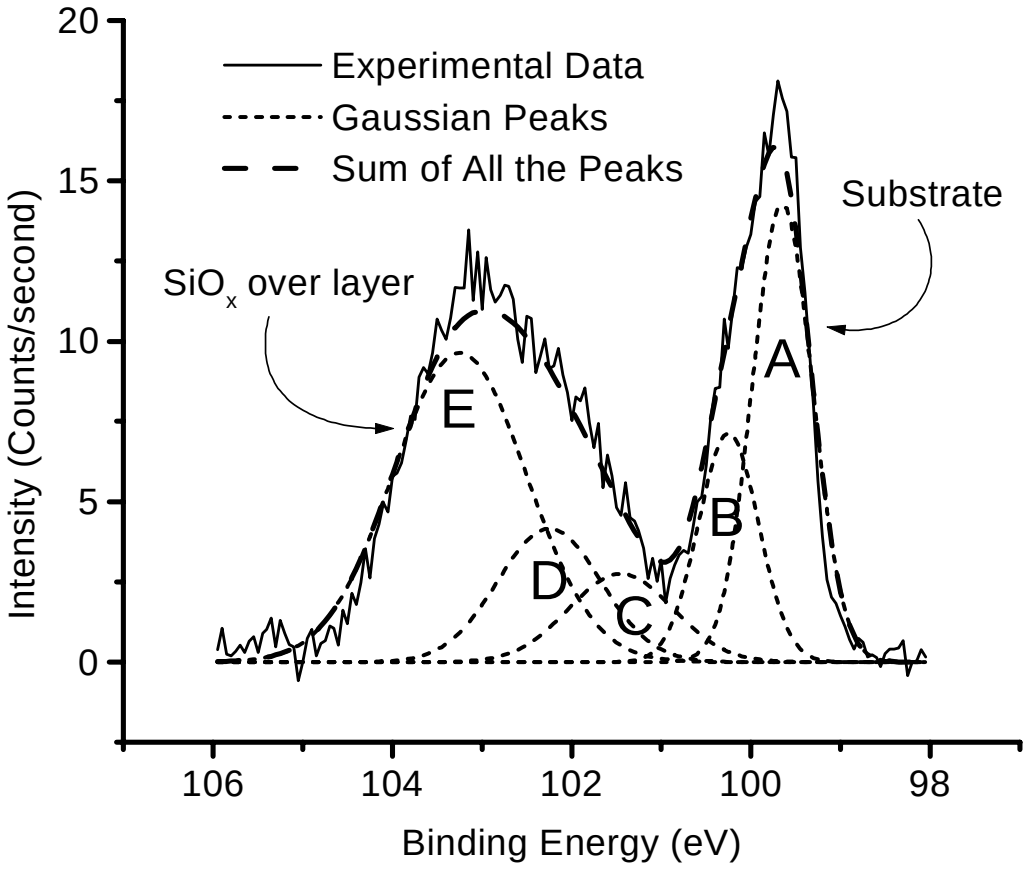


Figure 3.2: **An example of experimental data (solid line), fitted peaks (dots), and sum of the fitted peaks (dashes) for the Si 2p spectrum. The Si 2p peaks have been deconvolved into five Gaussian peaks A–E (see Table 3.2 for description of peaks, the Si^{+1} peak was too small to be included here). The heavy dashed line is the sum of the contribution of peaks A–E and hence represents an indication of the overall quality of the fit.**

Label in Fig. 3.2	A	B	-	C	D	E
Peak Name	$2p_{3/2}$	$2p_{1/2}$	Si^{1+}	Si^{2+}	Si^{3+}	Si^{4+}
Relative Shift (eV)	0	0.6	0.9	1.8	2.6	3.6

Table 3.2: **Assignment of the Si 2p Gaussian peaks. A and B are from bare Si. C–E are oxidized states of Si. It is to be understood that each individual oxidized peak represents a double peak similar to the substrate doublet. Chemical shifts are taken from Shallenberger [12].**

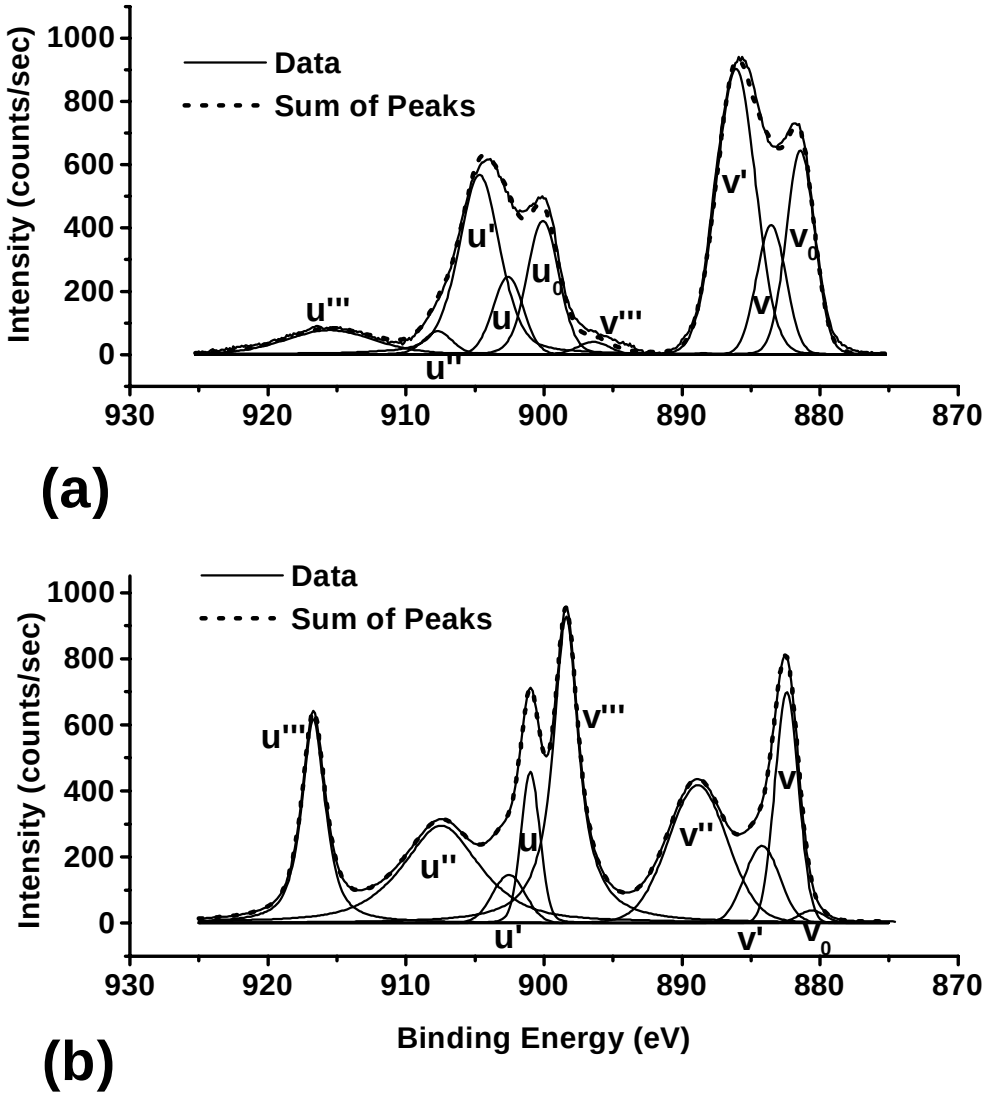


Figure 3.3: **Examples of experimental data of Ce 3d spectra and their deconvolution into Gaussian peaks as described in Table 3.3. The labeled peaks are the individual Gaussian peaks used to fit the spectrum and the dashed line is the sum of them which should simulate the experimental data. Figure (a) shows a reduced spectrum close to Ce_2O_3 and Fig. (b) shows a well oxidized spectrum close to CeO_2.**

comparing the relative spectral weights of these peaks in the Ce 3d spectrum, the average oxidation state can be obtained. This is essentially the method of Romeo et al. [19] It recently has been pointed out by Holgado et al. [20] that a more accurate way to accomplish this would be via factor analysis, but for the purposes of this study, Romeo's technique is simple and accurate enough. First the whole Ce 3d spectrum is fit to these 10 peaks. Hopefully the entire spectrum is well represented by appropriate weighting of just these 10 peaks. The heights, FWHM's, and positions of the peaks are varied to obtain the best fit. Once the least squares fit is obtained, the individual spectral weights are added to determine the total weight of Ce(III),

$$\text{Ce(III)} = v_0 + v' + u_0 + u', \tag{3.6}$$

and Ce(IV),

$$\text{Ce(IV)} = v + v'' + v''' + u + u'' + u'''. \tag{3.7}$$

Then the total percentage of the cerium in the CeO_x layer that is in the (III) oxidation state is determined using the relation

$$\%\text{Ce(III) in Ce 3d} = \frac{100 \times \text{Ce(III)}}{\text{Ce(IV)}+\text{Ce(III)}}. \tag{3.8}$$

3.4.3 Oxygen 1s Core Level

The oxygen peaks were used both as another source of information about the cerium oxidation state and as a source of information for the thickness of the cerium oxide. Compared to the cerium spectrum there has been considerably less study devoted to the oxygen spectrum in CeO_2 [21]. It is clear that the 1s peak from oxygen atoms in a SiO_2 matrix should be at higher BE than CeO_2 because silicon has a higher electronegativity (1.9 on the Pauling scale) than cerium (1.12). The higher electronegativity atoms pull more electron density away from the oxygen nuclei and hence reduce the shielding of the electrons from the nuclear charge, thus increasing the binding energy of the electrons. Further, one would expect that 1s electrons in

		v_0	v	v'	v"	v"'
Ce $3d_{5/2}$	Origin	Ce(III)	Ce(IV)	Ce(III)	Ce(IV)	Ce(IV)
	Shift (eV)	-36.1	-34.1	-30	-27.85	-18.3
		u_0	u	u'	u"	u"'
Ce $3d_{3/2}$	Origin	Ce(III)	Ce(IV)	Ce(III)	Ce(IV)	Ce(IV)
	Shift (eV)	-17.8	-15.65	-13.65	-9.25	0

Table 3.3: **Assignments of Ce 3d peaks with u,v notation from Burroughs [17]. Shifts represent initial guesses for positions of the peaks with respect to the highest BE peak for the fitting process.**

oxygen attached to cerium atoms in the (III) oxidation state would be more tightly bound than for cerium in the (IV) state. Thus it is possible to deconvolute the oxygen 1s spectrum into contributions from Si-O, Ce(III)-O, and Ce(IV)-O bonded species.

Figure 3.4 shows that indeed the spectrum can be well fit by three Gaussian peaks. The height, FWHM and position were all varied to obtain the fit. Peak C is attributed to SiO_{2-x} species, not surface adsorbed oxygen because it is slightly more shifted in BE than the peak attributed to chemisorbed hydroxide species identified by Sundaram et al. [21] and because it grows so substantially during oxygen annealing. Peak A is from Ce(IV) species, and peak B is from Ce(III) species. It was found that, in general, the peak due to Ce(III) was just over 1eV higher in binding energy that the Ce(IV) peak.

Examining the relative areas of peaks A and B gives another indication of the percentage of the CeO_x that is in the Ce(III) state,

$$\%\text{Ce(III) in O 1s spectra} = \frac{100 \times \text{B}}{\text{A+B}}. \tag{3.9}$$

Meanwhile, an equation similar to Eq. 3.4 can be used to determine the total CeO_x overlayer thickness:

$$t_{CeO_x} = \frac{\lambda_{CeO_x}}{\sqrt{2}} \ln\left(1 + \frac{\lambda_{SiO_{2-x}}\left[1 - \exp\left(-\frac{\sqrt{2}t_{SiO_{2-x}}}{\lambda_{SiO_{2-x}}}\right)\right]}{\mathbf{r} \times \lambda_{CeO_x}}\right), \tag{3.10}$$

where this time **r** is the ratio of the area of the silicon-related peak, C, divided by the sum of the areas of the Ce-related peaks, A and B.

3.4.4 Valence Band

Observations of the valence band are used as a third method of monitoring the oxidation state of the cerium. As CeO_2 is reduced to Ce_2O_3, one of the most obvious feature changes in the entire XPS spectrum is the appearance of the Ce 4f level at the edge of the valence band [22, 23, 24, 25]. Because of the extremely low cross section

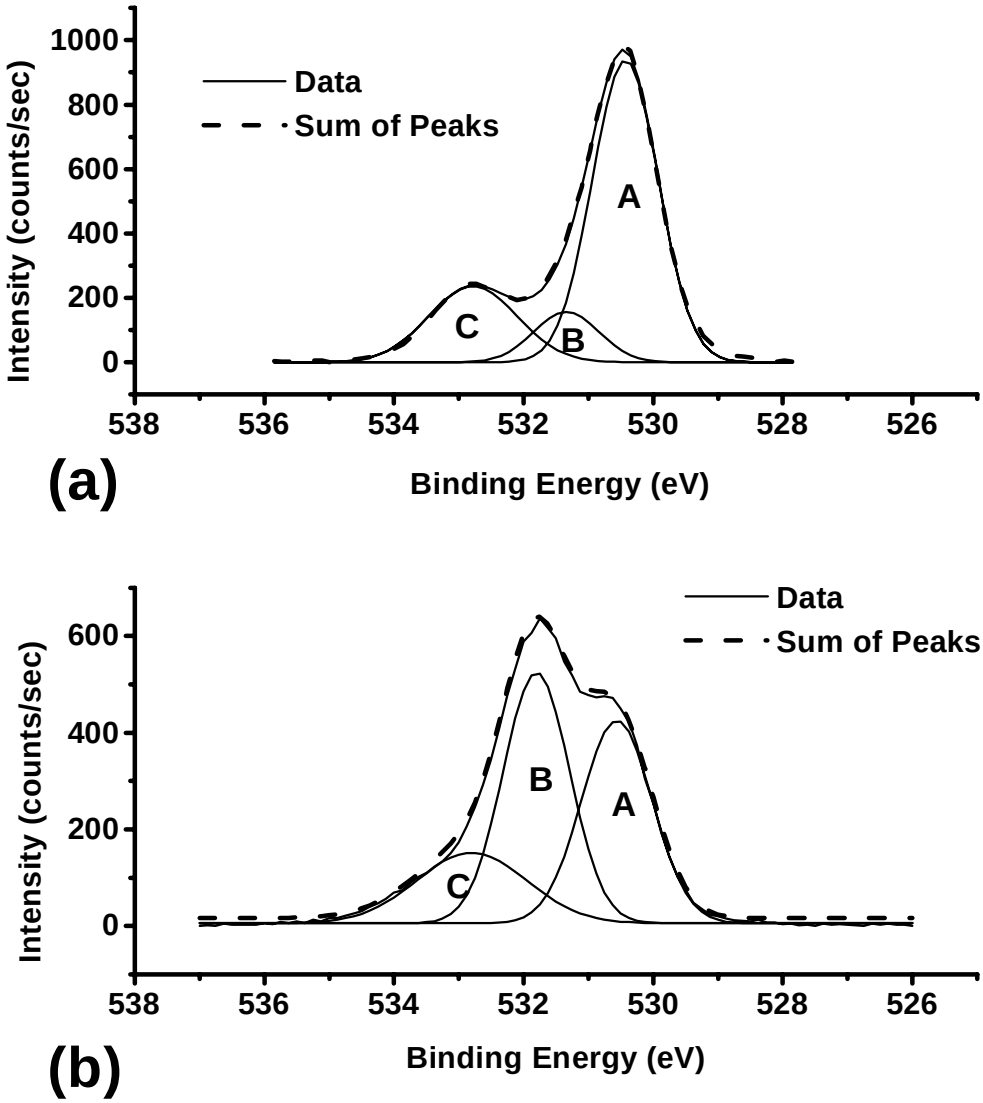


Figure 3.4: **Examples of experimental data of O 1s spectra and their deconvolution into Gaussian peaks. The labeled peaks are the individual Gaussian peaks used to fit the spectrum and the dashed line is the sum of them which should simulate the experimental data. Peak A is from oxygen bonded to Ce(IV) atoms, peak B is from oxygen bonded to Ce(III) atoms and peak C is from oxygen bonded to Si. Figure (a) shows a well oxidized spectrum after annealing at 700°C in O_2. Figure (b) shows a reduced spectrum after annealing at 620°C in vacuum.**

for creating valence band photoelectrons, our x-ray source was not intense enough to enable the resolution of fine features, but changes in the shape due to the mixing of this 4f level easily could be observed. While it is unclear that there is any direct linear correlation between the strength of this feature and the oxidation state of the cerium, the valence band spectra of the same samples discussed above are presented as a reality check.

Figure 3.5 shows normalized versions of all the valence band spectra taken for the sample annealed in vacuum. Only one of the spectra from the sample annealed in O_2 is shown because there was hardly any change throughout the annealing. It is shown with the others as a reference point for what the valence band should look like for almost fully oxidized CeO_2. Obvious features appear on either side of a steady central band, which is assumed to be due mostly to O 2p valence electrons. These features can be attributed to an increase in mixing of the Ce 4f state (which is in the middle of the bandgap for CeO_2 [22]) with the valence band and perhaps somewhat to Si-O bonding.

3.5 Results

The results are shown in Figs. 3.6-3.10. The spectrum has been interpreted in terms of the oxide thicknesses and oxidation state of the silicon and cerium after various anneals. All anneals were executed for a fixed time of 30 minutes. At the end a detailed model of the evolution of the interface is presented.

Equation 3.4 was used to determine the thickness of the SiO_{2-x} layer that forms between the cerium oxide and the substrate (see Fig. 3.1). A summary of the growth of this oxide is shown in Fig. 3.6 and the corresponding variation of the oxidation state is shown in of Fig. 3.7.

The trends for anneals done in oxygen are fairly clear. The SiO_{2-x} grows steadily underneath the CeO_x and the oxidation state moves towards (IV), indicating fully oxidized SiO_2. It should be noted that the rate of SiO_{2-x} growth is actually comparable to that reported for oxidation of bare silicon in dry O_2 [26]. In other words,

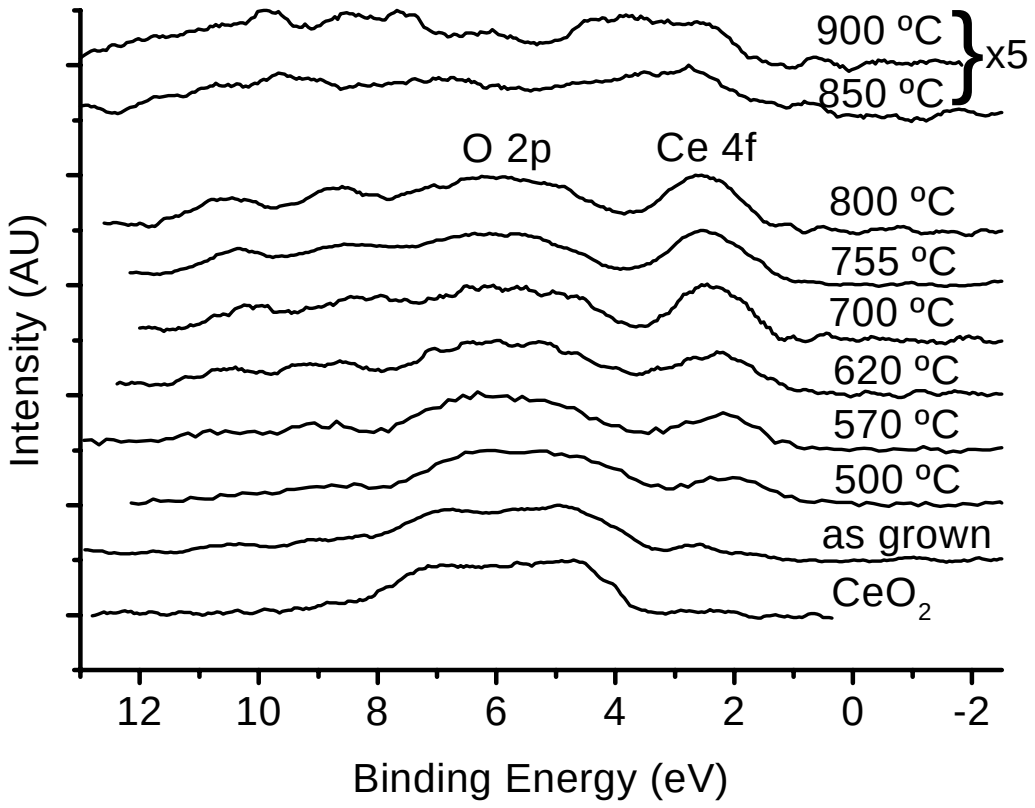


Figure 3.5: **The valence band spectra for the sample annealed in vacuum at all temperatures. One fully oxidized CeO_2 valence band is shown on the bottom for comparison. The top two spectra show the valence band after the desorption of oxygen and cerium that occurs for anneals in vacuum past 850°C. All anneals were for 30 minutes.**

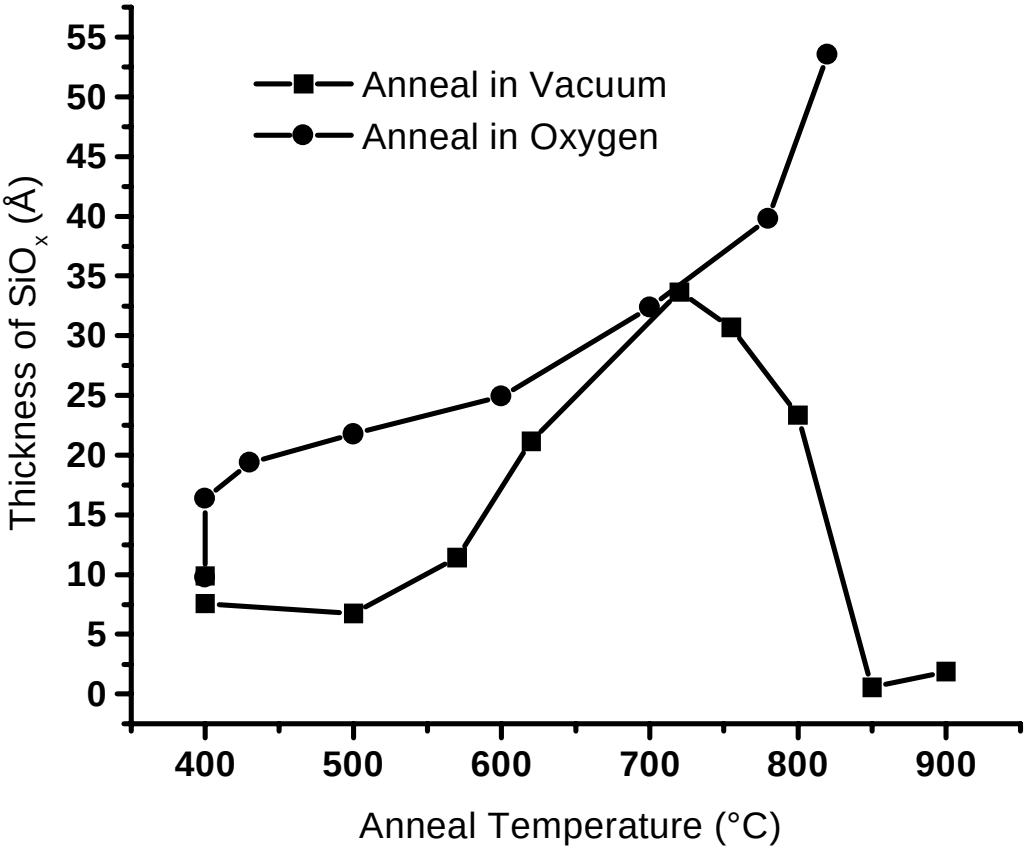


Figure 3.6: **The measured thickness of the SiO_{2-x} layer versus anneal temperature as determined from Eq. 3.4. The anneals were carried out both in vacuum (squares) and in O_2 (circles). All anneals were for 30 minutes.**

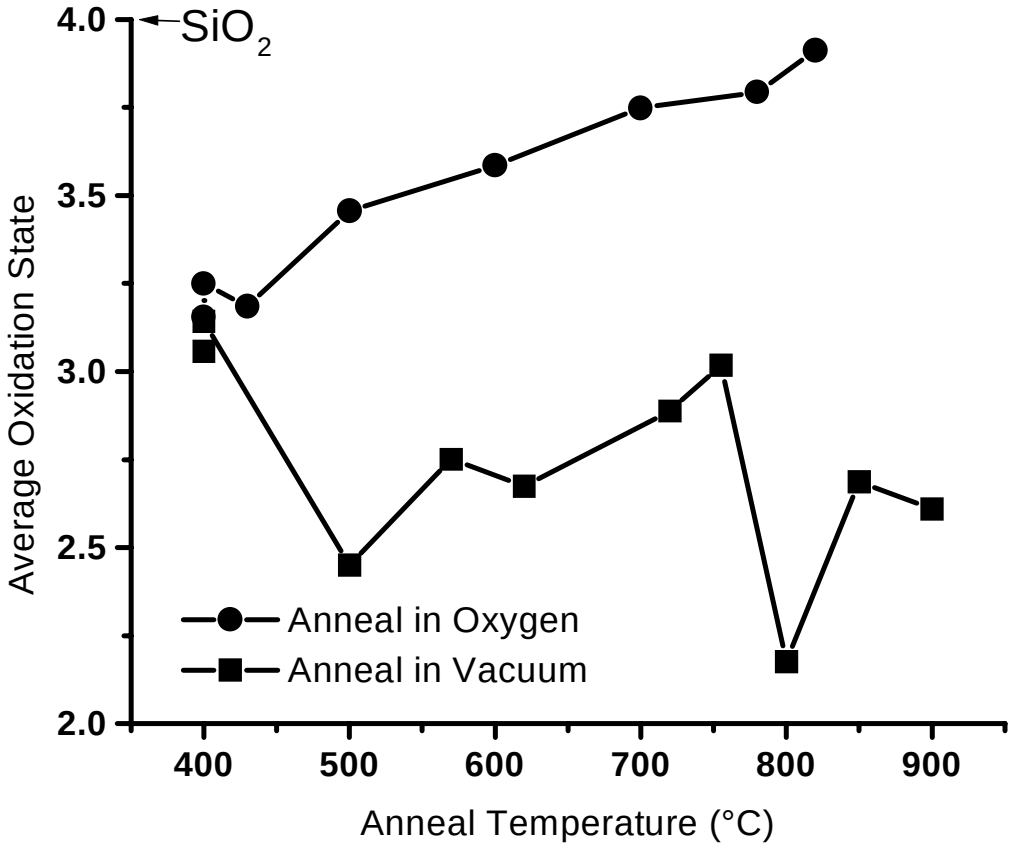


Figure 3.7: **The measured average oxidation state of the SiO_{2-x} versus anneal temperature as calculated from Eq. 3.5. The anneals were carried out both in vacuum (squares) and in O_2 (circles). All anneals were carried out for 30 minutes.**

the presence of CeO_x on the surface does not seem to increase the uptake of oxygen by the substrate nearly as much as for pure cerium metal as noted by Hillebrecht et al. [10]

The trends for the anneals done in vacuum, however, are very complicated. Three temperature regimes are apparent.

First, the thickness of the SiO_{2-x} stays constant or decreases slightly up to about 500°C. It will be shown later that this corresponds to a range in which the CeO_x overlayer is being reduced. The silicon oxidation state also decreases in this regime as shown in Fig. 3.7. It would thus seem that at this stage both the silicon and cerium oxides are losing oxygen. A possible explanation for this is that neither the cerium nor the silicon is in a fully oxidized or stable state. The silicon is, on average, in the (III) oxidation state (see Fig. 3.7) and cerium is somewhere in between (III) and (IV) (see Fig. 3.8). These are obviously metastable states; less thermodynamically stable than the full oxides, which might mean that some of the oxygen is not bound well enough to prevent outgassing.

Second, from 500°C to 720°C, the SiO_{2-x} grows rapidly. In this stage the silicon clearly is reducing the cerium because the SiO_{2-x} is growing at the expense of the oxidation state of the cerium (see Fig. 3.8). Except for perhaps an anomalous data point at 620°C, the oxidation state of the silicon increases as the layer grows but never reaches the full (IV) oxidation state of SiO_2.

Third, at 720°C and above, the SiO_{2-x} begins to recede until it totally disappears at 850°C. All of the oxygen and most of the cerium in the CeO_x overlayer also leave the surface in this range. This behavior has been reported earlier by Hirschauer et al. [4, 27], but it is unclear whether this is the same decomposition that Yamamoto et al. [28] mention occuring at 690°C for *ex situ* annealing of CeO_2 thin films. We also noted two RHEED reconstructions during this regime. At 850°C, the surface looks very much like a Si (111), 1×1 surface which could be the substrate with the unoxidized cerium that is left simply conforming to the underlying lattice. At 900°C, at least two distinct periodicities appear in the RHEED pattern which might be the Ce 2×2 and $\sqrt{3} \times \sqrt{3}$ reconstructions mentioned by Hirschauer [4]. More study

clearly needs to be done on this phenomenon to quantify it. It should be pointed out here that subsequent e-beam CeO_2 growth on the interface at this point was very successful, showing a strong CeO_2 1x1 RHEED pattern. So whatever species are left at the interface after this transition may actually facilitate further growth.

The oxidation state of the cerium is also important in giving a clear picture of the state of the layers. First the data from the Ce 3d peaks is used to determine the oxidation state. Figure 3.8 shows the percentage of material that is in the Ce(III) state, as calculated from Eq. 3.8, as a function of anneal temperature in both vacuum and O_2. If it is assumed that all Ce(III) comes from material with a stoichiometry of Ce_2O_3, and that all Ce(IV) comes from material that is CeO_2, a chart of CeO_x stoichiometry versus anneal temperature can be produced, which is the alternate ordinate in Fig. 3.8.

Again the trend for the sample annealed in O_2 ambient is clear. After one anneal at growth temperature (400°C) the stoichiometry improves dramatically to about $CeO_{1.93}$. From there it improves only slightly with further anneals, but it should be noted that even with the SiO_{2-x} growing underneath, there is no significant reduction of the cerium. The fact that it never reaches fully oxidized CeO_2 shows either that this method of determining oxidation state is slightly inaccurate or that even fully stoichiometric CeO_2 contains some finite amount of Ce(III) character. It also should be noted that the cerium saturates to a (IV) oxidation state faster than the silicon does (cf. Fig. 3.7), which indicates that underoxidized cerium perhaps has a higher oxygen affinity than underoxidized silicon.

The trend in the oxidation state of the CeO_x for anneals done in vacuum agrees with the oxidation state trend shown above for the SiO_{2-x} oxidation state. That is, very little change is noticed until after 500°C at which point the cerium reduces rapidly. One 30 minute anneal at 570°C reduces the cerium from 60% Ce(IV) to 32%. It continues to reduce only until 620°C where one would assume that the stoichiometry has reached Ce_2O_3. However, as a converse to the case above, the data suggests that even in fully stoichiometric Ce_2O_3 there is some Ce(IV) character.

The oxidation state of the cerium also can be calculated from the O 1s peaks using

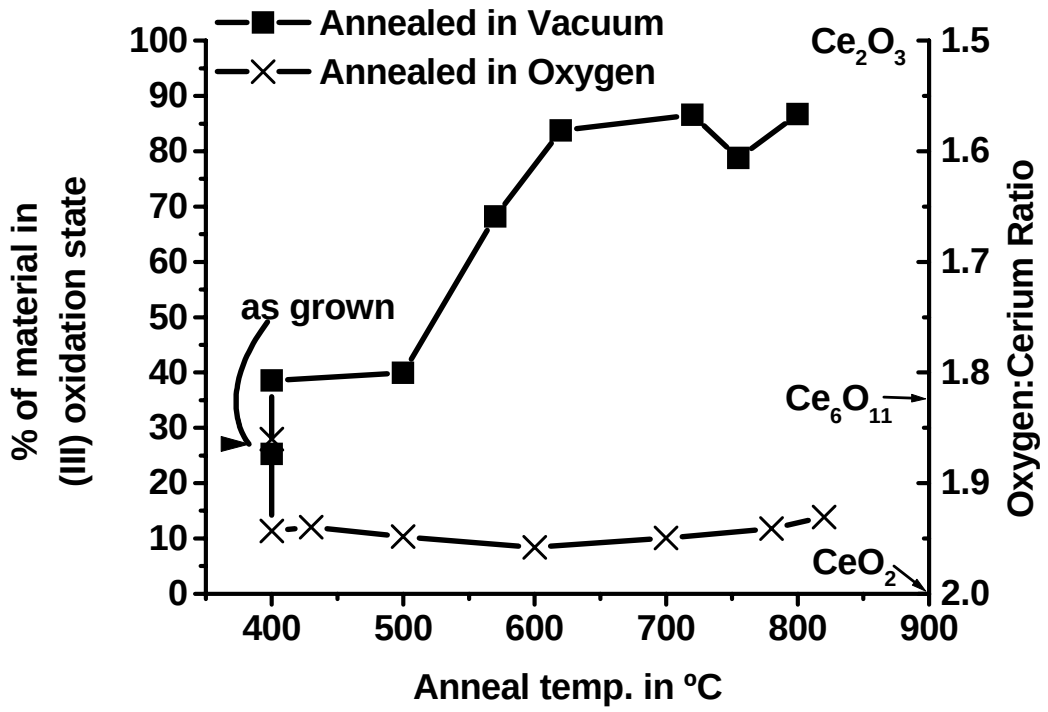


Figure 3.8: **The cerium oxidation state versus anneal temperature as calculated from the Ce 3d spectrum. The ordinate is $\frac{100\times \mathrm{Ce(III)}}{\mathrm{Ce(III)+Ce(IV)}}$ on the left and $\frac{1.5\ \mathrm{Ce(III)+2.0\ Ce(IV)}}{\mathrm{Ce(III)+Ce(IV)}}$ on the right, with quantities defined in Eq. 3.6 and Eq. 3.7. The squares are data from the sample annealed in vacuum. The crosses are for the sample annealed in O_2. All anneals were carried out for 30 minutes.**

Eq. 3.9. Figure 3.9 again shows the percentage of CeO_x that is in the (III) oxidation state as a function of anneal temperature in vacuum and in O_2. In this figure, the oxidation state as derived from the Ce 3d spectrum (Equation 3.8) is shown along with the oxidation state as derived from the O 1s spectrum (Equation 3.9). The upper panel labeled (a) is for the anneal in vacuum and the lower panel labeled (b) is for the anneal in O_2.

For the sample annealed in vacuum (Fig. 3.9 (a)), the two methods follow the same trend but shifted on the anneal temperature axis. An explanation for this is that the method of using the Ce 3d peaks is far more surface sensitive than the method using the O 1s peaks. This is because the IMFP of electrons emitted from Ce 3d levels is far smaller than for those emitted from O 1s levels (see Table 3.1). Thus there is both a surface sensitive and bulk sensitive probe for the oxidation state. This data would suggest that the cerium reduces from the *top* down rather than first at the interface then proceeding upwards towards the surface because the Ce 3d data shows the reduction before the O 1s data. Further, one can see that both methods show that the amount of cerium reduction saturates. This is most likely the point at which the material is completely reduced to Ce_2O_3 since there is no evidence for any stable cerium oxides with less oxygen [29]. The strange desorption of oxygen and cerium that occurs past 800°C obscures any further effects that might have been observed at higher temperatures.

For the sample annealed in O_2 (Fig. 3.9 (b)), it first should be pointed out that the scale is much smaller, i.e., that all the changes in oxidation state are on a smaller scale by almost an order of magnitude than the changes for the sample annealed in vacuum, so these curves are much more susceptible to error introduced by uncertainty in the fitting process. Nevertheless, it is striking that the two trends look to be almost completely opposite. This would seem to indicate that at first the surface is oxidizing while the region near the interface is being reduced slightly by the silicon. However, past 600°C the trend reverses and the surface begins to reduce slightly, possibly in the same fashion as the sample annealed in vacuum. Still there are several strange fluctuations along the two paths, so it is difficult to know what is going on in this

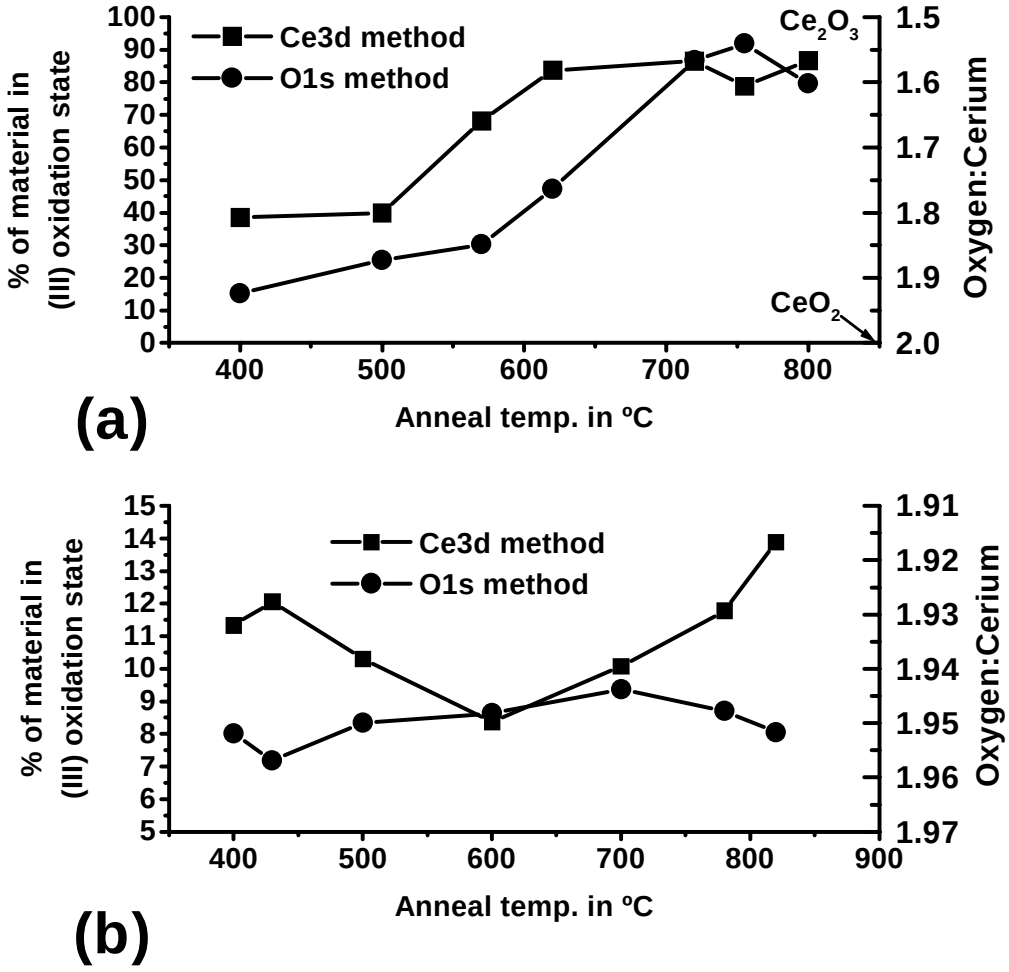


Figure 3.9: **The oxidation state as determined from the O 1s core levels (circles) versus anneal temperature. The ordinate is $\frac{100\times \mathrm{Ce(III)}}{\mathrm{Ce(III)+Ce(IV)}}$ on the left and $\frac{1.5\ \mathrm{Ce(III)+2.0\ Ce(IV)}}{\mathrm{Ce(III)+Ce(IV)}}$ on the right. The data from Fig. 3.8 is also plotted for comparison (squares). Figure (a) is from the sample annealed in vacuum, (b) is from the sample annealed in O_2. Please note the very different scale used on the ordinate in this plot (b) versus that in plot (a). All anneals were for 30 minutes.**

sample with the same confidence as with the vacuum-annealed sample.

Figure 3.10 shows real-space pictures of what happens at the interface for both samples during annealing. The total thickness of the CeO_x overlayer can be computed from the O 1s peaks using Eq. 3.10, and the thickness of the SiO_{2-x} has already been computed from Eq. 3.4. It is assumed that 45% of the SiO_{2-x} growth is down into the substrate, as in the Deal-Grove model. We use the information about both thicknesses combined with the information about oxidation states to create a picture of all the action happening at the interface during the anneals.

For anneals done in vacuum (Fig. 3.10 (a)), there is a maximum in the thicknesses of both oxides before they go through the desorption transition mentioned above, and the interface "regenerates" itself. The large thickness increase of the CeO_x may be an artifact of the method used to calculate the thickness, but since the density of Ce_2O_3 is slightly less than that of CeO_2, some thickness increase upon reduction is to be expected. Also, it has been shown recently that under-oxidized cerium oxide nanoparticles undergo a drastic lattice expansion [30] which may explain the phenomenon being seen here. At the highest temperatures all of the oxygen leaves the structure as discussed above, leaving the original silicon substrate and a thin layer of cerium metal.

For anneals performed in O_2 (Fig. 3.10 (b)), the SiO_{2-x} grows steadily and pushes up the overlayer, whose thickness and stoichiometry remain roughly constant. The oxidation state of the SiO_{2-x} increases steadily towards fully oxidized SiO_2. The CeO_2 is fixed in thickness due to the limited amount of Ce. On the other hand the SiO_2 layer continues to grow using up the silicon from the substrate.

Finally, more qualitative information is obtained about the oxidation state of the cerium from the valence band. Fig. 3.5 shows that as the cerium is reduced, the valence band edge takes on more and more Ce 4f character as one would expect from the other methods of determining the cerium oxidation state mentioned above. However, it is difficult to compare this method quantitatively with the other two, since there is no definite way of relating the spectral weight of the Ce 4f feature with oxidation state. The 4f character of the valence band increases in the region one would

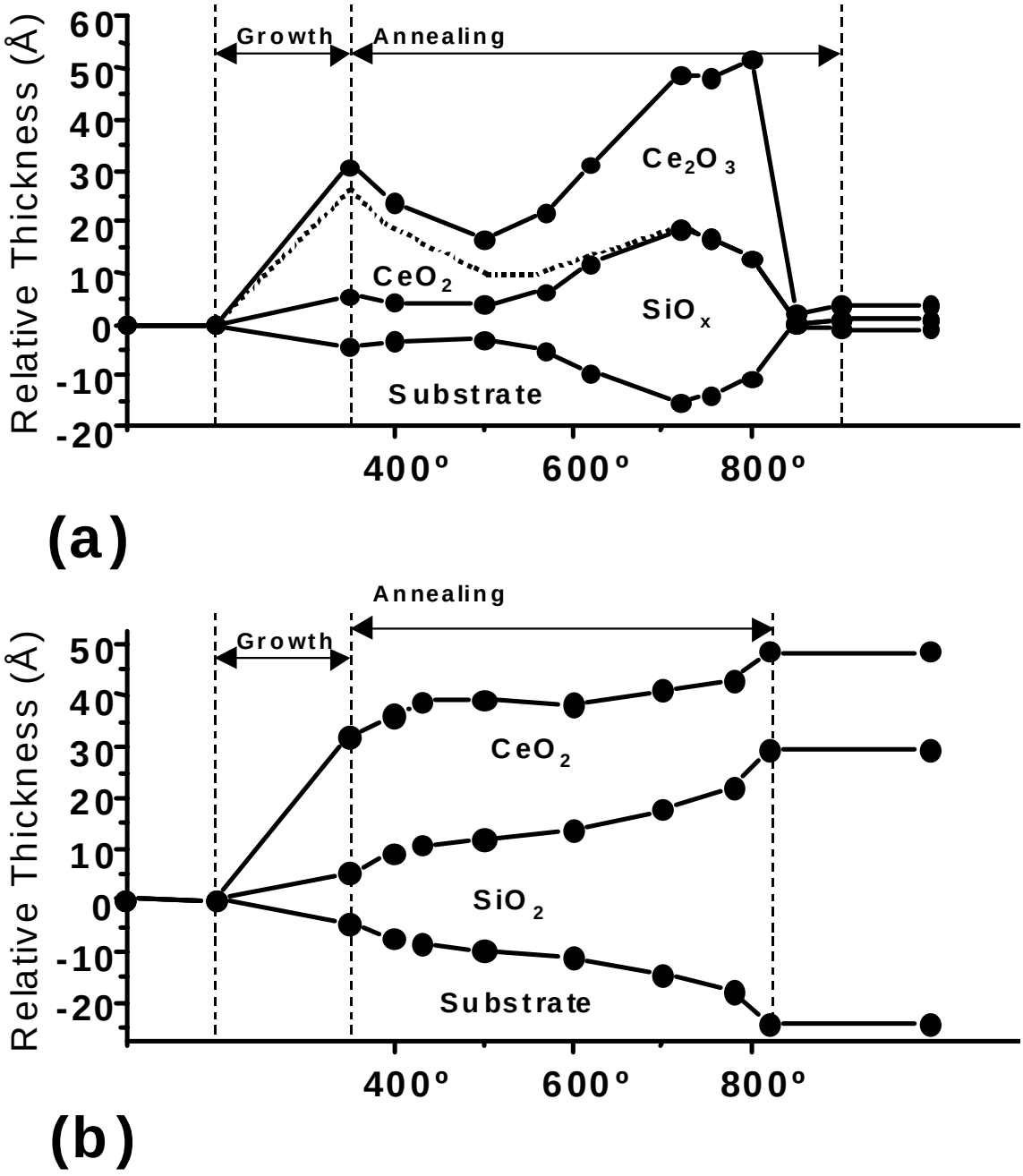


Figure 3.10: **The model of the interface obtained from analyzing all of the data. The relative thickness is computed by taking zero to be the surface of the original substrate. The abscissa can be thought of as time. First, the bare substrate is shown closest to the ordinate, then the growth takes place, then successive anneals are performed with the anneal temperature indicated on the abscissa, and finally the far right hand side of the abscissa is the final state after all annealing. Figure (a) is for the sample annealed in vacuum, Fig. (b) is for the sample annealed in O_2. All anneals were carried out for 30 minutes.**

expect it to based on observations of other spectra mentioned above. It saturates to a value where its height is roughly the same as that of the rest of the valence band, which is mostly derived from O 2p states. This saturation was also noted in the previous two methods and again seems to be the point of complete cerium reduction to Ce_2O_3. The valence band looks completely different past 800°C because of the desorption effect mentioned above. The spectrum resembles the structure that Hirschauer observed, attributing it to the combination 2×2 and $\sqrt{3}\times\sqrt{3}$ cerium surface reconstructions [4].

3.6 Conclusion

It is clear that complicated reactions occur at the cerium oxide/silicon interface. Some, such as the reduction of CeO_2 to Ce_2O_3 in the presence of silicon, are predicted by bulk thermodynamic arguments. However, several unique transitions occur such as the loss of oxygen from all the layers when annealed in vacuum past 800°C and the reduction of the CeO_x from the surface down to the interface that may not have been expected.

Extremely low temperatures are required to keep the surface completely free of SiO_{2-x} growth, which implies that a low temperature growth technique such as PLD might be the only way to achieve a true epitaxial interface for heterostructure applications. However, the strange high temperature desorption mentioned above may be a pathway to further high temperature growth as a sort of "self-cleaning" interface. Moreover, for buffer layer applications where the quality of CeO_2/silicon interface is not nearly as important as the quality of the CeO_2 surface, it is shown that a SiO_2 layer slowly forms underneath when annealed in oxygen but does not affect the CeO_2 stoichiometry. Finally, the emergence of the 4f states in the valence band does indeed echo the reduction of the cerium and saturates when the cerium is fully reduced to Ce_2O_3. The movement of this 4f level from a localized state in the bandgap towards an extended state in the valence band surely affects the electrical behavior of CeO_2 and thus is important to understand for MOS applications.

Bibliography

[1] T. Chikyow, S. Bedair, L. Tye, and N. El Masry, Appl. Phys. Lett. **65**, 1030 (1994).

[2] *CRC Handbook of Chemistry and Physics*, edited by D. R. e. Lide (C R C Press Llc, Boca Raton, Fl, 1994), Chap. 5.

[3] E. J. Tarsa, K. L. McCormick, and J. S. Speck, Mater. Res. Soc. Symp. Proc. **341**, 73 (1994).

[4] B. Hirschauer, M. Gothelid, M. Davila, and U. Karlsson, Surf. Sci. **464**, 117 (2000).

[5] E. Paparazzo, G. Ingo, and N. Zacchetti, J. Vac. Sci. Technol. A **9**, 1416 (1991).

[6] M. Seah, Surf. Sci. **420**, 285 (1999).

[7] M. Seah, I. Gilmore, and S. Spencer, Surf. Sci. **461**, 1 (2000).

[8] S. Tanuma, C. Powell, and D. Penn, Surf. Sci. **192**, L849 (1987).

[9] S. Tanuma, C. Powell, and D. Penn, Surf. Interface Analysis **11**, 577 (1988).

[10] F. Hillebrecht, M. Ronay, D. Rieger, and F. Himpsel, Phys. Rev. B **34**, 5377 (1986).

[11] Z. Wu, D. Huang, X. Yang, J. Wang, F. Qin, J. Zhang, and Z. Yang, Vacuum **49**, 133 (1998).

[12] J. Shallenberger, J. Vac. Sci. Technol. A **14**, 693 (1996).

[13] R. Alfonsetti, L. Lozzi, M. Passacantando, P. Picozzi, and S. Santucci, Appl. Surf. Sci. **70-1**, 222 (1993).

[14] A. Fujimori, Phys. Rev. B **28**, 4489 (1983).

[15] A. Fujimori, Phys. Rev. B **28**, 2281 (1983).

[16] A. Fujimori, Phys. Rev. B **27**, 3992 (1983).

[17] P. Burroughs, A. Hammett, A. F. Orchard, and G. Thornton, J. Chem. Soc. Dalton Trans. 1686 (1976).

[18] D. Mullins, S. Overbury, and D. Huntley, Surf. Sci. **409**, 307 (1998).

[19] M. Romeo, K. Bak, J. ElFallah, F. Lenormand, and L. Hilaire, Surf. Interface Analysis **20**, 508 (1993).

[20] J. Holgado, R. Alvarez, and G. Munuera, Appl. Surf. Sci. **161**, 301 (2000).

[21] K. Sundaram, P. Wahid, and O. Melendez, J. Vac. Sci. Technol. A **15**, 52 (1997).

[22] E. Wuilloud, B. Delley, W. Schneider, and Y. Baer, Phys. Rev. Lett. **53**, 202 (1984).

[23] J. Allen, J. Magn. & Magn. Mater. **47-8**, 168 (1985).

[24] T. Nakano, A. Kotani, and J. Parlebas, J. Phys. Soc. Jpn. **56**, 2201 (1987).

[25] A. Pfau and K. Schierbaum, Surf. Sci. **321**, 71 (1994).

[26] H. Massoud, J. Plummer, and E. Irene, J. Electrochem. Soc. **132**, 2685 (1985).

[27] B. Hirschauer, M. Gothelid, E. Janin, H. Lu, and U. Karlsson, Appl. Surf. Sci. **148**, 164 (1999).

[28] Y. Yamamoto, S. Arai, T. Matsuda, M. Satoh, and T. Inoue, Jpn. J. Appl. Phys. **36**, L133 (1997).

[29] *Binary Phase Alloy Diagrams*, edited by T. B. Massalski (ASM International, Materials Park, Ohio, 1990), Vol. 2.

[30] S. Tsunekawa, K. Ishikawa, Z. Li, Y. Kawazoe, and A. Kasuya, Phys. Rev. Lett. **85**, 3440 (2000).

Chapter 4 Metal-CeO_2-Silicon Capacitors

4.1 Overview

In any heterostructure system, the properties of the interface between the two materials may be even more important than the properties of the two materials themselves. In Chapter 3 the reactions that take place at the CeO_2/silicon interface in vacuum have been studied extensively. In this chapter, the electrical behavior of the interface in a real MOS device will be examined.

4.2 Introduction

Next to leakage current, the single most important metric for determining the usefulness of new gate dielectrics is the interface trap state density. The presence of interface states in a MOS device leads to several undesirable behaviors. First, the presence of any charge, whether it be at the interface or in the oxide itself leads to a shift in the capacitance voltage (C-V) characteristics of the device. In a MOSFET, this directly translates to a change in threshold voltage. Second, since interface traps fill and empty with some characteristic time constant, there will be a dispersion in the capacitance of the device as a function of frequency. Third, a large amount of trapped charge at the interface will degrade the channel mobility of a MOSFET via coulomb scattering of carriers. Finally, the subthreshold current swing in a MOSFET is adversely effected by interface trap states.

In the last chapter, the mechanical properties of the CeO_2/silicon interface were examined via *in situ* XPS measurements. In this chapter, the electrical properties of the interface are probed by examining the impedance characteristics of metal-CeO_2-

silicon capacitors. In addition, the charges present in the CeO_2 itself are measured from C-V characteristics.

These properties are examined for MOS devices made from as-grown CeO_2 layers and also from CeO_2 layers treated with various types of annealing in order to determine if *ex situ* processing might improve the device characteristics.

4.3 Experiment

Substrates consisted of (111) oriented silicon wafers doped 1×10^{14} cm^{-3} n-type. CeO_2 was grown at 400°C via pulsed laser deposition from a doubled Nd:YAG laser with a pulse energy of approximately 300 mJ/pulse at 533 nm. The pulse width was 5 ns and a 10 Hz repetition rate was used. The target was a 99.99% pure CeO_2 sputter target supplied by Kurt Lesker. For the first 30 seconds of growth ($\sim$20Å) the CeO_2 was ablated in vacuum at which point oxygen was slowly bled into the system at a partial pressure of 5×10^{-6} torr, similar to the growth technique of Yoshimoto et al. [1]. Growth was continued in this fashion for 15 minutes. The growth rate was calibrated using the method described in Section 2.5.1 to yield a final thickness of $\sim$700Å.

Some of the samples were left as-grown and others were annealed in either argon, oxygen, or 10% hydrogen in helium. All anneals were 10 minutes long at 450°C. This temperature was chosen both because it is similar to the temperature used for passivation anneals in traditional SiO_2 MOS [2] and because it is close to the highest temperature which the CeO_2 layers had already been exposed to during growth. Diodes were fabricated via direct current (DC) magnetron sputtering of aluminum through a photoresist mask followed by liftoff of the photoresist to form square pads with edges ranging from 30 μm to 500 μm.

Electrical characterization was performed in an electrically shielded dark box, purged with dry nitrogen. The pads were probed with large (50 μm diameter) tungsten probe tips to minimize series resistance and the backside contact was made with silver paint. All impedance measurements were performed with a Hewlett-Packard

4192 LF impedance analyzer.

The impedance of the wiring, test fixture feedthroughs, and probes was zeroed by measuring the series resistance and inductance with the leads short-circuited and the shunt capacitance with the leads open-circuited.

The C-V characteristics of all devices were measured from at least -2V to +2V. It was found that this range straddled the accumulation, depletion and inversion domains of the devices in a range over which the DC leakage current was below about 1 μA/cm^2, which is several orders of magnitude below the acceptable leakage current for state of the art MOSFET's [3]. All C-V measurements were performed using a 1 MHz, 10 mV root mean squared (RMS) alternating current (AC) test signal on top of the DC bias.

The conductance-frequency (G-ω) spectra were measured using the same bias range and same amplitude AC test signal. The frequency was swept from 100 Hz to 10 MHz.

4.4 Theory

In this section, an overview of the methods used to extract interface trap densities and flat band voltage shifts will be given.

The first method for determining interface state density is the high-frequency capacitance method, also known as the Terman method. In this method, the only measurement necessary is the high frequency (typically on the order of 1 MHz) capacitance versus voltage curve like those in Fig. 4.4. Even though interface traps cannot fill and empty on the time scales used to measure the capacitance, they can respond to the quasi-static bias sweep used to produce the C-V curve. Figure 4.1 illustrates that as the Fermi level crosses through the interface state distribution at a silicon surface, a sheet charge at the surface is created that is directly proportional to either the number of empty interface states for donor-like traps or filled states for acceptor-like traps. Since this sheet charge will depend on the position of the Fermi level at the interface, it will depend on the surface potential Ψ_s, and thus the applied gate

voltage. This voltage-dependent charge layer can have a pronounced effect on the C-V characteristics even though the traps at the surface are not filling and emptying in phase with the AC signal. What the trap states will do is change the gate voltage for which a given Ψ_s is obtained. From Appendix A.1 the following relationship is found for an insulator-semiconductor interface:

$$C_{ox}(V_g - \Psi_s) = -Q_{sheet} - Q_s(\Psi_s), \tag{4.1}$$

where C_{ox} is the oxide capacitance, V_g is the voltage applied to the gate electrode relative to ground, Q_{sheet} is any sheet charge located at the insulator-semiconductor interface and Q_s is the depletion/inversion charge in the semiconductor. In the case of a sheet charge generated by interface states, Q_{sheet} is also a function of Ψ_s as shown in Fig. 4.1. If Q_{sheet} is replaced by $Q_{it}(\Psi_s)$, V_g can be solved for, then differentiated with respect to Ψ_s to obtain

$$\frac{dV_g}{d\Psi_s} = 1 - \frac{1}{C_{ox}}\left(\frac{dQ_{it}}{d\Psi_s} + \frac{dQ_s}{d\Psi_s}\right). \tag{4.2}$$

The derivatives of the interface state charge and the semiconductor charge can be rewritten as capacitances, C_{it} and C_s, giving

$$C_{it}(\Psi_s) = C_{ox}\left(\frac{dV_g}{d\Psi_s} - 1\right) - C_s(\Psi_s). \tag{4.3}$$

For an interface state density that is slowly varying throughout the energy gap relative to k_BT the interface state density can be approximated simply as $D_{it}=qC_{it}$ [2]. The important parameter to determine for this calculation is $dV_g/d\Psi_s$, which is an indication of how the band bending changes with applied gate voltage. The Terman method calculates this parameter as a function of Ψ_s by comparing an ideal theoretical C-V curve as a function of Ψ_s with the measured C-V curve. For each value of capacitance at a given Ψ_s in the theoretical curve, there is the corresponding value of capacitance at a given V_g in the measured curve. Thus it is possible to construct a curve of V_g vs. Ψ_s such as the one shown in Fig. 4.2. The derivative of this curve

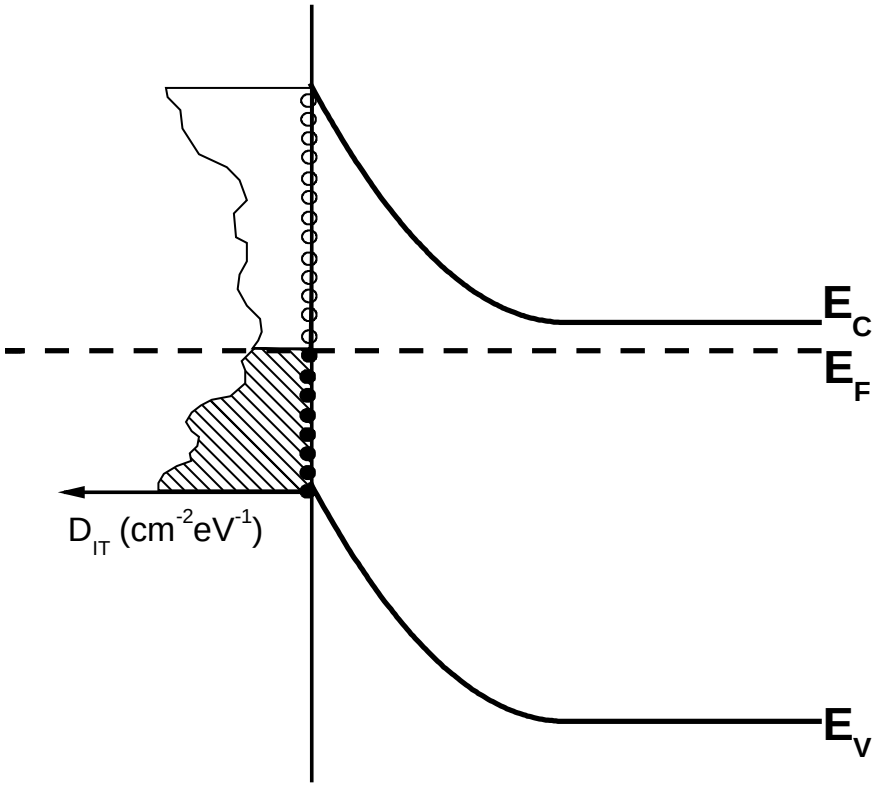


Figure 4.1: **Schematic of an insulator-semiconductor interface with interface states. Donor-like states are defined as positively charged when empty and acceptor-like states are defined as negatively charged when filled.**

along with the calculated curve for $C_s(\Psi_s)$ enables the construction of a curve of C_{it} and hence D_{it} as a function of the position of the Fermi level in the bandgap from Eq. 4.3. The form of the theoretical high frequency capacitance versus Ψ_s curve was taken from Brews [4].

In addition to using the V_g versus Ψ_s curve to determine interface state density via the Terman method, it was also used to relate the applied gate voltage to a given Fermi level position in the bandgap. Thus for the conductance method described below, it is possible to transform a plot of D_{it} versus gate voltage to a plot D_{it} versus position in the bandgap as in the Terman method.

Interface state densities were also calculated using the conductance method. In this method the frequency response of the interface traps is used to calculate their density. At low frequencies, all of the traps can fill and empty in phase with the changing AC signal. The capacitor is always in equilibrium, and thus no net energy is lost via the traps. At very high frequencies, the traps don't have time to fill or empty at all, and as far as the conductance is concerned, it is as if they are not even there. At intermediate frequencies, some traps can fill and empty but not fast enough to keep up completely with the AC signal. This leads to situations in which the Fermi level at the interface is higher than the highest filled interface state or, on the other half of the AC cycle, lower than the highest filled interface state. As it turns out, in both halves of the cycle, net energy is lost. If all the electrons in the semiconductor are viewed as one statistical ensemble at some average energy, then for the case where the Fermi level is above the highest filled interface state, electrons from the conduction band will fall into these states in order to restore equilibrium. As they do, they reduce the average energy of the ensemble. The energy is converted to phonons which heat the lattice. In the other half cycle, electrons in traps above the Fermi level are emitted into the conduction band. This energy is derived from absorbed phonons, increasing the average energy of the electron ensemble. But subsequently the ensemble of electrons must move in order to restore equilibrium, slightly reducing the energy of all the other electrons. Again this is done by emitting phonons that heat the lattice. These processes lead to a net energy loss in both halves of the cycle

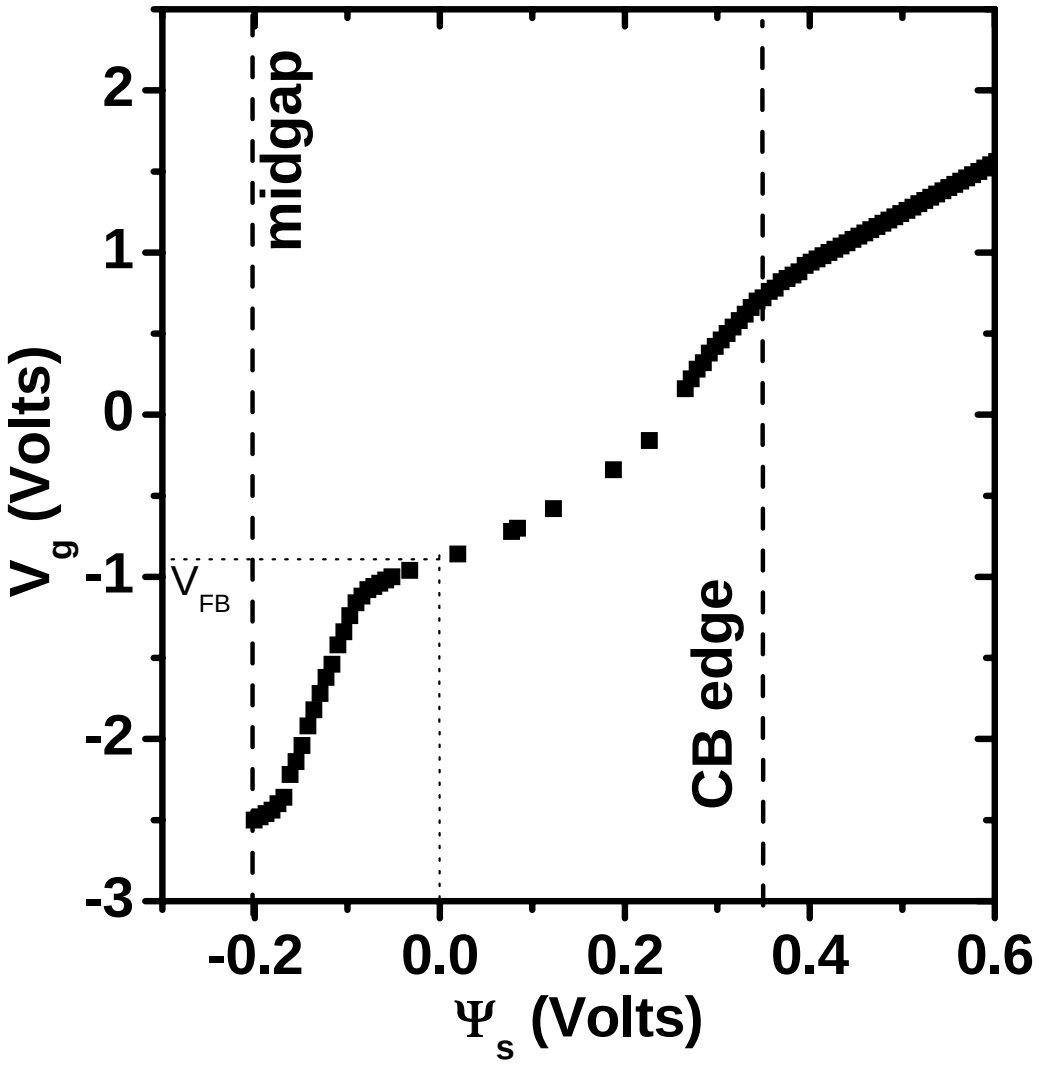


Figure 4.2: **Example of a measured curve of V_g vs. Ψ_s. The derivative of this curve is used to compute the interface state density in the Terman method. This curve is also used to relate the position of the Fermi level within the silicon bandgap to the applied gate bias in both the Terman method and the conductance method. Three volts of gate voltage moves the Fermi level by about half an eV and in this example the flatband voltage is about -1 volt.**

which can be modeled in a circuit as a conductance. This can be thought of as a situation in which the system is in a position in state space that is slightly perturbed from equilibrium and it relaxes to its thermodynamic equilibrium state by giving off heat.

In addition, there is a finite time delay associated with the capture and emission of electrons from these traps which leads to a phase delay in the AC response of the device. This can be modeled as an added capacitance. Figure 4.3(a) shows the small-signal equivalent circuit for a MOS capacitor in depletion with interface traps. The impedance bridge measures an equivalent parallel conductance and capacitance as shown in Fig. 4.3(b). It is easy to derive that for the model circuit shown, the measured values G_m and C_m will be

$$G_m = \frac{\omega^2 C_{ox}^2 G_t}{G_t^2 + \omega^2 (C_d + C_t + C_{ox})^2}, \tag{4.4}$$

$$C_m = \frac{\omega^2 C_{ox}(C_d + C_t)(C_d + C_t + C_{ox}) + C_{ox} G_t^2}{G_t^2 + \omega^2 (C_d + C_t + C_{ox})^2}. \tag{4.5}$$

One could theoretically fit the measured data to these equations, using a calculated C_d for depletion capacitance in the semiconductor. However, there is a much more elegant way to extract G_t. In Appendix A.2 it is shown that a formula can be derived for G_t as a function of G_m, C_m and C_{ox} alone:

$$G_t(\omega) = \frac{\omega^2 C_{ox}^2 G_m}{G_m^2 + \omega^2 (C_m^2 - C_{ox}^2)}. \tag{4.6}$$

C_{ox} can be measured directly in strong accumulation (see Eq. 4.12). A plot of G_t/ω will be peaked as shown in Fig. 4.3(c). This peak can be thought of as occuring at a frequency for which the trap states are in "resonance" with the applied AC signal. At this frequency the highest possible number of traps are contributing to the conductance as described above. Thus the peak value of G_t/ω, which we will denote as $(G_t/\omega)_p$, should be roughly proportional to the interface trap density. However, statistical variations in the nature of the charge distribution under the gate cause fluctuations in the band-bending at the semiconductor surface. This can smear out

the peaks and reduce their magnitude. The standard deviation of these fluctuations, σ_s, is related to the width of peak in the G_t/ω vs. ω curve and can be calculated from the data [2]. Then the interface trap state density can be related to $(G_t/\omega)_p$ times a factor that is a function of σ_s. Further, the frequency ω_p, at which the peak occurs can be related to the time constant, τ, of the traps. The derivation of these expressions is unnecessarily complicated for the purposes of this work but is described in detail by Nicollian and Brews [2]. The most useful results are presented here:

$$D_{it} = \frac{\left(\frac{G_t}{\omega}\right)_p}{q f_d(\sigma_s)} \tag{4.7}$$

$$\tau = \frac{\xi_p}{\omega_p} \tag{4.8}$$

$$f_d(\sigma_s) = \frac{\sqrt{2\pi\sigma_2^2}}{2\xi_p} \int_{-\infty}^{\infty} exp\left(-\frac{\eta^2}{2\sigma_s^2} - \eta\right) \ln\left(1 + \xi_p^2 exp2\eta\right) d\eta, \tag{4.9}$$

where ξ_p is defined by

$$\int_{-\infty}^{\infty} exp\left(-\frac{\eta^2}{2\sigma_s^2} - \eta\right) \left\{\frac{2\xi_p^2 exp2\eta}{1 + \xi_p^2 exp2\eta} - \ln(1 + \xi_p^2 exp2\eta)\right\} d\eta = 0. \tag{4.10}$$

The function $f_d(\sigma_s)$ only ranges between about 0.1 and 0.4 and ξ_p only ranges between about 2 and 2.6 for reasonable values of σ_s. Hence the important parameters are $(G_t/\omega)_p$ and ω_p.

For all measurements of conductance and capacitance, the effect of substrate resistance needs to be considered. In strong accumulation there is no depletion capacitance and the effects of interface states are shorted out. Thus the equivalent circuit for a MOS capacitor in accumulation is simply the oxide capacitance. However, if there is a finite resistance in the substrate, the equivalent circuit must be amended to have a resistance in series with the oxide capacitance (i.e., substitute the entire parallel network in Fig. 4.3(a) with a resistor R_s). From Appendix A.3, the relationships for the series resistance and oxide capacitance as a function of the measured parallel

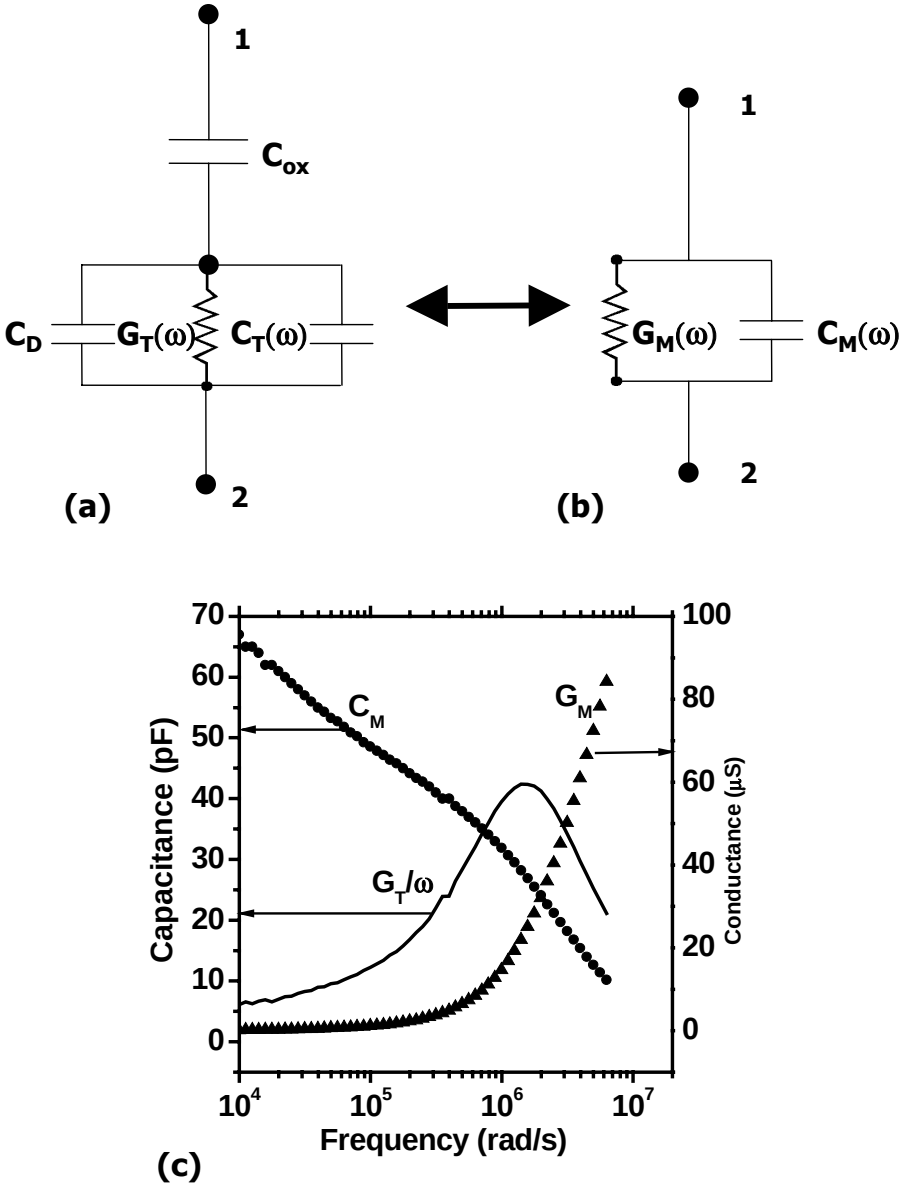


Figure 4.3: **The circuit in Fig. (a) shows an equivalent circuit model for a MOS capacitor with interface traps. The circuit in Fig. (b) shows the actual effective circuit that the impedance bridge measures. The plot in Fig. (c) shows an example spectrum of the two measured quantities, G_m and C_m along with the calculated spectrum of G_t/ω from Eq. 4.6.**

conductance and capacitance in accumulation are

$$R_s = \frac{G_{ma}}{G_{ma}^2 + \omega^2 C_{ma}^2}, \tag{4.11}$$

$$C_{ox} = C_{ma}\left[1 + \left(\frac{G_{ma}}{\omega C_{ma}}\right)\right]. \tag{4.12}$$

Once these values are determined from a measurement in strong accumulation, the adjusted values for the parallel MOS capacitance and conductance for the rest of the voltage range are calculated by

$$G_c = \frac{G_m - R_s(G_m^2 + \omega^2 C_m^2)}{(1 - G_m R_s)^2 + \omega^2 C_m^2 R_s^2}, \tag{4.13}$$

$$C_c = \frac{C_m}{(1 - G_m R_s)^2 + \omega^2 C_m^2 R_s^2}. \tag{4.14}$$

All capacitance and conductance data used in the following sections was treated in this way. As it turns out, the series resistance had a significant effect on the measurements since the silicon substrates were lightly doped ($\sim 10^{14}$ cm^{-3} n-type), leading to $R_s \sim$ 1kΩ for a 100μm square pad.

The oxide trap charges and the hysteresis charge were calculated from the flatband voltage shifts. In order to systematically determine the flatband voltage from the experimental C-V curve, the method of Schroder [5] was employed which involves taking derivatives of the C-V curve. The fixed charge density then was calculated from this measured flatband voltage using

$$Q_f = C_{ox}(V_{fb} - \Phi_{ms}), \tag{4.15}$$

where Φ_{ms} is the metal-semiconductor work function difference in volts. In this case, for aluminum gates and 1×10^{14} cm^{-3} n-type silicon, Φ_{ms} was calculated to be about -0.4 volts. In cases where hysteresis was present, the flatband voltage was taken as the average of the flatband voltage of the two curves. The hysteresis charge itself was taken to be

$$Q_{hyst} = \Delta V_{fb} C_{ox}, \tag{4.16}$$

where ΔV_{fb} is the difference of the flatband voltages for the forward and reverse bias sweeps.

4.5 Results

4.5.1 Capacitance-Voltage Analysis

Figures 4.4 and 4.5 show measured C-V curves and Table 4.5.1 lists the various charge densities calculated from the C-V curves. The C-V characteristics show clear evidence of strong inversion rather than deep depletion, asymptoting to a relatively constant minimum capacitance value after about a volt of negative bias. This indicates that the leakage current is not too severe to inhibit the formation of the minority carrier inversion layer by allowing (in this case) holes to tunnel through the insulator. In all but the argon-annealed case, the C-V characteristics also show a constant accumulation capacitance, again indicating good leakage characteristics. The dielectric constant can be calculated from the accumulation capacitances in Fig. 4.4 and was estimated to be about 10 for the as-grown sample.

The hysteresis observed in the as-grown case also has been observed in other studies of CeO_2-based capacitors [6, 7]. One possibility is that it is due to charges in the bulk CeO_2. However, another possibility is that it is due to interface states deep in the silicon bandgap, or states close enough to the interface to have a significant probability of having an electron tunnel into the silicon under bias. Figure 4.5 shows that decreasing the sweep rate reduces the hysteresis consistent with states with time constants on the order of seconds. While this is longer than the typical time constants for midgap interface trap states in silicon (see Fig. 4.8), it is quite probable that "near-interface" trap states in the CeO_2 could trap and emit at this slow of a rate. States in the bulk of the CeO_2 would have much longer time constants due to the large bandgap of CeO_2 so the hysteresis is most likely not due to bulk charges.

The C-V curves of devices treated in four different ways is shown in Fig. 4.4. Once again, all anneals were performed for 10 minutes at 450°C, and all of these samples

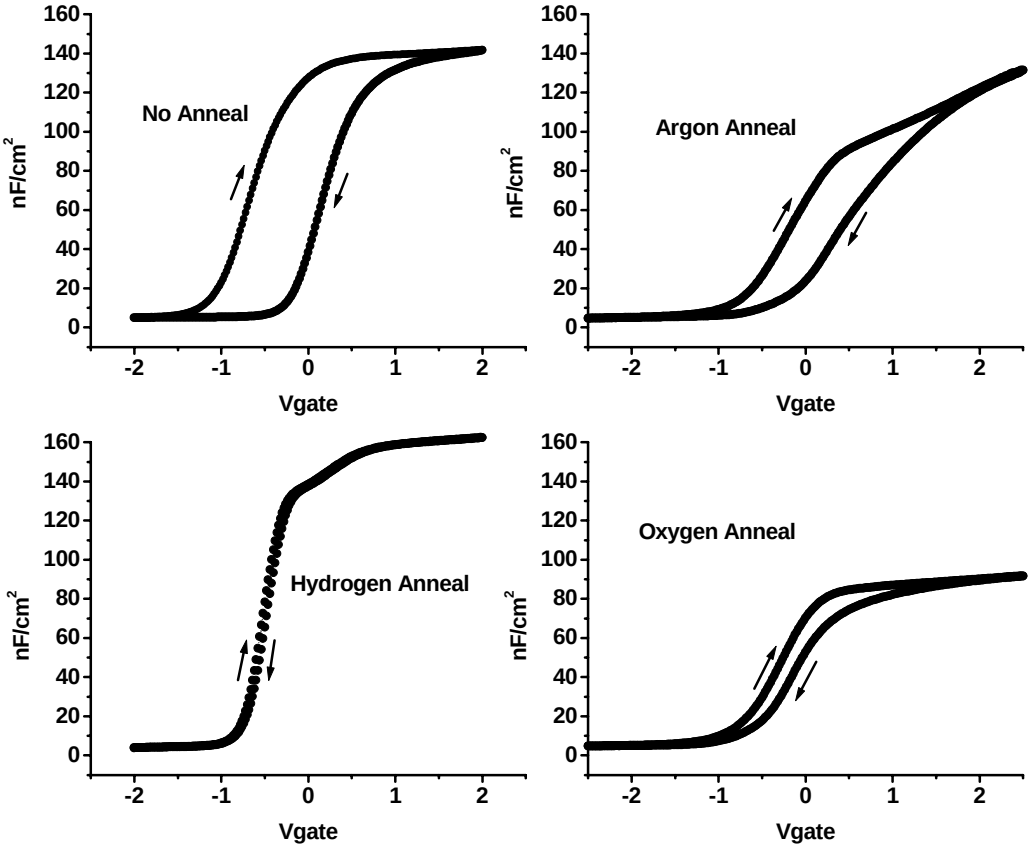


Figure 4.4: **C-V plots of metal-CeO_2-silicon capacitors annealed under different ambients. All the plots are in capacitance per unit area. The accumulation capacitance of the as-grown sample corresponds to a dielectric constant of about 10.**

Sample	$V_{fb} - \Phi_{ms}$ (volts)	C_{ox} $nFcm^{-2}$	Δ_{fb}^{hyst} (volts)	Fixed Charge cm^{-2}	Hysteresis Charge cm^{-2}
no anneal	-0.265	140	0.93	2.32×10^{11}	-8.14×10^{11}
H_2	-0.385	160	0.03	3.85×10^{11}	-3×10^{10}
O_2	-0.350	90	0.1	1.97×10^{11}	-5.63×10^{11}
Ar	-0.175	100	0.25	1.09×10^{11}	-1.56×10^{11}

Table 4.1: **Summary of bulk charges in metal-CeO_2-silicon capacitors**

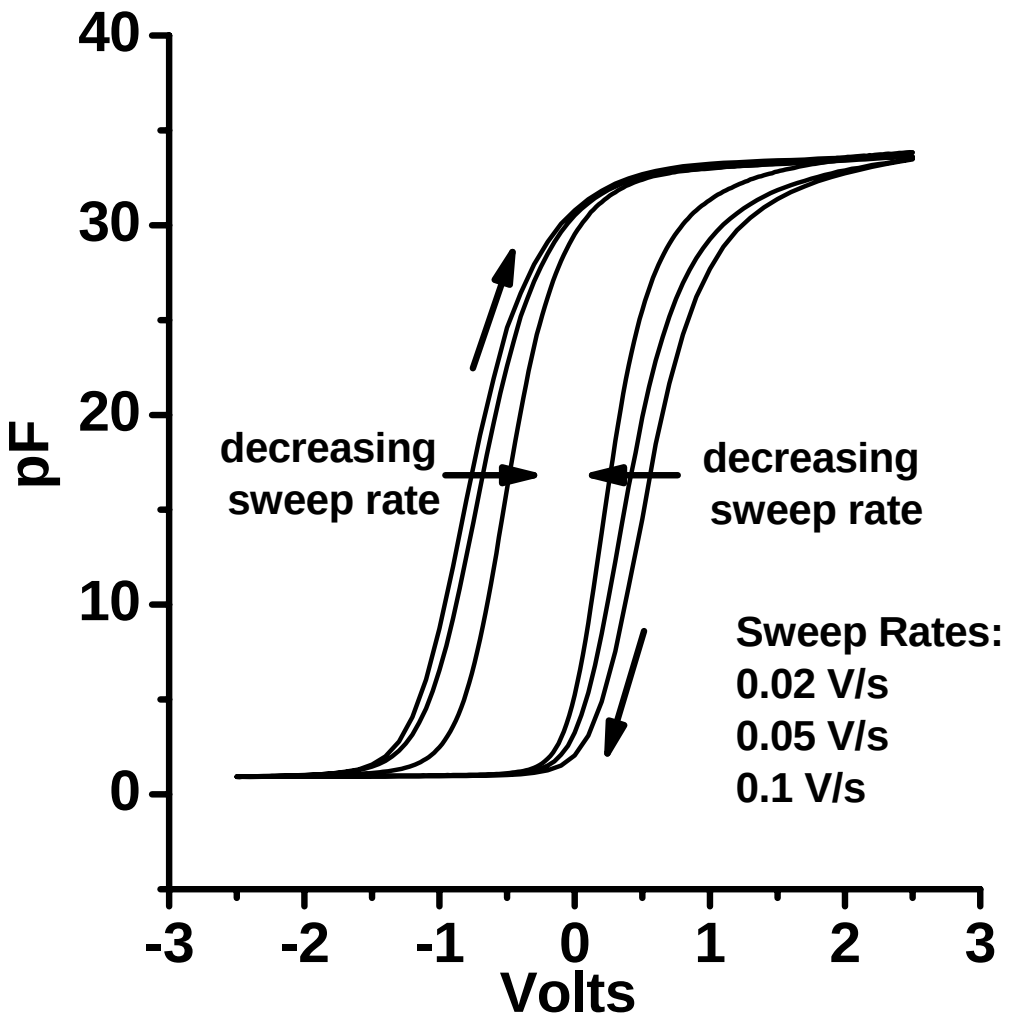


Figure 4.5: **Dependence of hysteresis on sweep rate. The fastest sweep shows the largest hysteresis indicating that the hysteresis comes from charge states that seem to fill and empty on the order of seconds. If the hysteresis were due to fixed charge deep in the CeO_2 the time constant would be much larger and if it were due interface trap states deep in the silicon bandgap it would be much shorter. Thus we are led to believe that this charge comes from near-interface states slightly into the insulator.**

were annealed before metalization. These curves represent an average over several devices on the same sample to eliminate the dependence of the results on thickness variations of the oxide film and other random variations. About 5% of all the devices behaved abnormally and are not included in this data because they are most likely related to extremely defective areas of the oxide (such as areas with pinholes that short to the substrate) and therefore are not representative of a true CeO_2 film.

It is observed that there are drastic differences in the behavior for the various annealing conditions. Any type of anneal seems to decrease the hysteresis but hydrogen annealing almost completely eliminates it. In another study on CeO_2 capacitors, Roh et al. [6] found that the hysteresis persisted and even switched direction after hydrogen annealing. However, all of their samples had a thick, intentionally oxidized SiO_{2-x} layer between the CeO_2 and the silicon so it is probably quite a different situation. The reduction of the hysteresis charge could be directly due to a reduction in interface state density which, as will be seen below, is quite substantial for hydrogen annealing. It could also be due to a change in the nature of the species at the interface. If it is assumed that the original hysteresis charge comes from oxygen vacancies in the SiO_{2-x} layer formed at the CeO_2/silicon interface during growth (see Chapter 3), it is possible that hydrogen annealing passivates the defects in this SiO_{2-x} layer. It also should be noted that the accumulation capacitance (and hence C_{ox}) actually *increases* relative to the non-annealed case. The slight increase in accumulation capacitance may be due to reduction of some of the cerium oxide to a suboxide (again see Chapter 3) which may have a slightly higher dielectric constant. Table 4.5.1 indicates that hydrogen annealing adds some fixed positive charge. This again may be due to reduction of the bulk CeO_2 leaving behind positively charged centers. These bulk charges must contribute to a fixed flatband voltage shift without contributing to the hysteresis because they exist in the wide bandgap CeO_2 where midgap states will have much longer time constants (on the order of hours or even days).

In the oxygen-annealed case, it is clear that some oxidation of the silicon has taken place because of the decrease in accumulation capacitance. One can calculate that

about 137 Å of silicon oxide has formed if one assumes that the oxide has the same dielectric constant as SiO_2. Further, this annealing reduced the charges responsible for the hysteresis but not as much as for the hydrogen anneal. If the interfacial SiO_{2-x} layer discussed above is indeed the source of the hysteresis, it would make sense that oxidizing this SiO_{2-x} would reduce the hysteresis charge. The flatband voltage shift is negative (indicating added positive charge) as in the case of the hydrogen anneal but clearly cannot be caused by the same bulk CeO_2 reduction. It is possible that new, positively charged states are created at the SiO_{2-x} -CeO_2 interface during the oxidation process that did not exist in the as-grown sample. These states would be deep enough (at least 137 Å into the insulating layer) to contribute only to a static flatband voltage shift, and not the hysteresis charge.

In the argon-annealed case, the situation is much less clear. The hysteresis has been reduced slightly but the curve never saturates in accumulation which is usually an indication of DC current leakage. The probable cause of this change is a reduction of the cerium oxide to a lower oxidation state with different dielectric constant and different band offsets to silicon, completely changing its electrical behavior. The surprising part is that no similar effects were seen in the case of the hydrogen anneal, which is usually considered to be a reducing agent. Further, the argon annealing was the only treatment that caused a reduction of the fixed positive charge relative to the as-grown sample.

4.5.2 Interface Trap State Density

The interface trap state density as a function of position in the bandgap as calculated by the Terman method is shown in Fig. 4.6. Because of the hysteresis, D_{it} was calculated from data for both up and down sweeps. However, there was very little difference in the resulting interface state distributions. Thus, to avoid clutter, only the C-V curves that were swept from inversion to accumulation (low voltage to high voltage) were used to generate the data in Fig. 4.6. Note that all the samples were grown on silicon (111) substrates which have about double the dangling bond density

of a (100) surface. Thus to compare to the values of D_{it} commonly quoted in the literature for (100) silicon, the absolute values of D_{it} shown in the figure should be divided by a factor of about two to compensate for the increased bond density on (111) surfaces. The minimum value of D_{it} is compared with that of other alternative gate dielectrics in Table 1.1. While the value for CeO_2 is nowhere near as low as for device quality SiO_2 on silicon, it is comparable to the values of the other alternative gate dielectrics reported in the literature.

Clearly hydrogen annealing is the most favorable process for lowering interface trap state density. The minimum interface trap state density is lowered by a factor of about five. The hydrogen and oxygen anneals both reduce the absolute value of the interface trap state density over some of the bandgap. In the case of the oxygen anneal the shape of the distribution throughout the gap also changes. Again this lends credence to the theory that the hysteresis charge is caused by a SiO_{2-x} layer formed at the interface in the as-grown sample. The large reduction in the minimum interface state density in the hydrogen-annealed case almost assuredly is due to passivation of silicon dangling bonds by atomic hydrogen. The reduction in the interface state density for the oxygen-annealed case may be due to the fact SiO_2/silicon interfaces show a lower surface state density than most other insulators in general. This would also explain the change in the shape of the distribution since the actual semiconductor/insulator interface after oxygen annealing is more like a SiO_2/silicon interface. However, SiO_2/silicon interfaces generally still need to undergo further annealing in H_2 in order to produce the extremely low values of D_{it} seen in state of the art MOS devices. In the argon-annealed case, the interface state density is *increased* across the whole energy range, with no substantial change in the shape of the distribution. Hence we are led to believe that the SiO_{2-x} layer has not been passivated or oxidized as in the other two cases, rather that more dangling bonds are somehow created as the interface is annealed in an inert ambient.

Figure 4.7 shows the interface state density for the as-grown and hydrogen-annealed samples measured by the conductance method. The conductance method only is accurate from depletion into weak inversion. Beyond weak inversion, bulk traps begin

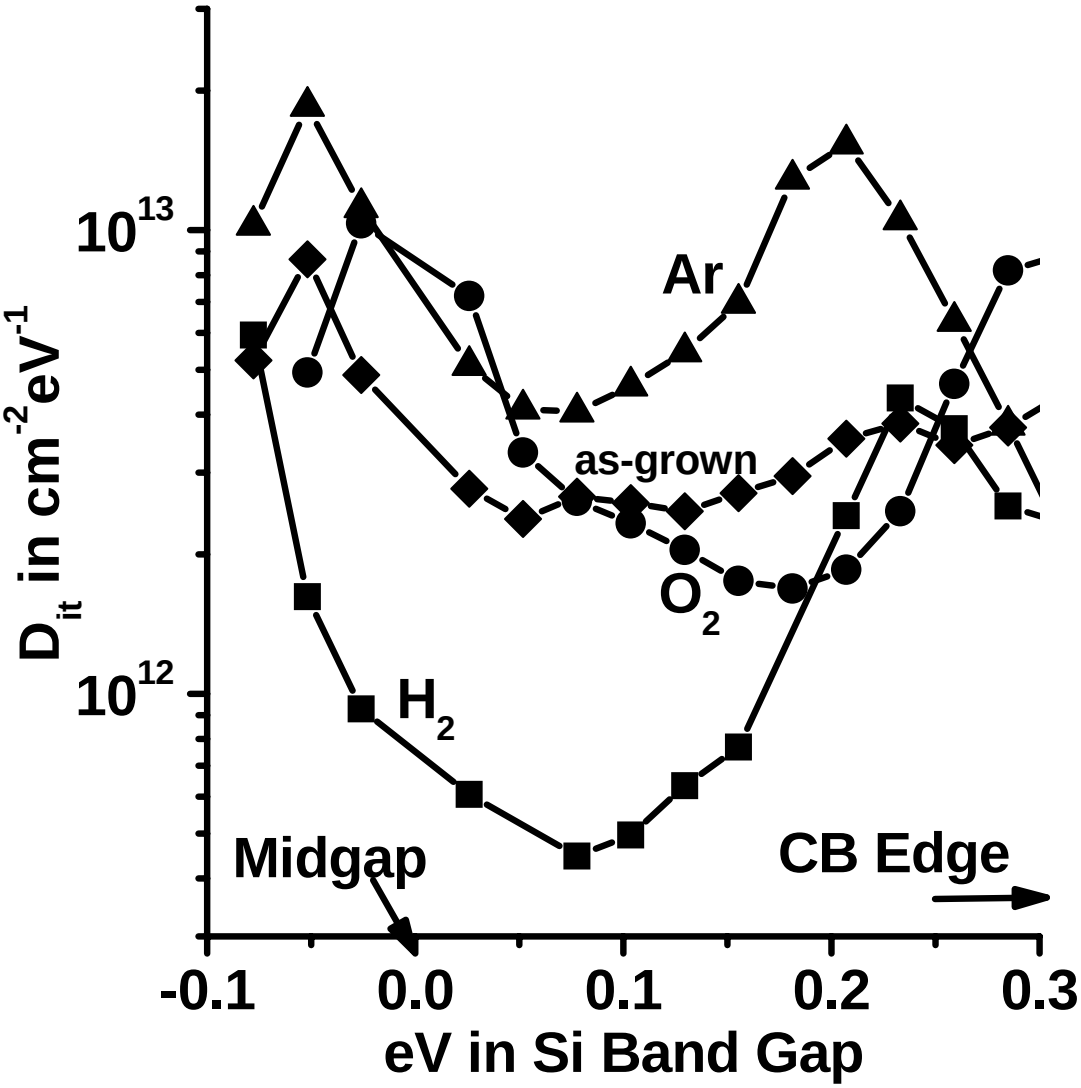


Figure 4.6: **Interface trap state density of CeO_2 MOS capacitors annealed under various conditions calculated by the Terman method. The diamonds are for the as-grown sample, the circles are for the oxygen-annealed sample, the boxes are for the hydrogen-annealed sample and the triangles are for the argon-annealed sample.**

to dominate the energy loss in the capacitor so that the measured conductance is no longer only due to the interface traps. The hatched areas of the figure represent the limits of the validity of this measurement. Near midgap, the values of the two curves are close to their minimum values obtained from the Terman method. In general, the as-grown curve shows a higher interface state density than for the Terman method. Again, it is shown that there is a significant improvement in the minimum interface trap state density after hydrogen annealing.

Figure 4.8 shows the time constants calculated from Eq. 4.8 of the various interface states plotted in Fig. 4.7. Both datasets appear linear on a log-linear plot near midgap indicating a simple exponential dependence of the time constants of the states within the gap. If the time constant was to follow an exponential dependence on the position in the bandgap with a slope of β=1/k$_B$T, the time constant should change by about nine orders of magnitude from midgap to the band edges! From the data in the figure, a slope equal to approximately $\beta/3$ is obtained for the as-grown sample but only $\beta/15$ for the hydrogen-annealed sample. It is not clear why the time constants are so much shorter for the hydrogen-annealed sample. As mentioned above, a significant amount of fixed positive charge is added in the oxide after hydrogen annealing. Perhaps these states can assist in emitting more quickly an electron from a filled interface trap state. The highest time constants near midgap are on the order of several milliseconds. This is right on the borderline of the values that might cause hysteresis. However, the idea that the hysteresis charge comes from charges placed slightly into the insulator is still the more probable explanation.

4.6 Conclusion

Hydrogen annealing has the most beneficial effect on the electrical characteristics of metal-CeO_2-silicon capacitors. Interface state density is reduced from about 2×10^{12} $cm^{-2}eV^{-1}$ to 4×10^{11} $cm^{-2}eV^{-1}$. The large hysteresis charge present in the as-grown samples is almost completely removed by the hydrogen annealing. Oxygen annealing creates a significant SiO_2 layer between the silicon and the CeO_2 but also helps to

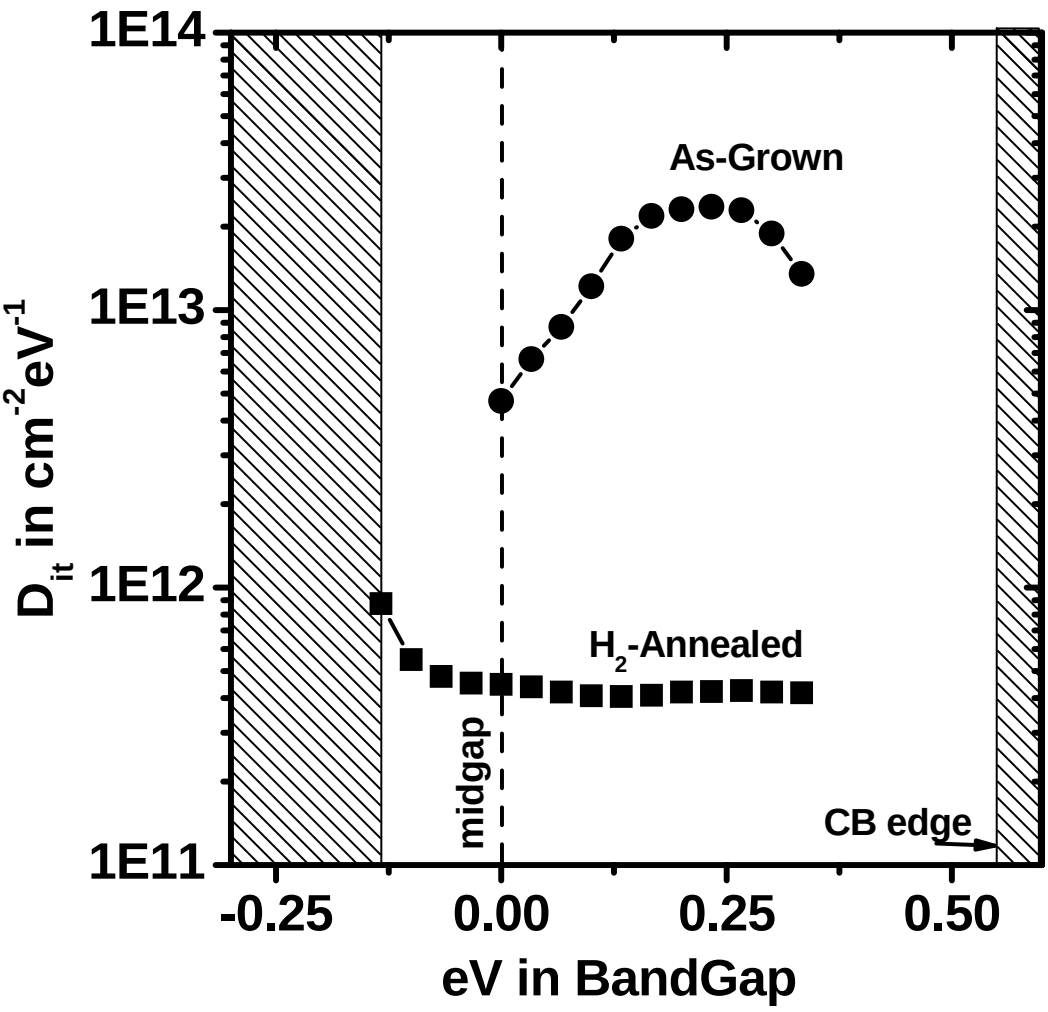


Figure 4.7: **Interface trap state density of CeO_2 MOS capacitors as measured by the conductance method. The hatched areas indicate the limits of the validity of the conductance method.**

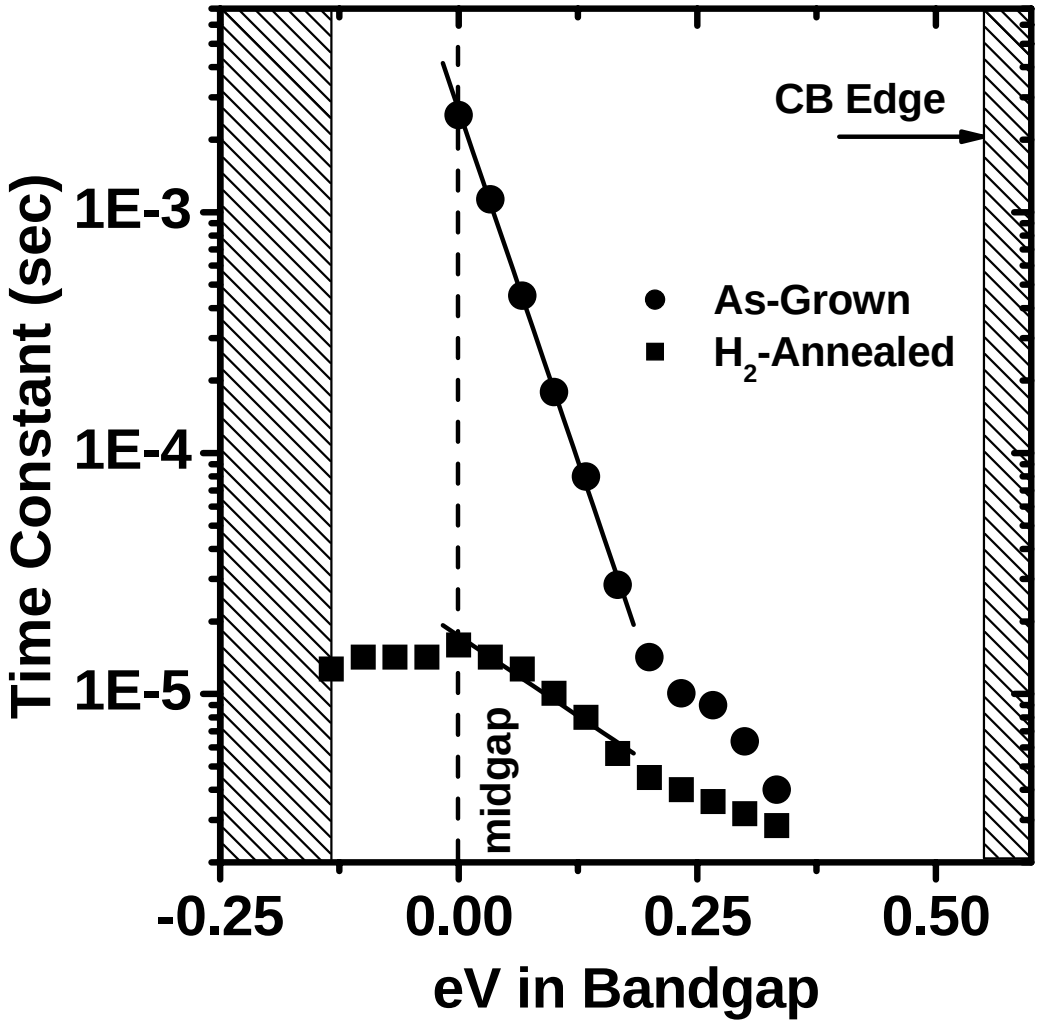


Figure 4.8: **Measured time constants of the distribution of interface traps. The hatched areas indicate the limits of the validity of this measurement.**

reduce interface states and the hysteresis effect. Annealing in argon has drastic effects on the behavior of the capacitors but they are much more difficult to explain. Argon annealing increases interface state density and has the least effect on the hysteresis charge. The source of the hysteresis is well modeled as originating in an SiO_{2-x} layer that forms between the silicon and the CeO_2 during CeO_2 growth.

Bibliography

[1] M. Yoshimoto, K. Shimozono, T. Maeda, T. Ohnishi, M. Kumagai, T. Chikyow, O. Ishiyama, M. Shinohara, and H. Koinuma, Jpn. J. Appl. Phys. **34**, L688 (1995).

[2] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology* (Wiley-Interscience, New York, NY, 1982).

[3] B. Doyle, R. Arghavani, D. Barlage, S. Datta, M. Doczy, J. Kavalieros, A. Murthy, and R. Chau, Intel Technology Journal **6**, 42 (2002).

[4] J. R. Brews, J. Appl. Phys. **45**, 1276 (1974).

[5] D. Schroder, *Semiconductor Material and Device Characterization*, 2 ed. (Wiley, New York, NY, 1998).

[6] Y. H. Roh, K. Kim, and D. G. Jung, Jpn. J. Appl. Phys. **36**, L1681 (1997).

[7] L. Kim, J. Kim, D. Jung, and Y. Roh, Appl. Phys. Lett. **76**, 1881 (2000).

Chapter 5 Part I Conclusion

Several aspects concerning the viability of CeO_2 for use in silicon-based electronic devices have been discussed. In particular, the nature of the interface between silicon and CeO_2 has been examined extensively both in terms of its stability under various annealing conditions and in terms of its electrical behavior.

It has been determined that growth of sharp interfaces between CeO_2 and silicon must necessarily be performed at low temperature (less than about 400°C). However, in order to obtain single-crystalline growth, traditional MBE growth techniques are not sufficient at these temperatures. A technique that produces much more energetic and reactive species, PLD, has been studied as a means for producing single-crystalline layers at low temperature. While a good degree of success was obtained by growing via PLD with a high power 533 nm pulsed laser source, the surface morphology of these layers was extremely poor. By lowering the laser wavelength, a drastic improvement in the surface morphology was seen but only at the cost of a decline in crystal quality. It is determined that extremely low wavelength and high power light such as that produced by excimer lasers is necessary to improve the surface morphology while still obtaining high crystalline quality at low growth temperatures.

The manner in which the CeO_2/silicon interface evolves with increasing temperature has been examined in great detail. Knowledge of the nature of interfacial reactions as a function of temperature is critical in determining whether thin films will survive device processing with their properties intact. It is found that the CeO_2/silicon interface is extremely reactive. At the temperatures used in this study, there is always an ultrathin layer of SiO_{2-x} that forms beneath the CeO_2. As the sample is heated in vacuum, this layer grows in thickness but never increases its oxidation state. At the same time, the CeO_2 layer on top decreases in oxidation state, first at the surface, then progressing towards the interface. Annealing in vacuum at temperatures greater than 800°C causes both the SiO_{2-x} and CeO_x species to become volatile and actually

desorb from the wafer surface. Annealing in oxygen increases the oxidation state of both the SiO_{2-x} and the CeO_x towards SiO_2 and CeO_2 respectively. The SiO_2 layer steadily grows with a CeO_2 layer of constant thickness and oxidation state riding up on top of it. It is found that in order to minimize interfacial reactions, an extremely low temperature growth technique is necessary and that processing past about 400°C may cause a degradation of the interface quality.

The electrical characteristics of CeO_2 thin films and of the CeO_2/silicon interface were examined by studying the impedance characteristics of metal-CeO_2-silicon capacitors. The capacitors show good C-V characteristics, with an as-grown dielectric constant of about 10 and evidence of strong inversion and accumulation. However, the as-grown capacitors show a high interface trap state density and a significant hysteresis effect in the C-V curves. The charges responsible for the hysteresis seem to have a discharging time constant on the order of seconds so they are attributed to charged defects in the insulator very near to the interface. The interfacial SiO_{2-x} layer discussed above is thought to be responsible for the existence of these charges. Hydrogen annealing is found to reduce the interface trap state density from 2-4$\times 10^{12}$ $cm^{-2}eV^{-1}$ to 4×10^{11} $cm^{-2}eV^{-1}$. Hydrogen annealing also completely eradicates the charges responsible for the hysteresis. In general, it is shown that CeO_2-based MOS capacitors can be made to have suitable characteristics for use as a gate dielectric for a silicon MOSFET.

In summary, growth of thin, stoichiometric, single-crystalline CeO_2 layers on silicon is found to be extremely difficult. Thus, for tunnel barrier applications such as in a resonant tunneling diode, use of this material may not be advised. However, thicker single-crystalline layers with good insulating properties and reasonable interface quality can be manufactured. This suggests that CeO_2 may prove useful as an insulator for SOI devices or as an alternative gate dielectric for a silicon MOSFET.

Part II

6.1 Å Semiconductor-Based Avalanche Photodiodes

Chapter 6 The Physics of Avalanche Photodiodes

6.1 Overview

The theory of impact ionization in semiconductors is reviewed, and connected with the phenomenon of avalanche multiplication. The origin of the excess noise that is created during the avalanche multiplication process is described and the dependence of this excess noise on the ratio of the impact ionization coefficients for holes and electrons is illustrated. Differences in the impact ionization coefficients for different semiconductor materials are discussed. Two systems which will theoretically have enhanced hole impact ionization coefficients are identified for further study in this work.

6.2 Introduction

As discussed in Section 1.2.2, avalanche photodiodes (APD's) operate by using impact ionization to create a current gain. In the gain medium, a high field is present which can accelerate carriers to energies sufficient to excite an electron in the valence band of the material to the conduction band. The newly created electron-hole pair subsequently can be accelerated along with the initiating particle and all three can initiate new impact ionization events. The physics that govern this basic process are discussed in detail in this chapter.

The impact ionization coefficient, which is a measure of how easily a carrier can initiate an impact ionization event, will be defined. This impact ionization coefficient will be proportional to the electric field and inversely proportional to a "threshold energy." The threshold energy will be seen to be a material-dependent parameter

that can vary greatly between the two types of carriers.

The ratio of the hole and electron impact ionization coefficients is shown to be a critical parameter in determining the performance of a given gain medium. APD's designed with gain mediums that show a large difference between the two coefficients have the best noise performance and highest bandwidth. Further, the device performance will depend on the ability to select the right type of carrier to inject into the gain medium. It is always desirable to have the signal be carried by the carrier with the highest impact ionization coefficient.

The threshold energy is derived for a specific band structure that is applicable to 6.1 Å semiconductors. It is shown that when the bandgap energy equals the energy separation between the valence band maximum and the spin-orbit split-off band maximum, the threshold energy will be at a minimum. Also, at this "resonance" condition, all the carriers involved in the impact ionization event have zero momentum, greatly increasing the event's cross section.

6.3 Threshold Energy and Impact Ionization Coefficient

Even though carriers that exist in the bands of a given material are in a basically continuous distribution of momentum and thus energy states, one can define a threshold energy for carriers, below which, no impact ionization events can occur. It is to be understood that this energy is not a simple "on-off" switch for carriers initiating an impact ionization event, but it should roughly determine the minimum energy for carriers to start creating a significant number of impact ionization events. It is clear that, in a semiconductor, this energy must at least be equal to the bandgap energy since, for energy to be conserved in the process, the initiating carrier must have at least enough energy to excite an electron from the valence band to the conduction band. In an electric field, the distance required for a ballistic carrier to achieve this

energy is simply

$$d_{impact} = \frac{E_{th}}{q|\vec{E}|}, \tag{6.1}$$

where $\vec{E}$ is the magnitude of the electric field and E_{th} the threshold energy. In a real material there is always some probability that the carrier will undergo inelastic scattering before it reaches the threshold energy. The probability that it will gain this energy will be denoted as $P(E_i)$. Combining these two factors, the impact ionization coefficient for electrons is defined as

$$\alpha = \frac{q|\vec{E}|}{E_{th}^{elec}} P(E_{th}^{elec}). \tag{6.2}$$

Qualitatively, α^{-1} is the average distance between impact ionization events. One can write a similar expression for β, the hole impact ionization coefficient:

$$\beta = \frac{q|\vec{E}|}{E_{th}^{hole}} P(E_{th}^{hole}), \tag{6.3}$$

where neither E_{th} nor the form of $P(E_{th})$ need be the same for electrons and holes as will be seen in Section 6.5. There are several proposed forms for $P(E_{th})$, the most common being the Shockley model for low fields. In this model, inelastic collisions are assumed to occur randomly in space with a characteristic distance between collisions called the inelastic mean free path, λ. Thus the distance between events is governed by a Poisson distribution and the probability of a particle going a given distance, d, before undergoing a collision is exp(-d/λ). This is called the "lucky electron model":

$$P(E_{th}) = exp\left(-\frac{E_{th}}{e|\vec{E}|\lambda}\right). \tag{6.4}$$

The lucky electron model is valid for low fields. There is another commonly used probability function called the Wolff theory which is valid for high fields and is given by

$$P(E_{th}) = exp\left(-\frac{3E_{th}E_p}{(e|\vec{E}|\lambda)^2}\right), \tag{6.5}$$

where E_p is the energy lost per inelastic collision. Finally, there is the Baraff theory that is a sort of hybrid of the two. The distribution function should also theoretically take into account the actual band structure of the material, thus more accurately expressing the energy distribution of electrons and the probability of interaction with various types of phonons. The three distributions mentioned here in addition to some more intricate ones that take band structure into account are described in detail by Capasso [1].

6.4 Noise Considerations

As one might guess, there are tradeoffs in employing the enormous gain offered by APD's. Current measurements have shot noise associated with them. This stems from the fact that electrons are discrete particles and arrive at a given point in a circuit randomly in time, with the arrival times governed by a Poisson distribution [2]. Ideally, upon amplification, the current noise gets multiplied by the same multiplication factor as the signal. However, impact ionization processes are themselves statistical in nature, obeying some sort of distribution in space. Thus instead of having current multiplied by a fixed value, "M," it is multiplied by a distribution of M's. It is the spread of this distribution that gives rise to the so-called excess noise factor, "F" [3, 4]. Further, it was shown by McIntyre [3] that this excess noise factor is extremely dependent on the relative values of the ionization coefficient for each type of carrier and on which type of carrier first initiates the chain of impact ionization events. For electron initiation,

$$F_n = M_n \left\{ 1 - (1-k) \left[\frac{M_n - 1}{M_n} \right]^2 \right\} \tag{6.6}$$

and for hole initiation

$$F_p = M_p \left\{ 1 - [1 - (1/k)] \left[\frac{M_p - 1}{M_p} \right]^2 \right\}, \tag{6.7}$$

where k$\equiv \beta/\alpha$ and

$$M_n = \frac{1}{1 - \int_0^W \alpha exp\left[-\int_0^x (\alpha - \beta) dx'\right] dx} \tag{6.8}$$

$$M_p = \frac{exp\left[-\int_0^W (\alpha - \beta) dx\right]}{1 - \int_0^W \alpha exp\left[-\int_0^x (\alpha - \beta) dx'\right] dx}, \tag{6.9}$$

where W is the width of the multiplication region.

Figure 6.1 shows plots of Eqs. 6.6 and 6.7 as a function of multiplication. One can see that in order to minimize the excess noise factor, k has to be very small for electron injection and very large for hole injection. So in a system where $\beta \gg \alpha$ there must be a preferential injection of holes into the multiplication region for the benefits of the high k value to be seen. Note that for k=1, the excess noise factor is exactly equal to the multiplication in either case.

This dependence on the ratio of the impact ionization coefficients can be intuitively understood as follows. Figure 6.2 (a) is a digram of several stages of multiplication initiated by one electron, for a material in which k$\sim$1. Impact ionization events create new carriers headed in both directions in space. These carriers initiate new impact ionization events once they have traversed identical distances since $\alpha = \beta$. Note that while the carriers travel the same distance between events, they may not necessarily take the same amount of time to reach the next event since in general the drift velocities of holes and electrons are not the same. Thus the two types of carriers "bounce" back and forth between impact ionization events creating an unrestrained feedback process. Figure 6.2 (b) shows the case where $\alpha \gg \beta$. While carriers are created that travel in both directions, multiplication only progresses in one direction. It is clear from these figures that when $\alpha = \beta$ the multiplication processes that take place after the injection of one electron sample a much broader section of space and time. Hence, even though the multiplication for the $\alpha = \beta$ case may be enormous, the chances of encountering a fluctuation in the carrier density such as a thermally generated electron-hole pair is far greater. And since these fluctuations are themselves multiplied along with the rest of the carriers, the noise is amplified.

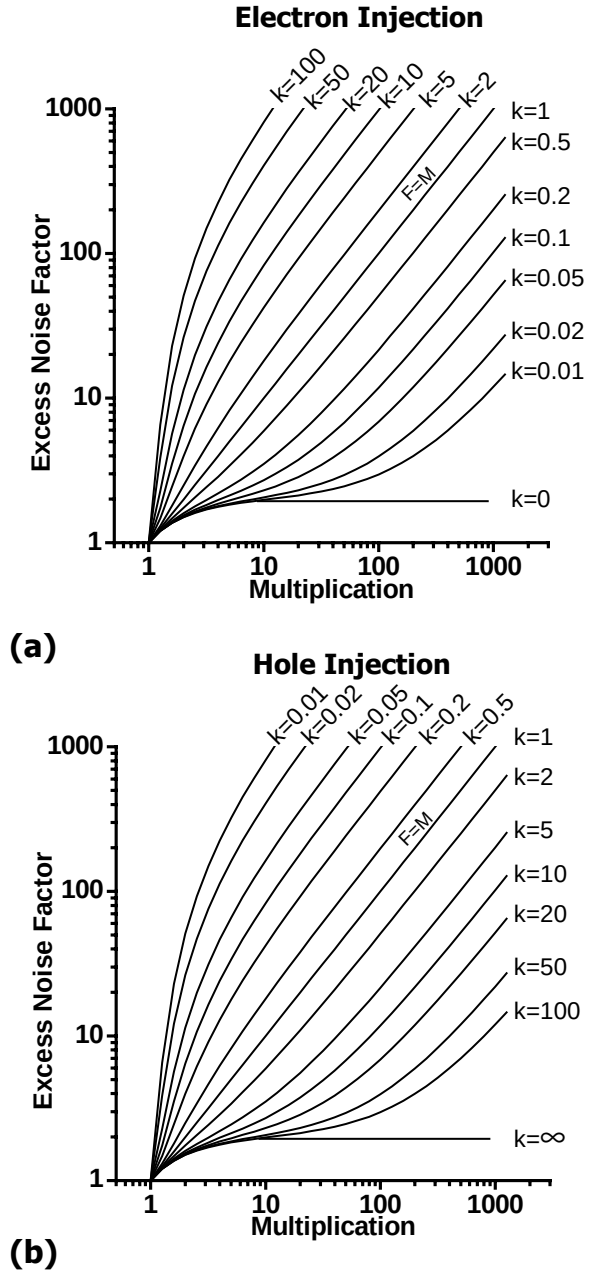


Figure 6.1: **Excess noise factor as a function of gain with the impact ionization ratio k=β/α as parameter. Figure (a) is for electron injection taken from Eq. 6.6 and Fig. (b) is for hole injection taken from Eq. 6.7. Note that for k=1, the excess noise factor is simply equal to the multiplication.**

For all cases in between the two shown in Fig. 6.2, one can imagine that the feedback process still exists but dies down faster in time. Thus the amount of space and time that a "pulse" initiated by one electron samples in the multiplication region is roughly proportional to the ratio of the impact ionization coefficients.

In addition to the degradation in noise behavior, another immediate consequence of having α close to β is that the bandwidth of the device will be greatly reduced. Again, the time it takes for an electron pulse to be completed is roughly proportional to the ratio of the impact ionization coefficients. However, it is not immediately obvious that the gain-bandwidth product will suffer. Emmons was able to derive an expression for the frequency dependence of the multiplication of an APD with arbitrary values of the impact ionization coefficient from direct solutions of the carrier transport equations [5]. The results of these calculations are summarized in Fig. 6.3 for the case of electron injection. One can see immediately that for $\alpha \gg \beta$, the 3 dB bandwidth hardly changes at all with increasing multiplication (top part of the figure). On the other hand for values of β close to α, there is a drastic change as multiplication increases. So having very different values of impact ionization coefficients does indeed improve the gain-bandwidth product as well.

In summary, the two most important factors in designing an APD seem to be picking a material with divergent values of the electron and hole impact ionization coefficients and taking advantage of this difference by preferentially injecting the appropriate carrier type into the multiplication region of the device.

6.5 Materials Considerations

As stated above, for optimal APD performance, one should choose a material exhibiting very different values of the electron and hole impact ionization coefficients. A table of some common semiconductor materials along with their values of k=β/α, and energy gap are given in Table 6.1. As mentioned in Section 1.2.2, use of the 6.1 Å semiconductors is most advantageous in the infrared spectrum, but it is interesting to note that all of the wider gap semiconductors seem to have very poor values of k

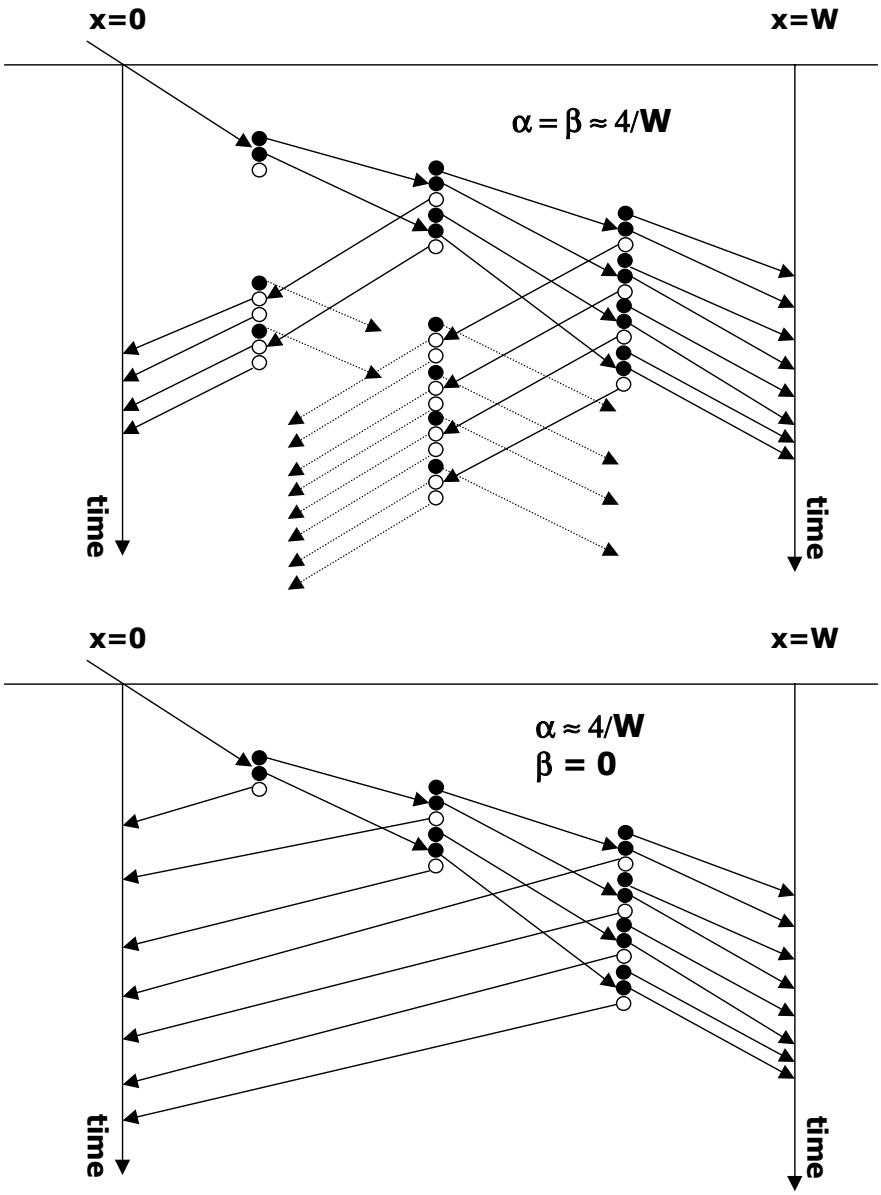


Figure 6.2: **Diagram of avalanche multiplication for $\alpha = \beta$ and $\alpha \gg \beta$.**

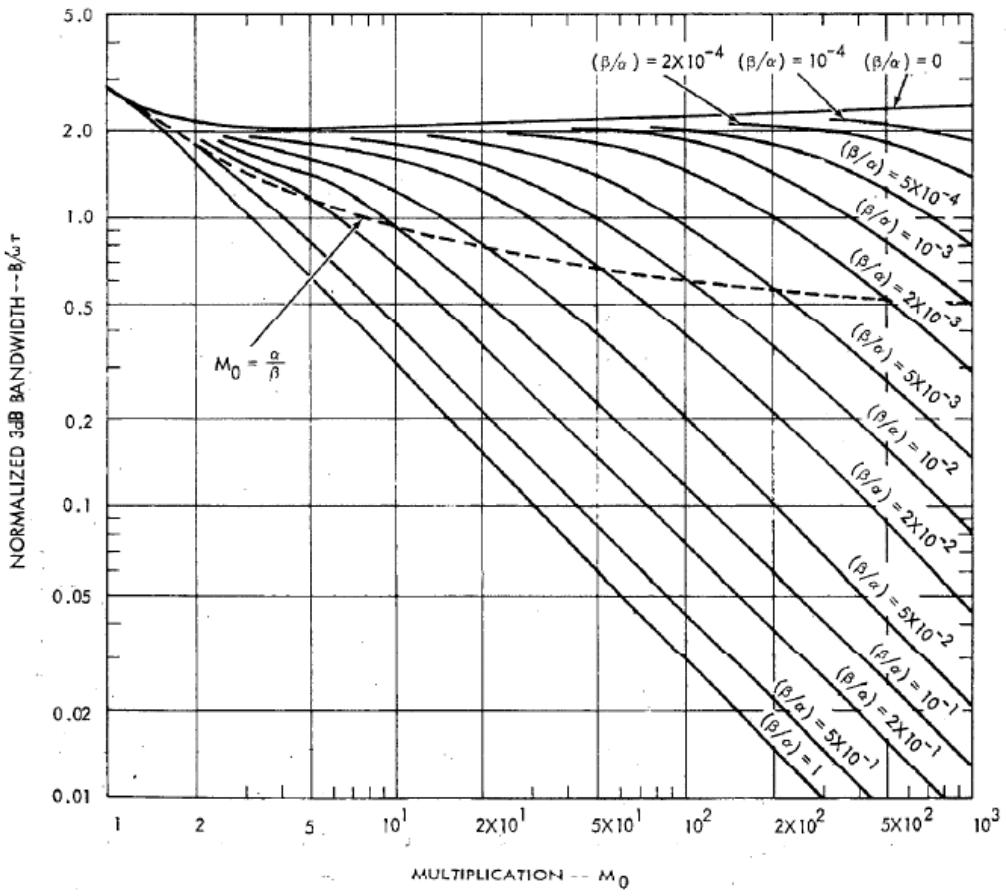


Figure 6.3: **3 dB bandwidth for electron injection in APD's with various values of k. Taken from Emmons [5].**

(close to one).

The impact ionization coefficients of various alloy combinations of all the III-V elements also have been measured, but, except for in two cases, the alloys show very little change in k as compared to the pure semiconductors [1]. The two exceptions are $Hg_xCd_{1-x}Te$ and $Al_xGa_{1-x}Sb$. In these systems, for very specific alloy concentrations, a large increase is seen in the hole impact ionization coefficient: k=20 for $Al_{0.06}Ga_{0.94}Sb$ [7] and k=10 for $Hg_{0.73}Cd_{0.27}Te$ [1]. As it turns out, at these specific alloy concentrations the bandgap of the material is close to the split-off band to valence band edge energy separation (Δ_{so}). This "resonance" enhances the impact ionization coefficient in two ways. Firstly, one can derive expressions for the threshold energy of a material with three parabolic bands, a conduction band and two valence bands separated by Δ_{so}. For an electron-initiated impact ionization event, the threshold energy is derived in Appendix B to be

$$E_{th}^{elec} = E_g \left(2 + \frac{m_e}{m_{hh} + m_e}\right), \tag{6.10}$$

where E_g is the bandgap of the material, m_{hh} is the heavy hole effective mass and m_e is the electron effective mass. On the other hand, for a hole-initiated impact ionization event

$$E_{th}^{hole} = E_g \left[1 + \frac{m_{so}(1 - \Delta/E_g)}{2m_{hh} - m_{so} + m_e}\right], \tag{6.11}$$

where m_{so} is the split-off band hole effective mass. It is evident from these equations that the hole threshold energy goes through a minimum equal to the bandgap energy when $\Delta_{so}=E_g$ whereas the electron threshold energy does not. Referring back to Eq. 6.2, one can see that lowering the threshold energy for a given carrier will clearly increase that carrier's impact ionization coefficient. While this minimum in the hole threshold energy certainly contributes to the resonance in the hole impact ionization coefficient, it is not enough to explain the large increase in k seen for the two cases mentioned above. In fact $E_{th}^{elec}/E_{th}^{hole}$ is at most equal to three (for $\Delta_{so}=E_g$ and $m_{hh} \rightarrow 0$). Further insight can be drawn concerning this resonance by examining the expression for the momenta of the particles involved. The momentum for hole-

initiated events is derived in Appendix B to be

$$k_{th}^{hole} = \left[(E_g - \Delta_{so}) \left(\frac{2m_{so}}{\hbar^2} \right) \left(\frac{2m_{hh} + m_e}{2m_{hh} - m_{so} + m_e} \right) \right]^{1/2}. \tag{6.12}$$

It is obvious that for the case of Δ_{so}=E_g, k_{th}^{hole}=0. In fact under these conditions the momenta of all the particles will be zero (see Appendix B, Eqs. B.12,B.13). Thus, as illustrated in Fig. 6.4, the transitions are completely vertical in momentum space at k=0 . This means that no phonons or other scattering mechanisms need be involved in the process. Also, the density of states will be highest in the split off band near k=0. In Eqs. 6.2 and 6.3 it was assumed that when a carrier reaches the threshold energy, it ionizes with unity probability. However, even past the threshold energy, the impact ionization event has a finite cross section, lowering α and β. This effect can be folded into the probability term P(E_{th}). Basically, the rest of the enhancement effect comes from the fact that this cross section is greatly increased for the case where Δ_{so}=E_g by the zero momentum condition.

6.6 Choosing Materials for Study in This Work

As mentioned above, specific alloys of $Hg_xCd_{1-x}Te$ and $Al_xGa_{1-x}Sb$ exhibit this resonance between the split-off band energy and the bandgap. It is also possible to engineer the resonance by constructing superlattices. El-Rub et al. [8] have calculated several superlattices based on III-V semiconductor systems that are predicted to show this resonance. These superlattices require several layers per period to obtain the appropriate energy spacings. One of their calculated superlattices (35Å InAs/20Å $In_{0.4}Ga_{0.6}Sb$/15Å AlSb) lends itself to the growth capabilities of our lab. Thus for the experiments described in the subsequent chapters, two multiplication regions were used for constructing APD's. In Chapter 7, $Al_xGa_{1-x}Sb$ alloys are used as the multiplication layer in p-i-n APD's to enhance the hole multiplication coefficient. In Chapter 8, an attempt is made to fabricate the three-stage superlattice described above for use as a multiplication layer.

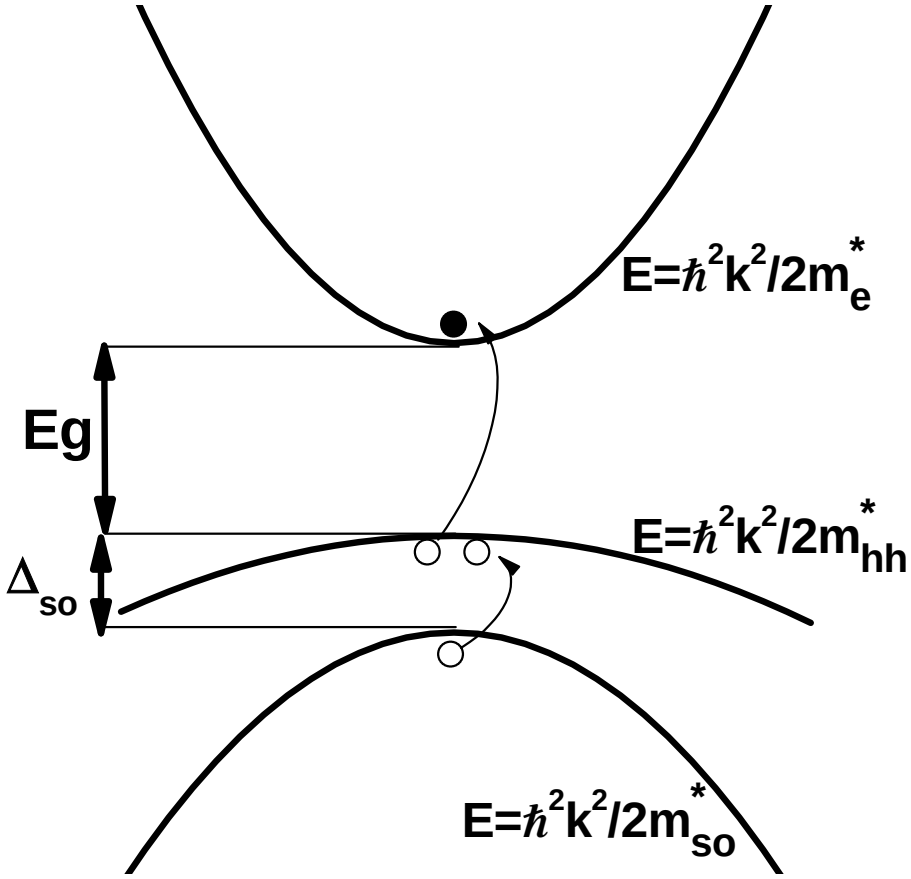


Figure 6.4: **Schematic of an impact ionization process for a semiconductor with three parabolic bands.**

Material	k = β/α	E_G (eV)	E_G (μm)
GaP	1	2.26	0.55
GaAs	0.5	1.42	0.87
InP	2	1.35	0.92
Si	0.05	1.1	1.13
Ge	2	0.66	1.88
InAs	0.1	0.36	3.4
$Hg_xCd_{1-x}Te$	10	≤ 0.83	≥ 1.5

Table 6.1: **Values of k and E_G for some common semiconductors. Data taken from Stillman and Wolfe [6] and Capasso [1].**

Bibliography

[1] F. Capasso, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beers (Academic Press, New York, New York, 1985), Vol. 22, Chap. 1, pp. 1–172.

[2] S. Sze, *Physics of Semiconductor Devices* (Wiley, New York, Ny, 1994).

[3] R. J. Mcintyre, IEEE Trans. Electron Devices **Ed13**, 164 (1966).

[4] R. J. Mcintyre, IEEE Trans. Electron Devices **Ed19**, 703 (1972).

[5] R. B. Emmons, J. Appl. Phys. **38**, 3705 (1967).

[6] G. E. Stillman and C. M. Wolfe, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beers (Academic Press, New York, New York, 1977), Vol. 12, pp. 291–392.

[7] O. Hildebrand, W. Kuebart, K. W. Benz, and M. H. Pilkuhn, IEEE J. Quantum Electron. **17**, 284 (1981).

[8] K. A. El-Rub, C. H. Grein, M. E. Flatte, and H. Ehrenreich, J. Appl. Phys. **92**, 3771 (2002).

Chapter 7 $Al_xGa_{1-x}Sb$ APD's

7.1 Overview

In Chapter 6, the nature of the resonant enhancement of the hole impact ionization coefficient in $Al_xGa_{1-x}Sb$ alloys was discussed. In this chapter, an attempt is made to fabricate APD's using this material in the multiplication region. Aspects of growth, materials characterization and device processing are discussed in the first sections. Then analysis of the electrical and photoresponse behavior of the finished devices is presented.

7.2 Introduction

Gallium-rich alloys of $Al_xGa_{1-x}Sb$ have been shown to greatly favor hole-initiated impact ionization over electron-initiated impact ionization because of a resonance in the energy spacings of the bandgap, E_g and the split-off band energy, Δ_{so} as demonstrated by Hildebrand et al. [1, 2]. In his experiment, it was found that values of k as high as 25 could be achieved by picking the alloy concentration such that Δ_{so}=E_g. Despite these promising results, the nature of this resonant enhancement in the hole ionization coefficient was a source of some controversy both in experiment [3] and theory [4, 5, 6]. More recently, Grein and Ehrenreich [7] have resolved some of the inconsistencies by describing the enhancement effect as a function of electric field in the multiplication region. They found that for high fields, where carrier energies are on the order of optical phonon energies, the hole ionization enhancement is reduced whereas for lower fields the effect does occur. This seems to explain why the experiments of Hildebrand et al. showed the enhancement whereas similar experiments by Kuwatsuka et al. did not [3]. Further, Grein and Ehrenreich's study identified the optimum alloy concentration for enhanced hole impact ionization to be a

slightly lower aluminum concentration than that used in previous experiments. Thus an alloy concentration of $Al_{0.05}Ga_{0.95}Sb$ was used in all of the experiments discussed below.

Molecular Beam Epitaxy (MBE) grown GaSb tends to have a high p-type background doping concentration because of Ga on Sb anti-site (Ga_{Sb}) defects [8]. For the MBE growths used in this study, a hole concentration of about 5×10^{16} cm^{-3} was found for unintentionally doped GaSb layers. At these doping levels, an n-i junction at a reverse bias of 10 volts will only have depleted 500 nm into the intrinsic region. In addition, it will see a maximum electric field of about 300 kV/cm which is in the high field regime as defined by Grein and Ehrenreich, where the enhancement effects are reduced. All previous studies of $Al_xGa_{1-x}Sb$ multiplication layers were grown with liquid phase epitaxy (LPE) which does not suffer from this background doping concentration. However, in order to integrate APD's with other 6.1 Å-based devices or utilize superlattices to tune the absorption wavelength of the detectors, MBE growth is necessary. It is clear that in order for MBE-grown antimonide p-i-n APD's to function properly and take advantage of the resonant enhancement of the hole ionization coefficient, a means for reducing the effective background doping concentration is needed.

Longenbach and Wang [9] showed a significant improvement in the characteristics of $Al_xGa_{1-x}Sb$ p-i-n structures by growing on GaSb (111)B surfaces because of an enhanced antimony sticking coefficient and thus a decrease in Ga_{Sb} defects. However GaSb (111)B substrates are much more difficult and expensive to obtain and all other technologically interesting structures in antimonide based semiconductors are traditionally grown on (001) substrates. A simple, straightforward counterdoping technique is suggested to reduce the effective background doping concentration in $Al_xGa_{1-x}Sb$ p-i-n APD's via the introduction of a small concentration of tellurium as an n-type counterdopant from thermal evaporation of GaTe. Tellurium doping of GaSb has been shown to exhibit not only a shallow donor level but also a midgap level [10]. While the nature of this midgap level is a source of some controversy [11, 12], the combination of donor-like behavior with midgap states provide a good basis for

compensating native p-type material.

Another stumbling block in creating a practical APD using $Al_xGa_{1-x}Sb$ alloys is surface leakage. In fact, one of the fundamental roadblocks to creating any practical devices in the 6.1 Å system has been the quality of the semiconductor surfaces [8]. As is the case in GaAs, 6.1 Å semiconductor devices are hindered by a high surface state density and oxidation upon exposure to air. In the case of mesa diodes, surface leakage current can dominate the device behavior [13, 14]. Several attempts have been made to find passivation techniques for GaSb [15, 16, 17], InAs [14] and various ternary [13, 18] and quaternary [19] alloys. While some of these efforts have been moderately successful in reducing dark currents in p-n junction diodes and reducing surface state densities, an all-inclusive passivation technique for the entire 6.1 Å system is still being sought.

Recently, plasma nitridation of GaAs has become of interest because of success in both growth of nitride semiconductors on GaAs substrates [20] and surface passivation [21, 22]. It is thus possible that plasma nitridation of the 6.1 Å semiconductors would yield similar beneficial results. The process of nitridation of 6.1 Å semiconductor surfaces is analyzed below in order to determine if it will be beneficial for mesa APD's. However, it is found that the reverse bias leakage current is reduced only by a factor of two under this treatment.

In order to obtain devices with good reverse bias characteristics, it is found that a temporary solution to the surface passivation problem is simply to reduce the total exposed surface area. This is accomplished by etching mesas down only to the intrinsic region in the p-i-n structure. This is just deep enough to isolate devices while leaving behind the lowest possible surface area exposed to the air.

Using the techniques described above, viable p-i-n devices can be fabricated. In the latter parts of this chapter, the performance of these devices is assessed using photoresponse measurements.

7.3 MBE Growth

The device layers for these studies were grown either on GaSb (001) or on GaAs (001) wafers. The wafers first were degreased in acetone, isopropyl alcohol, and deionized water. The GaSb wafers then were etched in a mixture of Br_2, CH_3COOH and HNO_3 to smooth the surface and reduce the pit density on the wafer surfaces [8], whereas the GaAs wafers were not etched in any way. After a final rinse in deionized water, the wafers were indium bonded to a solid molybdenum block. Finally, these blocks were introduced into vacuum through a load lock capable of reaching a pressure of less than 1×10^{-7} torr within two hours.

All the samples were grown in a Perkin-Elmer model 430 MBE system by traditional III-V MBE techniques. Gallium and aluminum were evaporated from Knudsen cells while antimony and arsenic were sublimated from cracking effusion cells to create Sb_2 and As_2 dimers for growth. Substrates were heated in a flux of the column V material until the native oxide desorbed. This occurred at around 550°C for the GaSb and around 600°C for the GaAs. After the oxide desorption, the substrate temperature was lowered to the growth temperature of approximately 500°C. Growth rates were calibrated by observation of RHEED intensity oscillations. The growth rates of GaSb and AlSb were calibrated separately on thick samples of each material. From these growth rates, antimonide materials with arbitrary aluminum to gallium ratios could be grown. In order to check the alloy concentration, thick layers were grown with a constant aluminum to gallium ratio. These layers were examined *ex situ* by x-ray diffraction (XRD) in order to verify the alloy concentration. Figure 7.1 shows two examples of XRD data for samples with $Al_xGa_{1-x}Sb$ layers of varying aluminum concentrations. The displacement of the main overlayer peak relative to the substrate peak gives a direct measurement of the overlayer's lattice constant in the growth direction. This lattice constant can be converted to a composition if the overlayer is assumed to be completely relaxed or completely strained. In our case, all of the $Al_xGa_{1-x}Sb$ layers were completely strained. Further, the interference fringes evident in both of the scans in the figure give information about the total thickness

of the overlayer, just as in the case of x-ray reflectivity mentioned in Chapter 2.

After growth, the samples were cooled in a flux of Sb_2 until a stable (1×5) reconstruction was visible with RHEED. Below about 350°C, the reconstruction could be maintained without the Sb_2 flux and it persisted down to room temperature indicating a smooth finished surface.

7.3.1 Doping

For n-type layers, tellurium doping was used, whereas for p-type layers, beryllium was used. The tellurium was evaporated from a solid GaTe source and the beryllium from a solid beryllium source. The dopant concentration as a function of cell temperature was calibrated by secondary ion mass spectroscopy (SIMS) measurements as shown in Fig. 7.2. To estimate the effect of Te counterdoping on $Al_xGa_{1-x}Sb$ layers, 1.5 μm layers of $Al_{0.05}Ga_{0.95}Sb$ were grown on semi-insulating GaAs (001) substrates with a thin AlSb buffer in between the two. Only the Te concentration in the epitaxial layer was varied from growth to growth. In this way four-point probe and hall measurements could be carried out without correcting for a contribution of the substrate to the conductivity.

Six samples with varying amounts of Te counterdoping were grown on semi-insulating GaAs substrates. A simple four-point probe technique was used to monitor the resistivity and room temperature Hall measurements were used to determine the mobility and carrier type. The results are shown in Fig. 7.3. The secondary abscissa represents an estimate of the total concentration of Te atoms/cm^3 derived from Fig. 7.2 for the various GaTe cell temperatures listed as the primary abscissa. It can be seen that the resistivity reaches a maximum for a Te doping level roughly equal to the background p-type carrier concentration. While the mobility goes down slightly with increased Te doping it picks back up again due to the much higher mobility of electrons as they become the dominant carrier type. Thus compensation is taking place, making the residual carrier concentration lower than that in the unintentionally doped case. An estimate of 7×10^{15} cm^{-3} *n-type* was obtained for the highest

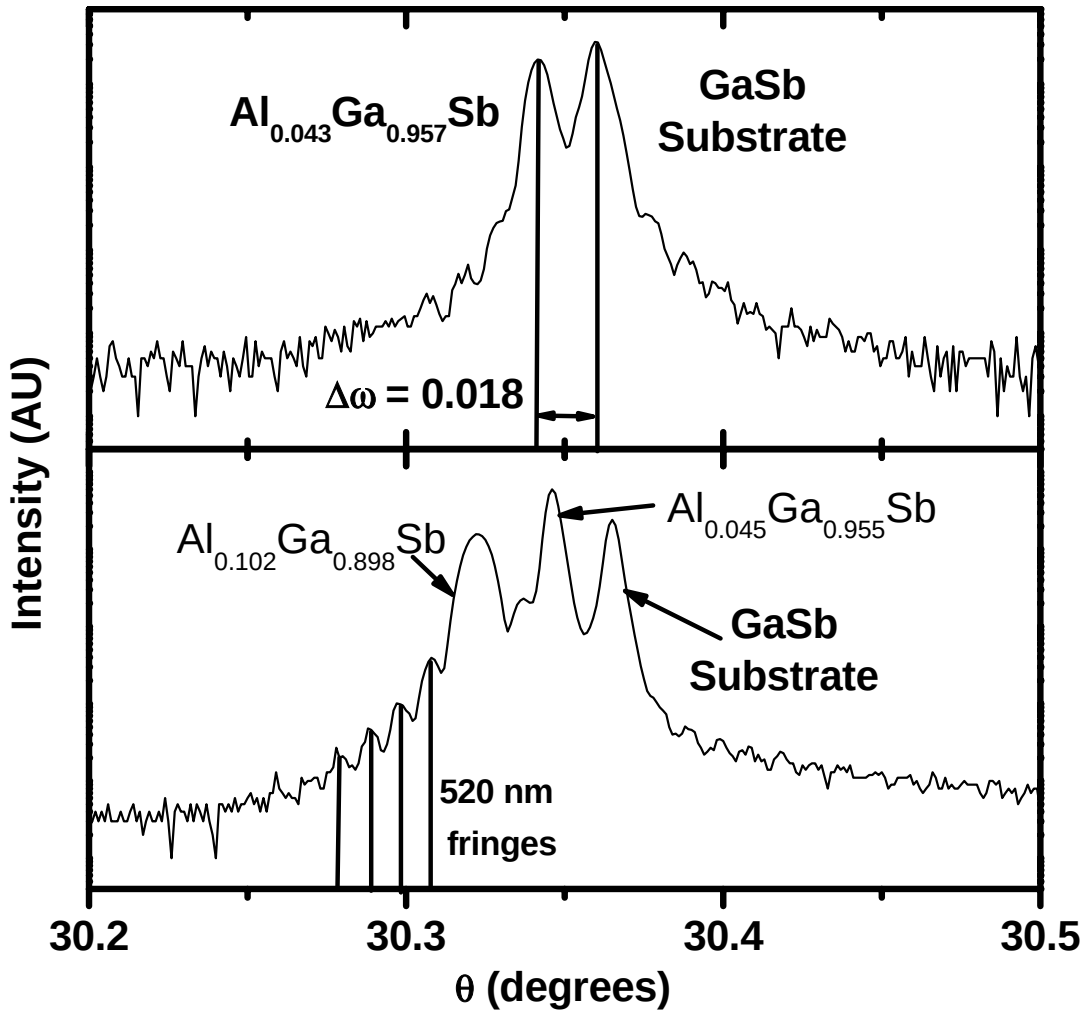


Figure 7.1: $\theta/2\theta$ **scans of** $\mathbf{Al_xGa_{1-x}Sb}$ **on GaSb. The top panel is from a single 500 nm layer of** $\mathbf{Al_{0.043}Ga_{0.957}Sb}$ **grown on GaSb and the top panel is from a sample with two successive layers of** $\mathbf{Al_xGa_{1-x}Sb}$ **grown on GaSb with different aluminum concentrations.**

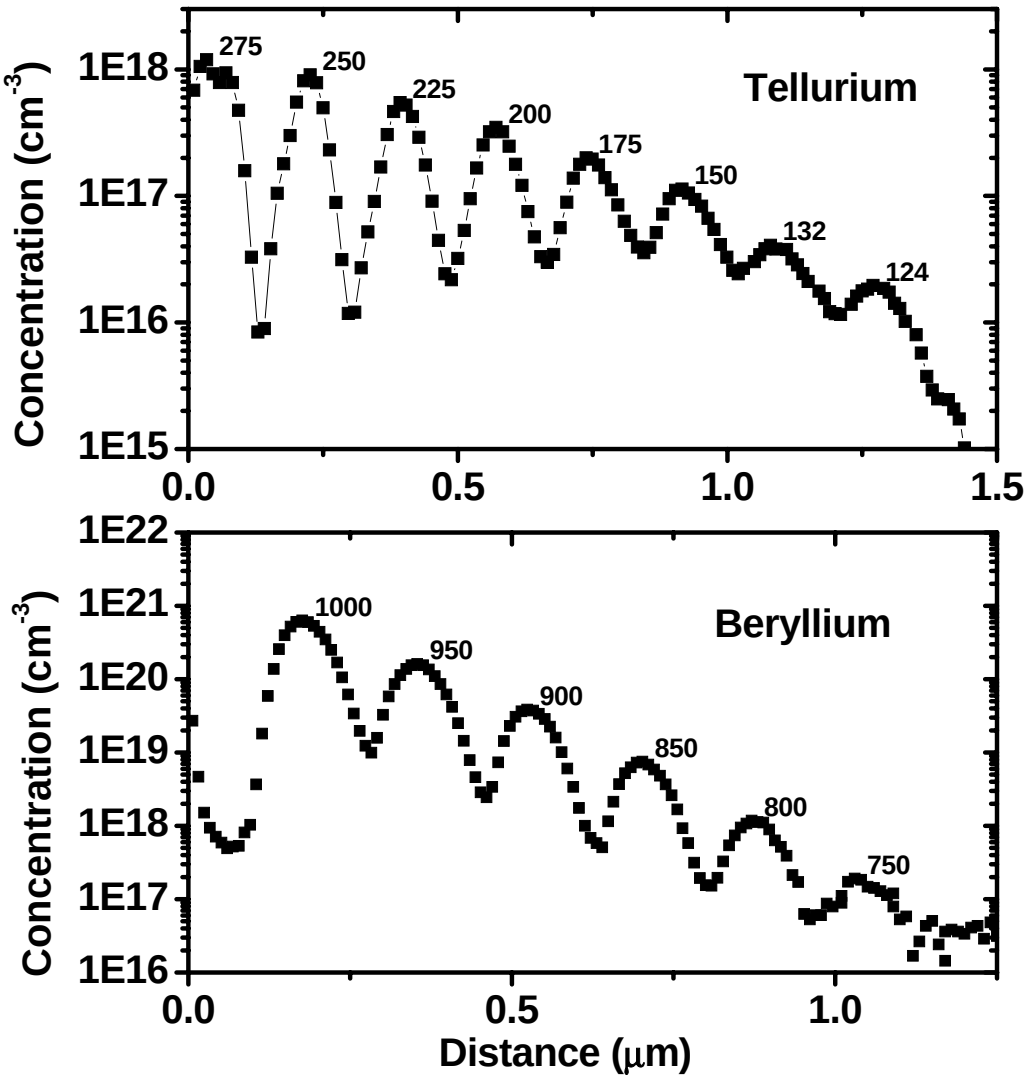


Figure 7.2: **SIMS Profiles of Te and Be doping in MBE growths. The samples examined for this data were grown with a chopped doping profile where the dopant cell shutter was opened for 100 nm of growth then closed for 100 nm of growth. Each successive layer was doped with the dopant cell at a higher temperature. Thus after some diffusion has taken place, the dopant profile as a function of depth exhibits the oscillatory features seen in the figure. The peak labels represent the cell temperature used to create the doping at the depth of the peak. These concentrations were obtained from a sample with a 2Å/sec growth rate. Data is courtesy of C. J. Hill.**

resistivity case. Further optimization of the counterdoping is difficult since each data point requires a long MBE growth, but the trend is certainly encouraging.

P-i-n structures were grown on n-type GaSb (001) substrates. From the substrate up, these structures consisted of 500 nm of roughly 1×10^{17} cm^{-3} n-type GaSb followed by 1 μm of $Al_{0.05}Ga_{0.95}Sb$ either undoped or counterdoped with Te, followed by 500 nm of 1×10^{17} cm^{-3} p-type GaSb. The cap layer was chosen to be p-type material because it is generally much easier to make ohmic contact to p-type GaSb as compared to n-type [8]. Since the entire back plane of the device could be used for making the n-type contact, the rectifying nature of metal/n-GaSb contacts could be avoided simply by having a large contact area.

7.4 Device Processing

The p-i-n structures were fashioned into mesa diodes by chemically assisted ion beam etching (CAIBE) using a xenon plasma and Cl_2 gas. The sample was covered with a photoresist mask during the etch such that etching left behind cylindrical mesas. It was found that the etch rate strongly depends on the doping of the layers; unintentionally doped layers etched the fastest followed by p-type, and n-type was the slowest. Etch rates needed to be carefully controlled in order to determine the optimum mesa depth to reduce surface leakage as will be seen below.

Following the etch step, the photoresist mask was removed. A second lithography step was needed to form recessed metal contact pads. An inverse photoresist mask was used to define holes that line up with the mesas through which metal could be deposited. Gold-germanium (AuGe) was thermally evaporated on the sample followed by lift-off of the mask to form the ohmic contact pads. The entire backside of the sample, which was covered in indium, was used as a contact to the n-type material. Figure 7.4 shows an optical microscope picture of the completed devices along with a schematic of the cross section of the device structure.

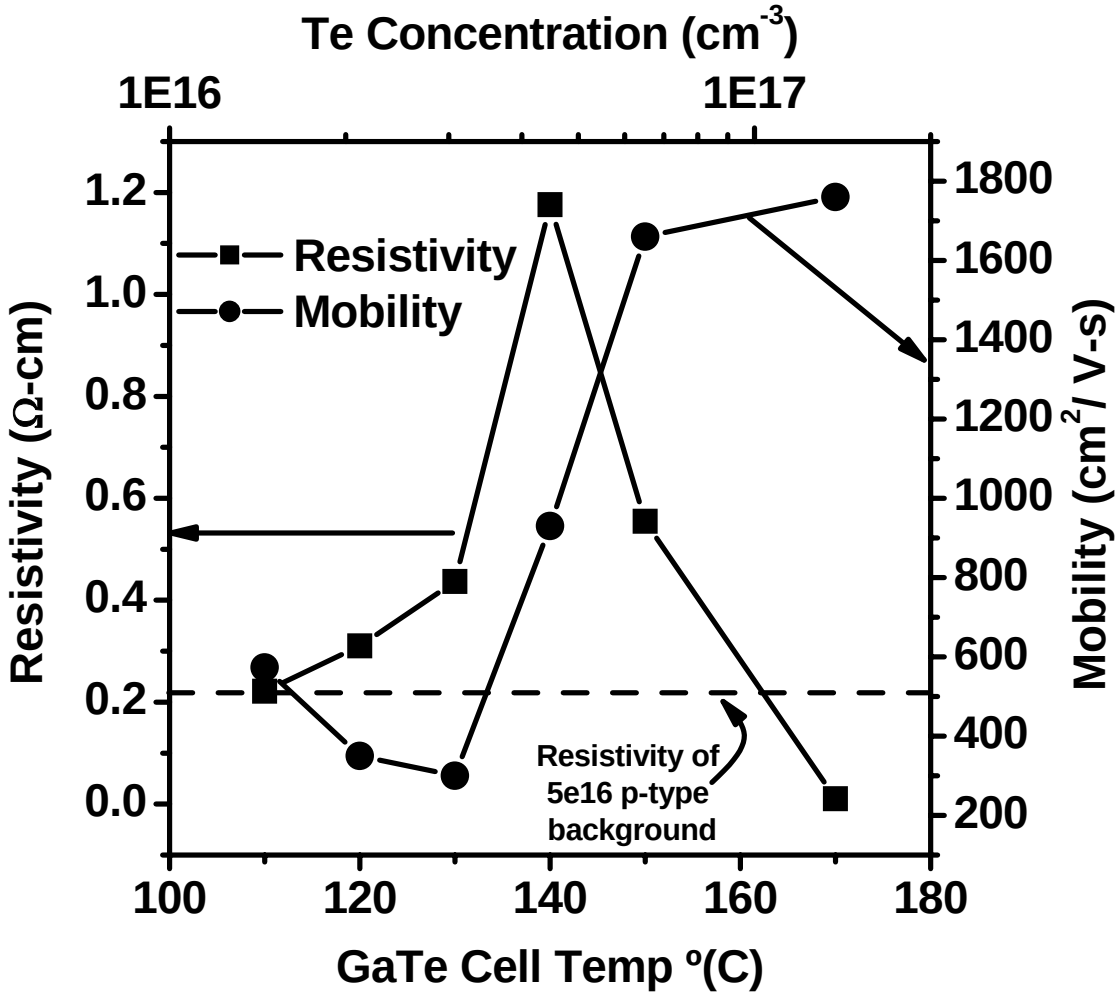
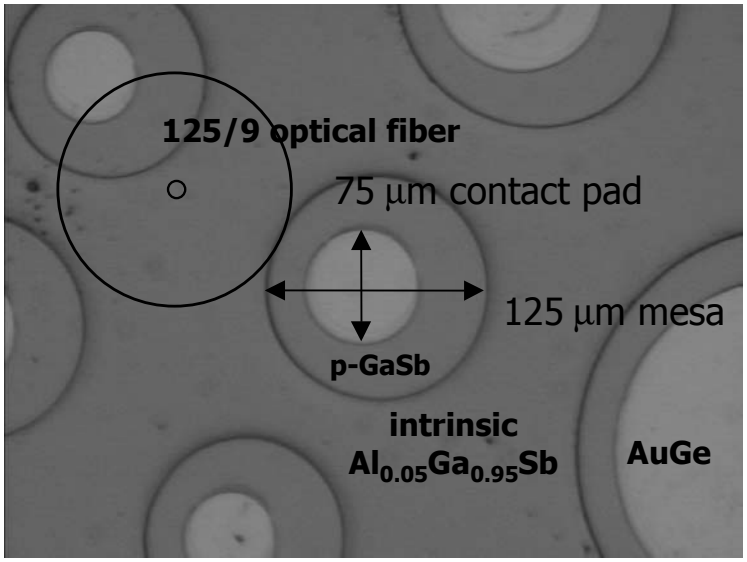


Figure 7.3: **Resistivity and mobility of counterdoped $Al_{0.05}Ga_{0.95}Sb$ layers. The primary abscissa is the GaTe cell temperature while the secondary abscissa is the approximate Te concentration derived from Figure 7.2.**



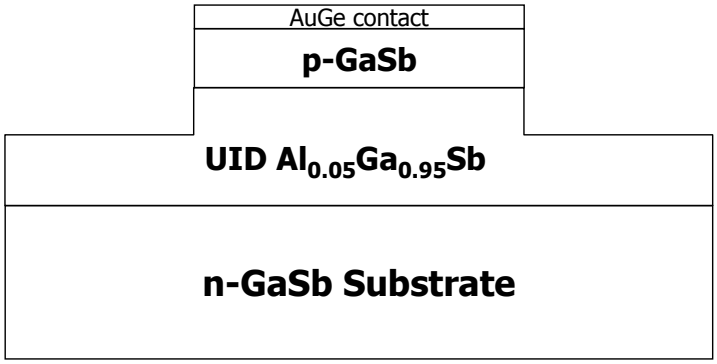


Figure 7.4: **Image of processed p-i-n mesa APD's and schematic cross section of mesas. The footprint of an optical fiber with inner core diameter and outer cladding diameter is shown next to a device in order to give an idea of the scale.**

7.5 Electrical Characterization

The effects of counterdoping were verified by performing current-voltage (I-V) and capacitance-voltage (C-V) measurements on samples with and without the tellurium counterdoping. The tellurium doping level that yielded the highest resistivity samples in the experiment discussed above was used to counterdope the intrinsic region. This sample was grown and processed identically in every other way to a control sample that had no counterdoping. From here on, the non-counterdoped sample will be referred to as sample A and the counterdoped sample as sample B. The data for 150 μm diameter mesas is shown in Fig. 7.5.

It can be seen immediately from the I-V plot that there is a stark difference between the two samples. Sample A breaks down at about half the voltage of sample B. This is consistent with a higher effective doping in the intrinsic region because higher fields will be present at the n-i interface for sample A. This leads to a greater tunneling current and greater impact ionization at a lower bias.

The inset of Fig. 7.5 shows a plot of $1/C^2$ vs. V/V_b, where V_b is the breakdown voltage of the devices as determined from the current-voltage characteristics. Assuming a simple one-sided p-n junction model for the depletion width, the slopes of the two curves near zero bias yield an effective doping of 1.24×10^{16} cm^{-3} and 5.24×10^{15} cm^{-3} for samples A and B, respectively. Hence counterdoping was deemed to be effective enough to be used in every subsequently fabricated APD.

In addition to the effects of counterdoping, it was found that other factors greatly influence the electrical characteristics of the diodes. It was quickly determined that processing plays a significant role in creating diodes with good I-V characteristics. Reverse bias current leakage was found to be extremely large for the first devices that were processed as described above in section 7.4. Various chemical surface treatments were applied but very little effect was noticed. Also, chemical treatments seemed to reduce the uniformity from device to device and the device behavior was not consistent from run to run. Nitrogen plasma treatment was also performed. Figure 7.6 shows the characteristics of mesa diodes with and without post-processing exposure to a

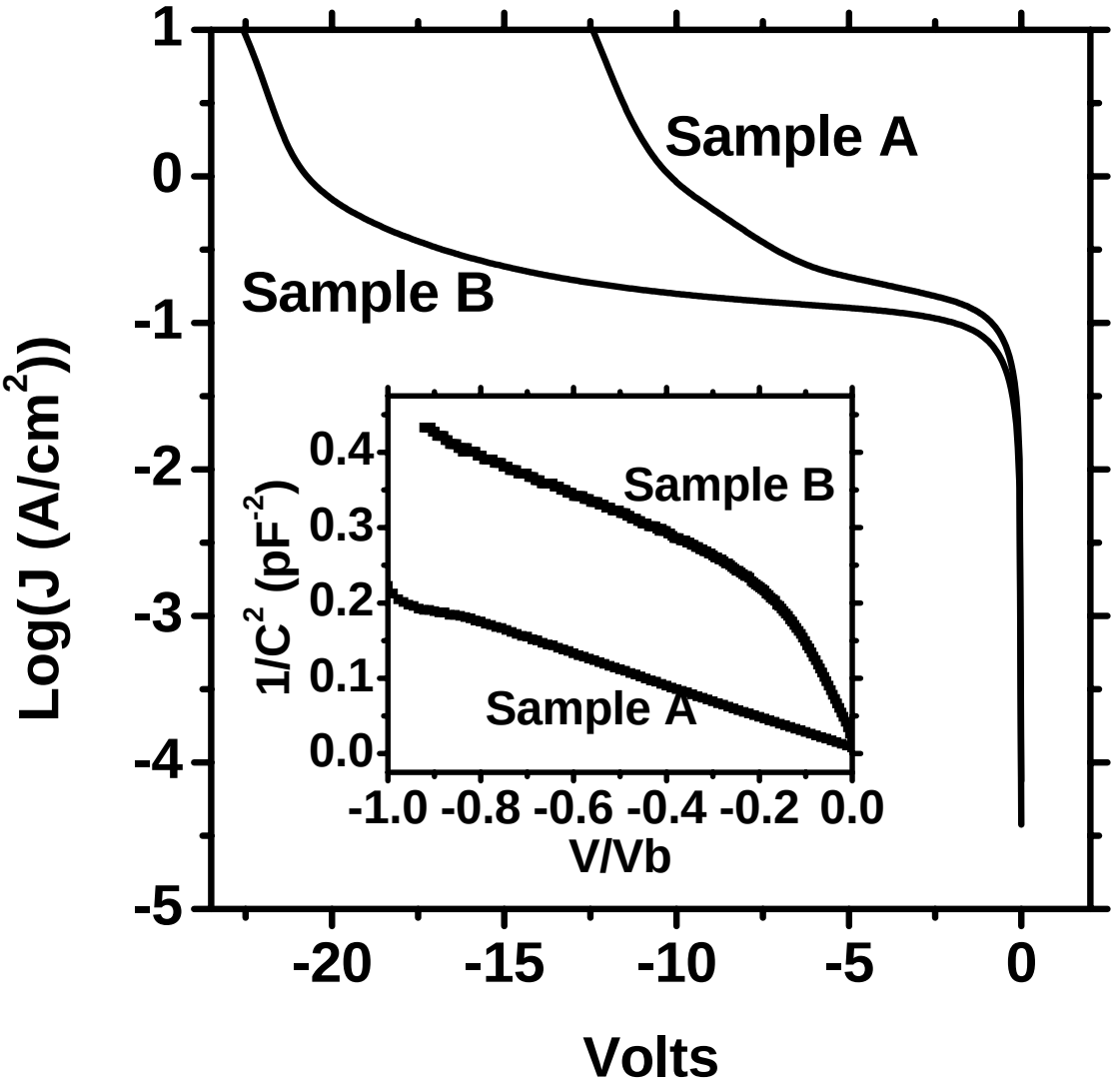


Figure 7.5: **Current-voltage and capacitance-voltage plots for samples A (non-counterdoped) and B (counterdoped). At low voltages, the lower effective doping in sample B can be seen clearly in the slope of the $1/C^2$ vs. V/V_b curve.**

nitrogen plasma. The top two panels show plots of reverse bias leakage current versus device perimeter, with reverse bias voltage as a parameter, plotted on a log-log scale. A slope of one means that the leakage current scales precisely with perimeter and a slope of two means that the leakage scales precisely with device area. The untreated samples show an exponent close to one whereas the nitrogen-treated samples show an exponent of about 1.45, about half way to scaling with area. Figure 7.6(c) shows the average current density over several devices each, for the untreated and nitrided sample. In summary, the nitrogen treatment improved the reverse bias leakage by about a factor of two, and caused the currents to scale somewhat more with the device area. However, this still was not a satisfactory solution.

Figure 7.7 shows I-V curves for p-i-n mesa diodes etched to various depths. It is clear that the reverse bias leakage current scales with the device perimeter instead of the device area and that the leakage gets worse the deeper the mesa is etched. These are both clear indications that the reverse bias leakage is dominated by surface effects. Drastically worse behavior is seen once the etch reaches into the n-type substrate indicating that n-type surfaces may be more defective than p-type. Devices that were not etched deep enough to reach into the intrinsic layer (not shown) showed the worst characteristics of all which is to be expected since, in this case, the devices are shorted together by the conductive p-type GaSb layer and the entire top of the sample behaves as one large device. Thus for optimum performance, it was determined that the mesas should be etched down to the middle of the intrinsic region. This minimizes the effects of surface leakage while still isolating the devices from one another.

In order to determine the breakdown mechanism, low temperature I-V was performed. Figure 7.8 shows I-V curves for a p-i-n mesa diode etched to an optimum depth, with $Al_{0.05}Ga_{0.95}Sb$ as the intrinsic region. At low voltages the reverse bias current is dominated by tunneling, thermionic emission, or surface leakage, all of which should increase with temperature. The only leakage mechanism that becomes stronger at lower temperatures is avalanche breakdown. This is because the mean free path of electrons and holes increases as the temperature is lowered due to decreased phonon scattering. From the discussion in Chapter 6 it is clear that longer mean free

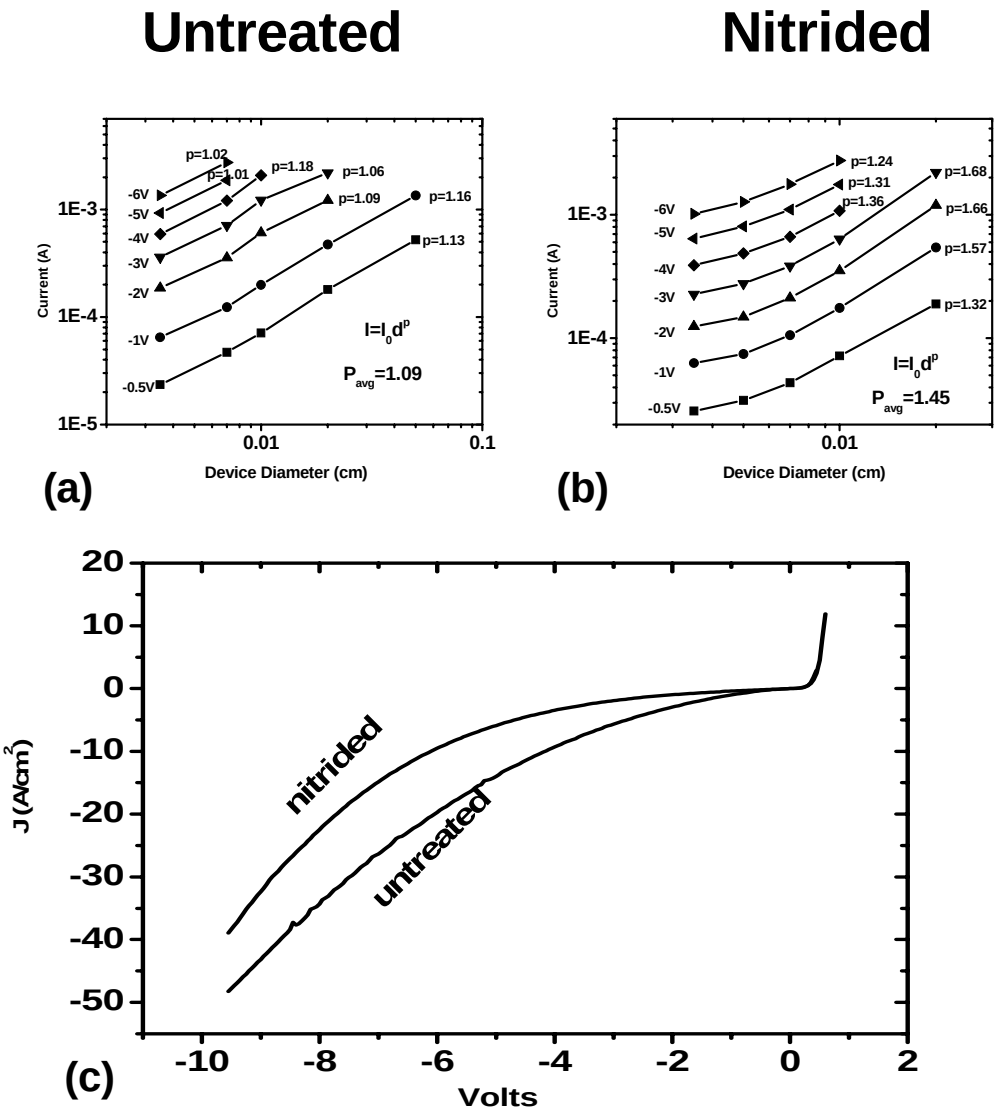


Figure 7.6: **Figures (a) and (b) show plots of reverse bias leakage current versus device perimeter, with reverse bias voltage as a parameter, plotted on a log-log scale. Figure (c) shows the average current density of several samples for nitrogen-plasma-treated and untreated diodes. The number next to the data for each bias point is the exponent obtained by fitting each dataset to $I=I_0d^p$ where d is the device diameter.**

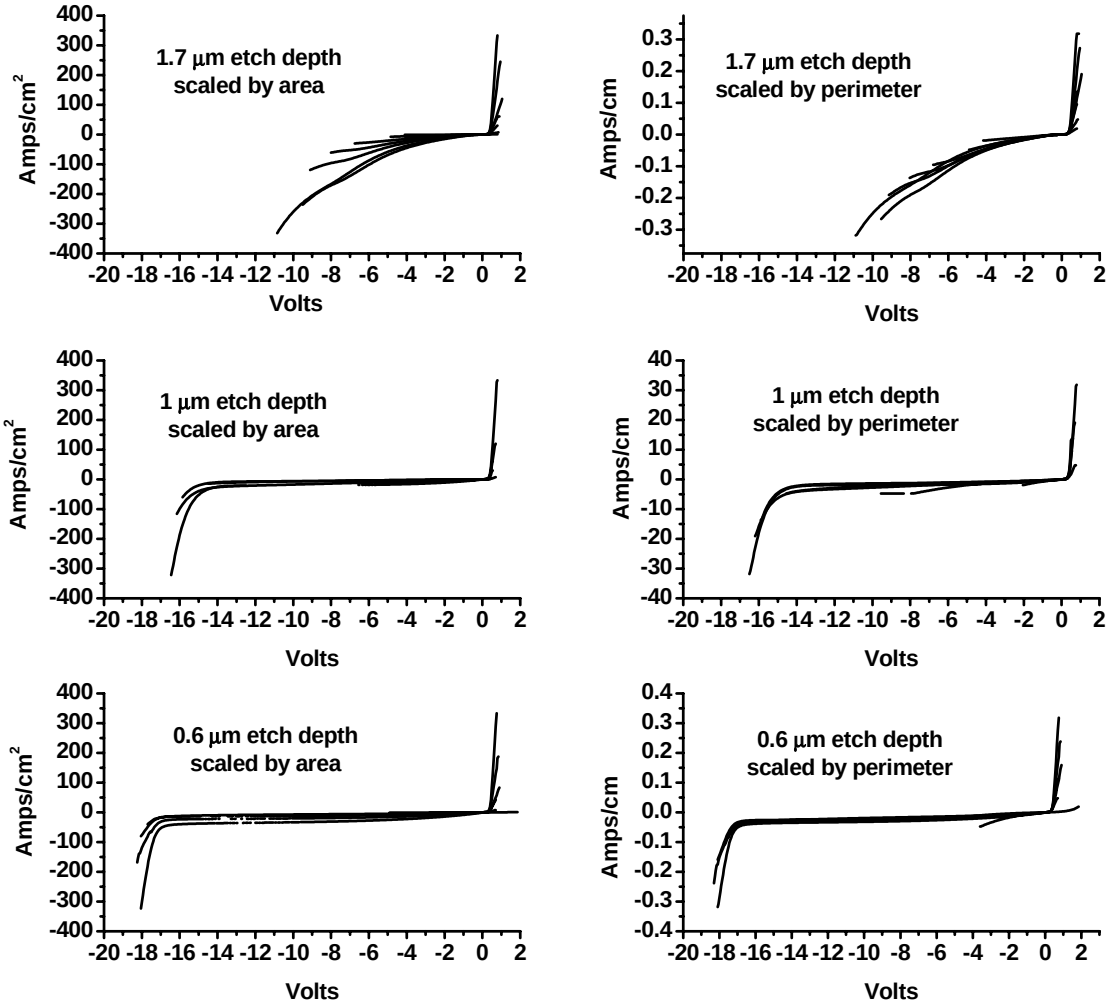


Figure 7.7: **I-V curves of p-i-n mesa diodes etched to different depths. The plots on the left show the current density for mesas of several different diameters. The plots on the right show the current scaled by perimeter instead of area. It is clear that most parts of the curves scale better with perimeter which is an indication of surface leakage. The top plots are for a 1.7 μm deep mesa which is etched all the way down to the n-type material. The middle plots are for a 1 μm mesa which is stopped in the middle of the intrinsic material. The lower plots are for a 0.6 μm mesa which is etched just barely past the p-type cap layer into the intrinsic layer.**

paths should enhance impact ionization. It is seen in the figure that the curve taken at lower temperature reaches a sharper breakdown at lower voltage. Thus it is clear that these samples are undergoing avalanche breakdown.

7.6 Photoresponse

In order to measure the photoresponse of these diodes, a fairly elaborate experimental setup was required. Figure 7.9 shows a schematic of the experiment. By positioning an optical fiber directly over a specific mesa, light can be injected only into the top p-type GaSb layer. Light from laser diodes of three different wavelengths was used for the experiment: 785 nm, 1645 nm and 1740 nm. These three wavelengths have very different penetration depths in GaSb. Figure 7.10 shows the absorption spectrum for GaSb at various temperatures [23, 24]. The shortest wavelength laser (785 nm) absorbs within the first 180 nm of the GaSb. Hence, at room temperature, this light will be absorbed completely in the p-type material. The 1645 nm light absorbs throughout the p-type, intrinsic and into the n-type. It will be much less absorbed in the intrinsic region because the bandgap of the intrinsic region is larger than that of GaSb because of the aluminum content. Finally, the 1740 nm light penetrates deep into the substrate such that the majority of the light is absorbed in the n-type region. These three wavelengths should excite very different photocarriers in the GaSb, with the 785 nm light injecting only electrons into the multiplication region, the 1645 nm light injecting some holes and some electrons, and the 1740 nm light injecting mostly holes. Although this technique does not provide a method for pure hole injection, it should be possible to determine whether there is a pronounced difference in the impact ionization coefficients especially if $\beta > \alpha$ [25].

The photoresponse was measured as follows. The laser diodes can be modulated electrically simply by applying an AC voltage signal between the laser cathode and anode, with the DC bias set such that the laser is past threshold and emitting light. This modulation signal is also fed to the reference input of a lock-in amplifier. The laser light is routed to the device by means of an optical fiber. All of the laser diodes

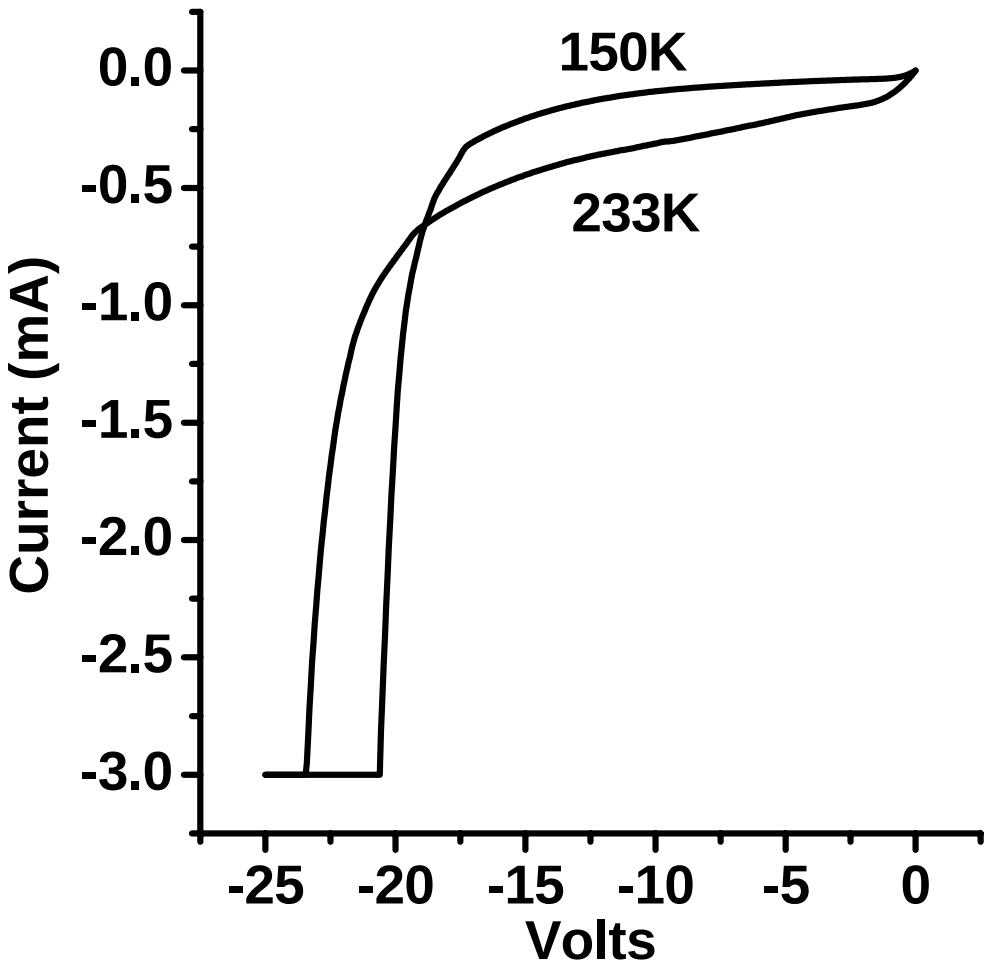


Figure 7.8: **Temperature-dependent I-V curves of p-i-n mesa diodes. The curve taken at 150 K undergoes avalanche breakdown at a lower voltage due to the longer mean free path of carriers at low temperatures. The pre-breakdown current is larger for higher temperature since all other leakage mechanisms are enhanced by increasing the temperature.**

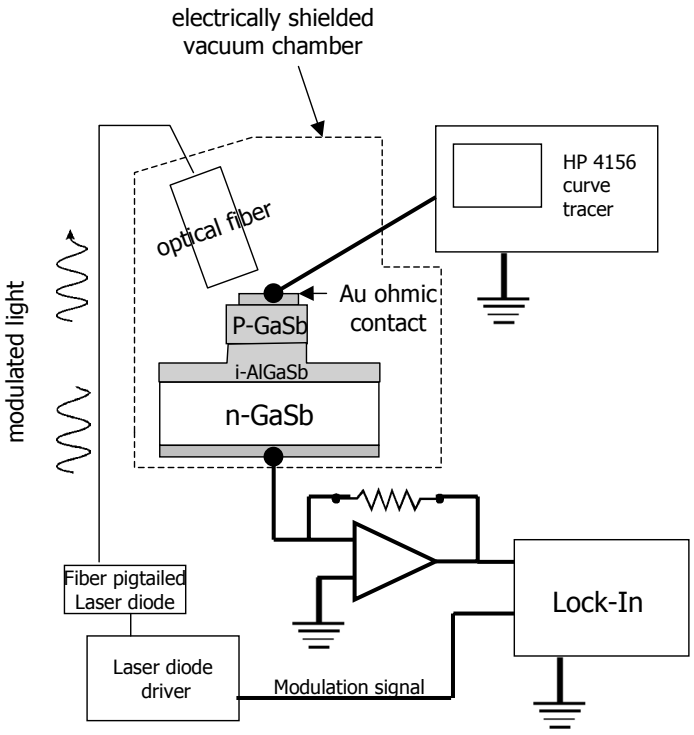


Figure 7.9: **Schematic of setup for photoresponse measurements.**

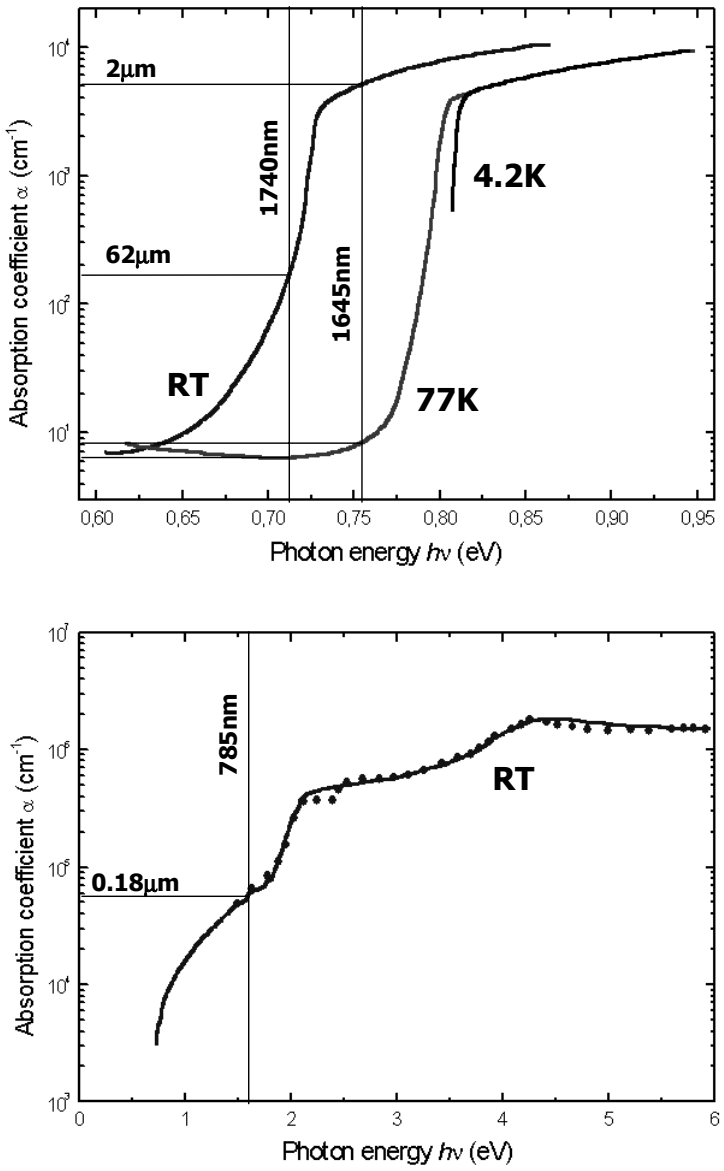


Figure 7.10: **Absorption spectrum of GaSb. The top plot is of the near-band-edge absorption in GaSb for room temperature, liquid nitrogen temperature and liquid helium temperature. The lines overlayed on the absorption data are the laser lines used in this experiment. The bottom plot is the absorption spectrum beyond the bandgap at room temperature. The 785 nm laser line is superimposed on this plot. Data is taken from references [23] and [24].**

used in this experiment came with an optical fiber pigtail installed by the manufacturer. Thus there is very little loss or decoherence of the optical signal between the laser and the device under test (DUT). The end of the optical fiber is mounted on a three-axis positioner so that it can be centered above the DUT and can be lowered to focus all the light in a specific area. As is demonstrated in Fig. 7.4, the actual inner diameter of the fiber is so small that absorption of stray light in other parts of the semiconductor is very unlikely. The DUT is wired as shown in Fig. 7.10, with a DC bias applied to the anode and the cathode wired to a transimpedance amplifier (TIA). The TIA acts as a virtual ground so that no parasitic load will ever appear at the terminals of the DUT. The current going into the TIA is converted to a voltage which then is sent to the lock-in amplifier. Since the optical signal is modulated, the component of the current from the DUT at the modulation frequency should be related directly to photocurrent. The DC current is measured at the same time, but DC leakage should not contribute to the current at the modulation frequency except as added shot noise. Low temperature measurements were performed by probing the DUT while bonded by silver paint to a cold finger. The cold finger is a miniature Joule-Thompson refrigerator supplied by MMR Technologies.

In order to produce the gain data, the effects of a non-unity quantum efficiency in the material need to be subtracted from the data. As the bias is increased, the depletion width expands until punchthrough is achieved (the entire intrinsic region is depleted). Up to this point, the photoresponse may seem artificially lower since some of the photoelectrons must diffuse across the non-depleted part of the intrinsic region instead of being swept across by the field. For most of these devices, punch-through occured at about negative one or two volts. After punch-through, the quantum efficiency should be constant except for two effects. First, the depletion width begins to extend slightly into the doped regions, decreasing the amount of distance photocarriers need to diffuse. The second effect is the Franz-Keldysh effect. Sub-bandgap light actually can be absorbed in a semiconductor if there are defect or exciton levels near the band edges. With zero applied electric field, a photon will just be re-emitted when the electron decays out of the trap state. However, in a large electric field,

there is a finite probability that, once an electron has been excited into the near-band-edge trap state, it will tunnel into the conduction band and thus register as a photogenerated carrier. This phenomenon in semiconductors was first described by Callaway [26]. The dependence of the absorption coefficient as a function of electric field can be approximated by [25]

$$\alpha(\lambda, E) = \frac{10^4 f (2\mu)^{4/3} E^{1/3}}{n m^{4/3}} \int_{\beta}^{\infty} |Ai(z)|^2 dz, \tag{7.1}$$

where β=1.1$\times 10^5$(E_g-$\hbar\omega$)(2μ/m)$^{1/3}$$E^{-2/3}$, f=1+m/$m_v$, n is the index of refraction, μ is the reduced mass, m is the free electron mass, m_v is the valence band effective mass, E_g is the bandgap, ω is the frequency of the light, E is the absolute value of the electric field and Ai(z) is the Airy function. Figure 7.11 shows the absorption coefficient as a function of applied voltage for the devices used in this experiment as derived from Eq. 7.1. One can see that in the high voltage regime, the absorption coefficient increases roughly linearly with applied bias. Although the curves in Fig. 7.11 were calculated for GaSb, this effect will be most prominent in the $Al_{0.05}Ga_{0.95}Sb$ multiplication region in the devices in question because most of the voltage drops across this layer. Nevertheless, the general trend of a linear increase in absorption coefficient is expected. Thus, in order to normalize the photoresponse data to a constant quantum efficiency, a linear fit was performed on the data for voltages well below the onset of avalanche breakdown. The data then was divided by the resulting equation of a line over the entire voltage range, such that the gain remains at a value of one until avalanche gain takes over. For the 785 nm laser, the photon energy is always much greater than the bandgap so the Franz-Keldysh effect is unimportant. In order to normalize the data from the 785 laser, the data was divided by a constant value, taken to be the photoresponse signal just after punchthrough.

Figure 7.12 shows the photoresponse and current-voltage characteristics for devices illuminated with all three laser wavelengths at room temperature, normalized as described above. There is a clear difference between the three wavelengths. The longer wavelength light, which injects more holes, shows a much greater gain, qual-

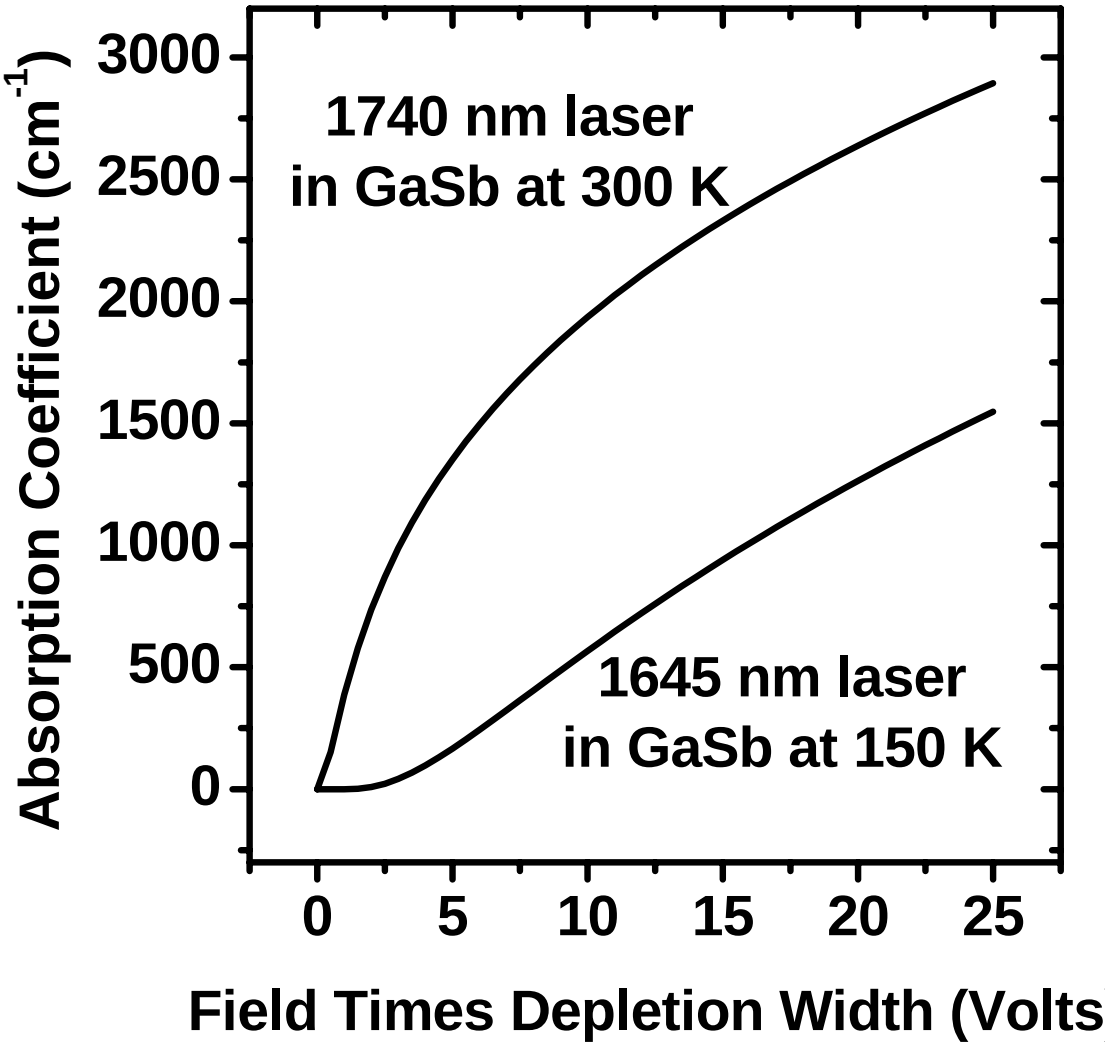


Figure 7.11: **Absorption coefficient due to the Franz-Keldysh effect as a function of applied bias in GaSb for two of the important situations in the experiments of this chapter. The 1740 nm laser has a sub-bandgap photon energy at room temperature whereas the 1645 nm laser has a sub-bandgap photon energy at 150 K. In the high voltage regime, the absorption coefficient increases roughly linearly with voltage.**

itatively indicating that $\beta > \alpha$. It should be noted that APD's fabricated without counterdoping showed no avalanche gain at all in similar measurements (not shown), again verifying the importance of a low effective doping in the intrinsic region.

At room temperature, the gains are fairly low and there is a lot of noise in the long wavelength measurements due to blackbody radiation from all the surfaces near the sample. Thus low temperature photoresponse also was measured. Below about 100 K neither of the long wavelength lasers have significant absorption so measurements were performed at 150 K. Using the formula from Wu and Chen [27] for the temperature dependence of the bandgap of GaSb, a bandgap of 0.796 eV was calculated for a temperature of 150 K. This value was used to produce the absorption data in Fig. 7.11. Figure 7.13 shows the photoresponse for the 785 nm and 1645 nm light at 150 K. There is now a drastic difference between the photoresponse of the two different wavelengths. It is even more clear here that hole injection is preferred. If it is assumed naively that all the signal from the 785 nm laser comes from injected electrons and that all the signal from the 1645 nm laser comes from holes, expressions can be derived for α and β. For a p-i-n diode where a uniform electric field is assumed throughout the intrinsic multiplication region [25]

$$\beta(E) = \frac{1}{W}\frac{M_p(E)-1}{M_p(E)-M_n(E)}\ln\left(\frac{M_p(E)}{M_n(E)}\right) \tag{7.2}$$

$$\alpha(E) = \frac{1}{W}\frac{M_n(E)-1}{M_n(E)-M_p(E)}\ln\left(\frac{M_n(E)}{M_p(E)}\right), \tag{7.3}$$

where W is the depletion width (roughly equal to the width of the metallurgical intrinsic region) and $M_{n,p}$(E) are the electron and hole gain as a function of electric field. The results of applying these equations to the gain data at 150 K are shown in the inset of Fig. 7.13. The logs of the impact ionization coefficients are fairly linear as a function of 1/E, which is close to the behavior one would expect from the simple theoretical forms for α and β described in Chapter 6. The highest value of k=β/α from this data is about 100 for low fields. The trends for α and β quickly converge as the field is increased, reaching a value of about k=5 for the highest fields. This is

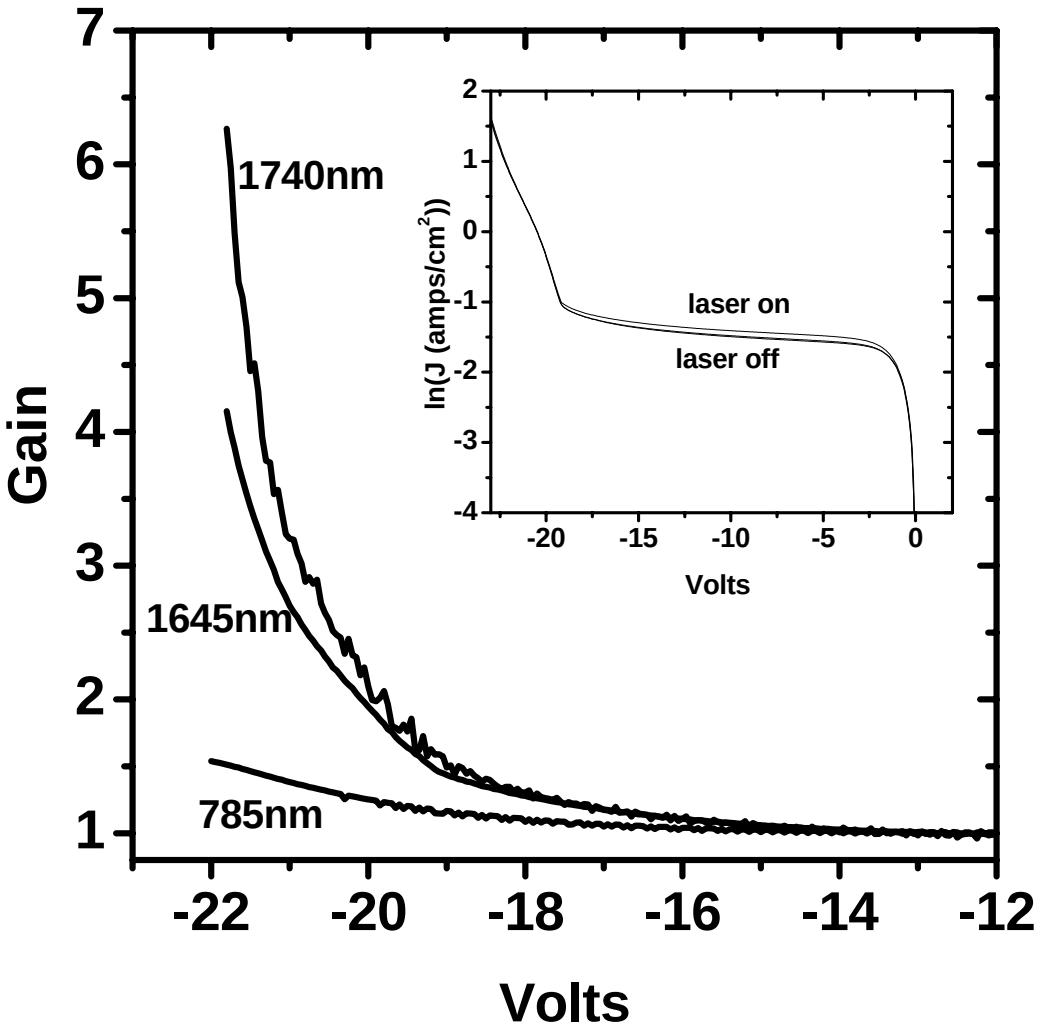


Figure 7.12: **Room temperature photoresponse and I-V for an $Al_{0.05}Ga_{0.95}Sb$ APD illuminated by three different laser wavelengths. The short wavelength tends to preferentially inject electrons into the multiplication region whereas the longer wavelengths inject both types of carriers.**

in excellent agreement with the predictions of Grein and Ehrenreich discussed above; that the resonance effect for the hole ionization coefficient is muted for high electric fields. It should be reiterated that the calculation is approximate since the gain curve from the 1645 nm laser is actually contaminated by some electron injection.

7.7 Conclusion

P-i-n APD's have been fabricated via MBE, using $Al_{0.05}Ga_{0.95}Sb$ as the intrinsic multiplication layer. It is found that counterdoping the intrinsic region with tellurium greatly enhances the electrical characteristics of the diodes. Further, it is found that etching mesas down only to the middle of the intrinsic region provides by far the lowest pre-avalanche reverse bias leakage current. The best APD's fabricated with this mesa depth and with a counterdoped intrinsic region showed a strong preference for hole ionization. At 150 K, the value of k was estimated to be as high as 100 for low fields and remains as high as 5 for high fields. Thus a significant enhancement of the hole impact ionization coefficient in $Al_{0.05}Ga_{0.95}Sb$ is seen using MBE grown material.

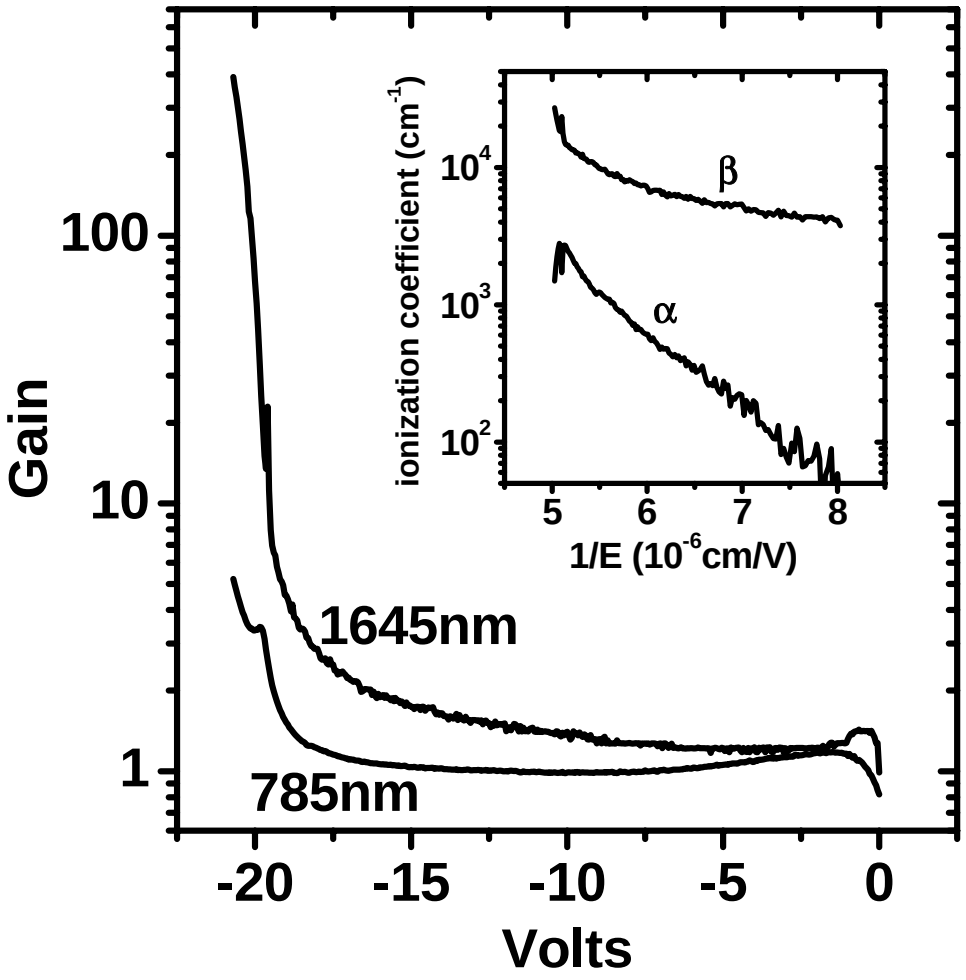


Figure 7.13: **Photoresponse and impact ionization coefficients at 150 K for $Al_{0.05}Ga_{0.95}Sb$ APD's illuminated by two different laser wavelengths. The impact ionization coefficients are calculated from Equations 7.2 and 7.3.**

Bibliography

[1] O. Hildebrand, W. Kuebart, J. Lutz, and M. Pilkuhn, IEEE Trans. Electron Devices **27**, 2188 (1980).

[2] O. Hildebrand, W. Kuebart, K. W. Benz, and M. H. Pilkuhn, IEEE J. Quantum Electron. **17**, 284 (1981).

[3] H. Kuwatsuka, T. Mikawa, S. Miura, N. Yasuoka, M. Ito, T. Tanahashi, and O. Wada, J. Appl. Phys. **70**, 2169 (1991).

[4] A. Rmou, H. Luquet, L. Gouskov, M. Perotin, and P. Coudray, Jpn. J. Appl. Phys. **33**, 4657 (1994).

[5] Y. Jiang, M. C. Teich, and W. I. Wang, J. Appl. Phys. **67**, 2488 (1990).

[6] K. Brennan, K. Hess, and Y. C. Chang, J. Appl. Phys. **57**, 1971 (1985).

[7] C. H. Grein and H. Ehrenreich, Appl. Phys. Lett. **77**, 3048 (2000).

[8] P. S. Dutta, H. L. Bhat, and V. Kumar, J. Appl. Phys. **81**, 5821 (1997).

[9] K. F. Longenbach and W. I. Wang, Appl. Phys. Lett. **59**, 1117 (1991).

[10] P. S. Dutta, K. Rao, K. S. Sangunni, H. L. Bhat, and V. Kumar, Appl. Phys. Lett. **65**, 1412 (1994).

[11] A. Baraldi, F. Colonna, G. Covucci, C. Ghezzi, R. Magnanini, A. Parisini, L. Tarricone, A. Bosacchi, and S. Franchi, Mater. Sci. Engineering B **44**, 70 (1997).

[12] L. Wende, R. Sielemann, and G. Weyer, Hyperfine Interactions **121**, 365 (1999).

[13] M. Perotin, P. Coudray, A. Etcheberry, L. Gouskov, C. Debiemmechouvy, and H. Luquet, Mater. Sci. Engineering B **28**, 374 (1994).

[14] R. M. Lin, S. F. Tang, S. C. Lee, and C. H. Kuan, J. Crystal Growth **227**, 167 (2001).

[15] P. S. Dutta, K. Rao, and H. L. Bhat, J. Appl. Phys. **77**, 4825 (1995).

[16] P. S. Dutta, K. S. Sangunni, H. L. Bhat, and V. Kumar, Phys. Rev. B **51**, 2153 (1995).

[17] P. S. Dutta, K. S. Sangunni, H. L. Bhat, and V. Kumar, Appl. Phys. Lett. **65**, 1695 (1994).

[18] M. Kodama, Solid-State Electron. **37**, 1567 (1994).

[19] B. H. Wu, G. Q. Xia, Z. H. Li, and J. Zhou, Appl. Phys. Lett. **80**, 1303 (2002).

[20] A. Kikuchi, H. Hoshi, and K. Kishino, Jpn. J. Appl. Phys. **33**, 688 (1994).

[21] S. Anantathanasarn, S. Ootomo, T. Hashizume, and H. Hasegawa, Appl. Surf. Sci. **159**, 456 (2000).

[22] M. Losurdo, P. Capezzuto, G. Bruno, G. Perna, and V. Capozzi, Appl. Phys. Lett. **81**, 16 (2002).

[23] W. M. Becker, H. Y. Fan, and A. K. Ramdas, J. Appl. Phys. **32**, 2094 (1961).

[24] S. Adachi, J. Appl. Phys. **66**, 6030 (1989).

[25] G. E. Stillman and C. M. Wolfe, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beers (Academic Press, New York, New York, 1977), Vol. 12, pp. 291–392.

[26] J. Callaway, Phys. Rev. **130**, 549 (1963).

[27] M. C. Wu and C. C. Chen, J. Appl. Phys. **72**, 4275 (1992).

Chapter 8 6.1 Å Superlattice APD's

8.1 Overview

In Chapter 7, $Al_xGa_{1-x}Sb$ layers were fabricated to serve as multiplication layers for APD's. In this chapter, APD's are designed that use various 6.1 Å superlattices to achieve optimal electrical and photoresponse behavior.

8.2 Introduction

One of the main advantages of the 6.1 Å semiconductor system is the rich array of superlattices that all can be grown on the same substrate. The use of superlattices to enhance the properties of APD's has been extensively studied in InGaAs/GaAs systems [1] and, more recently, in the 6.1 Å system [2, 3]. As mentioned in Chapter 6, El-Rub et al. [4] have calculated several 6.1 Å superlattices that should show a resonance in the split-off band energy with the bandgap energy, and hence should show a strong preference for hole multiplication. In this chapter, the current attempts in our group to build a practical APD that employs the 35Å InAs/20Å $In_{0.4}Ga_{0.6}Sb$/15Å AlSb superlattice as a multiplication layer will be outlined. The first section of this chapter will describe this superlattice and our attempts to grow it and fabricate an APD using it.

In trying to create a practical device using this superlattice, it will be shown that the entire structure around the superlattice multiplication layer will need to be carefully designed in order to achieve satisfactory electrical behavior from the diode. In the middle section of this chapter, the current-voltage (I-V) characteristics of p-i-n diodes made with this superlattice as the intrinsic region will be shown. The reverse bias behavior of these diodes is found to be greatly dependent on the band offset between the substrate and the superlattice.

In the last section of this chapter, methods for improving the diodes by inserting layers of different material between the substrate and the superlattice will be examined. It will be shown that supplementary superlattice layers which "clad" the multiplication superlattice should provide the optimum characteristics for creating diodes with good electrical behavior.

8.3 Description of Three-Stage Multiplication Superlattice

Figure 8.1 shows a schematic diagram of the three-stage superlattice in question. The band offsets between the different materials are shown as described by Yu et al. [5] The effective bandgap of the superlattice is shown as a pair of dotted lines crossing through the rest of the band diagram. The effective bandgap of the superlattice and its band offset relative to the bulk materials were calculated with a **k•p** code developed in our group [6].

Samples were grown via the MBE techniques described in Chapter 7. Several samples were grown with between 50 and 100 periods of the superlattice sandwiched in between thick GaSb layers of opposite doping in order to create a p-sl-n diode where "sl" is the superlattice layer. It is important to grow layers thick enough such that, on average, at least one impact ionization event would take place as a carrier travels through the region, but not so thick that the layers of the superlattice crystal would relax. The computations of El-Rub et al. assume that the superlattice layer is completely strained to a GaSb substrate.

Figure 8.2 shows the structure of the four samples used in these experiments. The undoped spacer regions between the p or n GaSb and the superlattice were originally added simply to keep the dopants from diffusing into the superlattice region, but as will be seen below, they have a profound effect on the electrical characteristics of the diodes.

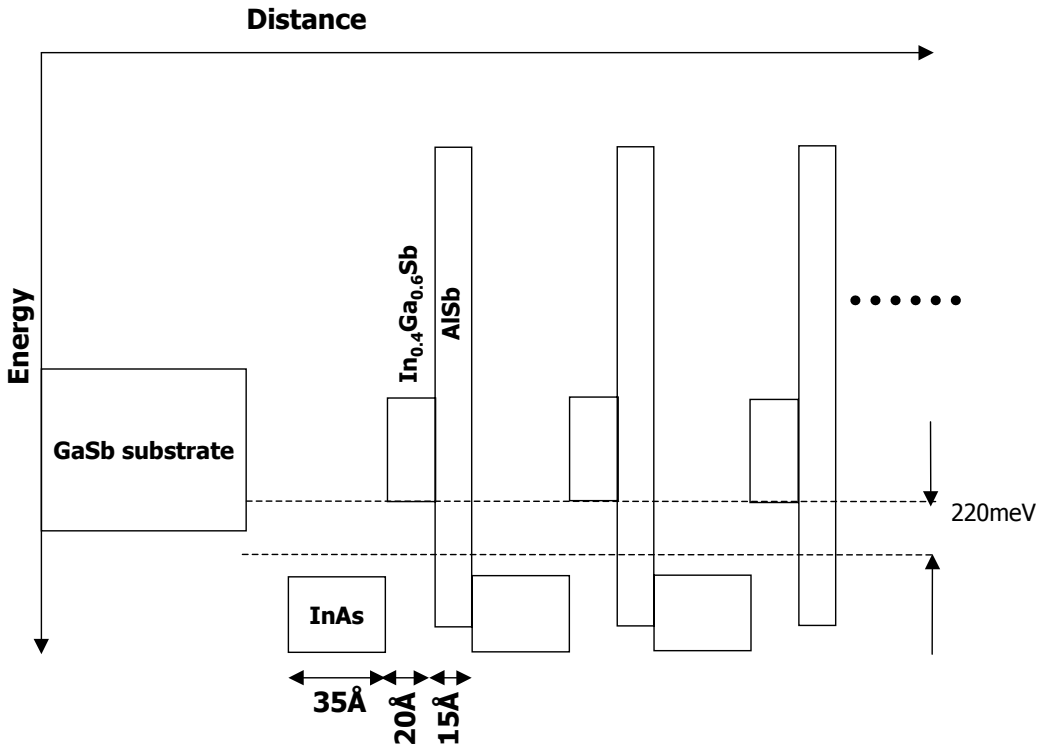


Figure 8.1: **Schematic of the three-stage multiplication superlattice. The dashed lines represent the effective bandgap of the superlattice region. Note that the effective superlattice valence band sits below the valence band of bulk GaSb.**

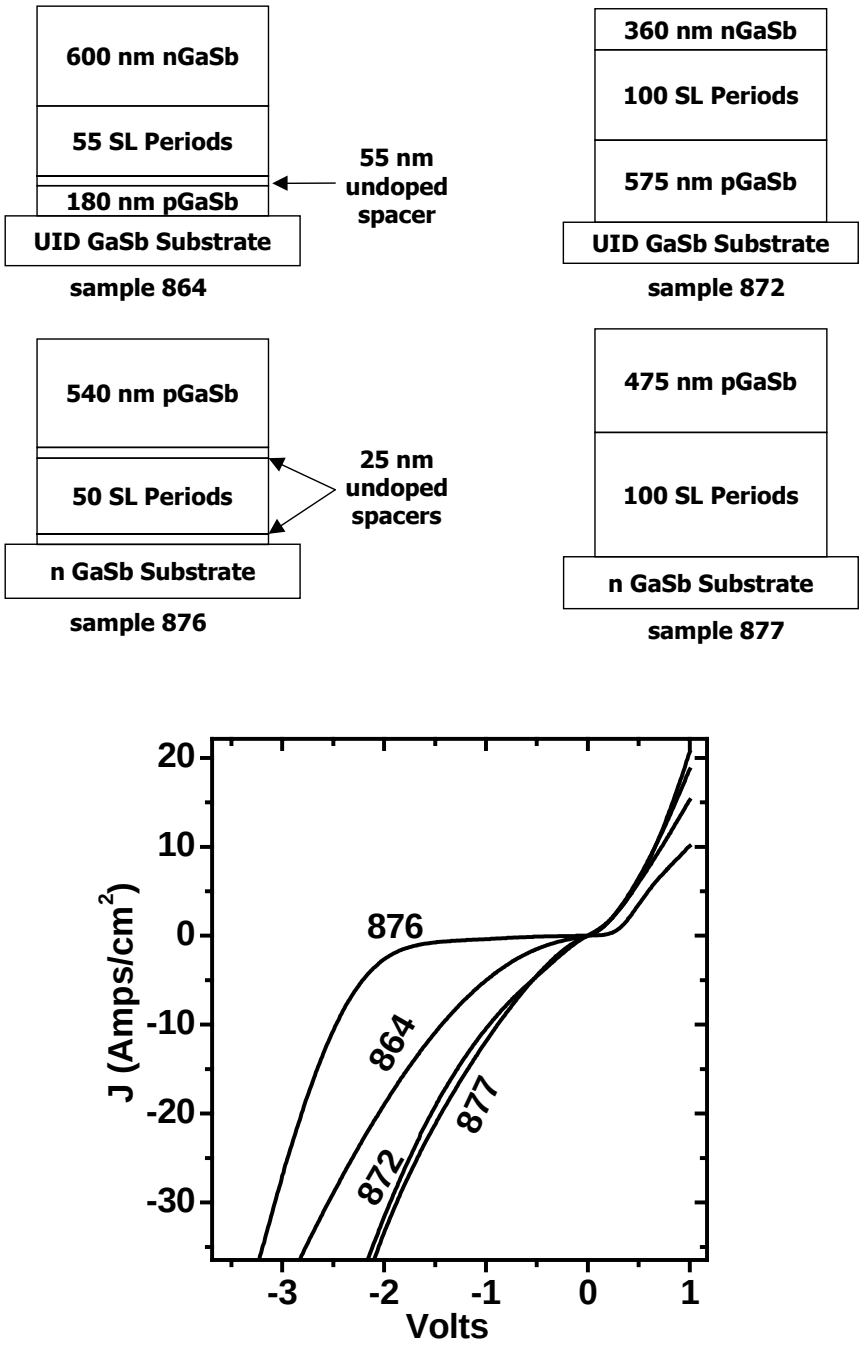


Figure 8.2: **Diagram of four superlattice APD samples along with their typical I-V characteristics. The samples grown with an undoped spacer region between the GaSb and the superlattice clearly have superior reverse bias leakage characteristics.**

8.4 Electrical Characteristics of Superlattice APD's

The MBE-grown structures were fabricated into mesa diodes using the process described in Chapter 7. It is seen in Fig. 8.2, that the four different structures had very different electrical characteristics. The superlattice has such a narrow effective bandgap (220meV) that one would expect breakdown to occur either via tunneling or avalanche at a very low reverse bias. However, as observed in the figure, there is a large discrepancy between the reverse bias characteristics of the various samples. Upon closer inspection, there is a correlation between the existence of the thin, undoped spacer regions and better reverse bias characteristics. In order to assess the importance of these regions, simulations were performed using the commercial device modeling software ATLAS. Four structures similar to the ones grown and described in Fig. 8.2 were simulated. The superlattice region was simulated by a bulk material with a 220 meV bandgap and a valence band offset of -150 meV relative to GaSb. In the simulation, the only reverse bias leakage mechanisms that were allowed to contribute to the current were tunneling and thermionic emission. Figure 8.3(a) shows the reverse bias leakage current of the four devices. The current is almost exclusively due to tunneling. The simulation accurately predicts that the sample with two undoped spacer regions will have the lowest leakage current. Figure 8.3(b) illustrates the origin of this phenomenon. When a small spacer region is placed between the bulk GaSb and the narrow gap superlattice multiplication region, a small barrier is created that inhibits tunneling of electrons from the valence band of the p-type material into the conduction band of the superlattice material. The barrier between the superlattice and the n-type GaSb also helps somewhat by reducing the local field in the superlattice near the interface. As can be seen from the figure, this reduces the leakage current somewhat but not as much as the spacer region between the superlattice and the p-type.

In order to see if the spacer region is satisfactory in inhibiting reverse bias leakage, the breakdown mechanism was probed by temperature-dependent I-V. Figure 8.4 shows the I-V curves of the best diode (sample 876) at various temperatures. It is

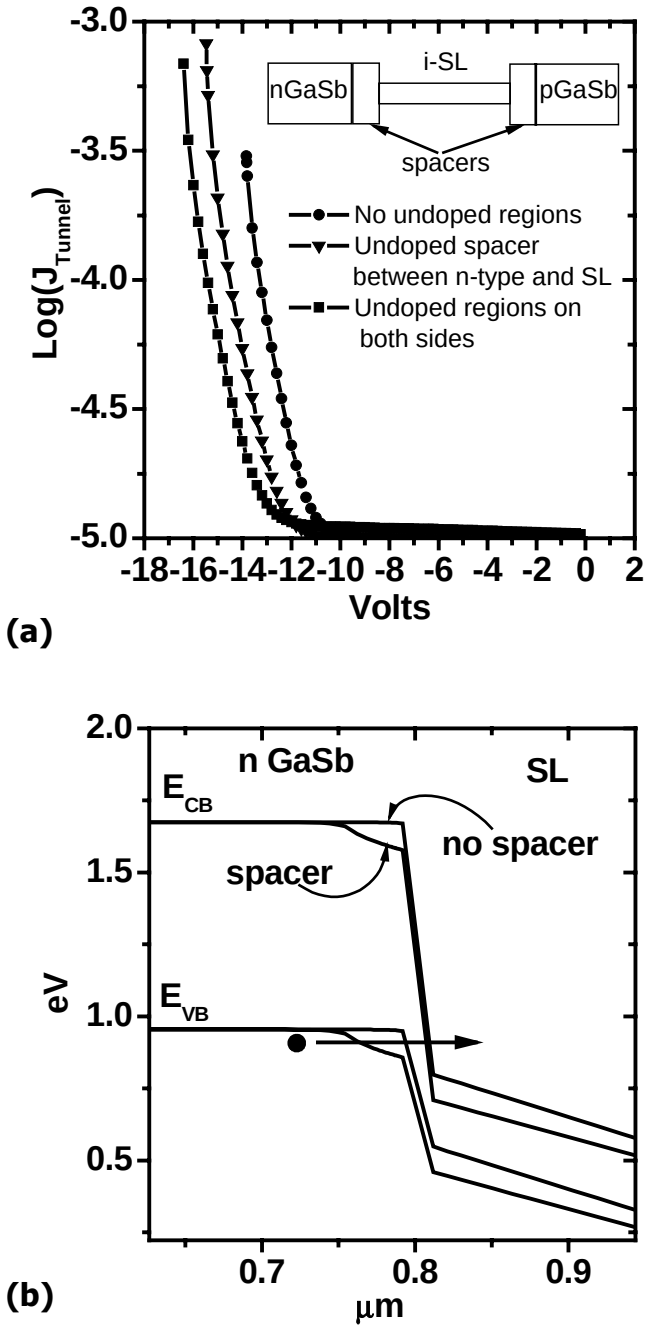


Figure 8.3: **Computer simulations of p-sl-n diodes with or without undoped spacer layers. Figure (a) shows the reverse bias current due to tunneling for the different structures. The samples with an undoped spacer layer clearly show the lowest current. Figure (b) shows the band diagram for devices with and without an undoped GaSb spacer region between the p-type GaSb and the superlattice. The addition of the undoped spacer region creates a slight extra barrier to reduce tunneling of electrons from the GaSb valence band into the superlattice conduction band.**

clear that, unlike the diodes in Chapter 7, the breakdown mechanism is tunneling, not avalanching. This is evident because the leakage current *increases* with increasing temperature. The leakage is also most likely due to tunneling instead of other thermal processes because the current is not increasing too rapidly with increasing temperature. If an Arrhenius temperature dependence is assumed, an activation energy of only 14 meV is obtained from the data in Fig. 8.4. The barriers to thermal activation of carriers in the band structure of this device are all substantially greater than 14 meV so tunneling is the most likely leakage mechanism. This again confirms the idea proposed above; that the leakage is dominated by tunneling from electrons from the p-type valence band into the intrinsic superlattice layer.

8.5 New Superlattice Device Designs

It is clear from the arguments of the preceeding sections that in order to get these superlattice APD's to function properly electrically, a different scheme will need to be created to minimize the interband tunneling. One could theoretically extend the undoped spacer regions until they were wide enough to inhibit tunneling to the point where avalanche breakdown could be observed. However, extending the undoped regions essentially means extending the high-field multiplication region into bulk GaSb material. If too large a percentage of the multiplication region is in the GaSb, the advantageous multiplication properties of the superlattice material will be minimized.

A better solution would be to engineer the p and n materials to have optimum band offsets with respect to the multiplication superlattice. In order to do this, a wide array of binary superlattices made from 6.1 Å materials were examined: AlSb/GaSb, AlSb/InAs and GaSb/InAs superlattices with varying thicknesses of each component. The bandgaps of these superlattices and their band offsets relative to a universal zero energy (taken to be the valence band of InAs) were calculated using the same **k•p** code mentioned in Section 8.3. Figure 8.5 shows a schematic representation of where the effective conduction and valence bands of the binary superlattices line up relative to the bandgap of the multiplication superlattice. To make an optimal p-type junction

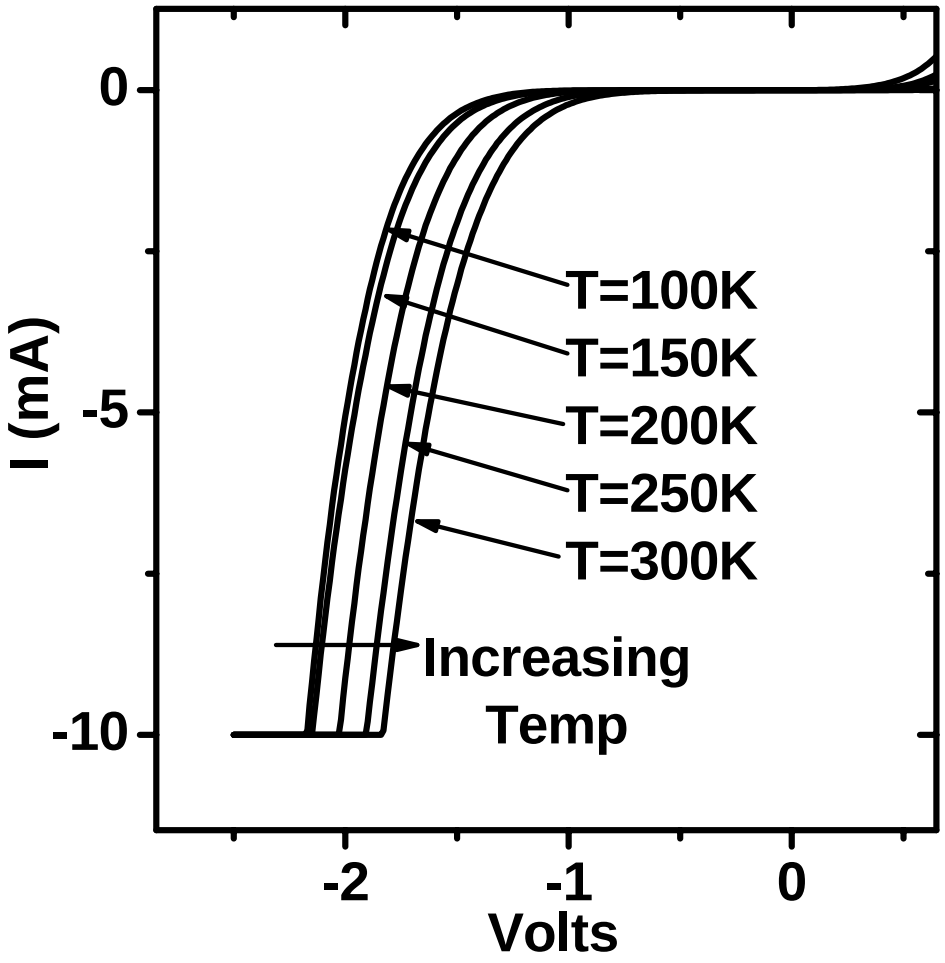


Figure 8.4: **Temperature-dependent reverse bias I-V curves of a p-sl-n APD. The leakage current clearly increases with increasing temperature indicating that the leakage mechanism is not avalanche breakdown.**

with the multiplication superlattice, it is desirable to have a positive valence band offset (valence band lower than that of the superlattice) which is not the case for bulk GaSb as illustrated in Fig. 8.1. However, the band offset should not be too large, otherwise photogenerated holes that drift across the multiplication region will encounter a barrier that they must tunnel through or thermally activate over in order to make it into the p-type region. Also, since the multiplication regions used in this study favor hole multiplication, one should seek to reduce electron injection into the multiplication region. Thus it would be beneficial to have a barrier for electrons between the bulk p-type material and the multiplication region. To make an optimal n-type junction with the multiplication region, there again should be no barrier to hole conduction and ideally, the material should be constructed such that there is a built-in field moving holes from the n-type region where they are photogenerated into the multiplication region.

As it turns out, a binary superlattice consisting of five monolayers each of AlSb and GaSb satisfies all these criteria as both an n-type and p-type contact. The band alignments of this particular superlattice are indicated at the top of Fig. 8.5 and a complete diagram of the theoretical APD structure using this binary superlattice is shown in Fig. 8.6. As can be seen from the figure, this superlattice has a very slight, but positive, valence band offset with respect to the multiplication superlattice. This should have a similar effect to the undoped spacer regions discussed above. There is a very large barrier to electrons created in the p-type material if this structure is to be grown on a p-type substrate. Finally, in order to align the Fermi level from the intrinsic multiplication region with the n-type binary superlattice contact, the bands must bend upwards as shown in Fig. 8.6, sweeping photogenerated holes into the multiplication region.

8.6 Conclusion

An attempt was made to fabricate APD's using an optimized 3-stage superlattice as the multiplication region. However, due to the intrinsic band alignments of this

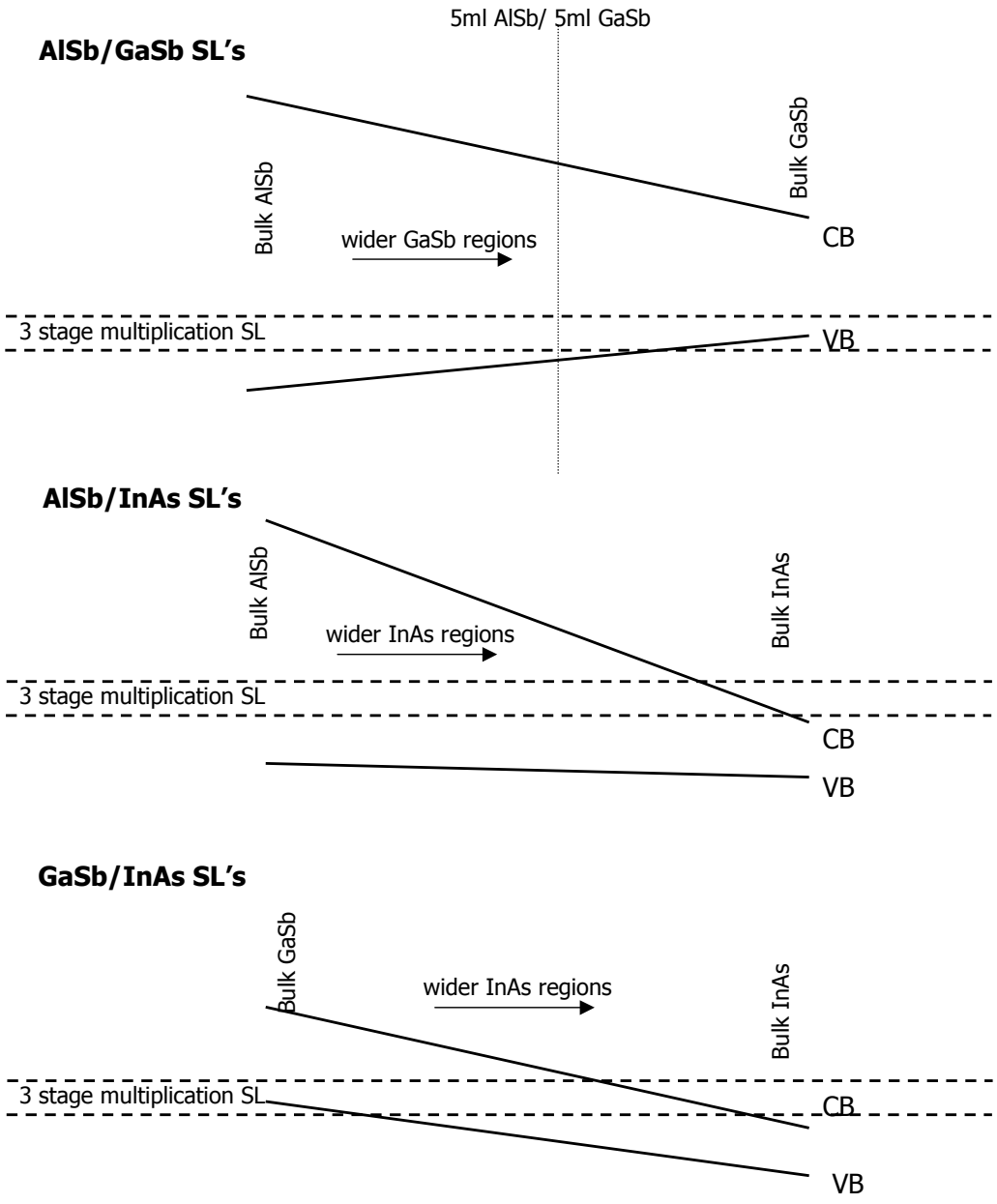


Figure 8.5: **Schematic diagram of 6.1 Å binary superlattice band structures with the energy scale drawn relative to the bands of the three stage multiplication superlattice. The right and left endpoints of the binary superlattice bands represent bulk material. As one moves to the center of the diagram, a higher percentage of the other material is represented in the binary superlattice.**

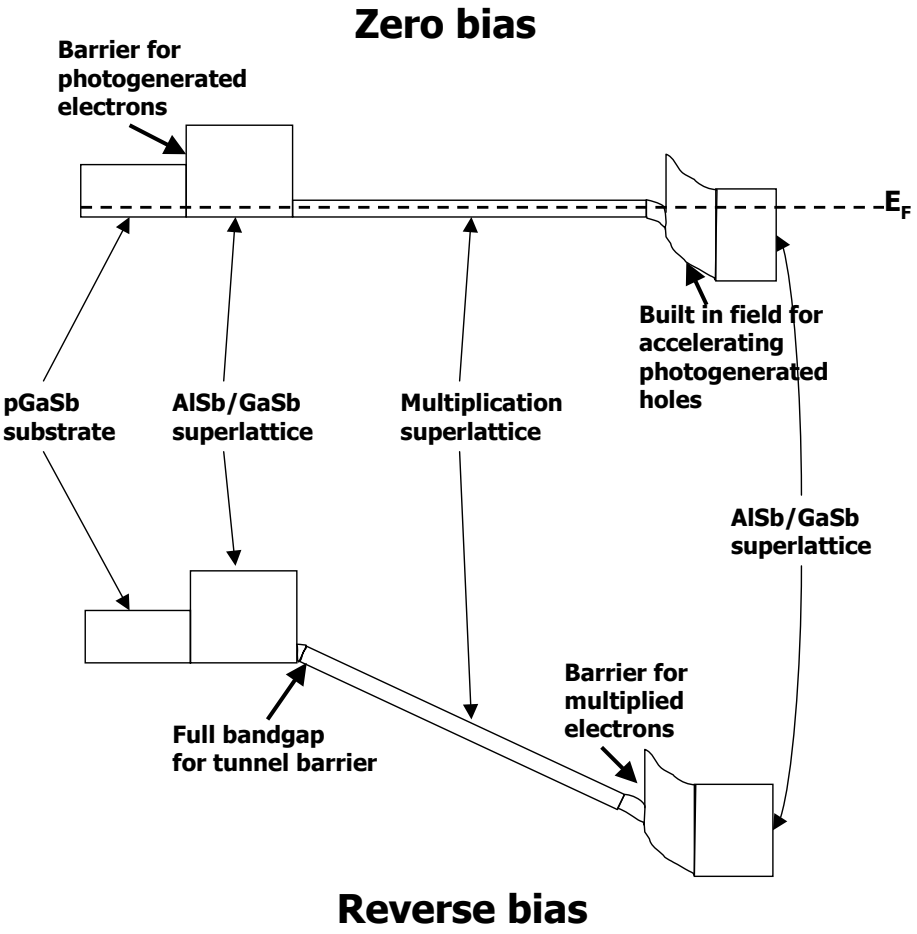


Figure 8.6: **Band diagram of a theoretical APD using superlattices for both the multiplication region and the n and p-type contacts.**

superlattice with respect to bulk GaSb, tunneling leakage overwhelmed the electrical behavior of the device such that no avalanche gain could be observed. An alternative device design is suggested. By using a new binary superlattice layer on each side of the multiplication superlattice, tunneling leakage should be reduced while the injection of photogenerated holes is enhanced. At the same time the injection of photogenerated electrons into the multiplication region is reduced providing optimum conditions for APD performance.

Bibliography

[1] K. F. Brennan and J. Haralson, Superlattices And Microstructures **28**, 77 (2000).

[2] X. C. Cheng and T. C. McGill, J. Appl. Phys. **86**, 4576 (1999).

[3] X. C. Cheng and T. C. McGill, J. Crystal Growth **208**, 183 (2000).

[4] K. A. El-Rub, C. H. Grein, M. E. Flatte, and H. Ehrenreich, J. Appl. Phys. **92**, 3771 (2002).

[5] E. T. Yu, J. O. Mccaldin, and T. C. Mcgill, Solid State Physics: Advances In Research And Applications, Vol 46 **46**, 1 (1992).

[6] X. Cartoixa, Ph.D. thesis, California Institute of Technology, 2002.

Chapter 9 Part II Conclusion

Several different approaches to using the 6.1 Å semiconductor system to create practical infrared APD's have been pursued. Two devices, utilizing 6.1 Å semiconductor heterostructure systems to create an enhancement in hole multiplication, have been fabricated. The two multiplication layer materials are $Al_{0.05}Ga_{0.95}Sb$ and a three-stage superlattice consisting of 35Å InAs/20Å $In_{0.4}Ga_{0.6}Sb$/15Å AlSb periods which should both greatly favor impact ionization of holes over electrons.

APD's with excellent properties, utilizing an $Al_{0.05}Ga_{0.95}Sb$ multiplication layer, have been fabricated by employing several new techniques to improve the electrical behavior of the diodes. It is found that counterdoping the unintentionally doped $Al_{0.05}Ga_{0.95}Sb$ multiplication layers is necessary in order to maximize the depletion width and minimize the electric field present in the multiplication region. This is critical because extremely high electric fields negate the enhancement of the hole impact ionization coefficient in these materials. A second technique was needed to minimize surface leakage which can be the dominant reverse bias leakage mechanism for p-i-n APD's in this system. It is found that by etching down to the intrinsic region only, and minimizing the exposure of n-type surfaces to the air, satisfactory reverse bias characteristics can be obtained, exhibiting strong avalanche breakdown past -20 volts. Diodes fabricated in this way show a strong preference for hole multiplication as determined by photoresponse measurements. When light was shined on the device in a manner that should maximize hole injection into the multiplication region, avalanche gains of as high as 300 at -20 volts were observed at 150 K. The k value of these devices seems to be about five for the highest electric fields measured and increases substantially as the field is lowered.

Superlattice multiplication layers were grown and used to fabricate p-sl-n APD's. However it was found that the intrinsic band offsets of this superlattice with respect to GaSb are extremely poor for preventing reverse bias tunneling leakage. The nature

of the leakage is discovered by device simulations and proven by low temperature IV measurements. In order to combat the problem of interband tunneling in the p-i-n diodes, new binary superlattices have been proposed which should not only reduce the tunneling current but should also help to selectively encourage photogenerated holes to enter the multiplication region as opposed to electrons.

Appendix A Derivation of Interface State Density Formulae

A.1 Charge-Potential Relationship at Insulator- Semiconductor Interface

At an insulator-semiconductor interface with dielectric constants, ϵ_{ox} and ϵ_s, and with a sheet charge δQ_{sheet} present at the interface, one of Maxwell's equations,

$$\nabla \bullet \mathrm{D} = \rho \tag{A.1}$$

leads to the boundary condition

$$\epsilon_{ox} E_{ox}^{\perp} - \epsilon_s E_s^{\perp} = -Q_{sheet}, \tag{A.2}$$

where the $\perp$ superscript implies the component of the electric field perpendicular to the interface. If Ψ_s is the potential at the semiconductor surface then boundary condition A.2 becomes

$$C_{ox}(V_g - \Psi_s) - \epsilon_s E_s^{\perp} = -Q_{sheet}, \tag{A.3}$$

where C_{ox}=ϵ_{ox}/t_{ox} and V_g is the applied gate bias (see Figure A.1). Now we apply Gauss' law to an area just below the interface, from the semiconductor surface to beyond the depletion layer. The electric field entering the Gaussian box at the top is just $E_s^{\perp}$. Since the other side of the box is outside of the depletion region, the field leaving the box on the bottom is just zero. Because of symmetry in the direction parallel to the interface, there are no field lines leaving the sides of the box. Thus,

Gauss' law gives

$$E_s^{\perp} = \frac{Q_s(\Phi_s)}{\epsilon_s}, \tag{A.4}$$

where the Φ_s is included in order to emphasize the point that the semiconductor space charge is a function of the band bending. Now plugging A.4 back into A.3 we have

$$C_{ox}(V_g - \Psi_s) = -Q_{sheet} - Q_s(\Psi_s). \tag{A.5}$$

A.2 Simple Equivalent Circuit for MOS Capacitor with Interface Trap Levels

We wish to extract from the circuit in Fig. 4.3(a) the values of G_t and C_t. The impedance bridge measures the total equivalent parallel conductance and capacitance across terminals 1 and 2 in Figure 4.3(b), giving two values, G_m and C_m. Hence the total admittance of the measured circuit is

$$Y_m = G_m + i\omega C_m. \tag{A.6}$$

If we convert this admittance to an impedance,

$$Z_m = \frac{1}{G_m + i\omega C_m} \tag{A.7}$$

it is then simple to subtract off the complex impedance of the capacitance C_{ox}:

$$Z_s = \frac{1}{G_m + i\omega C_m} - \frac{1}{i\omega C_{ox}} = \frac{G_m + i\omega(C_m - C_{ox})}{\omega^2 C_{ox} G_m - i\omega C_{ox} G_m}. \tag{A.8}$$

This is the impedance of everything below the oxide capacitance in Figure 4.3. In order to extract the values of G_t and C_t, we must subtract the admittance of the

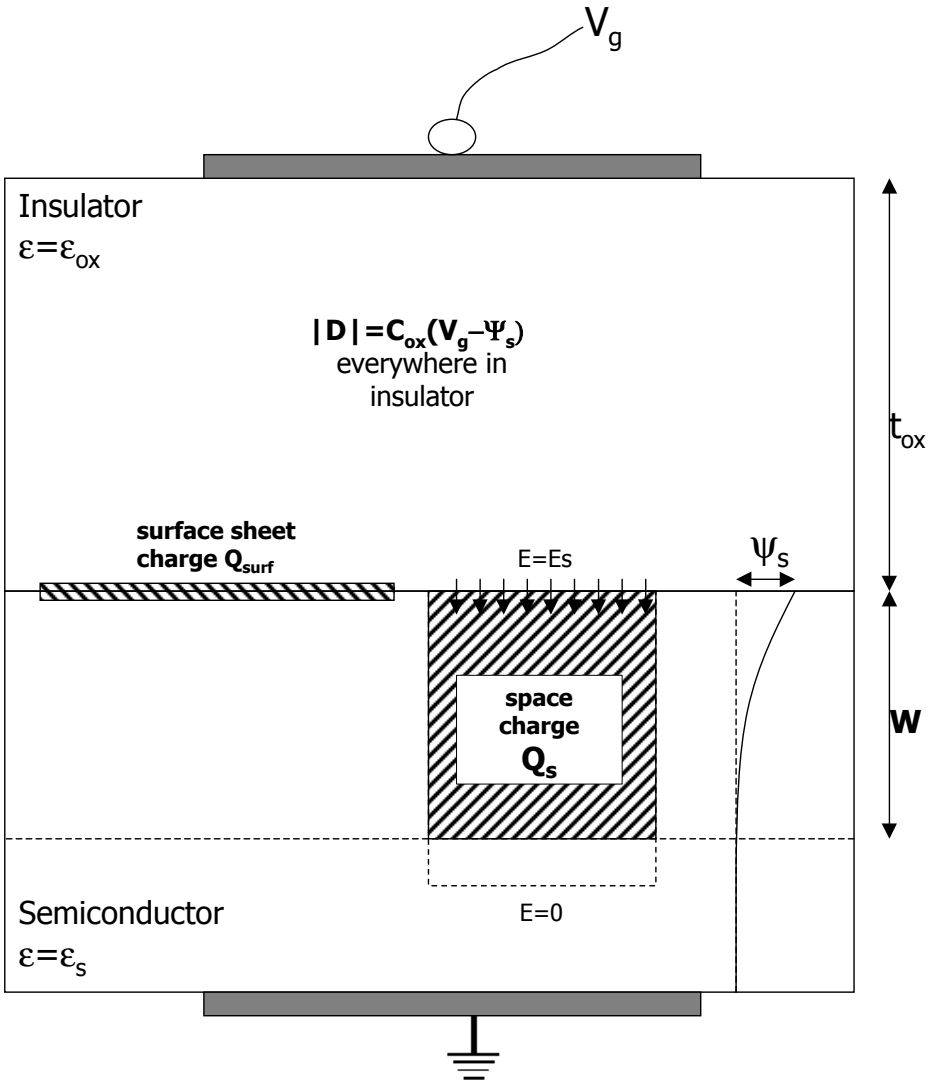


Figure A.1: **Schematic of the important charges, fields and potentials at the insulator semiconductor interface.**

depletion capacitance C_d. To make this easy, we convert Z_s back into an admittance:

$$Y_s = \frac{1}{Z_s} = \frac{\omega^2 C_{ox}^2 G_m - i\omega C_{ox}\left[G_m^2 + \omega^2 C_m(C_m - C_{ox})\right]}{G_m^2 + \omega^2(C_m - C_{ox})^2}. \tag{A.9}$$

Now the conductance due to interface traps is simply the real part of Y_s and the capacitance due to interface traps is the imaginary part of Y_s divided by ω:

$$G_t = \frac{\omega^2 C_{ox}^2 G_m}{G_m^2 + \omega^2(C_m - C_{ox})2} \tag{A.10}$$

$$C_t = \frac{-C_{ox}\left[G_m^2 + \omega^2 C_m(C_m - C_{ox})\right]}{G_m^2 + \omega^2(C_m - C_{ox})^2}. \tag{A.11}$$

A.3 Adjusting Measured Quantities for Bulk Series Resistance

Following nearly the same procedure as in the previous section, we wish to subtract the effects of an extra series resistance from the equivalent circuit in Figure 4.3(b). The impedance of the measured parallel capacitance and conductance minus the series resistance R_s is

$$Z = \frac{1}{G_m + i\omega C_m} - R_s = \frac{1 - G_m R_s - i\omega C_m R_s}{G_m + i\omega C_m}. \tag{A.12}$$

Now converting this impedance back into an admittance,

$$Y = \frac{G_m + i\omega C_m}{1 - G_m R_s - i\omega C_m R_s} = \frac{G_m - R_s(G_m^2 + \omega^2 C_m^2) + i\omega C_m}{(1 - G_m R_s)^2 + \omega^2 C_m^2 R_s^2}, \tag{A.13}$$

we can obtain the true parallel MOS capacitance and conductance

$$G_c = \frac{G_m - R_s(G_m^2 + \omega^2 C_m^2)}{(1 - G_m R_s)^2 + \omega^2 C_m^2 R_s^2} \tag{A.14}$$

$$C_c = \frac{C_m}{(1 - G_m R_s)^2 + \omega^2 C_m^2 R_s^2}. \tag{A.15}$$

To solve for R_s itself, one must convert the equivalent circuit in accumulation into the measured parallel conductance and capacitance. This amounts to converting the series combination of C_{ox} and R_s into G_m and C_m in accumulation. Equation A.7 gives the measured impedance. If we separate it into imaginary and real parts,

$$Re(Z_m) = \frac{G_m - i\omega C_m}{G_m^2 + \omega^2 C_m^2} \tag{A.16}$$

$$Im(Z_m) = \frac{-\omega C_m}{G_m^2 + \omega^2 C_m^2} \tag{A.17}$$

The real part of course is directly equal to R_s. The imaginary part is equal to -$1/(\omega C_{ox})$ giving

$$R_s = \frac{G_m}{G_m^2 + \omega^2 C_m^2} \tag{A.18}$$

$$C_{ox} = C_m \left[1 + \left(\frac{G_m}{\omega C_m}\right)\right]. \tag{A.19}$$

Appendix B Derivation of Energy and Momentum Relations for Impact Ionization Events

The impact ionization process involves three carriers. One initiating carrier, and an electron-hole pair created by the ionization event. It is obvious that total energy and momentum must be conserved in this process. If the items with subscript i's are the initiating carriers and the subscript 1, 2 and 3 are the final carriers then

$$E_i = E_1(k_1) + E_2(k_2) + E_3(k_3) \tag{B.1}$$

$$\mathbf{k}_i = \mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3. \tag{B.2}$$

Note that E_i is a scalar and only depends on the magnitude of the $\mathbf{k}_n$ which we have denoted simply as k_n. We will now proceed in one dimension but the following arguments are easily extended to three dimensions. Also, in most of the III-V semiconductors, the bands are very symmetric near k=0 so that a one-dimensional analysis is generally sufficient for the cases discussed in this thesis.

In general, for an impact ionization event to be possible, Eqs. B.1 and B.2 are all that is required. However, we seek to find the most likely events. To do this we will seek to minimize the energy of the initiating particle for a given k_i. The energy and momenta that satisfy this condition along with the conservation of energy and momentum will be the threshold values. Since E_i is a function of $k_{1,2,3}$ and picking k_i places a constraint on the values of the $k_{1,2,3}$, the problem of minimizing E_i will be precisely that of finding Lagrange multipliers. The generic formulation for Lagrange multipliers is

$$\nabla f = \lambda \nabla g, \tag{B.3}$$

where f is the function to be minimized and g is the constraint equation of form g=k where k is a constant. In our example, Eq. B.1 is the function to be minimized and Eq. B.2, limited to one dimension, is the constraint equation. Taking the gradient of Eq. B.1 we have

$$\nabla E_i = \frac{dE_1}{dk_1}\hat{k_1} + \frac{dE_2}{dk_2}\hat{k_2} + \frac{dE_3}{dk_3}\hat{k_3}. \tag{B.4}$$

Taking the gradient of the constraint equation we have

$$\nabla g = \hat{k_1} + \hat{k_2} + \hat{k_3}. \tag{B.5}$$

Plugging Eqs. B.4 and B.5 into B.3 we have

$$\frac{dE_1}{dk_1} = \lambda \tag{B.6}$$

$$\frac{dE_2}{dk_2} = \lambda \tag{B.7}$$

$$\frac{dE_3}{dk_3} = \lambda \tag{B.8}$$

More generally, this can be expressed as

$$v_1 = v_2 = v_3, \tag{B.9}$$

where the v_n are the group velocities of the final particles defined by $v_n=(dE_n/dk_n)/\hbar$. Equation B.9 turns out to hold in three dimensions as well.

Now we address the case of three parabolic bands (see Figure 6.4). Each band has a parabolic dispersion relation: $E=\hbar^2k^2/2m$. Take one conduction band having effective mass m_e and two valence bands having effective masses m_{hh} and m_{so}. The bandgap between the heavy hole band and the conduction band is E_g and the energy separation between the heavy hole band and the split-off band is Δ_{so}. We will consider two cases: an impact ionization event initiated by an electron in the conduction band and an event initiated by a hole in the split-off band. For an event initiated by a hole in the split-off band, the three final carriers will be two holes at the top of the heavy

hole band and one electron in the conduction band. If we assert the conservation of momentum and Eq. B.9, we can derive

$$k_{i,so} = k_{final,e} + 2k_{final,hh}, \tag{B.10}$$

$$\frac{\hbar k_{final,e}}{m_e} = \frac{\hbar k_{final,hh}}{m_{hh}}. \tag{B.11}$$

Combining these two equations we have

$$k_{final,hh} = k_{i,so}\left(\frac{m_{hh}}{2m_{hh}+m_e}\right), \tag{B.12}$$

$$k_{final,e} = k_{i,so}\left(\frac{m_e}{2m_{hh}+m_e}\right). \tag{B.13}$$

Now we deal with the energies. The total initial energy of the particle is

$$E_{i,so} = \Delta_{so} + \frac{\hbar^2 k_{i,so}^2}{2m_{so}} \tag{B.14}$$

and the total final energy of the three particles is

$$E_f = E_g + \frac{\hbar^2 k_{final,e}^2}{2m_e} + 2\frac{\hbar^2 k_{final,hh}^2}{2m_{hh}}. \tag{B.15}$$

Now by conservation of energy, we can set Eqs. B.14 and B.15 equal,

$$\Delta_{so} - E_g = \frac{\hbar^2 k_{final,e}^2}{2m_e} + 2\frac{\hbar^2 k_{final,hh}^2}{2m_{hh}} - \frac{\hbar^2 k_{i,so}^2}{2m_{so}}. \tag{B.16}$$

Then we use Eqs. B.12 and B.13 to relate the final momenta to the initial momentum. After some simplification we have

$$k_{i,so} = k_{th}^{hole} = \left[(E_g - \Delta_{so})\left(\frac{2m_{so}}{\hbar^2}\right)\left(\frac{2m_{hh}+m_e}{2m_{hh}-m_{so}+m_e}\right)\right]^{1/2}. \tag{B.17}$$

Then we can solve for E_f=E_i=E_{th} by simply plugging Eq. B.17 into the dispersion

relation E=$\hbar^2$k^2/2m and adding the offset Δ_{so}

$$E_{th}^{hole} = E_g \left[1 + \frac{m_{so}(1 - \Delta/E_g)}{2m_{hh} - m_{so} + m_e}\right]. \tag{B.18}$$

Now for an electron initiated event, the initiating carrier is an electron in the conduction band. The final carriers are two electrons in the bottom of the conduction band and one hole at the top of the heavy-hole band. From conservation of momentum and Eq. B.9 we have

$$k_{final,hh} = k_{i,e}\left(\frac{m_{hh}}{2m_e + m_{hh}}\right), \tag{B.19}$$

$$k_{final,e} = k_{i,e}\left(\frac{m_e}{2m_e + m_{hh}}\right). \tag{B.20}$$

As before, we now use conservation of energy to obtain

$$E_g + \frac{\hbar^2 k_{i,e}^2}{2m_e} = 2E_g + 2\frac{\hbar^2 k_{f,e}^2}{2m_e} + \frac{\hbar^2 k_{f,hh}^2}{2m_{hh}}. \tag{B.21}$$

Now plugging B.19 and B.20 into B.21, after some simplification,

$$k_{i,e} = \left(\frac{2E_g m_e}{\hbar^2\left[1 + \frac{m_e}{m_e + m_{hh}}\right]}\right)^{1/2}. \tag{B.22}$$

And finally, using the dispersion relation E=$\hbar^2$k^2/2m,

$$E_{th}^{hole} = E_g \left[2 + \frac{m_e}{m_e + m_{hh}}\right]. \tag{B.23}$$