



Black Silicon for Photodiodes: Experimentally Implemented and FDTD Simulated

Adam J. Williamson

► To cite this version:

Adam J. Williamson. Black Silicon for Photodiodes: Experimentally Implemented and FDTD Simulated. 2010. hal-03349103

HAL Id: hal-03349103

<https://amu.hal.science/hal-03349103>

Submitted on 20 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Black Silicon for Photodiodes: Experimentally Implemented and FDTD Simulated

Dissertation
Zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

vorgelegt der
Fakultät für Maschinenbau der
Technische Universität Ilmenau

von Herrn

M. Sc. Adam J. Williamson

geboren am 23.11.1980, London Ontario Kanada

Acknowledgements

Thanks goes to...

... X-FAB Semiconductor Foundries AG for financial support and the opportunity to experiment in their clean room. Special thanks to Dr. Andreas Voeckel for his time and for project oversight during my time at XFAB.

... Prof. Martin Hoffmann for the possibility to pursue a PhD externally and yet remain affiliated with his group, Micromechanical Systems (MMS), for support and advice at the Technical University of Ilmenau.

... Prof. Erich Runge for the opportunity to learn and implement Finite-difference Time-domain simulations within his group, Theoretical Physics 1, at the Technical University of Ilmenau.

And finally I would like to thank my wife, Mary, and my children, Noah, Daphne, Hazel, and Leontes. The time spent on the work presented here was not spent with the five of you.

Abstract

The experimental results presented here show the creation of black silicon through plasma processing and its subsequent implementation as an anti-reflective coating for photodiodes in the range of 200 nm to 800 nm. Finite difference time domain (FDTD) calculations of various nano-structure geometries are compared to integrating sphere reflection measurements of physical structures. The spectral sensitivity of manufactured black silicon photodiodes in the red region (675 nm to 750 nm) meets the set 0.5 A/W target, while in the blue region (375 nm to 425 nm) an improvement of +0.07 A/W over oxide covered diodes is achieved for center cites. Clear separation between nano-structure tips is experimentally demonstrated as the deciding factor to improve reflection suppression.

The appearance of nano-structures on the silicon surface when processed in the SF_6/O_2 plasma mixture is explained by generalizing the etch rate of silicon dependant on fluorine concentration (with a varying fluorine to oxygen gas ratio) into two regions, one of excess fluorine generation (weakly passivating) and one of excess fluorine consumption (strongly passivating). This is experimentally verified.

Inherent non-uniformities of the plasma processing chamber which yield a small process window of extremely anisotropic nano-structures and a poor center-to-edge wafer uniformity are eliminated by the introduction of a nano-loading step. The silicon wafer surface is first etched with a set of highly uniform nano-pillars. The nano-loading mask is subsequently plasma etched utilizing gas compositions that do not by themselves result in the appearance of nano-structures on the silicon wafer surface. This yields new less-anisotropic geometries with properties more desirable for reflection suppression and decouples the dependence of attainable nano-structure geometry on plasma chamber geometry. An extended nano-loading approach is successfully applied to transfer plasma processes from blank wafers to masked wafers.

FDTD results show that the majority of light absorption in the spectrum from 200 nm to 800 nm takes place within the silicon nano-structure antireflective coating, in contrast to the situation encountered with a standard anti-reflective coating. The periodicity of nano-structures is investigated with FDTD and clear diffraction grating properties are demonstrated. The ideal width of nano-structures for reflection suppression is calculated to be between 100 nm and 200 nm for the light spectrum of interest (200 nm to 800 nm). Any periodic structure with widths over 200 nm will begin diffracting light into non-specular angles. The ideal structure height is calculated to be (at least) 500 nm. However, the periodic structure simulations do not prove to be the most accurate way to evaluate the measured specular and diffuse components of real physical structures. The problem being that overlapping nano-structures (for example unseparated through the plasma etch process) disturb the conditions of periodicity (diffraction into component angles) corresponding to the structures overall width. FDTD simulations are best correlated to the optical properties of physical self-organized nano-structures (integrating sphere measurements of specular and diffuse components) when a non-periodic surface is simulated.

Boundary conditions for the FDTD are discussed in the process of simulating non-normal incidence light on 2D structures. Dispersion in the injection angle for a simulation creates difficulties in simulating broadband spectrums in a single simulation, as only the center wavelength possesses the correct theta, while the smallest and largest wavelength in the incident experience the most extreme error in incident theta. Simulating each wavelength is not possible as it yields “incalculable” computational times. The solution provided breaks the desired spectrum into several smaller spectrums and gives satisfactory results while limiting the amount of dispersive error. Furthermore it is shown here that the implementation of gradient-index in front of perfectly matched layers (PMLs) improves their angle-dependent performance.

Additionally it is experimentally demonstrated here that reflection suppression is not always correlated to electrical performance as plasma etching to modify surface topology

drastically reduces carrier lifetimes, due to an increase in recombination resulting from lattice damage and an overall increase in surface area.

Kurzfassung

Die hier vorgelegten experimentelle Ergebnisse zeigen die Herstellung von schwarzem Silizium zur Anwendung in Photodioden im Bereich zwischen 200 nm und 800 nm durch Plasmabearbeitung. Berechnungen verschiedener Nano-strukturgeometrien mit der Finite-Differenzen-Methode im Zeitbereich (FDTD) werden mit Ulbricht-kugel Reflexionmessungen von physikalischen Strukturen verglichen. Die spektrale Empfindlichkeit der schwarzen Siliziumphotodioden im roten Bereich (675 nm – 750 nm) treffen das 0.5 A/W Ziel. Im blauen Bereich (375 nm-425 nm) wird eine Verbesserung von Photodioden in der Wafermitte von +0.07 A/W im Vergleich zu oxide-bedeckten Photodioden erreicht. Klare Unterschiede zwischen verschiedenen Spitzen der Nanostrukturen werden experimentell und theoretisch belegt. Sie sind entscheidend, um die Reflexion gering zu halten.

Das Entstehen der Nanostrukturen in unserer SF_6/O_2 Plasmamischung wird durch die Siliziumätzrate in Abhängigkeit von der Fluorkonzentration (mit einem veränderlichen Fluor-zu-Sauerstoff Verhältnis) in zwei Bereichen erklärt. In einem Bereich gibt es überschüssige Fluorradikalerzeugung (schwache Passivierung) und im anderen überschüssige Fluorradikalverbrauch (starke Passivierung). Dies wird experimentell gezeigt.

Inhärente Ungleichmäßigkeiten in der Plasmaanlage, die ein kleines Prozessfenster mit extrem anisotropischen Nanostrukturen und nachteiliger Mitte-zu-Rand Waferuniformität ergeben, werden durch die Einführung von einem Nanoloading Schritt beseitigt. Die Siliziumoberfläche wird am Anfang mit einer Reihe von sehr uniformen Nanopfeiler geätzt. Diese Nanoloading-Maske wird anschließend mit Chemikalien, die allein keine Nanostrukturen ergeben, plasmageätzt. Die entsprechenden Ergebnisse sind isotropischer und haben Eigenschaften, die für Reflexionsabsenkung erwünscht sind. Dadurch entwickeln wir eine Abkoppelung zwischen die Abhängigkeit der erreichbaren Nanostrukturgeometrien und Plasmaanlagengeometrie. Erweitertes Nanoloading wird benutzt, um die Prozesse erfolgreich von blanken zu maskierten Scheiben zu transferieren.

FDTD Ergebnisse zeigen, dass der größte Teil der Lichtabsorption im Bereich vom 200 nm zu 800 nm direkt in der Antireflexschicht der Nanostrukturen stattfindet. Dies ist im starken Vergleich zu der Situation mit einer normalen Antireflexschicht. Weitere FDTD Simulationen untersuchen die Periodizität von Nanostrukturen, und zeigen deutliche Gittereigenschaften. Die ideale Breite der Strukturen für Reflexionsabsenkung wird zwischen 100 nm und 200 nm im interessanten Spektralbereich (200 nm - 800 nm) berechnet. Periodische Strukturen mit Breiten über 200 nm reflektieren Licht in verschiedene Gitterordnungen. Die ideale Strukturhöhe für vernachlässigbare Reflexion wird in diesem Bereich mindestens 500 nm. Allerdings, erweisen sich die periodischen Struktursimulationen nicht als die richtige Methode, um die Ulbrichtkugel Messungen von physikalischen Strukturen zu beurteilen. Das Hauptproblem ist eine Überlappung zwischen Nanostrukturen (z.B. ungetrennt während eines Plasmaprozesses) stört die Zustände für Periodizität (die Diffraction in die Komponentwinkeln) entsprechend der Breite der Strukturen, und stört auch die Gradient-brechungsindex (Dadurch ist Reflexion erhöht. FDTD Simulationen werden am Besten in Übereinstimmung mit den Eigenschaften der physikalischen selbst organisierten Nanostrukturen gebracht, wenn eine nichtperiodische Oberfläche mit einer simuliert wird.

Randbedingungen für FDTD in Bezug auf Simulationen von nicht-senkrecht einfallendem Licht in 2D werden diskutiert. Dispersion im Injektionswinkel führt zu Schwierigkeiten, wenn man Rechnungen mit einer Breitbandquelle durchführen will, weil nur einfallendes Licht mit der zentralen Wellenlänge den vorgegebenen Einfallswinkel hat. Die größte und kleinste Wellenlänge im Quellspektrum zeigen die stärkste Abweichung. Alle Wellenlängen einzeln nacheinander zu simulieren würde zu einer nicht akzeptablen Rechenzeit führen. Die hier gezeigte Lösung teilt das komplette gewünschte Spektrum, in kleinere simulierbare Bänder, und ergibt ausreichende Ergebnisse und schränkt gleichzeitig Dispersionsfehler ein. Weiterhin wird es hier gezeigt, dass

eine mögliche Verbesserung der winkel-abhängigen PML Leistung erreichbar ist, wenn ein Gradientindex vor der PML eingeführt wird.

Es wird experimentell bewiesen, dass Reflexionsabsenkung nicht immer mit ausreichenden elektrischen Eigenschaften verbunden ist, weil das Plasmaätzen um die Oberflächentopographie zu ändern auch stark die Lebensdauer der Minoritätsträger wegen eines Anstieges von Rekombination durch Gitterschädigung und einer allgemeinen Vergrößerung in Oberfläche ändert.

Table of Contents

1. INTRODUCTION	9
1.1 WHAT KIND OF BLACK SILICON DO WE WANT?	9
1.2 ADDITIONAL PROCESS CONSIDERATIONS	13
2. TOOLS AND FUNDAMENTAL PRINCIPLES	14
2.1 THE BASIC PLASMA PROCESSING CHAMBER	16
2.2 PLASMA ETCHING	23
2.3 TRANSFORMER COUPLED PLASMA LAM REACTOR	25
2.4 REFLECTION MEASUREMENT AND CHARACTERIZATION TOOLS	32
2.5 CONCLUDING COMMENTS	36
3. MORPHOLOGY	37
3.1 THEORY OF NANO-MASKING AND THE PROPOSED ORIGIN OF OUR NANO-STRUCTURES	37
3.1.1 INFLUENCING A NANO-MASK: RADICAL ABSORPTION, DESORPTION, AND PASSIVATION	37
3.1.2 INFLUENCING A NANO-MASK: THE INTERACTION OF IONS AND THE WAFER SURFACE	40
3.1.3 INFLUENCING A NANO-MASK: CHEMISTRY OF THE FLUORINE PLASMA	44
3.2 THEORY INTO REALITY: SILICON SAMPLE PROCESSING	51
3.2.1 TEXTURING SILICON INTO NANO-STRUCTURES, PART I: NANO-LOADING	51
3.2.2 TEXTURING SILICON INTO NANO-STRUCTURES, PART II: PILLARS TO PYRAMIDS	64
3.2.3 EXPERIMENTAL SUPPORT FOR THE THEORY OF OUR NANO-MASKING	67
3.2.4 MODIFYING THE PERCENTAGE OF OPEN AREA	70
3.3 CONCLUDING COMMENTS ON MORPHOLOGY	80
4. OPTICAL PROPERTIES	81
4.1. REFLECTION FROM A SURFACE	82
4.2. MODELING THE ELECTROMAGNETIC FIELD WITH FDTD	85
4.2.1 THE DISCRETIZATION PROCESS AND THE LEAPFROG METHOD	85
4.2.2 FDTD SIMULATION RESULTS OF OPTICAL PROPERTIES	88
4.3 SIMULATIONS COMPARED TO ACTUAL PHYSICAL STRUCTURES	118
4.3.1 THE LACK OF PERIODICITY IN SELF-ORGANIZED NANO-STRUCTURES	124
4.4 CONCLUDING COMMENTS ON OPTICAL PROPERTIES	129
5. ELECTRICAL PROPERTIES	131
5.1 PROCESS INTEGRATION	131
5.1.1 DEFECTS THROUGH ETCHING AND DOPING AND SUBSEQUENT ANNEALING	131
5.1.2 PROCESS INTEGRATION RESULTS BY MICROWAVE PHOTOCONDUCTIVE DECAY (μ PCD)	137
5.2 DIODES	142
5.2.1 ILLUMINATED PN JUNCTIONS	142
5.2.3 SPECTRAL SENSITIVITY RESULTS	147

5.3 CONCLUDING COMMENTS ON ELECTRICAL PROPERTIES	152
6. SUMMARY AND OUTLOOK	153
RESOURCES	155
APPENDIX A: IMPROVEMENT OF PMLS BY GRADIENT-INDEX	162
APPENDIX B: THESEN ZUR DISSERTATION	166
APPENDIX C: ERKLÄRUNG GEMÄß ANLAGE 1 DER PROMOTIONSORDUNG	167

1. Introduction

Black silicon is a nano-structured silicon surface formed normally by a self-aligning, maskless process with structured surfaces often discernible by their black color and tremendous anti-reflective properties. The experimental results presented here will show directly the creation of black silicon through plasma processing and its subsequent implementation as an anti-reflective coating for photodiodes in the range between 200 nm to 800 nm. Integrating sphere reflection measurements of physical structures are compared to Finite difference time domain (FDTD) calculations of various modelled structure geometries and measured spectral sensitivities of finished black silicon photodiodes are discussed. Further topics addressed include nano-loading surfaces to decouple the dependence of attainable structure forms from plasma etch tool geometry and process integration steps to improve the carrier lifetime within manufactured nano-structures. As the majority of this work was done in partnership with industry, the intention is to have the results presented here appear in as modular a form as possible. This leaves the reader the opportunity to select the relevant section for his or her area of interest, as it should be possible to read any chapter independently of the others.

This introduction will cover in section 1.1 the basic motivation for desiring to implement black silicon in photodiodes and in section 1.2 some process consequences derived from the first section that should be kept in mind throughout the reading. Together both sections are meant to provide the reader with the general scope of the complete work presented here. Chapter 2 is general information discussing basic plasma etching and the main tools used here for plasma etching and reflection measurement. Chapter 3 covers in detail the creation of black silicon on blank and masked silicon wafers and the manipulation of the corresponding morphological properties of the resultant nano-structures. Included in this chapter is a discussion of the origin of the self-masking phenomenon observed here and the use of nano-loading, a process in which sets of nano-structures are built to influence the behavior of subsequent etching steps with excellent experimentally observed benefits to the uniformity and to the sculpting of the resultant nano-structures (relatively) independent of the plasma chamber geometry. Chapter 4 discusses optical properties with FDTD simulations of numerous structured surfaces and compares the results to experimental reflection measurements. Chapter 5 deals with the electrical properties of the devices, namely minority carrier lifetimes within nano-structures characterized by microwave photo-conductivity decay and the spectral sensitivity of real fabricated black silicon photodiodes.

1.1 What kind of black silicon do we want?

The answer to this question is very interesting but not particularly short. Clearly at first glance the black silicon of interest for this dissertation must be the observed nano-structured silicon surface with tremendous anti-reflective properties mentioned above. However in general all black silicon is a micro- or nano-scale set of conical silicon structures formed normally by a self-aligning, maskless, process, and these structures frequently have a very discernible black color. Therefore our current definition fits structures observed at Harvard University after irradiating a silicon surface with a train of femtosecond laser pulses shown in figure 1.01 and which are currently being investigated for use in laser diodes [X2.1, 2]. Another fitting set of needles, this time fabricated with a deep reactive ion etch (DRIE), are those created with the now classic "Black Silicon Method" developed by Henri Jenson and pictured in figure 1.02 [X2.6, 7]. This method uses micro-grass as a gauge of etch anisotropy in deep trench etching of silicon. To achieve directional etching in a fluorine-based plasma an inhibitor is added, either from an oxide-

forming gas (O_2) or a fluorocarbon containing gas (C_4F_8 or CHF_3). This inhibitor can be introduced together with the etch gas, which is named a mixed-mode DRIE process, or the inhibitor can be added in a time-multiplexed manner, which is termed a pulsed-mode DRIE process creating the very familiar “scalloped” side walls. In both pulsed and mixed mode these needles can be very long, each in excess of $10\ \mu m$. An extraordinary use of such pulsed-mode DRIE needles is in wafer bonding at the Technical University of Ilmenau [X2.8]. The needles from one sample are pressed into the needles of a second sample creating a definite joint with the scalloping of both sets of needles gripping one another and resulting in the two samples quite solidly married together. These are interesting nano-structures with interesting applications and a person so inclined can find numerous articles discussing very small micro- or nano-scale needle structures manufactured in a wide variety of ways in numerous materials and many of the structures appear dark in color.

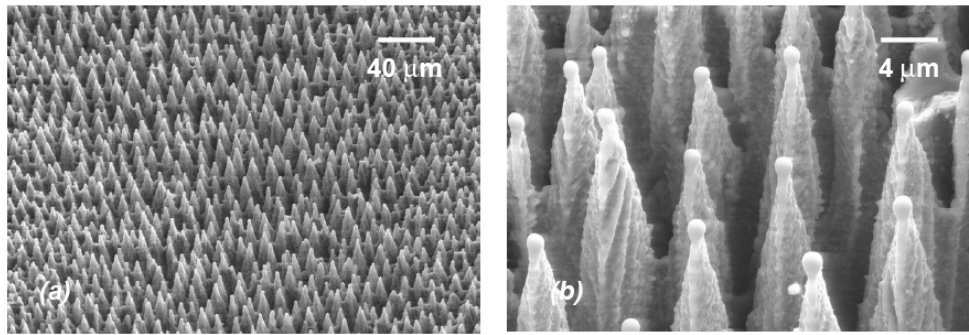


Figure 1.01: Black silicon created with femtosecond laser pulses [X2.6, 7]

It is very important then to distinguish the desired nano-structures for use in this dissertation from other forms of black silicon by establishing a stricter set of criteria. The application of interest here is for implementation in photodiodes as an anti-reflective coating (ARC) and as such the targets should be in terms of surface reflection and spectral sensitivity of the photodiodes. Both of these characteristics will be covered in more detail in chapter 3 and 4 respectively, however, it is at this point sufficient to introduce the ideal targets. A graph in figure 1.03 shows simulated percent reflection values as a function of wavelength. The curve shown in pink featuring the much larger values is reflection from bare silicon substrate while the approximately flat line at the 2% value represents the reflection value from an oxide substrate.

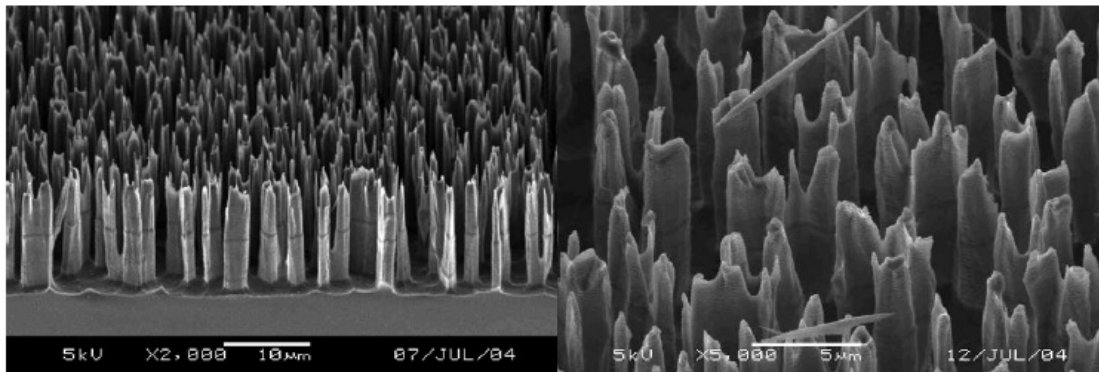


Figure 1.02: Black Silicon created with the Bosch process [X2.6, 7]

The green line is reflection from a nitride covering on a silicon substrate, a single layer ARC. Reflection from the bare substrates is simply a single reflection from the interface between air and the bulk substrate. With a covered substrate there are two reflections, a first from the air/covering interface and a second from the covering/substrate interface. Due to the interference between the two reflections, oscillating maximums and minimums are observed in the measured reflection. In a reflection minimum a transmission maximum is expected and thus for photodiodes applications an ARC is designed to provide a minimum at a desired wavelength of operation. For example the nitride ARC shown would be quite appropriate for detecting 425 nm light. Naturally a device with a wider reflection minimum or multiple minimums would be effective over a wider range of wavelengths.

It is conceivable that a perfectly structured substrate would remove the second interface completely and thus remove the second reflection and any subsequent interference. This would be the case if the set of nano-structures on the surface behaved as an ideal gradient index, a continuous transition in the index of refraction between the covering and the substrate. Such an ideal transition yields zero reflection as all light incident at the second interface is passed into the substrate [X3.2]. The resulting system would optically show as little reflection from its surface as a substrate made of only the covering, quite small in the case of oxide, and no visible oscillations in reflection values. However the same system would electrically retain the properties of the silicon substrate. This is critical as the band gap of SiO_2 is about 9 eV in contrast to 1 eV for silicon and clearly this would create some rather difficult problems for conventional photodiode applications. It is the target of this work to create a system which functions optically as one enormous reflection minimum between 200 nm and 800 nm. A comparison of the optical properties of experimentally created structures and simulated geometries using FDTD is performed in chapter 4, showing how close the fabricated structures come to performing as an ideal gradient-index.

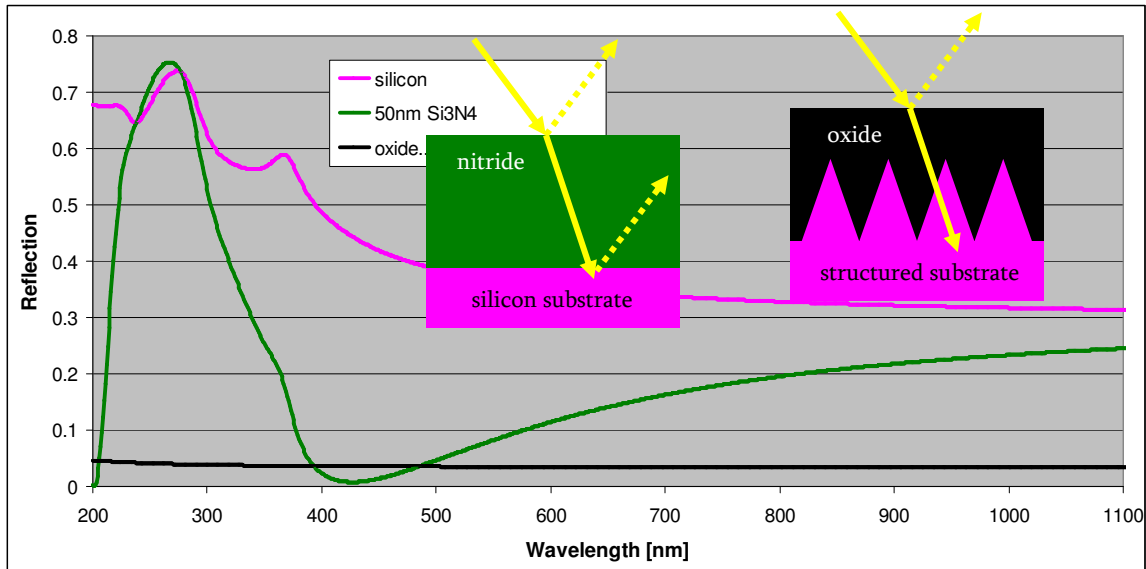


Figure 1.03: Simulated reflection from several surfaces [X4.4]. Bare silicon substrate in pink, oxide substrate in black, and a nitride ARC on a silicon substrate in green. It is conceivable that a perfectly structured substrate would remove the second interface completely. This would yield measured reflection values from the covering only, quite small in the case of oxide.

The simulated consequences for the electrical properties of the reflection curves from figure 1.03 are depicted in figure 1.04. The ideal quantum mechanical spectral sensitivity is

shown in yellow, representing a state in which every incident photon creates a unique electron-hole-pair (EHP) which subsequently separates and contributes to the photocurrent. The three additional curves are the ideal sensitivities of 10 μm thick silicon photodiodes (diodes having a total well depth of 10 μm) with the optical properties from figure 1.03. The bare silicon (no ARC) again in pink, the nitride ARC again in green, and in black, labeled “Si with BSi” (black silicon), is the calculated sensitivity of a silicon diode having experienced only the small reflection loss of the oxide substrate. The bare silicon photodiode demonstrates quite well the need for an ARC as it spectrally performs very poorly resulting from the loss of as much as 60% or more of incident light due to reflection depending on wavelength. We clearly see that the nitride curve meets the ideal performance at approximately 425 nm, as expected, as this was a transmission maximum in the reflection spectrum (a reflection minimum). For other wavelengths the performance of the device reduced correspondingly with reflection loss. The black line follows the ideal line minus only a few percent. This black line should ideally represent the oxide covered silicon nano-structures implemented as an ARC in a silicon photodiode. With such theoretical reflectivity and sensitivity this system could be very exciting as a real antireflective coating and would provide a remarkable technological achievement. However a few obstacles in the path should be discussed.

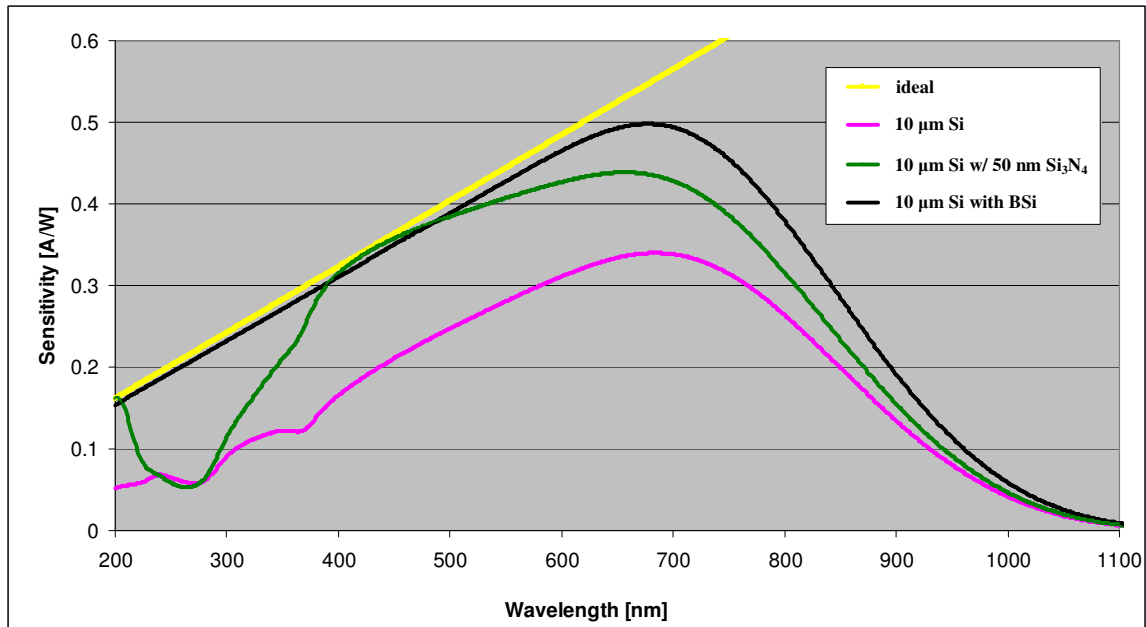


Figure 1.04: Simulated photodiode spectral sensitivity (for diodes 10 μm thick) [X4.4]. The ideal situation is shown in yellow, each incident photon generates an EHP which separates and contributes to photo-current. The bare silicon and nitride ARC are again in pink and green respectively. Due to the substantial amount of reflection from bare silicon its corresponding electrical performance is quite poor. In contrast the nitride meets the ideal performance line at 425 nm, its reflection minimum from figure 1.03, but nowhere else. The black line, labeled “Si with BSi” (black silicon), represents electrically a silicon substrate however with the only the small reflection loss of the oxide substrate. Thus the black line follows the ideal line minus only a few percent. This black line should ideally represent the oxide covered silicon nano-structure system.

1.2 Additional process considerations

The necessity for finding the correct nano-structure form (representative of an ideal gradient-index) was introduced, however the nano-structures must also be able to meet the production demands of a large scale CMOS fabrication environment. It is necessary then to select a tool for creating black silicon that is capable of handling wafer cassette loading and the rigorous requirements of standard mass production. This document deals directly with the evaluation and creation of black silicon in just such a production machine, namely the LAM TCP. The physics of the tool are treated in the following chapter along with general plasma processing information and the complete black silicon process results of this work are presented in chapter 3. These etch processes and the resulting nano-structures will meet the physical requirements of having a structure shape that is coverable with a thermal and PECVD (plasma enhanced chemical vapor deposition) oxide and of having an overall consumed silicon less than $1.5\text{ }\mu\text{m}$.

Pictured below in figure 1.05 is a SIMS doping profile measured on a fully prepared photodiode. Carrier concentration is shown down to a well depth of $4\text{ }\mu\text{m}$, however past the $1.5\text{ }\mu\text{m}$ mark (shown with a red dashed line for reference) the concentration of carriers begins to fall off quite sharply. It is immediately apparent then that silicon consumption values greater than perhaps $2\text{ }\mu\text{m}$ would appear to etch the diode away in its entirety and the implications of this are shown in figure 1.06 by a typical p-n junction.

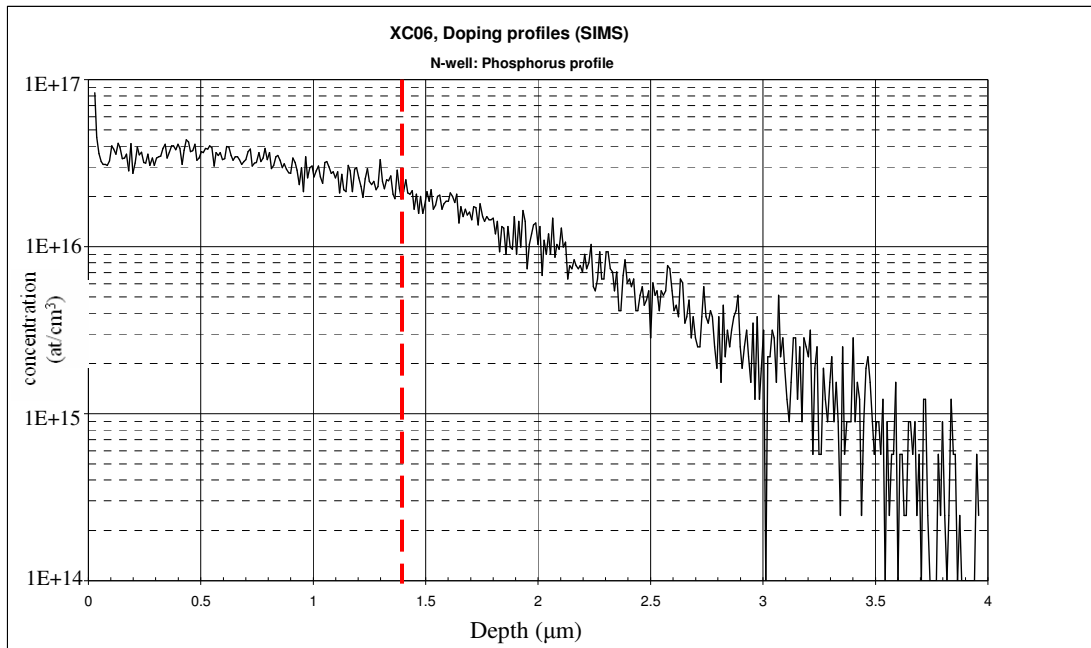


Figure 1.05: SIMS doping profile of an n-well [X4.5]. Structures are directly etched into the diode well making the minimization of consumed silicon a priority. A dashed red line is shown on the graph as a reference at $1.5\text{ }\mu\text{m}$. After this depth the concentration of dopants begins to decrease quickly. For this work we will attempt to limit consumed silicon, measured from the base of structures to the original surface of the diode well, to no deeper than $1.5\text{ }\mu\text{m}$.

The p-substrate is left unchanged however the Fermi-level of the junction (discussed in detail in chapter 5) on the n-side, E_{FN} , is reduced and will be further reduced as the structures are etched deeper and deeper into the n-well. Effectively this alters the intrinsic junction potential, lowering the overall field strength and thus the ability to separate any generated electron hole pairs. Eventually further etching would leave only a nothing-silicon interface, no junction potential.

Indeed the accompanying SEM picture seems to have done just this, showing the results of a very aggressive anisotropic etch and its massive 3 μm tall structures. Left unstopped such a trenching process could likely carry straight on through the wafer (or at least until the oxide mask is fully consumed). Necessity then dictates that this be considered a mandatory prerequisite for a functional black silicon etch.

The complete content of this work has been summarized in this introduction, the creation of a set of nano-structures meeting the above discussed criteria: Structures possessing a gradient-index reflective behavior and a corresponding maximization of spectral sensitivity over a wide range of wavelengths.

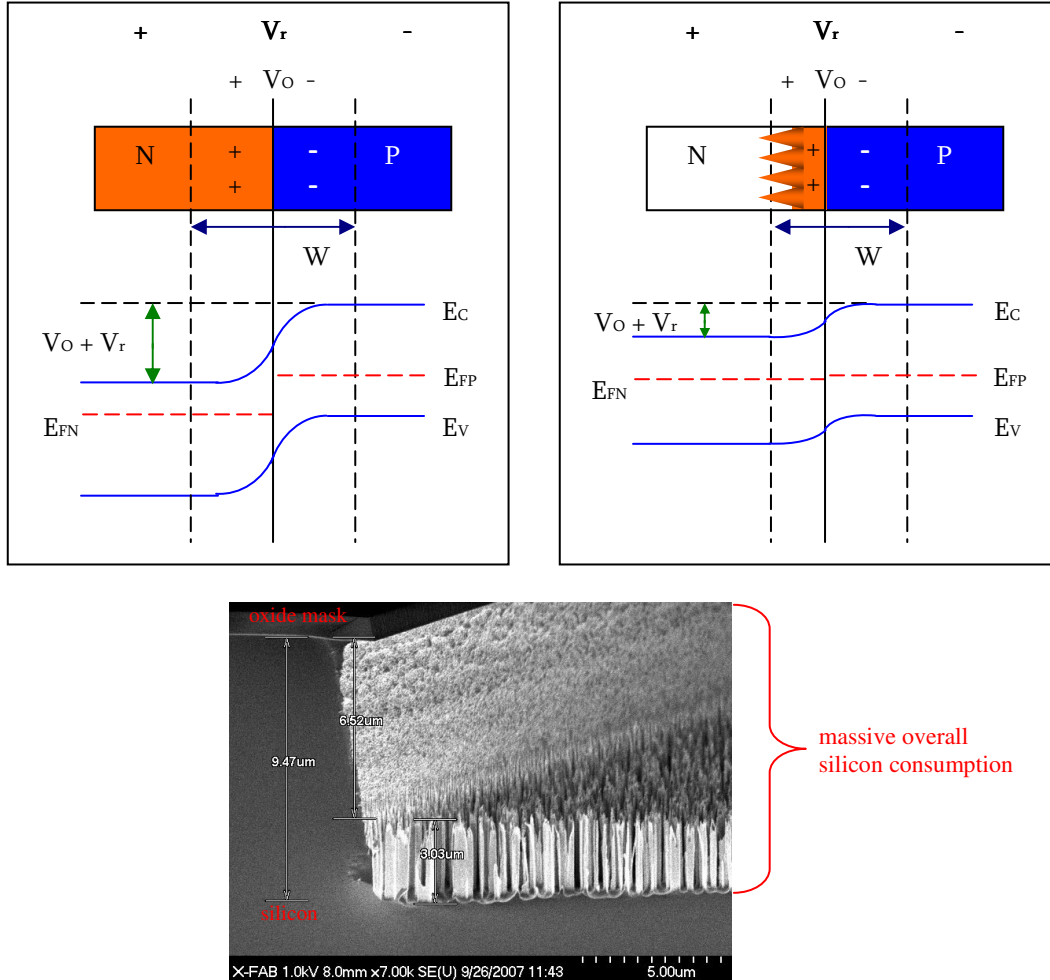


Figure 1.06: Etching the well of a diode completely away. The junction's Fermi-level (discussed in detail in chapter 5) on the n-side, E_{FN} , is reduced and as structures are etched deeper into the n-well. This changes the intrinsic junction potential, lowering the overall field strength and thus the ability to separate generated EHPs. Additional etching leaves only a nothing/p-silicon interface, no junction potential, no diode. An extreme SEM example is shown. Structures are over 3 μm tall (already violating our desired depth of no more than 1.5 μm), however total consumed silicon is in excess of 9 μm ! Control of the consumed silicon on masked wafers is discussed in chapter 3.

The latter point additionally guaranteed by limiting the amount of overall consumed silicon and minimizing the impact of plasma etch damage on the minority carrier lifetime within the structures. Additional discussion relevant to this work is the reason for the resultant black silicon

as a phenomenon in and of itself. No evidence of special contaminants or particulates for a masking effect have been discovered, rather initial emergence of structures and their corresponding properties seem to be related to the ratio of etchant to passivant present in the plasma, making the processes and appearance of the nano-structures themselves all the more interesting. Attention will be given this nano-structure formation process itself, a discussion of the self-masking effect, throughout chapter 3.

2. Tools and fundamental principles

This dissertation implements a unique etching effect observed in an SF_6/O_2 plasma mixture, often also in combination with CF_4 , resulting in the creation of a nano-structured silicon surface with tremendous anti-reflective properties, referred to as black silicon. For this reason a discussion of both the specific tool used in the plasma etching processes of this work and the tools used in the characterization of reflection from the nano-structured surface is necessary, as an understanding of the benefits (and limitations) of the specific tools themselves will be used to assess the results of the next chapters. This chapter discusses in sections 2.1 and 2.2 basic plasma processing chamber and basic plasma etching respectively. The information of these first two sections is expanded in section 2.3 to the more complex TCP (Transducer coupled plasma) from LAM Research, the actual tool used to create the nano-structures of this work. A number of chamber specific issues (location of gas ring, method of power coupling etc) discussed here relate directly to non-uniformities and process limitations observed in the morphological results of the next chapter. Section 2.4 introduces the integrating sphere and the NanoSpec, the two machines used to measure reflection from the surfaces etched in the TCP. The discrepancies between measurements of the two machines when characterizing the same sample are shown to be very sample dependant.

2.1 The basic plasma processing chamber

The processing chamber during a dry etch process contains plasma, matter heated beyond its gaseous state. The plasma bulk can be described as a fluid-like mixture of gas molecules, electrons, ions, and radicals and is also said to be “quasi neutral”, having approximately the same number of electrons and ions. However ordinary fluids such as gases are in thermal equilibrium, particles in the system have a range of different velocities and the velocity of any single particle is constantly changing due to collisions with other particles, but overall the collisions have stabilized the system’s particle distribution function to that of a Maxwell distribution [X1.6]. This then allows a single temperature and pressure to be attributed to the whole system. Plasma however is not in thermal equilibrium, the electrons and ions have very different Maxwell distributions resulting in a much higher average electron temperature. Densities and energies of various species in a low-pressure capacitively coupled discharge along with the energy distribution of electrons in plasma used in semiconductor processing are shown below in figure 2.01. With the exception of the electrons the main elements of the bulk all share a similar energy, although differing substantially in concentration. Thus processing plasma is always denoted as weakly ionized with only $\approx 1\text{-}10\%$ of the plasmas overall density made up of electrons and ions with the plasma itself sustained by the very high temperature electrons through their collisions with the other particles present in the plasma. At high energies, E_{diss} , electron collisions with gas molecules and radicals result in electron impact dissociation processes creating active etching components and other species not normally available in a simple gas. At still higher temperatures the energies, E_{iz} , of these electrons will cause impact ionization through subsequent collisions which break down gas molecules and radicals further into ions and electrons and thereby further sustain the plasma [X1: 3]. Not considered part of the bulk plasma is the highest energy species contained in the chamber, the bombarding ions, and their presence is a direct result of the structure and power coupling regime of the plasma processing chamber itself. The location and behavior of the bombarding ions becomes clear as the basic design of the chamber, the symmetric capacitive RF discharge, pictured in figure 2.02 a) with one upper grounded electrode and one

powered bottom electrode, is explained. As shown, the plasma bulk is separated from the top and bottom electrode by a thin sheath region in which the density of ions is greater than that of electrons. Almost all the RF voltage (normally at 13.56 MHz) is dropped across these sheaths with their corresponding electric field vectors directed from the bulk plasma towards the electrodes. The plasma is therefore at a positive (higher) potential V_p with respect to the electrodes and the normal parameter of representation for V_p in a plasma chamber is the chamber bias voltage, V_{bias} . A description of V_{bias} yields a slightly awkward result for the symmetric chamber as the two parameters, V_{bias} and V_p , are not exactly the same, however as will be seen with the asymmetric discharge the values begin to approach one another and notwithstanding the insight provided will correctly anchor the location of the bombarding ions. An equivalent circuit diagram of the chamber illustrates the bias voltage in figure 2.02 b) and we immediately notice that a plasma processing chamber is simply two capacitors in series representing the two sheaths and the plasma bulk is considered completely fluid and having no resistance (a negligible voltage drop).

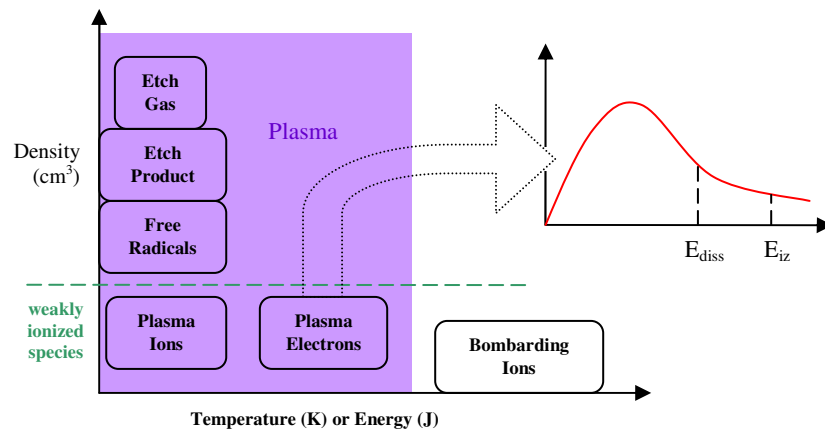


Figure 2.01: Contents of the basic plasma chamber [X1.2]. The density is shown for the four main components comprising a plasma dependant on temperature/energy. Bombarding ions are also shown, although, as we will discuss, these can be thought of as outside the discharge accelerating toward the wafer. We see that the majority of the plasma is made up of neutral contents, free radicals, etch products and etch gas. As a result the processing plasma is said to be a weakly ionized plasma (the green line in the figure separating the charged and neutral contents for clarity). The second item to note is that the contents of the plasma is not in thermal equilibrium. The temperature of ions and electrons in the plasma is significantly different. A Maxwell-Boltzman distribution is represents the energy of the electrons present in the plasma. At high energies, E_{diss} , electron collisions with gas molecules result in impact dissociation creating active etching components. At still higher temperatures the energy, E_{iz} , of these electrons will cause impact ionization breaking down gas molecules and radicals further into ions and electrons.

Clearly then in a symmetric chamber the measured bias must necessarily be zero because the chamber bias is taken as the time averaged value measured from ground to the powered bottom electrode across the whole chamber, both sheaths, while the real plasma potential V_p is the time averaged voltage drop over just one sheath. At an RF voltage of 13.56 MHz the ions, as a result of their size and weight, do not respond to the signal's frequency. They respond only to the time averaged fields within the chamber meaning that ions which diffuse through the bulk to the sheath edge and leave the plasma are accelerated across these sheaths bombarding the upper and lower electrodes with an energy equivalent to V_{p-avg} , the upper or lower sheath voltage [X1.2].

The electrons leaving the plasma accelerating towards the electrode are, as a rule, always reflected back into the plasma bulk by the expanding and contracting RF sheath. Ions have a much greater mass and a correspondingly much greater transit time to the electrode. Thus ions begin to respond significantly to the oscillations in the sheath at low RF frequencies, when transit time has the opportunity to be less than one RF cycle, subsequently resulting in a spread of incident ion energies [X1.4]. At high RF frequencies it takes ions several periods to cross the sheath (the ions see a time averaged sheath). Correspondingly a time-averaged sheath delivers ions to the electrode with a very narrow distribution of energies, and clearly increases the overall density of ions as the higher frequency sheath reflects more electrons per unit time into the plasma creating more ionized species. Both cases are depicted below in figure 2.03 a).

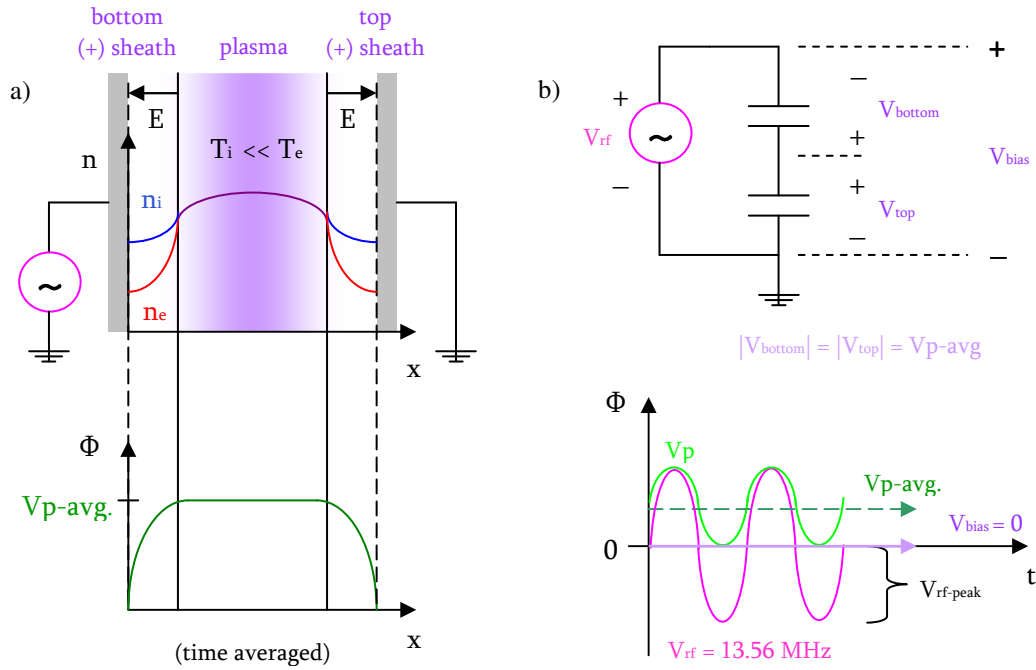


Figure 2.02: The a) symmetric capacitive RF discharge and b) chamber bias [X1.2, 3, 4]. The symmetric plasma discharge is thought of as a plate capacitor, with each plate having an equal size, driven by an AC source. The plasma bulk is separated from the top and bottom plates by a sheath region in which the density of ions, n_i , is greater than that of electrons, n_e . The majority of RF voltage is dropped across these sheaths with their corresponding electric field vectors directed from the bulk plasma towards the electrodes. The plasma is therefore at a positive potential V_{p-avg} with respect to the electrodes, as shown. An equivalent circuit diagram illustrates the chamber bias voltage (the voltage from plate to plate). We notice that the upper and lower sheaths are simply two capacitors in series and the plasma bulk is considered completely fluid and having no resistance (a negligible voltage drop). It follows that in a symmetric chamber the measured bias must be zero, depicted beneath the equivalent circuit.

The switch between regions of high and low RF frequency regarding ion behavior is referred to as the “ion transition region”, and is certainly dependant on the species of gas in the chamber (Xenon is quite a bit heavier than Helium for example), but as a rule of thumb it occurs between 1 and 3 MHz. The most often encountered operational frequency mentioned above, 13.56 MHz, is quite safe inside the high RF region.

Regardless, in both the low frequency and high frequency regions of operation, ion current and electron current to the electrodes are an equal amount (area) per time. However for

the high RF situation ion flux occurs continuously while electron flow only occurs during sheath collapse, pictured in figure 2.03 b). The interesting information to be gathered here is that during each RF period all of the voltage is at some point dropped over one sheath while the other sheath collapses to zero resulting in each sheath experiencing a collapse once per period (for the interested reader Appendix I provides some magnificent pictures of this), with the continuous plasma bulk potential V_p swinging from zero to the RF peak [X1.8]. At this point ion bombardment of each electrode is equal but the wafer to be etched is located at the bottom electrode only. The introduction of an asymmetric chamber, pictured below in figure 2.04, can greatly increase the efficiency of ion bombardment at the powered electrode and thus decrease the unnecessary and wasteful bombardment of the exposed upper electrode.

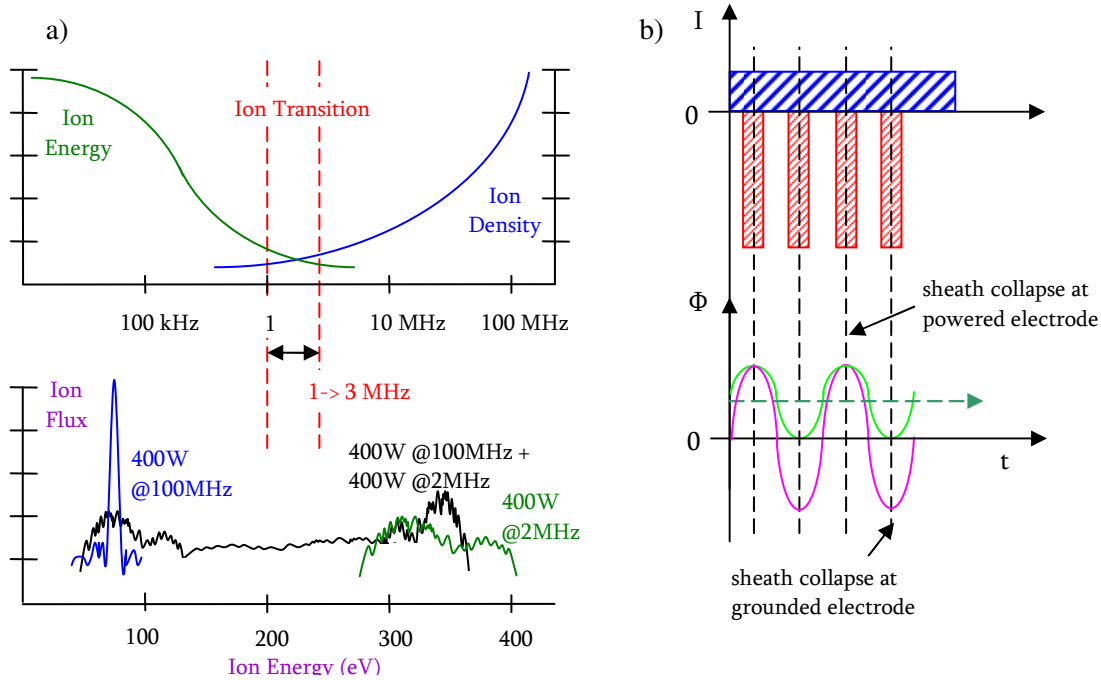


Figure 2.03: a) Pictured above is ion energy and ion density as a function of lower electrode driving frequency, and below is ion flux (to the electrode) as a function of ion energy. At low frequencies ions transit time can be less than a single RF period yielding the possibility for much greater energies and a much greater spread in energies of delivered to the electrode. At high frequencies the ions experience a time-averaged sheath yielding very consistent (uniform) incident energies and interestingly a much greater overall density as with each sheath expansion (many more per unit time for higher RF values) sends a new beam of reflected electrons into the plasma further ionizing species [X1.2, 3, 4]. In b) equal amounts (areas) of current flow are seen to the electrodes, however (shown) in the high RF region ion flux takes place continuously while electron flow occurs once per period during sheath collapse.

The symmetric capacitively driven discharge, as the symmetric discharge, has one grounded and one powered electrode however the size of the upper electrode is much greater than the size of the bottom electrode. The bottom sheath voltage is larger than the sheath voltage between the grounded electrode and the plasma. A much greater E-field is created between the small plate (bottom electrode) and the large plate (plasma bulk) of a capacitor. This clearly means that there will be a larger sheath voltage and thus larger ion energies to the wafer. Ions reaching the grounded electrode have energy V_p while those reaching the powered electrode have energy

$|V_p| + |V_{bias}|$. In reality the fluidity of the plasma itself (its resistance) will also contribute a voltage drop to the chamber bias. This can be seen when leaving all power parameters, and gas flows the same but simply changing the chemical composition of gas and watching the bias voltage change. Unfortunately, also seen in figure 2.04, different chamber geometries will yield different V_{bias} values making a comparison between to different types of machines quite meaningless. In practice the method of V_{bias} measurement is very dependant on tool manufacturer and is itself often not done in straight forward manner. However it is clear that the positive plasma potential V_p directly influences the energy of ion bombardment as it is the size (strength) of each sheath regions, setup by V_p , across which the ions accelerate.

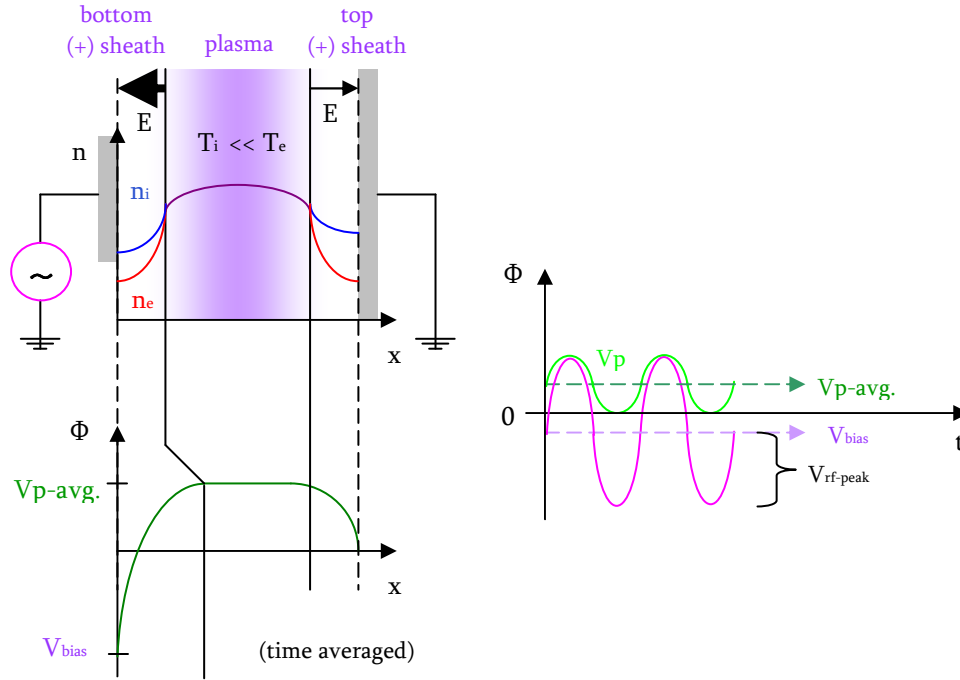


Figure 2.04: Asymmetric capacitive discharge [X1.2, 3, 4]. This chamber is again thought of as a plate capacitor. However, in comparison to figure 2.08, the plates are different sizes. The lower plate driven by the AC source is smaller. The effect on the electric field vectors directed from the bulk plasma towards the electrodes is shown. The lower field is much greater in comparison and thus the voltage drop across the lower sheath is greatly increased. The subsequent value for the chamber bias will be negative. One can see this by returning to the equivalent circuit of 2.02 however this time imaging that $|V_{bottom}|$ is much greater in comparison to $|V_{top}|$.

Power absorption from the RF source into the plasma is what maintains the discharge. In the plasma chamber this takes place through the coupling of RF power in 4 main ways [X1.8, 9, 10, 12]: 1) Ohmic heating: electron heating due to electron-neutral collisions in the bulk plasma 2) Stochastic heating: the principal source of electron heating in low pressure discharges is a result of electron collisions with the oscillating sheaths show below in figure 1.11 and constant source of new radicals. 3) (time-averaged) Sheath voltage acceleration of ions (The most efficient means of power coupling!) 4) Secondary electron emission (discussed in detail in the morphology section).

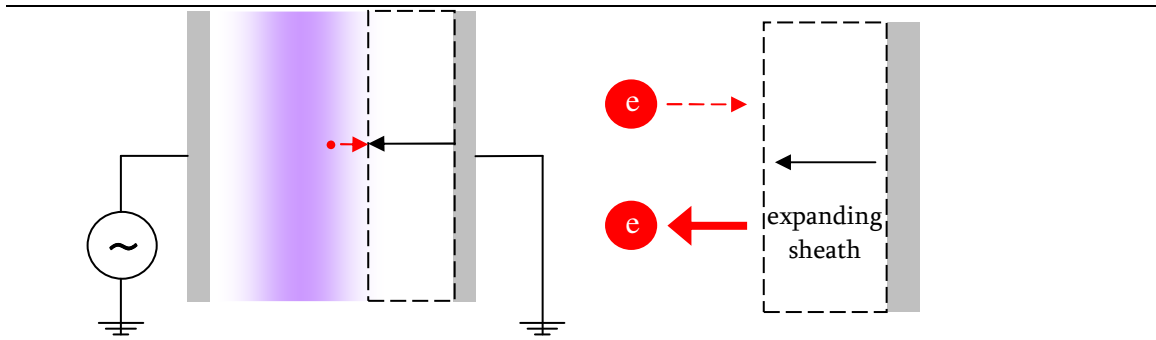


Figure 2.05: Stochastic heating [X1.2], a main mechanism of power coupling and the principal source of electron heating (the difference in temperature between ions and electrons in the processing plasma seen in figure 2.01), is a result of electron collisions with the oscillating sheaths. This sheath expansion and contraction process can be observed very well with phase-related optical emission spectroscopy (PROES), discussed later in this section

The overall point is that the RF voltage source is not only responsible for maintaining the plasma but also for dictating the chamber bias voltage (and therefore ion bombardment energy). The real resistive drop across the plasma itself will increase and decrease with plasma fluidity which is related directly to the amount of power coupled to the plasma.

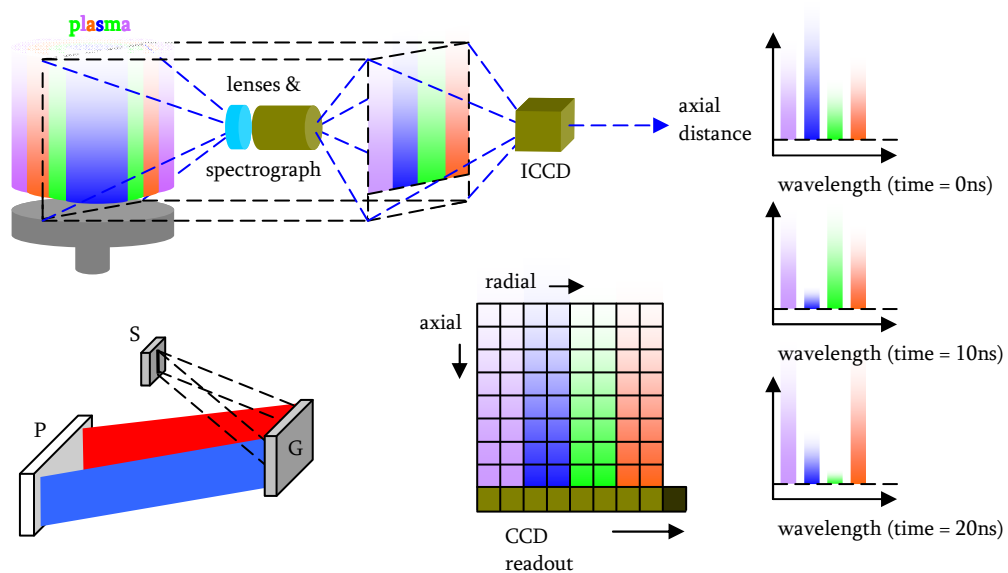


Figure 2.06: Phase-related optical emission spectroscopy (PROES) measurement with Intensified CCD (ICCD) used to monitor stochastic heating processes [X1.25-28]. A spectrograph placed in a chamber viewport with lenses used to focus the desired image plane (the center of the chamber for example) onto the entrance slit (S). The grating (G) of the spectrograph separates the light into the component wavelengths and images them onto the CCD. With a fast enough-gate rate, emission intensities can be read throughout a single 74 ns RF cycle (time = 0 ns, 10 ns, 20 ns, etc).

The unfortunate outcome is that plasma density and ion bombardment energy remain in the capacitively coupled discharge relatively dependant. We will see in the discussion of the TCP chamber exactly how this problem is overcome.

Such power coupling through stochastic heating in low-pressure processing plasma has in an asymmetrical capacitively coupled RF discharges with a negative self-bias voltage in front of the powered electrode (as we have) can be visualized with phase-related optical emission spectroscopy (PROES). The method spatio-temporally measures the modulating sheath voltage throughout the complete 13.56 MHz (74 ns) RF cycle from a few volts up to RF peak-to-peak voltage.

This shows the periodic sheath collapse and expansion in front of the wafer chuck and more importantly the oscillation of the electron front in the positive ion background. As discussed previously during the sheath expansion phase the electrons at the sheath edge gain considerable kinetic energy through the process of sheath electron heating, and those electrons with adequate kinetic energy contribute significantly to electronic excitation and ionization of species in the plasma. These spatio-temporally resolved measurements can be of emission intensities axially and radially in front of and across the wafer chuck respectively and provide insight into the locality and dynamics of the excitation processes.

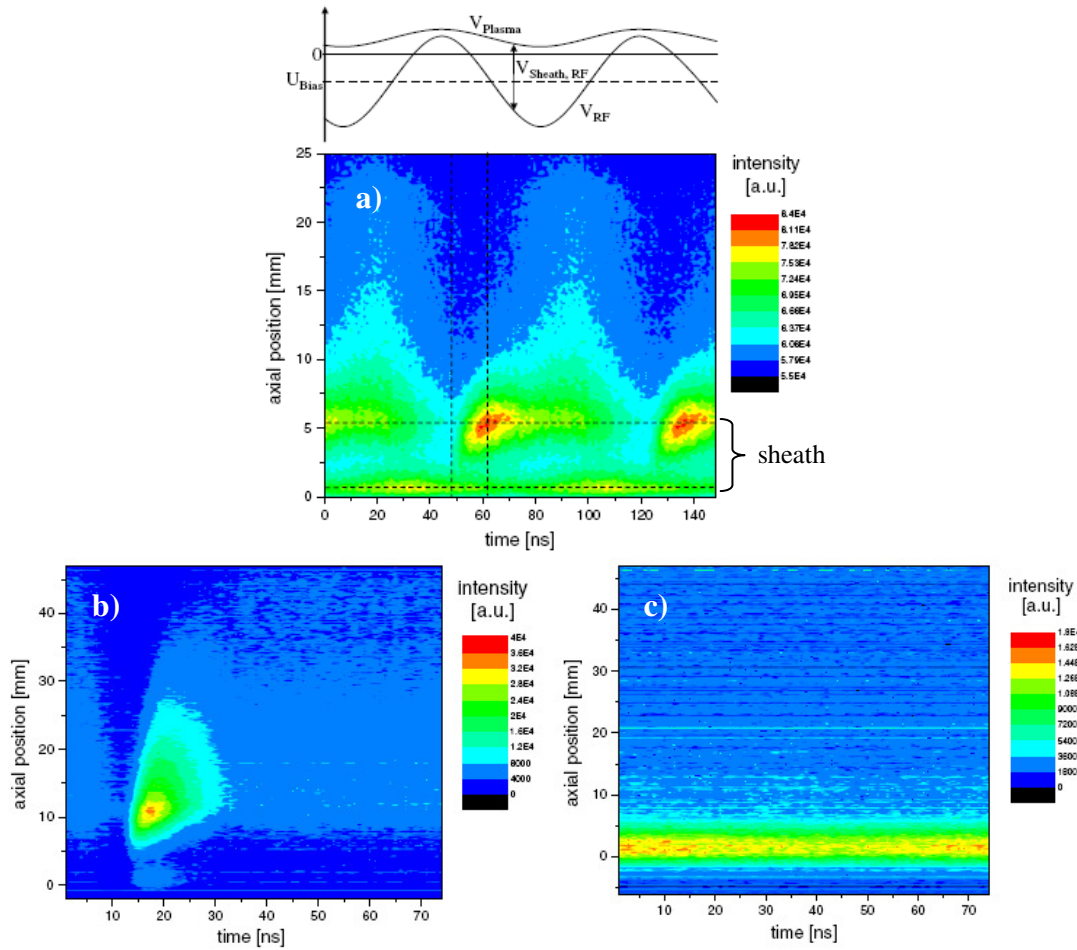


Figure 2.07: Spatio-temporally resolved emission pattern of atomic oxygen experimentally measured at a) 844 nm b) 250 nm and c) atomic carbon in CF_4 plasma at 193 nm. [X1.36]. Two RF periods are shown in part a), with the accompanying potential of the lower electrode dependant on time (this is the potential we introduced in figure 1.10, the asymmetric chamber). We see as the electrode potential swings negative electrons are accelerated away from the lower electrode, generating radicals. The main feature between parts a), b) and c) to be understood is

that the choice of observed radical (wavelength) can reveal a considerable difference in observed result. It is seen that the mean lifetime of the excited oxygen states in b) is small compared with a). The lifetime of the carbon is, at this wavelength, much longer than the RF cycle and in fact appears invariant with sheath expansion and contraction, an issue which we will address in the following chapter when CH₄ is added to our SF₆/O₂ plasma.

Performed with an ICCD Camera, shown in figure 2.06, the technique enables the user to resolve and characterize in situ plasma processes non-intrusively by optically observing the emission spectra of the plasma discharge. A spectrograph is placed outside the chamber view port, noting that a focusing lens between the view port and spectrograph is necessary to image the desired plane (the center as shown below for example) of the plasma discharge on to the entrance slit (S) of the spectrograph. The grating (G) of the spectrograph separates the light into the component wavelengths and images them onto the CCD. Emission intensities can be dynamically monitored through a single RF cycle with a high enough CCD gate-rate. Figure 2.07 shows an example of the axially resolved spatiotemporal emission pattern for atomic oxygen at 844.6 nm ($3p^3P \rightarrow 3s^3S^0$) in an O₂ rf plasma (60 W, 60 Pa). The experimental results reveal the characteristic dependence of the emission pattern in front of the powered electrode on plasma process parameters such as self-bias voltage and pressure. The x axis shows time and as the length of one period of a 13.56 MHz source is approximately 74 ns, two such periods of sheath expansion and contractions are shown.

Figure 2.07 also shows the emission of atomic carbon at 193 nm ($3s^1P^0 \rightarrow 2p^1D$) in a CF₄ rf plasma (50 W, 10 Pa). The main feature between the figures to be understood is that the choice of observed radical (wavelength) can reveal a considerable difference in observed result. It is seen that the mean lifetime of the excited oxygen states in b) is small compared with a) while the lifetime of the carbon is, at this wavelength, much longer than the rf cycle and in fact appears invariant with sheath expansion and contraction. These figures provide a very concise picture of the “real process” occurring in front of the wafer chuck (for example the plasma itself extends normally no more than a few centimeters into the chamber above the wafer chuck) and as such will be referenced numerous times in the coming chapters as we attempt to visualize the plasma process over specific points on a wafer in reference to changes in process parameters and their influence on resultant nano-structures.

2.2 Plasma etching

As previously mentioned, the nature of the high temperature electron collisions of a plasma is such that it provides an environment for the presence of not only ions for bombardment but also radicals and molecules not normally found in a simple gas. Figure 2.08 a) illustrates a simple chlorine discharge etching a silicon surface. Radicals and gas molecules diffuse to the surface and are involved in either chemisorption or physisorption processes bonding with the silicon. This bonding process produces one of two results. Either a stable passivating surface film resisting further radical etching and ion bombardment etching or a volatile gaseous etch product that diffuses away from the surface and is pumped away (or deposited on the chamber wall).

In addition to volatile chemical bonding, the bombarding ions, seen below in part b), can themselves physically sputter particles from the surface. These incident ions can also result in secondary electron emission, photon emission, or lattice displacement and therefore possibly a collision cascade resulting in the emission of photons/electrons/surface-particles at a location a distance away from the ion striking point. It is of course also possible for the ion to implant itself in the surface, and although deep ion implantation (forming an n-well for example) is usually accomplished at ion energies much higher than those found in standard plasma etching equipment, chapter 5 will discuss the possibilities of implanted ion damage to nano-structures.

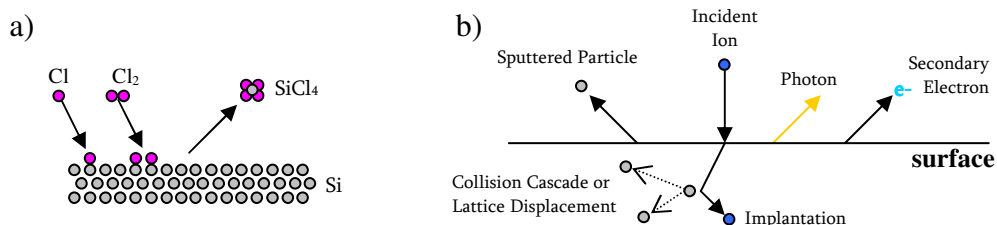


Figure 2.08: Simple a) radical (sometimes also called chemical) and b) physical plasma etching [X1.4]. In radical etching the reactive species (species without a charge!) generated by the aforementioned sheath expansion and contraction process diffuse to the substrate and bond to the surface (through chemisorption or physisorption, two topics discussed quite thoroughly in the next chapter within the context of nano-masking). Volatile products formed on the surface, normally containing silicon, are released and pumped out of the chamber. Part a) depicts this process with Cl. Charged species do not diffuse to the surface, rather they are accelerated across the sheath region and strike the substrate surface, physically sputtering surface atoms, and possibly damaging the substrate in the process. Part b) depicts this process, along with additional. Secondary electron emission and its possible role in nano-masking are discussed in the next chapter. Mechanisms of plasma damage, such as displacement, collisions, and implantation within the lattice, and their consequences for the carrier lifetimes within manufactured nano-structures are discussed in chapter 5.

The combination of both radical and ion etching combined, in figure 2.09, results in a much higher etch rate than either process alone.

Pictured in Figure 2.10, the etching cycle begins with electron collisions and the creation of radicals followed by a series of surface processes starting with adsorption and subsequently building into nucleation and island growth (masking) upon the surface of desorption in the case the generation of a volatile product.

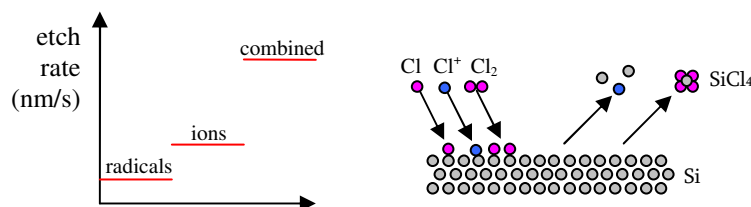


Figure 2.09: Free radical and ion etching combined [X1.4] yields the highest etch rate. This is understandable, as radicals etch the surface by forming volatile products and ions physically sputter surface atoms. Together incident ions can physically sputter surface passivation opening the opportunity for more radical etching.

Additionally there are unexpected and non trivial ion and surface interactions, not limited to simply sputter etching the surface. A complete understanding of the black silicon process begins with a much more rigorous treatment of these surface processes and passivating films and indeed such a detailed handling of some of the aforementioned etching processes will be necessarily given in the chapter on morphology when discussing the causes of nano-masking.

For now, this is a complete explanation for the (basic) plasma etching process, in reality however there is a great orchestra of processes happening within the chamber all of which are occurring continuously and simultaneously. The two last sections have armed us well with the

correct understanding and vocabulary to move on in discussing the tool implemented in this dissertation

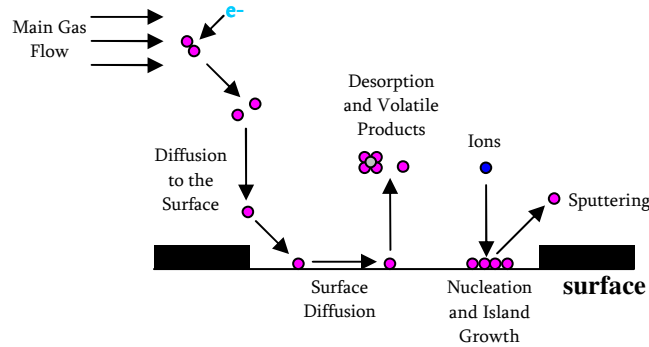


Figure 2.10: Complete picture of basic plasma etching [X1.4]. Gas enters the chamber and is broken into radicals and ions from collisions with the stochastically heated electrons. Radicals diffuse to the surface and form volatile products (etching the surface upon desorption, hence these radicals are called etchants) or form stable products (called passivation, hence these radicals are called passivants). The stable products on the surface begin to form islands overtime with the flux of more passivants to the surface. Ions physically sputter surface atoms away, including the passivation. This is important as a stable passivation is not normally as readily etched by radicals as the substrate surface atoms. This relationship between etching the surface and etching of passivation will be included in the next chapter in discussing the origin of the observed nano-structures in the dissertation.

2.3 Transformer Coupled Plasma LAM Reactor

The LAM TCP 9400 plasma etcher, shown in figure 2.11 and 2.12, consists of a load lock chamber located in a clean room, a transfer chamber, and an etch chamber with a high density TCP planar coil plasma source. The chamber is physically an asymmetric design (the bottom electrode is much smaller than the top) powered with two separate upper and lower RF systems both at 13.56 MHz. Both upper and lower systems include an RF signal generator and a match-network which provides feedback tuning for better power delivery. The RF generators deliver power most efficiently to a 50 Ohm load at 0 degrees phase error. The match systems compensate for changing chamber/coil and plasma conditions so that the generators will see only this 50 Ohm load at a 0 degree phase error. This means the total impedance of the system will equal 50 Ohms. Figure 2.13 shows a plot of the plasma and match impedances representative for an RF system such as this. If these impedances are plotted in the imaginary plane, the real resistive impedance is on the X-axis (resistance) and the imaginary phase component is on the Y-axis (phase). Here, the overall plasma impedance is a combined -20 Ohms imaginary and 25 Ohms real. Thus the compensating impedance from the match network would be +20 Ohms imaginary and +25 Ohms real. This gives the complete system a sum of 50 Ohms real and zero Ohms imaginary (0 degrees phase error). Since plasma can have a continuously changing load value based on its density (pressure) and the chemical nature of the species in the plasma (gas mixture), the match system must account for these various changes in phase and resistance. For the lower RF system the load, as seen by the RF delivery system, looks very much like capacitors (and some resistance) in series.

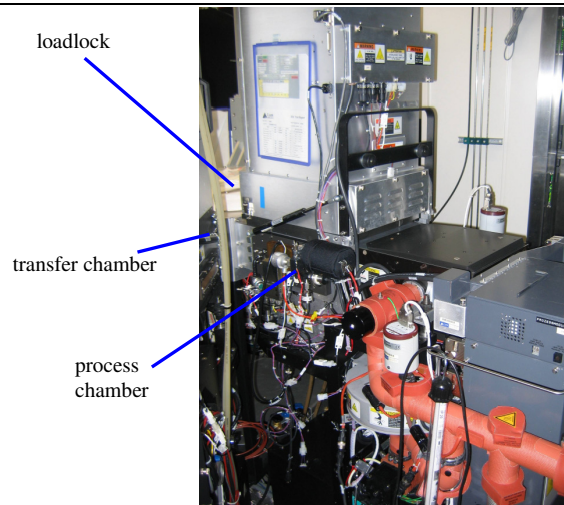


Figure 2.11: TCP system overview I (External view). Wafer cassettes are placed in the loadlock from the clean room environment. Everything behind the loadlock (process chamber, transfer chamber) is located in a maintenance area separate from the clean room. Each wafer is moved automatically to the transfer chamber and from the transfer chamber to the desired process chamber. It is possible to have as many as four process chambers connected to a single transfer chamber. This tool is not a laboratory tool. It is located in an industry clean room involved in large scale CMOS fabrication and meets the standards for tool selection mentioned in the introduction.

Since capacitors create a negative phase error (current leads the voltage) it is necessary to bring the phase error back to zero using a matching system that includes an inductor, or coil as an inductor creates a positive phase error (voltage to lead the current). The lower match on the LAM uses such a tunable coil to generate a compensating positive (inductive) phase and additional resistance to enable the sum of the system to equal 50 ohms and 0 degrees phase. Simply explained, the sensing/tuning circuitry for the RF generator sends signals to motors that adjust the position of the load coil in the matching network to balance, and thus balance out the effects of the plasma making the circuit tuned (resonant) to the RF generator. The upper RF system is a Transformer Coupled Plasma (TCP) configuration. The load, as seen by the upper RF delivery system, looks very much like one large inductor, as it is one very large inductor coil that magnetically couples to the plasma discharge through a quartz window. As opposed to the capacitive load in the lower RF system, the voltage in the large upper inductor coil leads the current to produce a positive phase shift. Again since the RF generator prefers to see a load with a resistive impedance of 50 ohms and a phase of 0 degrees, the match system must provide a compensating negative phase to balance the positive effect of the load coil and the smaller negative effect of the plasma. Capacitors provide negative phase shifts (current leads voltage), so this match uses a variable capacitor to sum all resistance and phase components in the system to 50 ohms and 0 degrees. The upper RF system is also responsible for plasma ignition which means that every TCP process must have the upper RF system on. Figure 2.14 shows a simplified circuit diagram of the upper and lower RF systems and introduces the idea of an “induced coil” in the plasma itself.

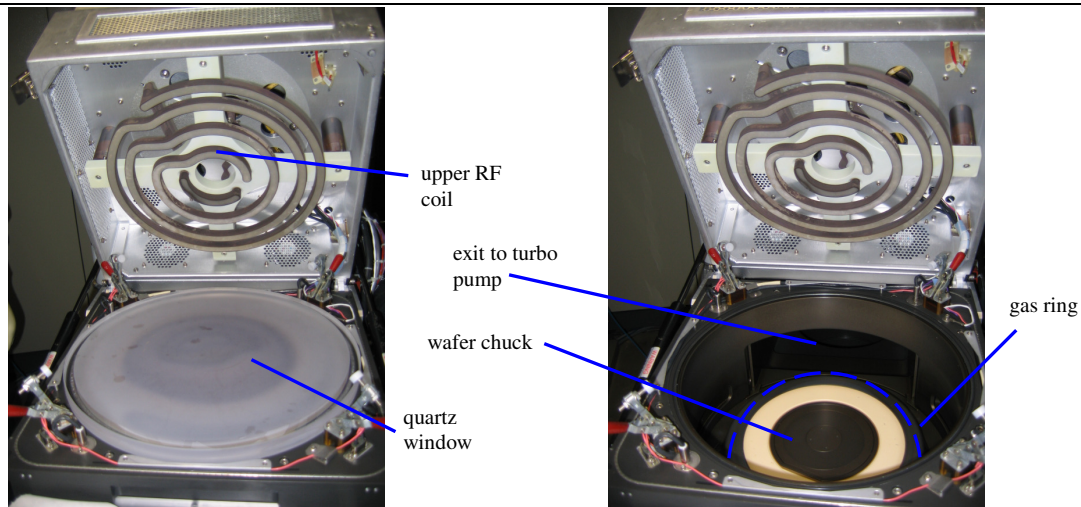


Figure 2.12: TCP system overview II (Internal view). In comparison to the basic chamber of the previous section this tool has two AC sources. The lower capacitively coupled electrode (wafer chuck) and an upper inductively coupled source, appearing as a large metal coil in the picture. This coil is separated from the processing chamber during a process by a large quartz window. Gas is delivered to the chamber from a perforated ring around the outside of the chamber at its base. This gas delivery system and upper coil geometry create challenges in maintaining nano-structure uniformity. This will be dealt with by the introduction of “nano-loading” discussed in detail in the morphology chapter.

Obviously there is no physical coil in the plasma but it is helpful to understand the notion of magnetically coupling energy into the plasma. So essentially the upper and lower RF systems are independent of each other but depend upon the same plasma.

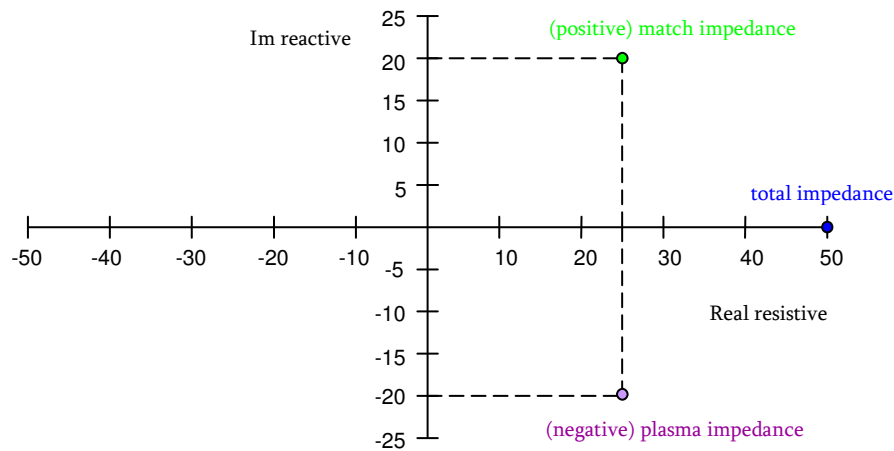


Figure 2.13: Example impedances for a Plasma RF system [X1.15]. The match systems compensate for changing chamber/coil and plasma conditions so that the generators will see only a real 50 Ohm load. Here, the overall plasma impedance is a combined -20 Ohms imaginary and 25 Ohms real. Thus the compensating impedance from the match network is +20 Ohms imaginary and +25 Ohms real, giving a sum of 50 Ohms real.

This and the “induced coil” will certainly require an explanation in more detail. As mentioned in the third section of this introduction, power absorption into the plasma from the bottom capacitively coupled source occurs in 4 main ways: ohmic heating (electron heating due to electron-neutral collisions in the bulk plasma), stochastic heating (electron heating resulting from electron collisions with the oscillating sheaths), simple sheath voltage acceleration of ions, and secondary emission of electrons from the silicon into the plasma from very close proximity positive ions to the wafer surface.

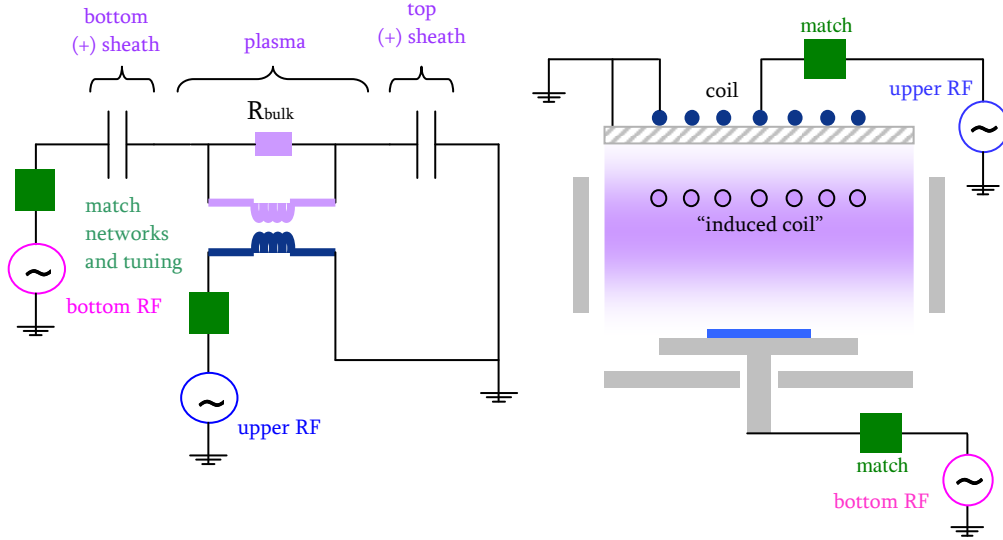


Figure 2.14: Simplified diagram of upper and lower TCP RF systems [X1.2, 16]. The equivalent circuit of the TCP plasma chamber retains the original properties of the simple capacitively coupled chamber, the upper and lower sheath are taken to be capacitors. The idealization of a negligible voltage drop across the plasma is not made. A resistor and transformer are in its place, with one side of the transformer representing the upper coil and the other side of the transformed representing an induced coil within the plasma (also pictured in the block diagram on the right hand side). Increasing the upper RF coil power increases the ionization of the plasma, and thus the etch rate, (almost) independently of chamber bias. Clearly from this picture a reduction of R_{bulk} would slightly change the chamber bias.

The problem encountered in basic plasma chamber discussion was that one RF voltage source maintains the plasma and sets the chamber bias voltage (ion bombardment energy), and the plasma fluidity (density of available species). The unfortunate outcome was that linking the chamber bias and plasma fluidity to one single RF source makes it impossible to alter one without altering the other. The upper inductive TCP coil shown will correct some of this problem to some extent.

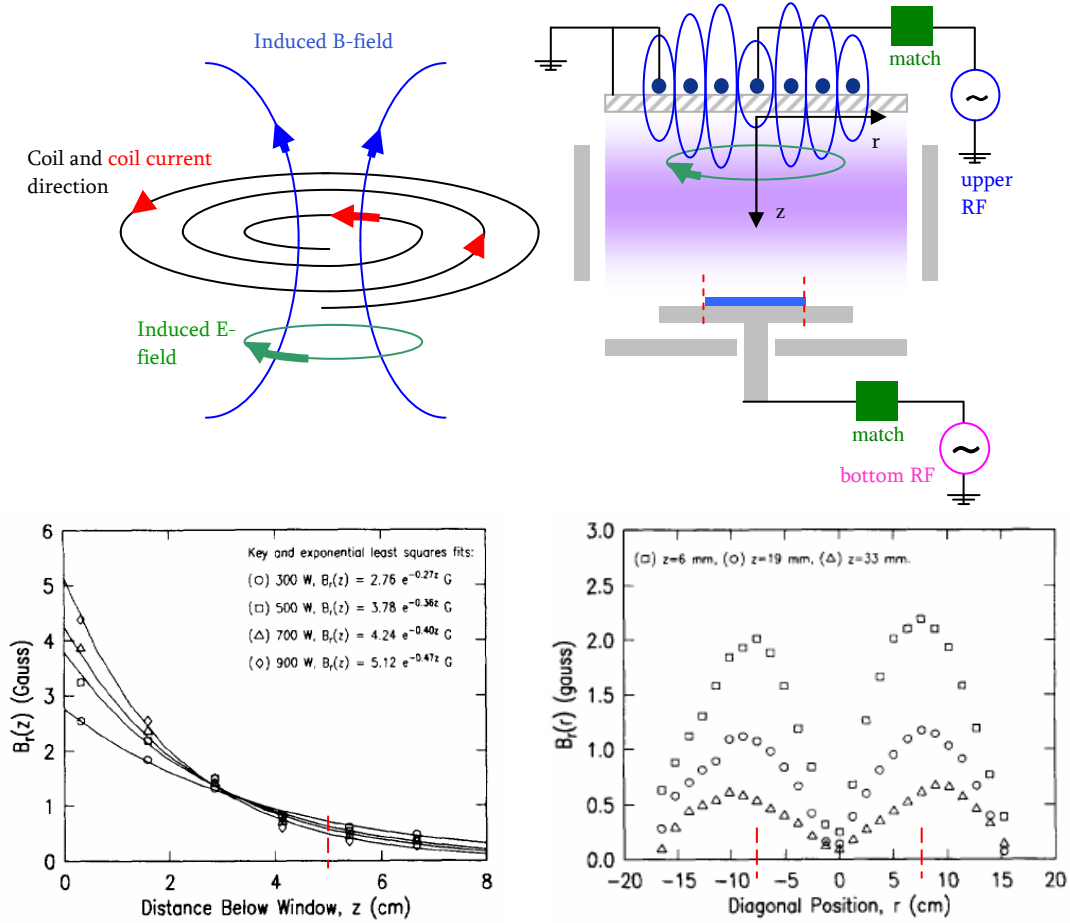


Figure 2.15: The induced B-field in the TCP chamber showing a non-uniform distribution in the z chamber direction [X1.2, 16]. The field strength below the quartz window decays with distance (bottom left figure), however when measurements are taken radially in the r direction (bottom right figure) this decay is not uniform. Although this radial non-uniformity is less pronounced with increasing distance, we believe that substantial non-uniformities in resultant nano-structures are related to this (discussed in detail in the next chapter). The approximate position of the wafer in this work is noted by the dashed red lines in the bottom two graphs.

As previously stated the TCP coil is located directly above the upper electrode, which is a large quartz window plate. The upper RF source drives a current through the coil and this current in turn induces a perpendicular B-field that penetrates the plasma [X1.17, 18, 19]. Within the plasma an induced E-field is generated perpendicular to the coil induced B-field. This is depicted in figure 2.14. This E-field stimulates electron movement essentially creating a current, the “induced coil” current, in the same direction as the coil current but within the plasma, and therefore directly couples power into the plasma through the above mentioned method of ohmic heating. Hence the name Transformer Coupled Plasma, the coil couples power into the bulk plasma as in a simple circuit where one transformer coil induces a current in the other coil non-intrusively.

As such, a greater upper RF power will create a greater plasma density (almost) independently of chamber bias and thus the ion bombardment energy remains unaffected. Obviously this increased plasma fluidity will change the resistivity of the plasma and hence the voltage drop across the plasma itself, subsequently altering the bias voltage, but for all intents and

purposes this value, as mentioned earlier, is considered small and in most cases the plasma is assumed to have no voltage drop. What is possibly problematic in regard to this increased plasma density is the possibility for non-uniform distributions of induced B-field and E-field intensities within the chamber itself as this could have significant consequences on the distribution and flux density of radicals and ions, and therefore the uniformity of surface structures, delivered to the surface. The B-Field distribution in the z and r chamber directions of a planar upper coil system is pictured in figure 2.15. It is seen then with an O₂ plasma the B-field decreases exponentially in intensity from the quartz window down to the bottom electrode however it penetrates non-uniformly in the r direction through the chamber. This very well leads one to believe that the chamber has a concentric ring-shaped E-field intensity gradient and possibly a very pronounced toroid shaped density distribution of radicals and ions. The theoretical location of the wafer in for the experiments in this dissertation is also noted on the figure.

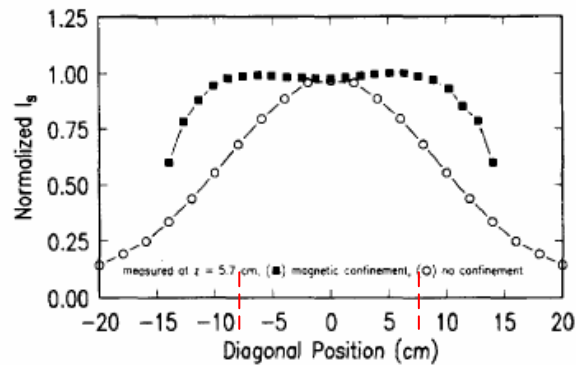


Figure 2.16: Magnetically confined and unconfined plasma uniformity [X1.2, 16]. A study measuring flux of ions (I_s , ion current) to the surface in a simple Argon discharge has shown that magnetically unconfined plasma (white circles in the graph) in an asymmetric chamber with an upper RF planar coil, the same type of chamber used in this work, has a very bell shaped distribution. Again the approximate position of our wafer is shown with red dashed lines. In view of the other non-uniformities discussed, it is not immediately evident that a greater etch rate would be observed in the wafer center. It is certainly possible that a much greater abundance of gas at the chamber edge, due to the gas ring delivery, could inhibit ion flux to the surface.

Other research has shown this to be certainly true (presented in figure 2.16). Further study using a Langmuir probe in a simple Argon discharge has shown that unconfined plasma in an asymmetric chamber with an upper RF planar coil has a very bell shaped ion distribution (like an “anti-toroid”). The chamber used in these experiments had no possibility for the addition of magnetic confinement rings to increase plasma uniformity. Note that the diameter and position of the wafer for the set of experiments included in this paper is denoted with dashed lines on both Figure 3 and 4. A visible center to edge ring uniformity signature created on unmasked silicon wafers etched in the processes to be discussed looks suspiciously like the distributions from these two previous studies.

Two further parameters influencing this center to edge uniformity are Helium backside cooling, provided to achieve a thermal contact between the wafer chuck and the sample (typically helium is used due to its inertness and thermal conductivity – see figure 2.17 part b), and the electrostatic chuck, providing an electrostatic force which clamps the wafer in place during processing. These two parameters are linked as the regulation of center-to-edge thermal uniformity of a wafer can be greatly improved with an ESC in comparison to simple clamping processes. The legs of the mechanical clamp resting directly on the wafer will render an area (the

size of which is process and tool specific) unusable. With this clamping force only on the wafer edge it is possible during plasma processing that the wafer may bow to some degree, the center is attracted (pulled up) towards the plasma. Furthermore the mechanical clamping process can be quite dirty (generating particle contaminants) in comparison to the ESC as each use of a mechanical clamp during plasma processing will result in the deposition of some film on the clamp components which may drop to a wafer surface in a subsequent process.

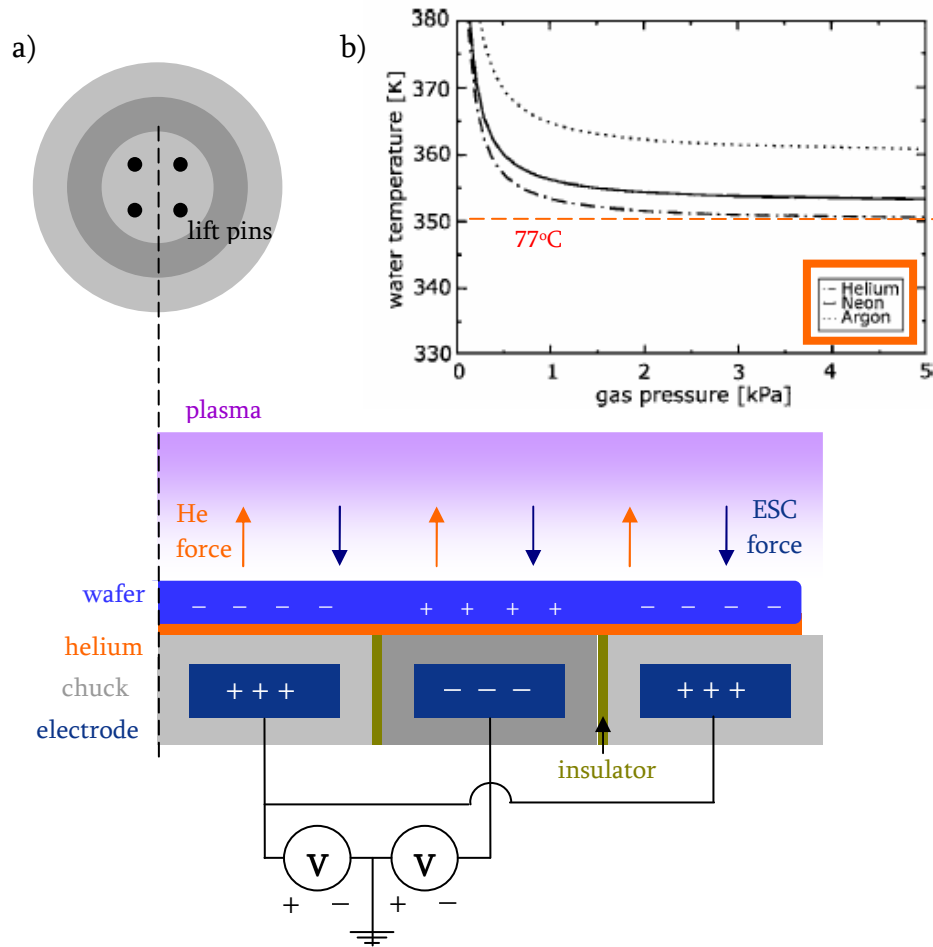


Figure 2.17: a) Bipolar electrostatic chucks eliminate the need to use the plasma as an upper ‘chucking’ electrode and thus reduce the frequency of wafers sticking to the chuck as a result of residual charge left over after the plasma is turned off [X2.22]. The force from Helium backside cooling is compensated by the ESC chucking electrical force. Lift pins attempting to raise a stuck wafer will shatter the wafer. b) As the pressure of the Helium backside cooling is reduced the temperature of the wafer becomes uncontrolled. This graph shows Helium most effectively maintaining the 77 degree wafer temperature target [X2.24].

A criticism of the ESC is its known possibility to shatter wafers after processing. A single polar chuck uses the plasma as a counter electrode. When the plasma is shut off residual charge may remain on the wafer surface causing the wafer to tend to stick in place. During the automated wafer removal process lift pins pressing against a still-clamped wafer will shatter it, dictating that the plasma etch tool be taken off line, opened, cleaned, and reconditioned before it can be returned to operation [X2.23]. The ESC used in this work is a bipolar design (see figure 2.17 part

a) which partially solves the problem by providing two electrodes with voltage applied between them.

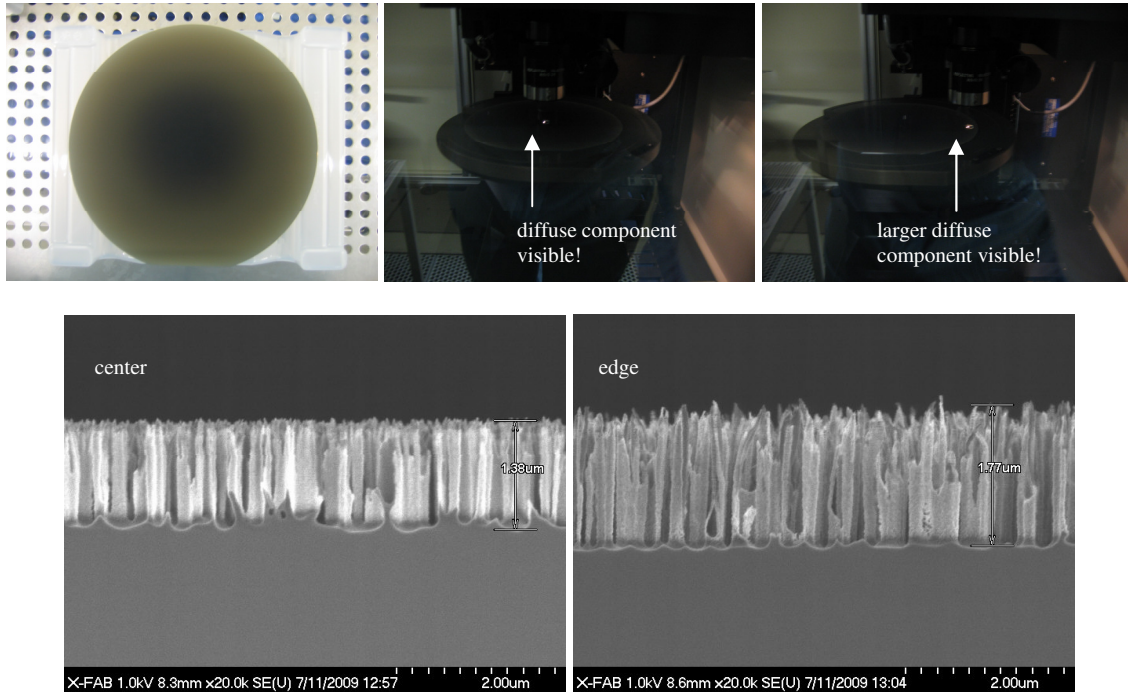


Figure 2.18: a) Top down digital camera picture of a nano-structured wafer (appearing quite brown with a blacker center). This sample is also seen on the chuck of the NanoSpec during measurement at b) the wafer center and c) the wafer edge. (Bottom) Corresponding SEM pictures of center and edge are shown. Notice that the measurement beam is visible at the wafer center (and even more so at the edge) during measurement.

The force from Helium backside cooling is compensated by the ESC chucking electrical force. The necessity of using the plasma as an upper electrode for ‘chucking’ and will reduce the frequency of wafers sticking to the chuck (and thus lift pins shattering the stuck wafer while attempting to raise it) as a result of residual charge left over after the plasma is turned off [X2.22].

2.4 Reflection measurement and characterization tools

Reflection measurements were performed using both a Nano-spec film thickness measurement tool and an integrating sphere (also known as an Ulbricht sphere). A comparison of the two instruments with regard to measuring black silicon shows very important differences and a brief description of both tools is crucial for understanding the reflection results. This is especially important given the necessity to characterize black silicon etched in a process flow quickly and accurately. NanoSpec machines are a manufacturing standard in all semiconductor foundries while the integrating sphere is a slightly more laborious process and certainly not available within a production flow.

The NanoSpec is normally a tool used for measuring the thickness of thin (10 to 4000 nm) optically transparent films on silicon wafers. Since these film thicknesses are comparable to the wavelength of the incident (unpolarized) measurement light, usually 700 nm, the tool function is based on the fact that the reflected intensity of this monochromatic beam will depend strongly on the film thickness and the corresponding interference between film and substrate. From a broad band source, a Deuterium lamp for wavelengths between 190 nm and 350 nm and a Halogen lamp for wavelengths between 350 and 800 nm, the tool utilizes a grating

monochromator to create the incident beam and a photomultiplier tube detector to measure the reflected optical spectrum from a bare silicon internal reference chip and from the wafer being tested. Taking the index of refraction for the thin film (user specified) and the two measured spectrums, the computer will evaluate the interference pattern to establish the film thickness.

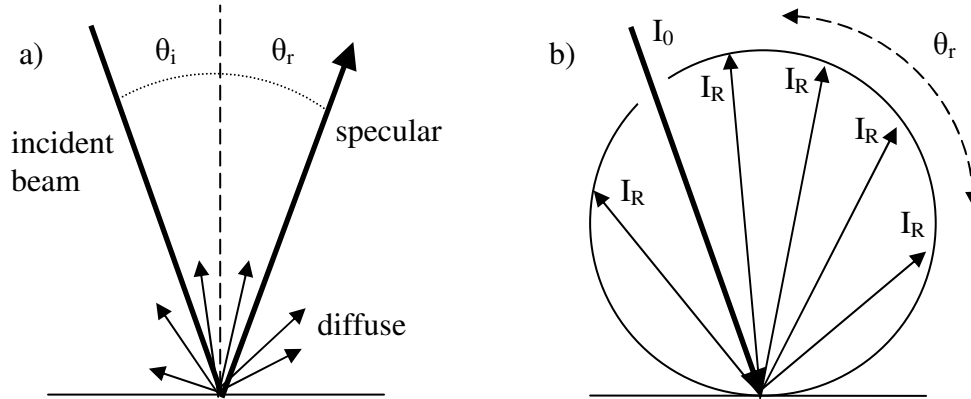


Figure 2.19: a) Specular reflection where the angle of incidence equals the angle of reflection ($\theta_i = \theta_r$) and diffuse reflection where the angle of reflection is not correlated to the plane of incidence ($\theta_r = ?$). b) The integrating sphere where reflected intensity from an incident beam I_0 is evenly distributed across the entire sphere interior meaning the reflected beam intensity I_R is the same at all values of θ_r . There are no unique specular and diffuse components (all angular properties of reflection are eliminated).

The tool can also very easily measure total reflectance of light from the surface by simply comparing the measured intensity of light at its detector with the intensity that was initially sent. For relatively flat surfaces this is not a problem unfortunately, in figure 2.18, we can observe a rather problematic phenomenon. In a) we see a top-down digital picture of a nano-structured wafer and the same sample being measured inside the NanoSpec in b) at the wafer center and c) at the wafer edge. The measurement laser is clearly visible (the normal measurement spot size is $30 \times 30 \text{ um}$) on the wafer surface from our point of observation outside the plane of incident, in violation of Snell's Law. For this beam to be visible the surface must be reflecting light at angles other than the angle of incidence. This is good example of the difference between specular and diffuse reflection, both shown in figure 2.19 a).

Specular reflection is mirror-like behavior in which light from a single incoming direction is reflected into a single outgoing direction (with respect to the surface normal, the angle of incidence equals the angle of reflection, commonly stated as $\theta_i = \theta_r$). Diffuse reflection is the light reflected from a surface that one would describe as rough, uneven or granular.

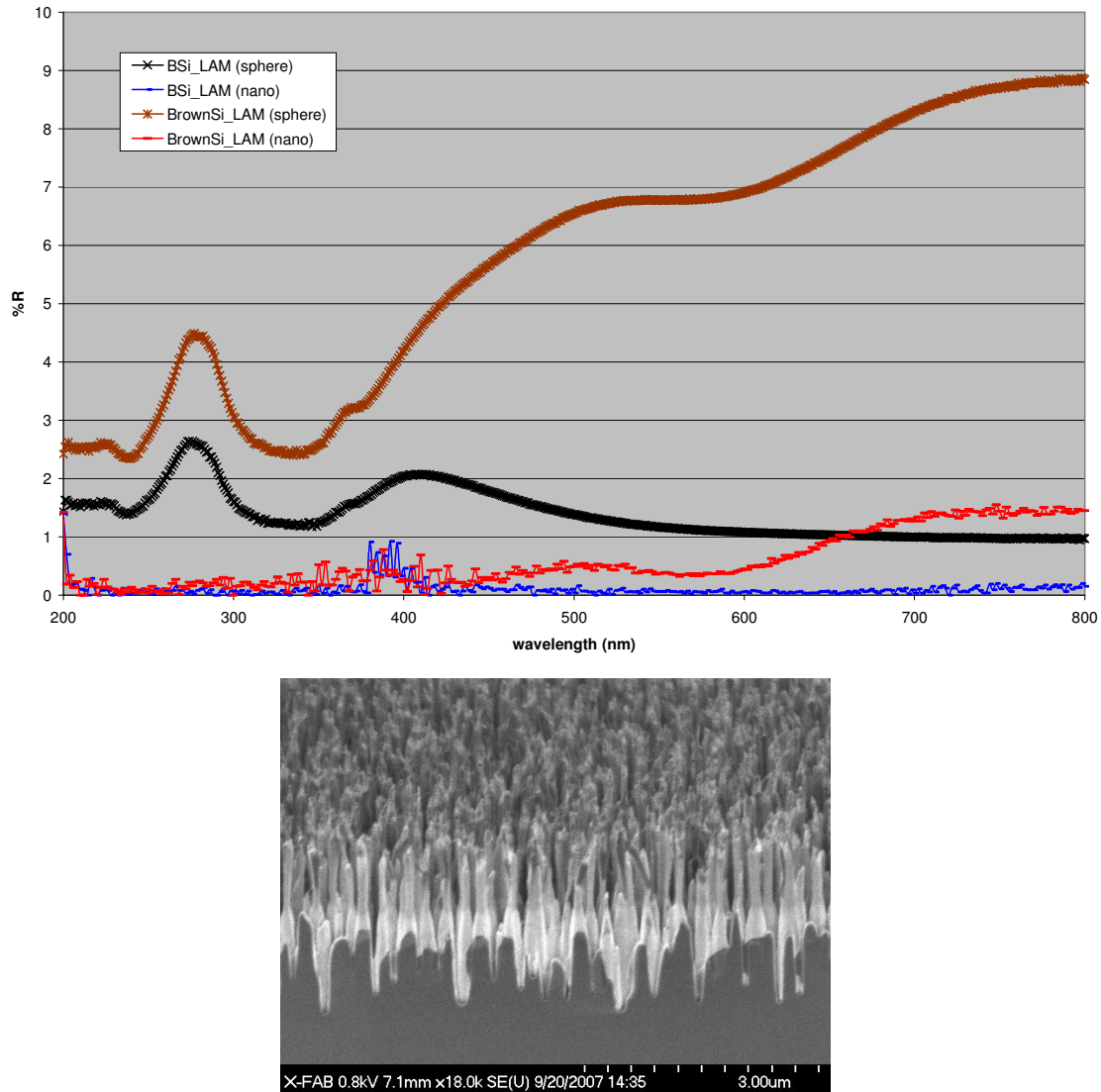


Figure 2.20: Integrating sphere reflection (both specular & diffuse components combined) versus NanoSpec reflection. The nano-structures of the edge site in the previous figure, labeled here “Brown SI”, are compared to the nano-structures of the sample shown in the SEM above, labeled here “BSI”. The latter showing smaller reflection values in both machines with the black line representing the integrating sphere measurement and the blue line representing the NanoSpec. The differences between the two samples are more difficult to see in the NanoSpec measurements (blue and red lines), as the diffuse component is not present. Chapter 4 will discuss the geometry of nano-structures and their corresponding optical properties in more detail, however we can already see here that the structures of the previous figure are slightly more anisotropic and have a very clear “floor”.

The result is such that a beam incident on the surface is reflected at a number of angles, as we are observing here with the measurement beam radiating off the structured wafer surface in directions not in the plane of incidence.

A further point with diffuse reflection which would certainly increase absorption and given the relatively large height of the nano-structures in reference to the wavelengths of

incidence is called diffuse interreflection. This is a process in which light reflects from one nano-structure, is reflected, strikes another nano-structure, and is reflected with the process repeating but each time with a reflected and transmitted component. However in all cases characterizing and measuring this diffuse component of light with an integrating sphere is absolutely necessary as the diffuse component will be incorrectly measured by the NanoSpec as light absorbed by the surface.

An integrating sphere is specifically made for just such a measurement. Made of a hollow sphere with its interior coated for Lambertian reflectance, the light incident on the interior of the sphere is scattered such that the intensity of reflected light is the same in all directions. This removes any unique specular and diffuse components, essentially all the angular properties of reflection are removed. The intensity from the first reflection off the sphere interior which illuminates every other location on the sphere would simply be the input intensity divided by the surface area of the sphere. An illuminated surface location is also reflective and will again reflect light evenly to all locations of the sphere interior. The relationship to determine the intensity of reflected light at a measurement port then becomes highly manufacturer specific as the calculation will depend on the size and shape of all non-reflective surfaces (ports, detectors) within the system.

To characterize the specular and diffuse components of a sample a port is opened in the sphere at the at the point of initial incidence within the sphere and replaced with the sample, exactly as is done here with a Cary 5000 spectrometer and its accompanying integrating sphere. The incident beam is unpolarized (although one could use a polarizer). For wavelengths between 175-350 nm the source is a Deuterium lamp and for wavelengths between 350-3300 nm a Mercury lamp is used. The beam itself is not exactly continuous, a chopper mirror (@ 30 Hz) diverts the beam for 10 ms to the sample and then 10ms to the reference blocking the beam for 10 ms. The angle of the incident light is 3° for the system when measuring only specular reflection and 0° (completely perpendicular to the surface removing the specular component) when using the integrating sphere to measure diffuse reflection. The two components can then be combined to give the (real) total reflection of the sample.

The activity of interest is to compare the values of both tools, NanoSpec and integrating sphere, and discern whether the loss experienced by the NanoSpec is significant. Figure 2.20 shows a comparison between the total reflection of two samples measured with the integrating sphere (specular plus diffuse) and the simple direct reflection measured with the NanoSpec. The two samples chosen were a black wafer, the corresponding SEM is shown and is labeled “BSI” in the graph, and a brown wafer, the edge site of the SEM sample shown in figure 2.18 and labeled “Brown SI” in the graph. Both samples were made in the TCP. The “BSI” sample shows a smaller reflection value in comparison to the “Brown SI”, regardless of machine used, with the black line representing the integrating sphere measurement and the blue line representing the NanoSpec. However the differences between the two samples are more difficult to see in the NanoSpec measurements (the blue and red lines) as the diffuse component is not present. Chapter 4 will discuss the geometry of nano-structures and their corresponding optical properties in more detail, however we can already see here that the structures of the previous figure are slightly more anisotropic and have a very clear “floor”.

Figure 2.21 shows that the relationship between the integrating sphere measurement and the NanoSpec measurement of the same sample is completely dependant on the sample itself. This is not entirely strange, as a sample which is shown to have a near zero specular and diffuse reflection in the integrating sphere should have a NanoSpec measurement which correlates very well to the integrating sphere. Since the sample has a near zero diffuse component, both machines are essentially measuring the same property and the error arising from any diffuse component not measured by the NanoSpec is negligible, as there was very little diffuse component to begin with. This does not however render the NanoSpec useless.

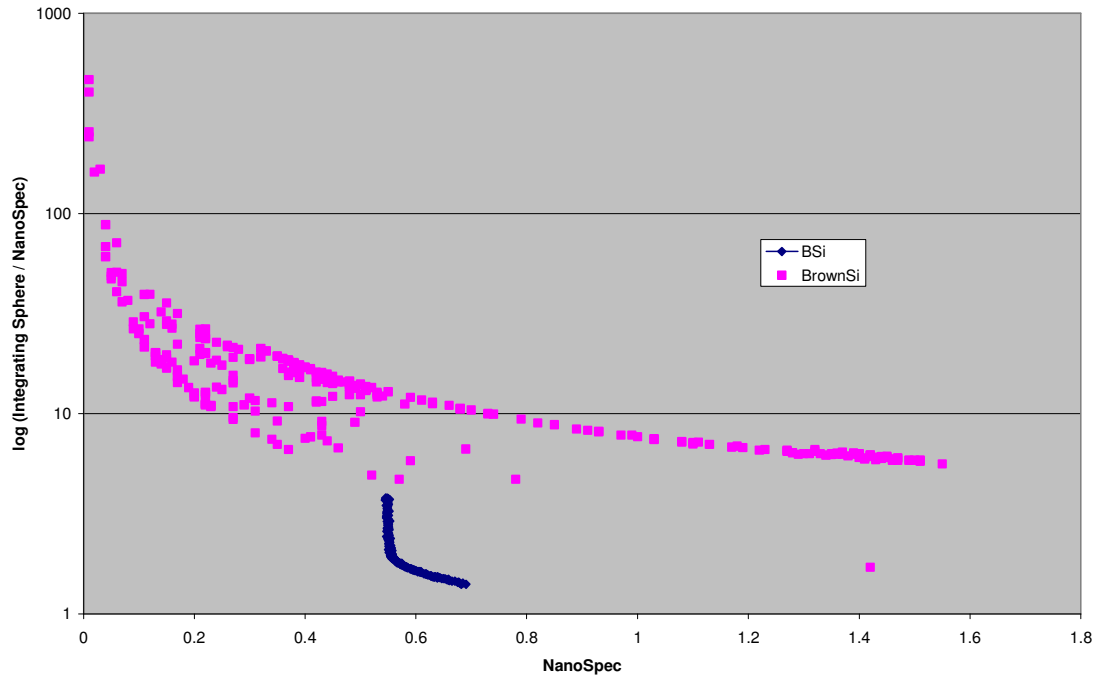


Figure 2.21: The relationship between the integrating sphere and NanoSpec measurement is sample dependant. Saying it another way, a sample measured in the integrating sphere that is shown to be nearly the ideal black color (having a near zero specular and diffuse reflection) will have a corresponding NanoSpec measurement which correlates very well to the integrating sphere measurement. This is because there was never any diffuse component loss in the first place.

A NanoSpec measurement that is non-zero does indicate that the specular component is non-zero. Additionally a NanoSpec measurement that is zero indicates that the structures are at least worth SEM investigation to evaluate the geometry of the surface. It will also be shown in chapter 3 that the NanoSpec is quite valuable in evaluating nano-structure uniformity across a wafer. This method of NanoSpec + SEM was invaluable in evaluating structures created in the work here as both tools were (and are normally) readily available in the CMOS industry environment. The integrating sphere is slightly more laboratory-oriented and was located at the University. Being able to (at least to a first approximation) evaluate the properties of manufactured structures on site is very convenient, especially when dealing with engineering.

2.5 Concluding comments

This chapter has been meant to give the reader a firm understanding of plasma for semiconductor processing from the process engineering stand point, the various aspects of the LAM TCP etch tool, and the tools used to characterize the optical properties of the nano-structures in the next chapters. The topics introduced in this chapter will be revisited and built upon in greater detail in the next chapters.

3. Morphology

Morphology could be succinctly described as the study of a given structure's form, such as its dimensions in terms of heights, widths, or angles. However, with regard to nano-scale structures morphology is not simply a mundane cataloguing of the physical properties of size and shape, it can dictate very strongly the electrical, chemical, and optical properties of structures (transport, wettability, and index of refraction for example). It is quite evident that such additional properties of nano-scale structures, in contrast to those of large substrates, would be strongly connected to form, as such structures are achieved often by self-assembling processes governed by the interaction of the atomic and molecular arrangement of deposited materials, directional nature of chemical bonds, and precise location of sputtered or passivated material, with resultant structures measured on scales not entirely different than those of the interactions which created them.

The framework of this dissertation separates the morphology of created nano-structures into two sections: Section 3.1, a theoretical discussion of the probable causes of nano-masking observed in this work, and section 3.2, a detailed description of experimental processes and parameters influencing structure size, shape and uniformity. Additionally, section 3.2 will reference the ideas and theory presented in section 3.1 in an attempt to understand plasma etching from the position of physical chemistry and not simply the arbitrary variation of tool parameters.

3.1 Theory of nano-masking and the proposed origin of our nano-structures

The needle phenomenon discussed in this dissertation will be concisely split into three sub-sections: Section 3.1.1, dealing with the role of plasma radical absorption, desorption, and passivation at the vacuum/substrate interface, section 3.1.2, treating the importance and influence of the interaction between incident ions and the substrate surface on absorbed species (not simply physical etching). Finally in section 3.1.3 a discussion of the SF_6/O_2 plasma discharge etching silicon is presented and combined with previous sections to generate a general picture of nano-masking leading to structure formation.

3.1.1 Influencing a nano-mask: Radical absorption, desorption, and passivation

Absorption and the reverse process desorption make up the set of reactions involved in the chemical etching or passivating of a surface in a plasma by radicals, and thus the very beginnings of nano-masking. Absorption itself is separated into two different categories, physisorption and chemisorption. Basic absorption can be thought of as due to the weakly attractive van der Waals forces between an atom (or molecule) and a surface, namely those of permanent dipoles and their corresponding induced dipoles, pictured in figure 3.01, or those of the London dispersion forces, instantaneously induced dipoles exhibited by nonpolar molecules as a result of probabilistic electron density movement about a molecule. As shown the dipole is taking to point from the negative charge to the positive charge and the oscillating dipole moment at $\pm x_1$ direction is then given by

$$p(t) \approx (r_{Bohr} e) \cos(\omega_{at} t) \quad (3.01)$$

where e is the unsigned electron charge and ω_{at} is the atomic frequency (the electron orbiting the nucleus). This means the electric dipole vector is in the direction from the negative charge to the positive charge. Standard electric field line notation from negative to positive is indicated in the opposite direction, but there should not be a contradiction as an electric dipole has to do with charge position and not the field lines themselves. From Coulombs law the force is attractive [X3.5]

$$F_x = -\frac{6 \langle p^2(t) \rangle}{4\pi\epsilon_0 (2x_1)^4} \approx -\frac{3 r_{Bohr}^2 e^2}{4\pi\epsilon_0 (2x_1)^4} \quad (3.02)$$

The van der Waals potential is found from $F_x = -e(dV/dx)$

$$V(x) = -\frac{r_{Bohr}^2 e^2}{64\pi\epsilon_0 (x_1^3)} \quad (3.03)$$

One might conclude that the attractive force experienced between the atom and image charge will pull the atom into direct contact with the surface.

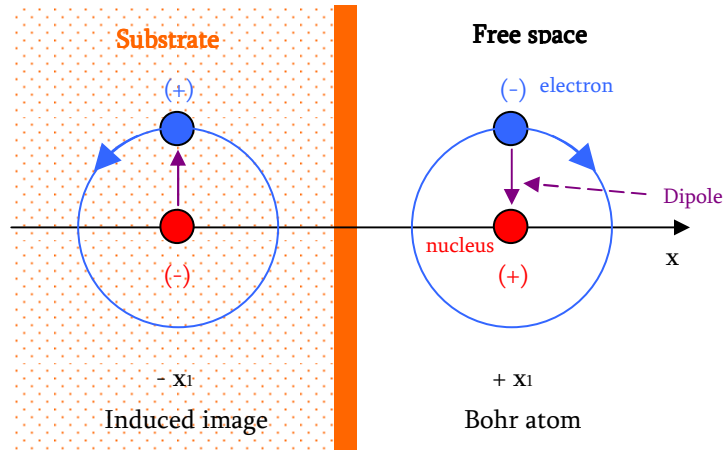


Figure 3.01: Van der Waals force between an atom and a surface [X1.2]. Notice the opposite polarity of the induced image charge within the substrate. Attractive forces between the image and the atom hold the atom in place (this is basic absorption).

However when the atom comes within a distance of $\sim 1-3 \text{ \AA}$ the clouds of the atom and the surface interact repulsively leading to a potential well at the surface, show in figure 3.02. Both chemisorption and physisorption can then be understood by looking at these potential energy curves for an atom, or better yet a molecule AB, in reference to a surface \S along our x axis. Considering first the dissociative chemisorption in a) at a great distance from the surface \S the elements A and B exist in a lower energy state together as a single molecule AB than as two separate elements $A + B$, the difference shown by the energy $E_{diss}/2$. Following the solid line of AB in the direction of \S we see that AB at first enters a lower (stable) potential energy state, E_{physi} in the physisorbed state. However past a certain distance in the direction towards the surface \S the lowest potential energy position is E_{chemi} where, without any necessary addition of external energy above the zero mark (only a brief reduction in AB's internal molecular potential), molecule AB separates into elements A and B and these chemisorb separately to the surface, hence the title of

a), dissociative chemisorption. The difference between a) and b) should then be quite evident. Again we see in b) at a large distance from § the elements A and B exist in a lower energy state together as a single molecule AB and again reach a lower energy state by physisorbing to the surface. However past this point, to reach the chemisorbed state, would require the addition of extra energy E_{ads} . In a situation such at this the molecule will very likely remain where it is and is said to be simply physisorbed. Lastly we see in c) that molecule AB itself can move through the physisorbed state into the chemisorbed state without the excess amount of additional energy E_{ads} . Markedly then, physisorption represents bonds with greater bond lengths (between 150 and 300 pm) and lower potential energies ($E \approx 0.01$ to 0.25 V) while chemisorption represents the opposite, smaller bond lengths and higher potential energies (typically 100 pm to 150 pm and $E \approx 0.4$ to 4 V).

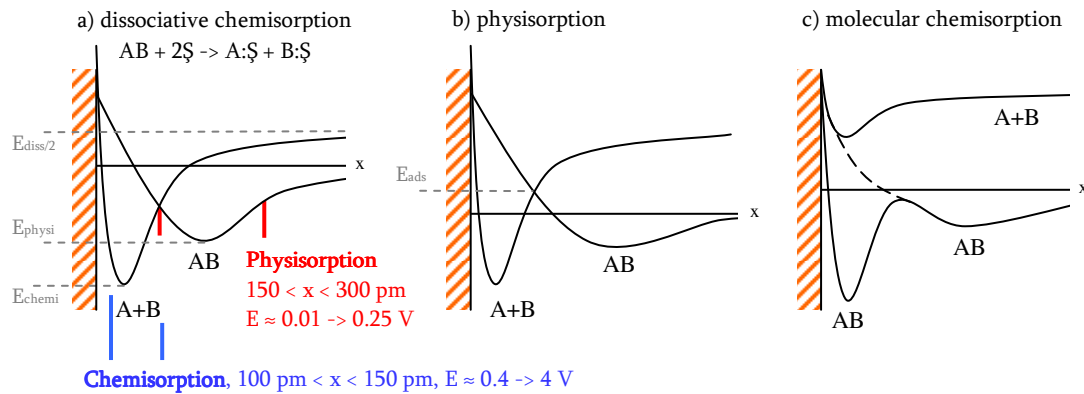


Figure 3.02: Potential energy curves of physisorption and chemisorption [X1.2]. The electron clouds of molecule AB and of the surface atoms interact repulsively leading to a potential well at the surface. The main point is that the addition of external energy, E_{ads} , (present in part b but not in part a) may be necessary to facilitate the chemisorbed state for some species (and the formation of their corresponding volatile products). See text for a complete detailed description of parts a), b) and c).

It is now possible to consider the SF_6 absorption/desorption radical etching process in more detail, pictured in figure 3.03. An SF_6 gas molecule is struck by a high energy electron in the bulk discharge and broken into an SF_2 molecule, four Fluorine atoms, and an electron. Subsequently a free F atom diffuses to the surface and is physisorbed as $\text{F}:\text{Si}$ corresponding to the physisorbed state shown in the graph. With the addition of energy E_{ads} the F atom is then chemisorbed as SiF_x to the silicon surface. This energy can come in the form of an absorbed photon (possibly from a secondary electron emission process discussed in the next section) or thermal energy absorption (through the chuck or plasma heating of the surface, discussed in the context of figure 3.17). The process of F atom diffusion and absorption repeats until the volatile etch product SiF_4 is formed.

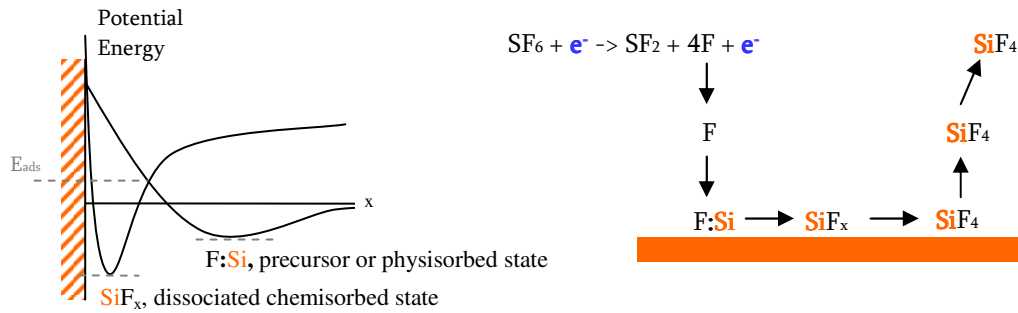


Figure 3.03: Potential energy curves for free radical etching [X1.4]. A free F atom, created by a collision between SF_6 and an electron, diffuses to the surface and is physisorbed as F:Si. With the addition of energy E_{ads} (possibly from secondary electron emission in the next section or thermal energy absorption, see figure 3.17) the F atom is then chemisorbed as SiF_x to the surface. This process continues until the volatile etch product SiF_4 is formed.

3.1.2 Influencing a nano-mask: The interaction of ions and the wafer surface

As mentioned earlier ions can sputter material from a substrate by physically striking the surface atoms with enough force to knock the atoms free, however a more complex explanation is necessary to understand the manner in which this process and the process by which the energy states of elements absorbed on the surface are modified by incident ions. Additionally the presence of these ions near the surface to be etched result in their own neutralization by the surface and occasionally the expulsion of additional electrons by the surface into the plasma further coupling power into the plasma and thus further sustaining the plasma. The latter two processes are called ion neutralization and secondary electron emission respectively, and can also effect the energy state of absorbed elements on the sample surface.

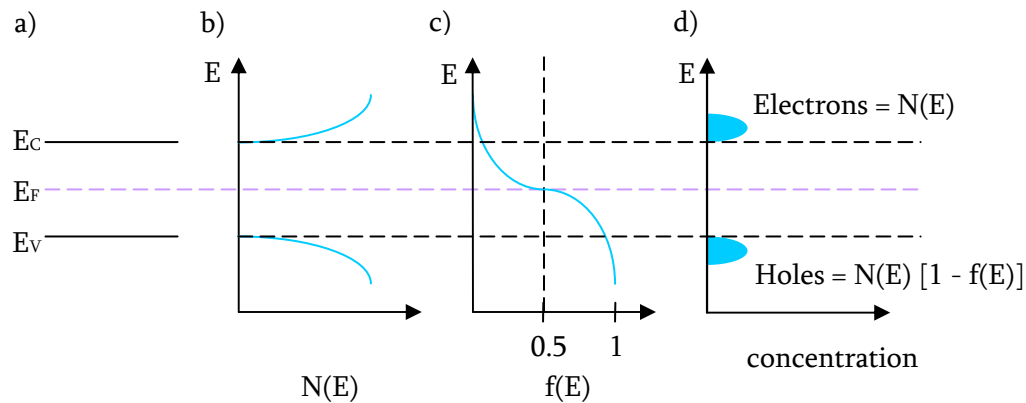


Figure 3.04: Fermi-Level of intrinsic silicon [X4.1] a) the schematic band diagram showing the conduction band E_c , Fermi-level E_F , and valence band E_v of silicon. b) The quantum mechanical density of states (not of carriers!). c) The Fermi-Dirac distribution at some non-zero temperature and d) the carrier concentrations. The band gap, the region between E_c and E_v , contains no electron energy states and hence no electrons.

It will be necessary to remind the reader of the Fermi-level of the silicon surface, pictured in figure 3.04, for our further description of potential energy levels at the surface. In a) the

schematic band diagram showing the conduction band E_c , Fermi-level E_F , and valence band E_v of silicon. Part b) shows the quantum mechanical density of states (not of carriers!). Part c) shows the Fermi-Dirac distribution at some non-zero temperature and d) the carrier concentrations. The band gap, the region between E_c and E_v , contains no electron energy states and hence no electrons. Placing the Fermi-level of the substrate next to free space (and considering the vacuum plasma chamber to be free space) we attain the first step in creating a model for the release of electrons from the surface. We see this illustrated in Figure 3.05. As we did in considering the notions of absorption and desorption, we again consider the idea of induced image charges. The difference is that in absorption we discussed the attraction of an atom to the surface within the neighborhood of a Bohr atomic radius to the surface, while here we first consider separating an electron from the surface by having an electron within the neighborhood of a Bohr atomic radius to the surface.

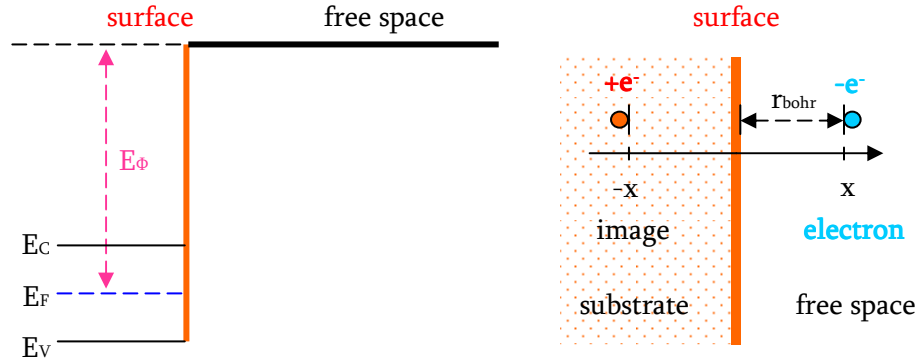


Figure 3.05: Potential barrier between the silicon surface and free space [X1.2] to determine the work function of the surface and subsequent release of electrons from the surface. Electrons are confined to the solid by a potential barrier equal to the silicon work function E_Φ . The attractive force on an electron a Bohr radius outside the surface also represents the attractive force holding an electron a Bohr radius within the surface which will be taken as an approximation to the work function E_Φ .

The electrons are confined to the solid by a potential barrier approximately equal to the silicon work function E_Φ (depending on their position – conduction, valence etc). From Coulomb's law we find F_x , the force between the image and the electron

$$F_x = -\frac{e^2}{4\pi\epsilon_0(2x^2)} \quad (3.04)$$

Subsequently, including our assumption that this attractive force F_x on the electron a Bohr radius outside the surface also represents the attractive force holding an electron approximately a Bohr radius inside the surface, the work function E_Φ , the work done moving an electron from within the surface to a point at infinity away from the surface, is

$$e E_\Phi = -\int_{r_{Bohr}}^{\infty} F_x dx \quad (3.05)$$

and evaluating we find,

$$E_{\phi} = \frac{e}{16\pi\epsilon_0 r_{Bohr}} = 6.8 \text{ V} \quad (3.06)$$

The work function E_{ϕ} for Si is normally taken to be between 4.52 eV and 4.9 eV [X2.1]. Equation 3.06 yields a close enough result.

With the introduction of the Fermi level and calculation of the work function of silicon complete it is now possible to visualize what occurs when a positive ion is present in close proximity to the surface. From figure 3.06 we observe a phenomenon known as a quantum (potential) well. When a positive ion is within an atomic radius of the surface, bending of the Band diagram can take place. This results in a lower overall escape potential for the surface electrons, especially those in the conduction band. We see that the normally required electron energies E_{e1} and E_{e2} for the two pictured electrons to surmount the potential barrier between free space and substrate will be reduced by some energy E_b .

Under normal gas phase conditions the strongly exothermic ion neutralization reaction:



is totally forbidden because momentum and energy are not both conserved during the creation of one body from two. However, directly near a surface, the possibility exists for a three body neutralization reaction to occur in one of three ways, where A^* is an excited atom:



These reactions are quite fast resulting in the neutralization of all positive ions at the surface. Most interesting is that the origin of the neutralizing electron e_1 is not from the plasma, rather from the surface itself, seen tunneling toward the ion in figure 3.06. The effect of electrons directly traversing the very thin barrier width created by the close proximity positive ion, without the normally required sufficient energy to surmount it, is called quantum mechanical tunneling. [X2.1].

3.1.3 Influencing a nano-mask: Chemistry of the fluorine plasma

There are a total of 47 main reactions given in the Ryan/Plumb SF_6/O_2 silicon plasma etch model including all main electron impact dissociation processes which occur during the active sheath expansion into discharge along with all main association reactions involving silicon are (when silicon is present in the discharge region) [X1.21]. Fortunately the authors of this study reduced the overall process to a series of only 6 small reaction sets that accurately describe the complete SF_6/O_2 silicon etching process based on their calculated reaction rates. The main gas phase reaction set for an SF_6/O_2 plasma is illustrated in figure 3.07. The model is quite straightforward. Reaction 1 is the decomposition of O_2 gas into two oxygen atoms by means of a high energy electron collision. Reaction set 2 shows the decomposition and subsequent recombination of SF_6 with decomposition again through high energy electron collisions, however here it is important to note that the creation of reactive free F atoms for the system is primarily the result of SF_6 breaking into SF_2 and 4 free fluorine atoms. Reaction 3 is the replacement reaction of gaseous SF_2 into gaseous SOF , while reaction set 4 represents the transition of this SOF molecule into SOF_4 , all products of which remain in the gaseous phase. Finally, like reactions 3 and 4, reactions 5 and 6 represent the replacement reaction and subsequent transitions of purely gaseous phase reactants.

As discussed previously for the processes of absorption and desorption, in an SF_6 plasma F atoms are transported to and successively fluorinate the silicon surface. The Ryan/Plumb model agrees with what has been already considered in this dissertation to be standard fluorine radical etching of silicon through the reactions

1)	Physisorption	2)	Chemisorption	3)	Desorption
	$\text{F}(\text{g}) + \text{S} \rightarrow \text{F}:\text{S}$		$\text{F}(\text{g}) + \text{SiF}_x(\text{s}) \rightarrow \text{SiF}_{x+1}(\text{s}),$		$\text{SiF}_4(\text{s}) \rightarrow \text{SiF}_4(\text{g})$
	$\text{S} = \text{Surface}$		$x = 0, 1, 2, 3$		

The above three phenomena, physisorption, chemisorption, and desorption have already been treated in some detail. The Ryan/Plumb model leaves the basic silicon etching unchanged. Fluorine is physisorbed and subsequently chemisorbed to the surface. At the point the volatile product SiF_4 is formed, a desorption reaction takes place and the product is removed. The key point of the Ryan/Plumb model was in the demonstration of reactant concentration as a function of distance away from the silicon surface being etched. This is depicted in figure 3.08. All the products of the 6 main reaction sets are clearly observed. One notices, highlighted for better intelligibility, the concentration of fluorine etchant and oxygen passivant and their variation in concentration near the surface depending on gas ratio.

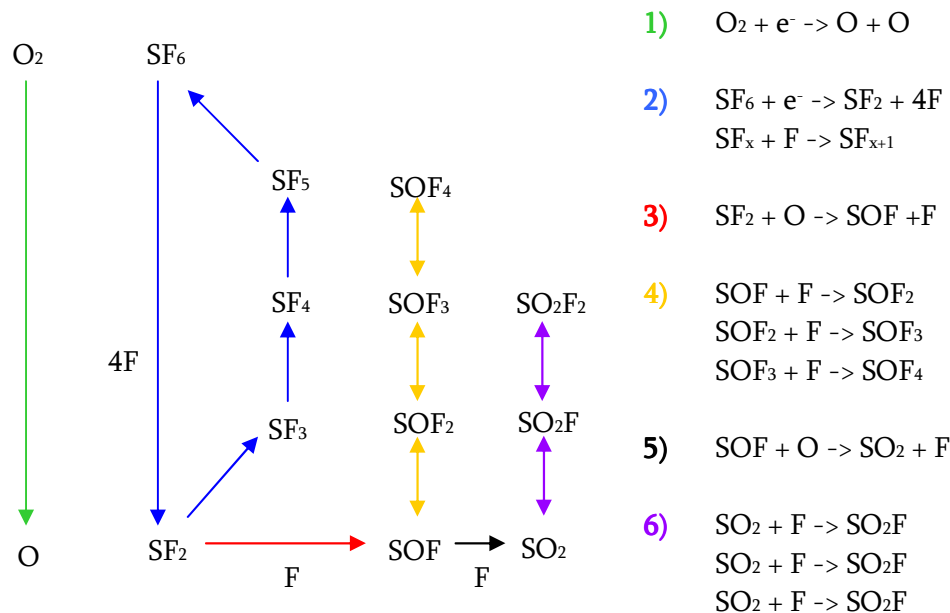


Figure 3.07: Main gas phase reaction set for SF_6 / O_2 plasma [X1.21]. Reaction 1 is the decomposition of O_2 gas into two oxygen atoms through high energy electron collisions. Reaction 2 is the decomposition SF_6 through high energy electron collisions, however here it is important to note the creation of free reactive F atoms. Subsequent recombination of SF_6 is also shown. Reaction 3 is the replacement reaction of gaseous SF_2 into gaseous SOF , notice the creation of another free reactive F. Reaction 4 is the transition of SOF molecule into SOF_4 . Reactions 5 and 6 represent the replacement reaction and subsequent transitions of purely gaseous phase reactants. Notice the creation of yet another free reactive F in reaction 5.

This concept will be exploited numerous times in following sections as we texture the silicon surface. The implications of reactions 3 and 5 from figure 3.07 will complete our plasma chemistry picture.

This final detail is added by introducing the actual macroscopic etch rate of silicon in the presence of the main gas phase reaction set for SF_6/O_2 and products just discussed. Shown in figure 3.09 are the results of a study in a CF_4/O_2 plasma, the presence of carbon in this study will become of paramount importance later in section 2.2 when carbon is introduced as a second passivant, and as is seen in the figure CF_4 combustion yields the same set of single replacement reactions occurring in SF_6 making it, for now, simple enough to generalize both the SF_6 and CF_4 gases as the fluorine radical source and the oxygen as passivation source. In general it is no problem to assume that black silicon will appear somewhere on this locus of points (as indeed it will be shown). The right hand side of the figure shows the etch rate of silicon dependant on the fluorine radical concentration. Beginning at the point (0, 0) and traversing the curve while holding the total gas flow constant (for example always 100 sccm) it is seen that as the percentage of oxygen is increased the etch rate begins to drastically increase with it. This is labeled region 1 in the figure and is a result of the now present oxygen atoms in the plasma combining with fluorocarbons (or fluorosulfates) in single replacement reactions and thus yielding additional fluorine radicals to take part in the etching. Region 2 sees an even greater increase in the number of fluorine radicals however the etch rate begins to substantially drop off as a result of oxygen radicals, now numerous in number, competing with the fluorine radicals for surface sites with which to react. Region 3 sees yet a further drop in etch rate, eventually approaching the etch rate of silicon in pure oxygen plasma, and now a significant drop in fluorine radical concentration.

The plasma in region 3 shows fewer properties of a fluorine containing plasma, such as radical etching of silicon, as the plasma is with each step less and less composed of a fluorine containing gas whatsoever.

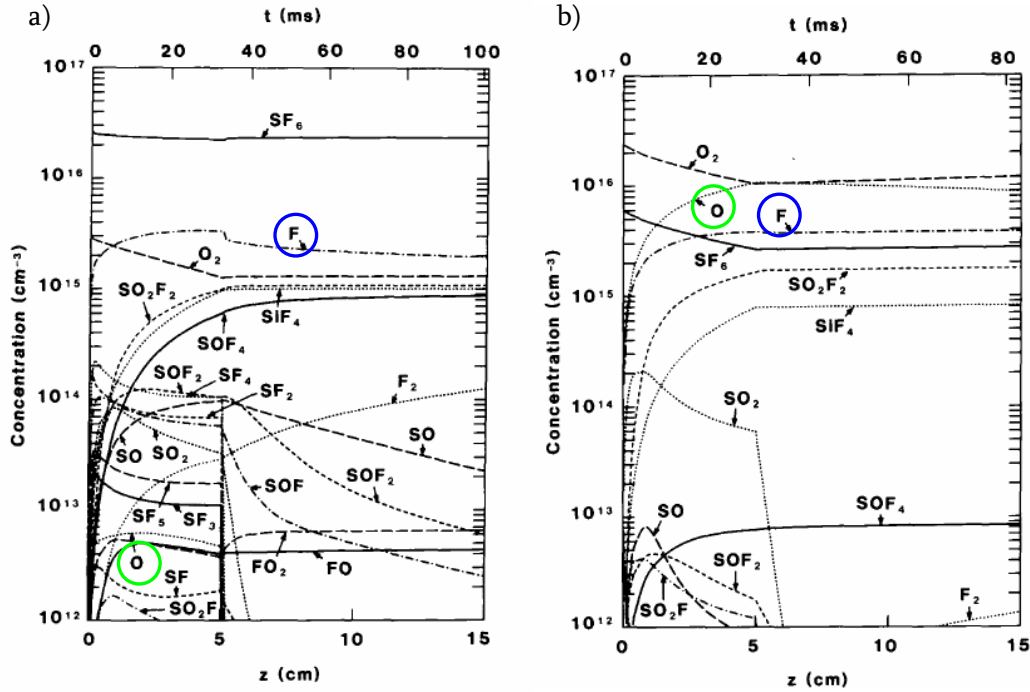


Figure 3.08: Computed concentrations with silicon samples for a) 90% SF_6 / 10% O_2 plasma b) 20% SF_6 / 80% O_2 [X1.21]. The concentration of the main species of the gas phases reactions shown in figure 2.07 with regard to distance and time away from the surface, with the concentration of radical (etchant) fluorine and radical (passivant) oxygen highlighted for clarity. Notice the drop in concentration of etchant and massive increasing in passivant for the second chemistry. This will be utilized in following sections for surface texturing.

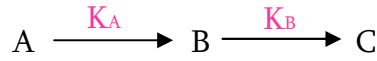
Having looked at the species involved it is now possible to make some theoretical predictions about how and from where the nano-masking can arise. Thus taking these reactants and products into account the relation of a surface being etched in a plasma discharge to the transient kinetics and the reaction rates involved can be proposed. Consider the consecutive time-varying 1st order reactions

$$\frac{d}{dt}n_A = -K_A n_A \quad (3.11a)$$

$$\frac{d}{dt}n_B = K_A n_A - K_B n_B \quad (3.11b)$$

$$\frac{d}{dt}n_C = K_B n_B \quad (3.11c)$$

Overall these equations forming the creation of a product C from a reactant A but dependant on the intermediate product/reactant B and altogether represented by the following reaction:



At the beginning of a process, for example a reaction in a standard plasma processing chamber, it is clear that we can let n_A the concentration of reactant A equal n_{A0} its initial concentration and self evidently then let n_B and n_C the concentration of product/reactant B and final product C respectively equal to zero as they have not yet been created.

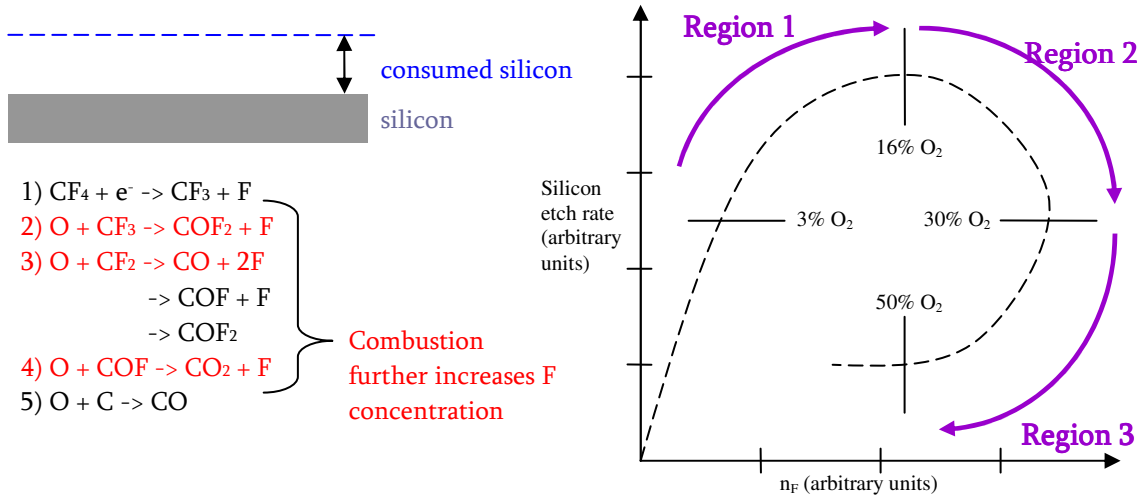


Figure 3.09: Modified silicon etch rate in fluorine radical containing gas with the addition of O₂ [X1.2]. With the addition of oxygen to a CF₄ plasma, combustion yields the same type of single replacement reactions occurring in SF₆ which further increase the concentration of fluorine. The graph shows the etch rate of silicon dependant on F radical concentration. Three regions have been noted. In region 1 oxygen is added to the CF₄ plasma and the concentration of F increases dramatically as does the etch rate. In region 2 the amount of oxygen has further increased, again increasing the concentration of F, however the decrease in etch rate is attributed to the abundance of oxygen in the plasma and thus the abundance of oxygen at the silicon surface competing for surface sites. In region 3 the percentage of oxygen in the gas mixture is so great that both the F concentration and etch rate decrease.

Thus, integrating equation 3.11a, the simple time-varying concentration of n_A , we obtain the time-varying concentration of reactant A

$$n_A = (n_{A0})e^{-K_A t} \quad (3.12a)$$

Substituting the resulting equation 2.11b into equation 2.11c and integrating we further obtain

$$n_B = n_{A0} \left(\frac{K_A}{K_B - K_A} \right) (e^{-K_A t} - e^{-K_B t}) \quad (3.12b)$$

And subsequently substituting equation 2.12b into 2.11c we find

$$n_C = n_{A0} \left[1 + \left(\frac{K_A}{K_A - K_B} \right) (K_B e^{-K_A t} - K_A e^{-K_B t}) \right] \quad (3.12c)$$

Taking the three above equations and selecting two different values for each rate constant K_A and K_B we obtain the graphs, pictured in figure X below. In part a) we select K_A to be larger than K_B and in part b) we select the reverse. From the graphs it is clear that the two different assumed values for the rate constants create two very different reaction situations.

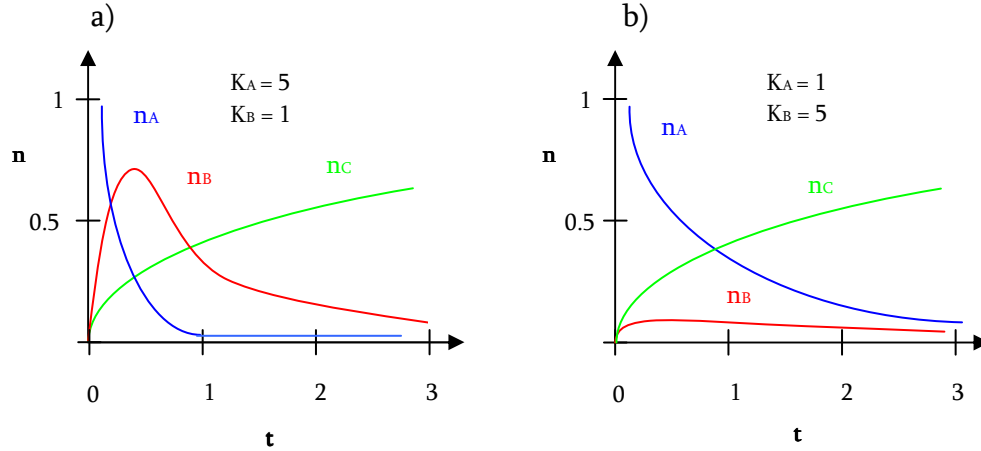


Figure 3.10: Time-varying densities of a process with a) $K_A > K_B$ and b) $K_A < K_B$. If we consider n_A to be an initial surface reactant concentration (blank silicon), n_B to be an intermediate surface product concentration (initial passivation), and n_C to be a final surface product concentration (etched silicon) we can see part a) as a description of nano-masking leading to nano-structures and part b) as standard surface etching. K_A and K_B would be determined primarily by plasma chemistry, thus in part b) we imagine ourselves working in region 1 or the beginning of region 2 of figure 3.09 and in part a) at the end of region 2 and into region 3.

This is of particular interest if we continue to assume that the overall reaction set represents a process in a plasma chamber. For example if we take n_A to be a gaseous reactant concentration, n_B to be an intermediate surface product/reactant concentration, and n_C to be a final surface product concentration. In part a) the gaseous reactant A is rapidly used and as a result its concentration n_A decreases quite quickly between time 0 and 1. This creates a rapid rise in the product/reactant B and obviously results in a similar increase between time 0 and 1 of its concentration n_B . During this time between 0 and 1 it's clear that the system is also creating product C, but the reaction rate K_B which creates product C is so much slower than rate K_A that it's not until after time unit 1 that we begin to see a significant increase in the concentration of product C. Furthermore because n_A is (nearly) completely used up after time unit 1 no new product/reactant B will be created. This means the system will (appears to) balance out to a constant concentration of product C as the remainder of product/reactant B is used. Part b) shows that, just by reversing the two reaction rates, we obtain a completely different set of circumstances in the assumed plasma processes. The primary significance is that because rate K_A (the first rate) is the slower of the two rates this will leave the slow creation of product/reactant B as the bottle-neck of the complete process, since product C is very quickly created from any available product/reactant B. Essentially rate K_A (the creation of product/reactant B) will be the rate limiting step and subsequently the creation of product C will occur at the same rate as the

consumption of reactant A. This is straightforwardly seen in the basic inverse relationship between concentration n_A and n_C .

With only a few modifications it is possible now to expand the above described general plasma process reaction directly to a passivating and etching process resulting in the nano-structuring of a silicon surface, pictured in figure 3.11. It should be apparent already to the reader that one modification in the above reaction described is certainly that in a true plasma chamber the concentration of a reactive gas n_A is not likely to ever decrease as new feed gas is always being pumped into the chamber.

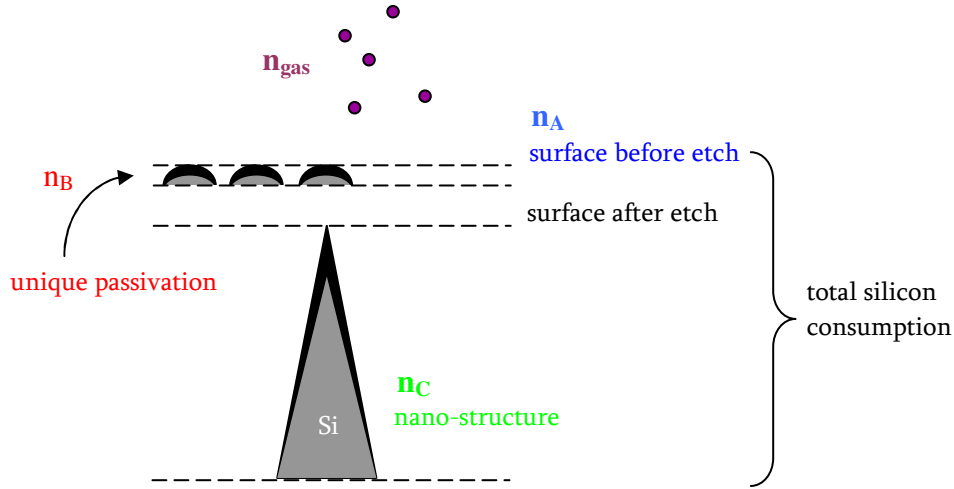


Figure 3.11: Surface nano-structuring process. Graphical representation of figure 3.10 part a).

In this nano-structuring process it is much more logical to assume that reactant A is in fact the pure unetched silicon surface. So this process could be interpreted as follows: rate K_A is the rate at which the surface is passivated, turned from pure silicon into nano-islands, and K_B is the rate at which pure silicon is etched (implicitly assumed here that little to no passivation is etched – resulting in a very small difference between the surface height before and after etch). Important to remember, as mentioned in the introduction, is that the overall silicon consumption is of tremendous importance and as pictured in the above figure, it would not be the difference in height between the surface before and after etch, rather it would be the difference in height between the surface before etch and the base of the nano-structure. In effect the size of the actual structure must be accounted for as it is itself imbedded into the photodiode well.

Pictured in figure 3.12 is a modified version of figure 3.09 combined with a modified version of the previously discussed main gas phase reaction set, together providing a model to describe the appearance of nano-masking. The pure silicon surface of concentration n_A is passivated creating the small nano-island etch mask, which is the product/reactant concentration n_B , on the surface very quickly. This process K_A is done much more quickly than K_B , the etching process of the pure silicon (around the islands). After the reactant concentration of available unetched silicon n_A is completely used up the etching process of product/reactant n_B becomes the dominant process and the system stabilizes to a constant “concentration” of fully etched needle-structures n_C .

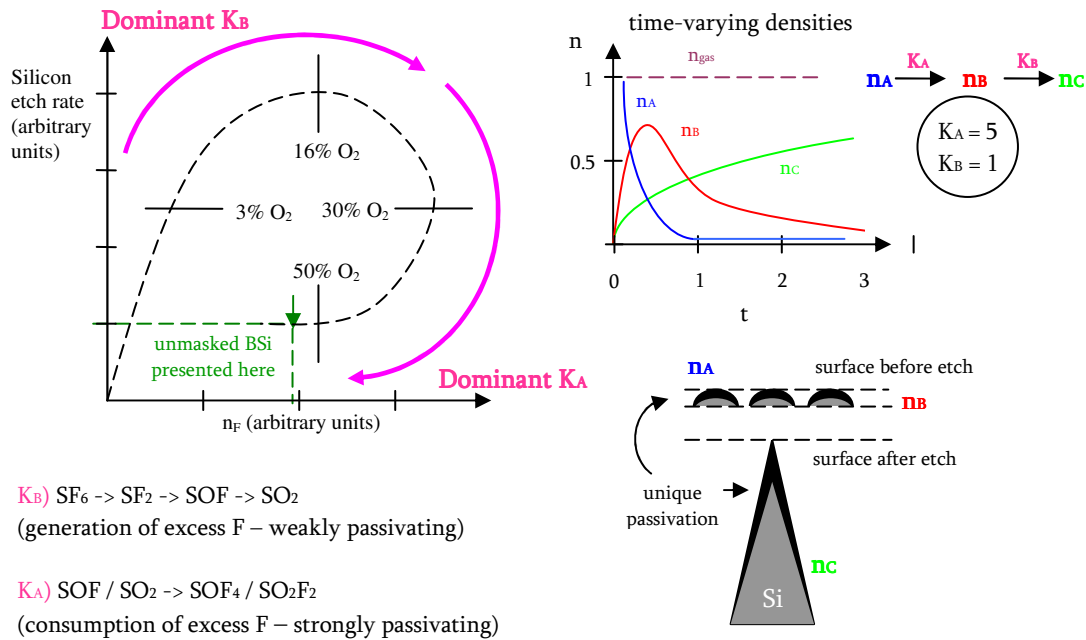


Figure 3.12: Compiled theory of nano-masking and nano-structure appearance. We generalize the 3 regions of figure 3.09 into K_A and K_B dominant regions. With a dominant K_A the reactions of figure 3.07 which consume excess fluorine are considered dominant (also building an abundance of the surface product n_B), the time-varying densities of this situation are shown again for clarity. With a dominate K_B the reactions of figure 3.07 which generate excess fluorine are considered dominant. We conclude from this that nano-structures should be observed in the dominant K_A region, and indeed in the next section this will be experimentally verified.

The graph of silicon etch rate in fluorine radical containing gas is split into two regions, a highly etching and highly passivating regime (without a really hard value between the two). The dominant K_B chemistry region of $SF_6 \rightarrow SF_2 \rightarrow SOF \rightarrow SO_2$ with a generation of excess fluorine radicals (thus only weakly passivating) and the dominant K_A region of $SOF / SO_2 \rightarrow SOF_4 / SO_2F_2$ with a strong consumption of excess fluorine, greater dominance of O_2 and thus much more strongly passivating. Furthermore, this set of curves could indicate several predictions (which will be validated in section 3.2.3) about nano-structure shape and possibilities for texturing. The first is that sharply defined nano-structures (very anisotropic) should appear in the early part of the K_A region, while more tapered (pyramid-like) structures should appear later on in the K_A region, both dependent on the rate of at which nano-islands appear and further passivation takes place in comparison to the rate at which the silicon around the islands and passivation is etched. This and all the ideas presented in this section will be revisited in throughout the following sections. In summary here the attempt was to intentionally speak about the plasma chemistry without introducing generalized notions of available process parameters such as chamber pressure or upper electrode power because it is clear that an experienced user will be able to modify the various machine parameters of their given tool to yield the desired chemistry (or something as close as possible).

3.2 Theory into reality: Silicon sample processing

The general process begins with a wafer cassette (up to 25 individual wafers) being placed into the loadlock of the LAM etch tool. The loadlock is then pumped down to vacuum pressure. A single wafer can then be transferred to the etch chamber from the cassette via the transfer chamber. The etch chamber itself is the LAM 9400 Alliance TCP process module providing, in theory, a uniform, planar, low-pressure, low-bias, high-density plasma for sub-half-micron etch control and high etch rate requirements. Other features include: separately-controlled and automatically-tuned top and bottom radio frequency (RF) power for independent control over ion energy and ion density, low-pressure chamber vacuum generated by a 1000 liter per second turbo pump, and an electrostatic chuck (ESC) with helium backside cooling (a topic that will be covered in detail here) to hold the wafer during processing with no mechanical clamps. Once a cassette is loaded automatic indexing of wafers is provided by the system software. Wafers are selected for processing from the left and right loadlock stations by the robotic arm inside the transfer chamber, which subsequently transfers the selected wafer for alignment, centering, and finally hand-off to a 4-pin lifter in the process chamber. After processing wafers are automatically returned to the loadlock which remains at vacuum until the user executes an open door command, which vents the loadlock, opens the door, and allows the user to unload the cassette after all the wafers in the cassette are processed.

Initial etch process trials used SF_6/O_2 plasma, and later the addition of CF_4 , which as described in the previous section populates the surface of the silicon with a process dependant distribution of (sub)nano-sized passivation sites while simultaneously beginning the etching of the surface. This subsequently creates a disparity in local sub-nanoscale etch rates across the wafer and results in the phenomenon referred to as nano-masking and the observed self-aligned property of the black silicon surface. Indeed these initially sub-nanometer disparities in local etch rates are the reason for the black silicon appearance at all. Section 3.2.1 deals with the structuring of unmasked silicon wafers with various pillar and pyramid geometries made with single and multi-step etch recipes and furthermore addresses the process problems by introducing an extremely novel solution to uniformity and sculptability utilizing a nano-scale loading effect which is, to the author's knowledge, not seen anywhere in the literature. Section 3.2.2 treats the issue of masked wafers and the problems encountered in transferring the recipes to wafers with much smaller open areas of silicon.

3.2.1 Texturing silicon into nano-structures, part I: Nano-loading

For silicon etching it is well known in the literature, without ever needing to provide a formal theoretical exposition of the root cause, that in fluorine-based plasmas (SF_6) the addition of an oxide-forming (O_2) or fluorocarbon (CF_4) inhibitor can create the conditions for nano-masking and the appearance of nanoscale structures [X2.7]. First experiments were here performed on unmasked blank 6 inch (15.24 cm) silicon wafers and showed very little tolerance for process variation with the needle phenomenon seeming to be within only a very specific window containing just the right balance of pressure, gas flow values, and temperature. Once the needles were observed very little variation in the above mentioned parameters (and additionally the upper and lower RF power settings) was possible to improve uniformity and structure density. Only the time parameter was observed as flexible and related directly to nano-structure height.

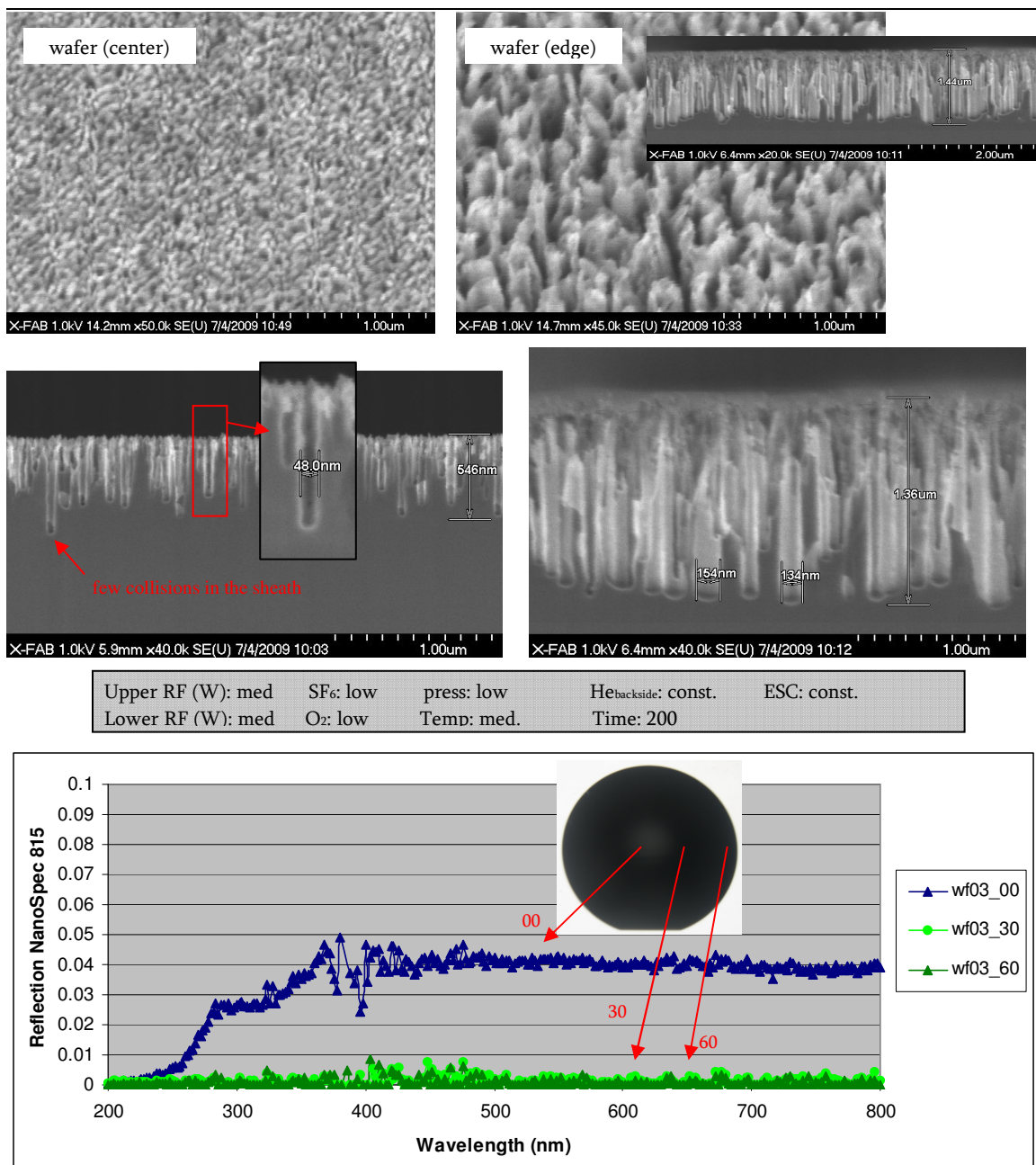


Figure 3.13: Structures etched without a separate nano-masking step showing direct reflection from black silicon and visible non-uniformities. Although discussed in much greater detail later, the reader is reminded that these structures appear black in the direct reflection measurement as most light is scattered into diffuse directions (noise at 400 nm is due to the machine switching sources).

A further serious problem encountered was the relatively low gas flow values of the recipe in combination with a rather high chamber pressure in comparison to other etch processes normally run in machines as the one used here. An example of such structures along with the recipe is shown in figure 2.13. The uniformity is quite terrible and the shape of the structures is, as will be seen, not optimal for covering with an oxide layer. Both pictures undesirable, but the

center shows signs of a collisionless plasma sheath, a sheath with few radicals and gas molecules diffusing through it to the surface and thus a relatively strong perpendicular ion trajectory. Essentially there is very little stopping the ions from traversing the sheath unimpeded and certainly given the small sccm (standard cubic centimeters per minute) gas flow values and low pressure value (a high rate of out-pumping) this is a logical conclusion to draw seeing how pitted the wafer surface appears (also a decrease in the amount radical etchant).

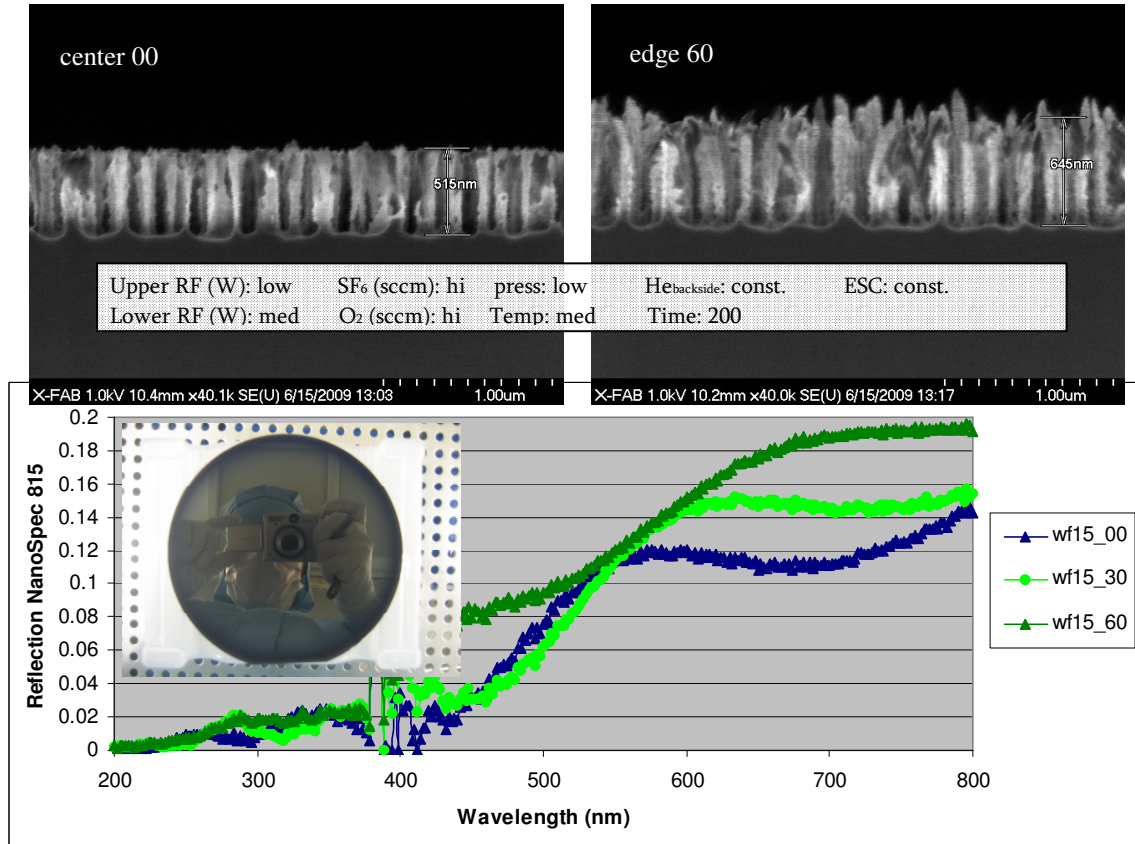


Figure 3.14: Direct reflection from black silicon and visible non-uniformities (noise at 400nm is due to the machine switching sources).

As discussed in chapter 2, gas is delivered to the wafer from a gas ring around the outside of the chamber. If the gas flow is small and the turbo pump is working hard to empty the chamber at a very low pressure then the conditions above the wafer, particularly in the center, will be thinner plasma.

For this dissertation the application is maximization of antireflection and although the topic is thoroughly covered in the chapter Optical Properties some basic, albeit extremely functional, direct reflection measurements from the NanoSpec discussed in the introduction, are introduced here within the framework of uniformity as well as the rather useful notion of, “Yes... that wafer looks black”, when initially searching for the correct operating point. Hence phrases such as “the color is (less/more) black” will refer in general to the visible observations of material exhibiting less or more reflection within the 200 nm to 800 nm range. Figure 3.13 also shows a top down digital picture of the wafer along with direct reflection measurements dependant on wavelength at 00 millimeters (wafer center), 30 mm and 60 mm (wafer edge) and the y axis maximum at 10%.

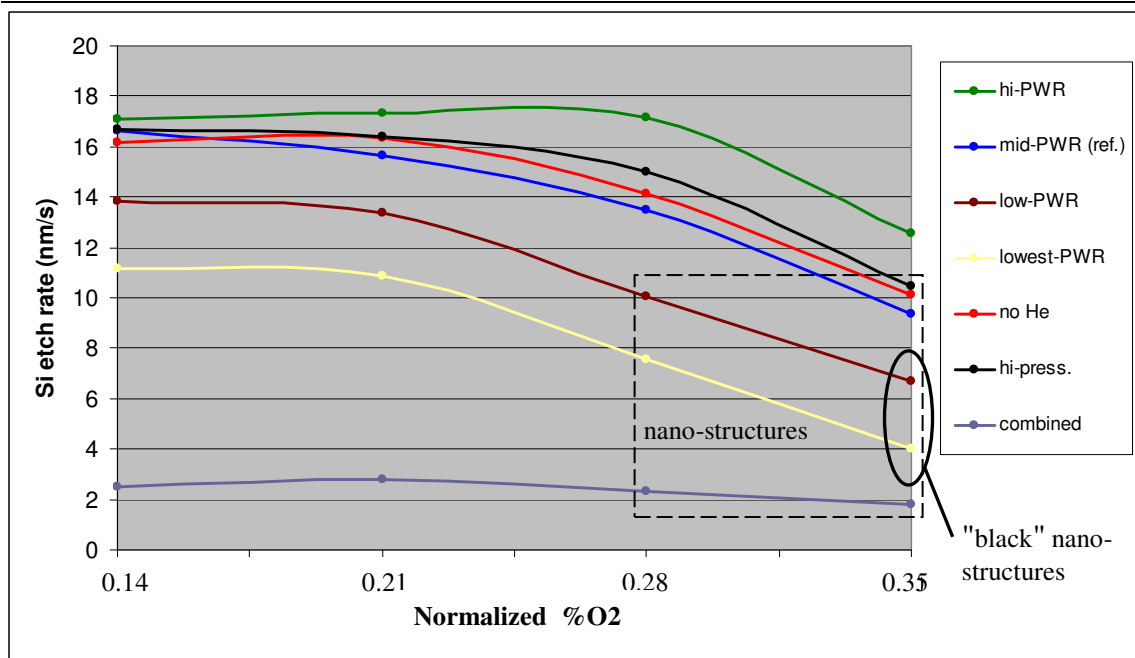


Figure 3.15: Silicon etch rates (wafer edge) for samples dependent on percentage of O₂ (tool parameters normalized to the mid-PWR samples). A group of 4 samples (mid-PWR) were etched with all tool parameters held constant, except gas ratio (amount of O₂ compared to SF₆) which was swept from 0.14 to 0.35 with structures appearing between 0.28 and 0.35 however only the 0.35 ratio yielded structures which appear “black”. The etch rate predictably decreases with increasing O₂ (Phase 2 into Phase 3 from figure 2.09). A subsequent set of 4 samples (hi-PWR) were etched at a higher upper RF power, as expected the etch rate is for all samples increased, raising the 0.35 gas ratio sample out of the “nano-structures” window. Instead, decreasing the power for 4 samples (low-PWR) lowers the curve further into the window and a further reduction in etch rate is seen for a further reduction of upper RF power (lowest-PWR). Modifying other parameters such as no Helium backside cooling, higher chamber pressure, and combined (high gas flow and lower power) shifts the curve, however the “black” process window is in all cases small and results in very awkward surfaces (although it should be noted that etch rate combined with gas ratio is not a bad relationship to observe when hunting initially for structures within a new tool).

The wafer is indeed very black however as mentioned earlier modifying the recipe to improve uniformity or more importantly the structure shape from anisotropic pits to pyramids, a shape which one can more readily cover, does not improve the situation. In fact as shown further in figure 3.15, with the y axis maximum at 20%, modifications that one would normally perform to reduce anisotropy and thin plasma, namely decreasing upper electrode power and increasing gas flow while maintaining the same gas ratio, with regard to the color, worsen the result (albeit comparatively improving uniformity). Additionally the structures of figure 3.14 remain quite gruesome looking when one considers conformance of a deposited oxide. It should also be noted that the noise in both reflection measurements is noticed around the 400 nm area and is due to the switching of light sources within the reflection measurement tool.

In fact all single step samples etched in the LAM tool showing structures exhibited a less than desirable uniformity and massive irreducible structure anisotropy. One explanation could be given by some inherent non-uniformities of the tool itself, as discussed in chapter 2, and the tools limitations when forced into a small processing window well outside of the normal or planned

operating range. This range is presented in figure 3.14. Etch rates for silicon wafers are shown (edge only) dependent on the percentage of oxygen in the SF_6/O_2 gas mixture. The tool parameters for this graph are normalized to the mid-PWR samples. An initial 4 samples (mid-PWR) etched holding all tool parameters constant, except gas ratio (amount of O_2 compared to SF_6) are shown in blue. The gas ratio was swept from 0.14 to 0.35 with nano-structures only appearing for the samples between 0.28 and 0.35 however only at 0.35 were “black” surfaces observed. As was demonstrated in the previous studies we have cited, one can predicate that increasing O_2 (Phase 2 into Phase 3 from figure 2.09) will result in a decreasing etch rate. A further 4 samples (hi-PWR, shown in green) etched with a higher upper RF power yielded an increased etch rate, as one would expect. Notice that there is a window located on our graph in which nano-structures were observed on the wafer surface. Increasing the etch rate has moved the entire data series, including the 0.35 O_2 sample, above the “nano-structures” window. In contrast moving the RF coil power to a lower (low-PWR) value and still lower (lowest-PWR) moves the curve down further (and still further) into the nano-structures window. Indeed, now the 0.28 O_2 sample shows surface structuring. Other parameters from the mid-PWR setting can be modified influencing the curves position relative to the nano-structures window, albeit upper coil power has obviously the most significant effect on shifting etch rates. Complete absence of Helium backside cooling (no He) is shown in red expectedly with a slightly elevated etch rate (implicitly derived from figure 2.02, the energy of the silicon atoms are greatly increased with the uncontrolled substrate temperature resulting in an increased etch rate).

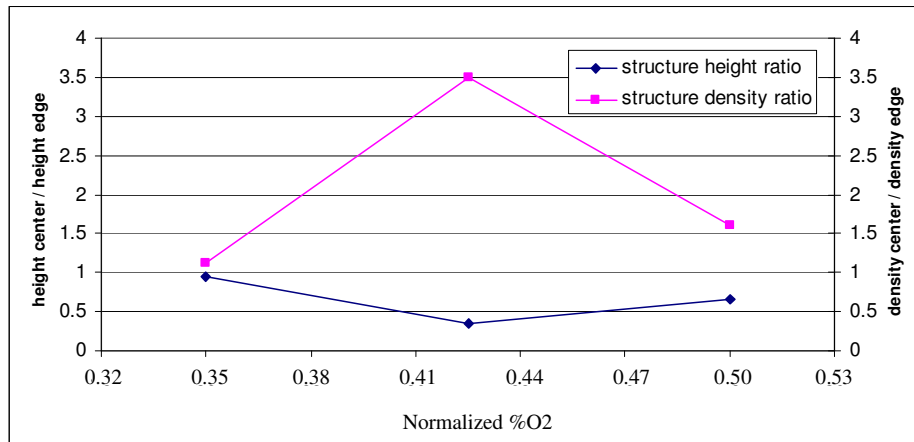


Figure 3.16: Center-to-edge ratios of structure height (left axis) and density (right axis) dependant on normalized $\% \text{O}_2$ in gas mixture. A value of 1 for both pink and blue lines corresponds (approximately) to perfect uniformity. This is seen for a value of 0.35, which in figure 3.12 corresponds to the approximate mid-point of the dominant K_A region.

A higher chamber pressure (hi-press.), “normally” considered to decrease the etch rate by decreasing the turbo-pump’s chamber evacuation speed and thus resulting in a increased density of species (etch products and gas) over the wafer surface which inhibits ion and radical delivery to the surface, however from the thin plasma of mid-PWR (already at a very low pressure – very aggressive chamber evacuation) an increase in pressure allows a more normal delivery of radical to the surface by giving gas enough time to diffuse to the wafer and become ionized slightly increasing the etch rate. Combining a high gas flow and the lowest power (combined) shifts the curve drastically down, as certainly a chamber full of barely ionized gas will not etch the surface in a particularly aggressive manner regardless of gas ratio. In all cases however the “black” process window is small and results in very awkward structuring of surface surfaces.

There are actually a two major positives in these results, firstly that the etch rate combined with the gas ratio (percentage of passivant to percentage of etchant) has proved to be not a bad relationship to monitor when a user is hunting for first signs of structures within a plasma tool. Secondly, the shapes of all curves conform to the shape of the modified silicon etch rate in fluorine radical containing gas with the addition of O_2 shown previously in figure 3.09 and as designated in our theoretical summary in figure 3.12.

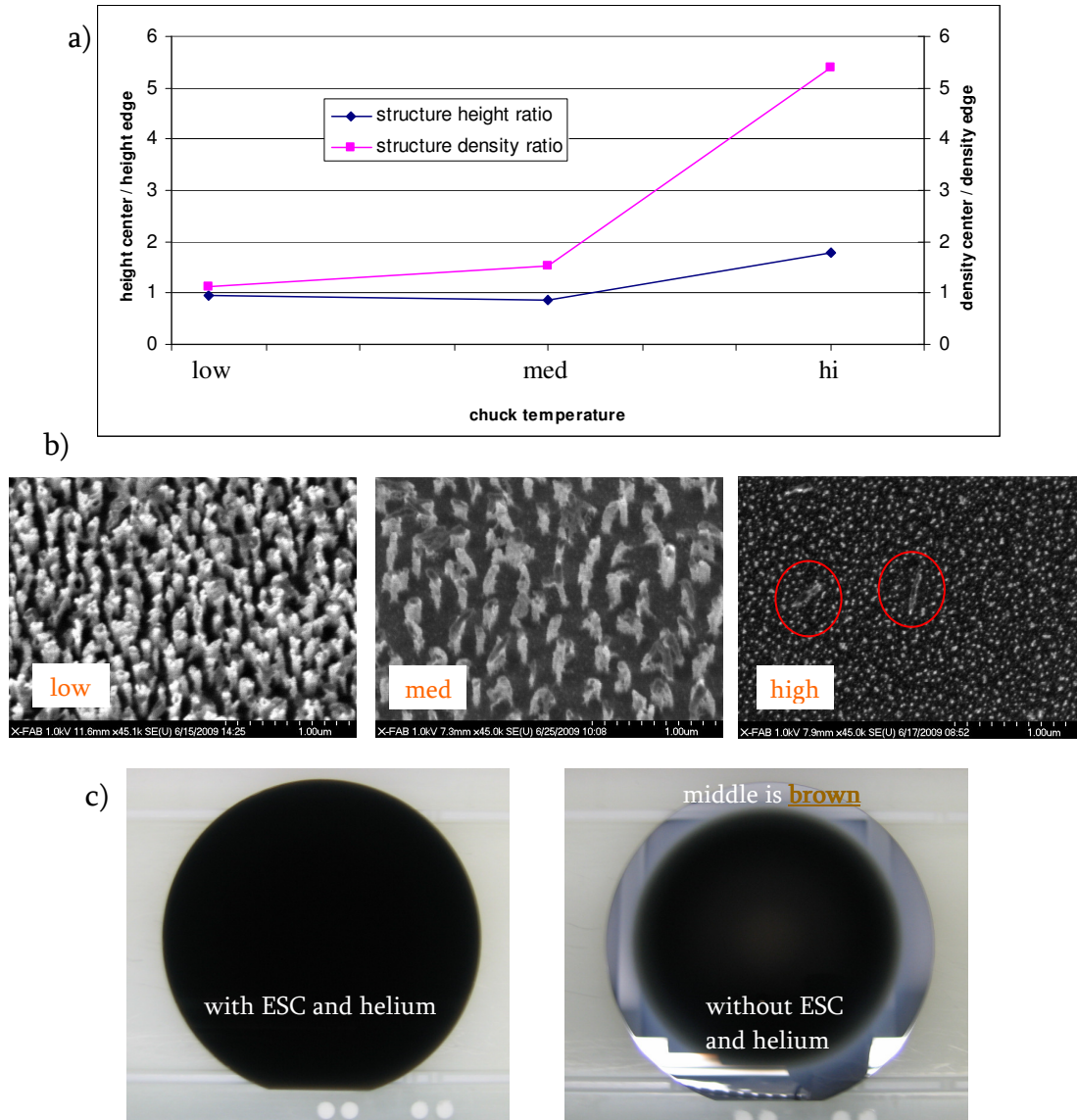


Figure 3.17: a) Ratios of structure height and density dependant on chuck temperature. b) Various regulated wafer temperatures and their structural impact on the wafer surface. SEM pictures show edge sites and the disappearance of nano-structures at higher wafer temperatures. c) The most extreme case, no helium backside cooling (uncontrolled wafer temperature). The “low” temperature setting provided the best center-to-edge structure uniformity.

This is certainly not air-tight evidence, however additional experimental information supporting our initial theoretical assumptions will be provided in section 3.2.3. For now we can say that this

is an (unacceptably) inflexible process window. The solution in the case could be to move to another tool. However this dissertation deals with the integration of nanotechnology into standard foundry tools and the solution is to simply be smarter than the needles.

The target is then to create a nano-scale loading effect with a set of structures of uniform density and height to be used as an etch mask and furthermore as it will be shown to decouple much of the dependence of texturing and form optimization from the initial masking without having to add exotic deposition and lithography. Several sets of samples were etched showing the two parameters with the strongest observed influence on controlling uniformity depicted in figure 3.16 and 3.17, the dependence on the overall percent of O₂ in the gas mixture, exactly related to its concentration relative to the fluorine source as provided in the exposition of section 3.1 on plasma chemistry, and the wafer chuck temperature, a concept derived directly from section 3.1 dealing with potential energy at the surface and its influence on activation energy. On the left y axis is the ratio of structure height at the wafer center to structure height at the wafer edge and on the right y axis is the ratio of structure density at the wafer center to structure density at the wafer edge. Naturally the target is in both cases a value of 1 and graphed together it is clear that a ratio 50% O₂ and the low temperature yielded the best results. Outside of the shown x domain no structures were observed. An unoptimized step is shown in figure 3.18 and the differences in uniformity compared to figure 3.19, the optimized results, are stark.

SEM picture evaluation in figure 3.17 b) taken at an angle of 20 degrees away from the top-down perpendicular direction at the wafer centers show the samples processed with the same SF₆/O₂ plasma recipe except for a variation in the value of the regulated wafer chuck temperatures.

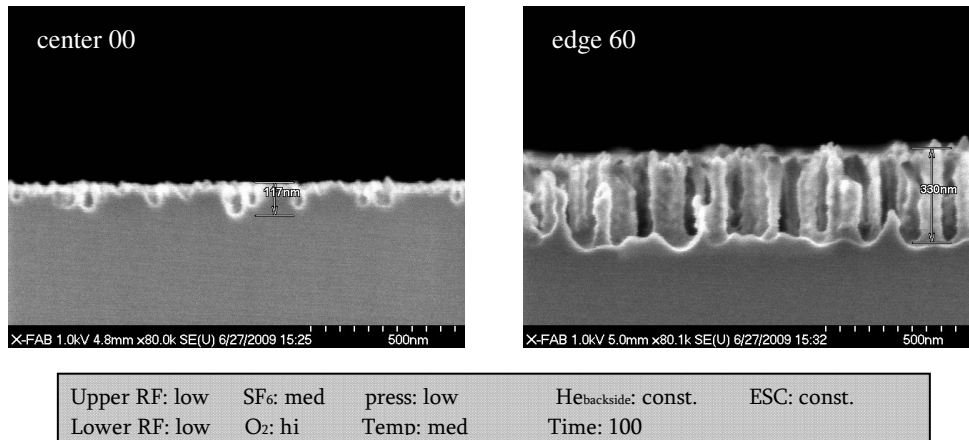


Figure 3.18: Unoptimized nano-structures with poor uniformity.

Without needing to count the number of nano-structures per μm^2 (a rather tedious activity) it is plainly clear that not only the density but in the case of the high temperature sample the very presence of nano-structures are dependant on temperature. However one should be careful not to generalize too much with process parameter names. When possible etch phenomena should be reduced to the vocabulary of quantum chemistry presented in section 3.1. For example here the increase in surface temperature is modifying the potential energy curves of figure 3.02 and thus the activation energy influencing absorption and desorption of radical species from the plasma on to the silicon wafer surface. The absorbed foreign species and the silicon surface atoms experiencing an increase in thermal excitation with increasing temperature etch more readily with the physical impact of incident ions. Furthermore, as stressed in section 3.1, the energy, E^* , depicted in figure 3.02 and assumed now to be possibly taken up by an absorbed monolayer of foreign atoms on the silicon surface will now even more readily reach activation energies to

changes states of absorption or outright desorb and return to the plasma. Obviously an optimal operating point is then dictated by the specific application requirements of the nano-structures (perhaps not desired at all) and determined by experimentation.

It should go without stating that simply increasing etch time will not improve a situation if the results in the first few seconds show tremendous discrepancies structure separation. Adding subsequent steps will also not improve the situation, however the additional steps do give insight into the behavior of the initial mask, for example the case of the two different surfaces in figure 3.20 and figure 3.21, having been etched with different nano-masking recipes (shown above), but subsequently receiving the same series of steps. Multiple steps and texturing will be discussed later, for now it is only important to observe the properties of the initial mask held throughout the process independent of the following steps. Also apparent in the two figures is the relationship between percentage of O_2 versus SF_6 in the discharge and also perhaps the dependency of a process on physical system construction are seen.

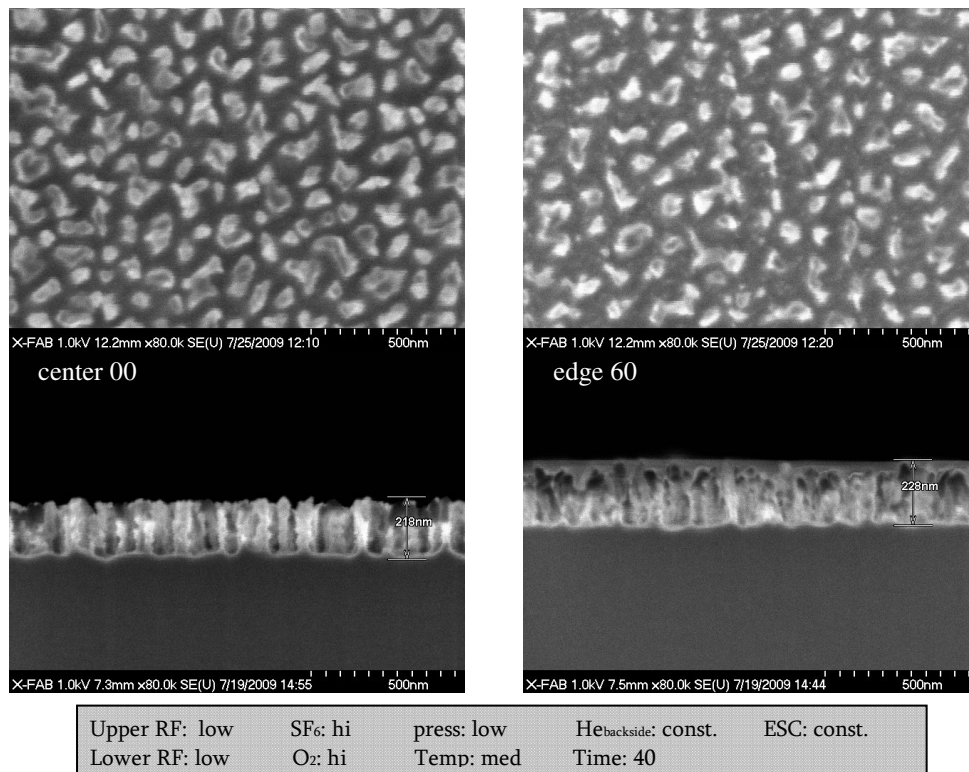


Figure 3.19: Optimized nano-structures for nano-loading. Separation between center and edge structures is very similar.

Gas is delivered from the gas ring around the outside of the chuck diffusing inward towards the center of the chuck. In figure 2.20 it appears that not enough SF_6 is reaching the center of the wafer, most having reacted at earlier portions of the silicon wafer. As the percentage of SF_6 is increased more SF_6 reaches the wafer center and as seen clearly the surface in the center begins to etch and a greater separation between structures is observed. The ratio is further increased leading eventually to the excellent nano-structures of figure 2.19 with a very consistent spacing. Actually proof of this, the concentration of SF_6 dependant on wafer location, could easily be observed with the PROES method proposed in the introduction. Furthermore, it does not take a keen eye to notice that the form of the pyramid shaped edge structures in figure 2.20 and

2.21, having both been treated by same subsequent recipes, have been degraded through the attempt at modification of structure separation in the wafer center. Such is life when dry etching.

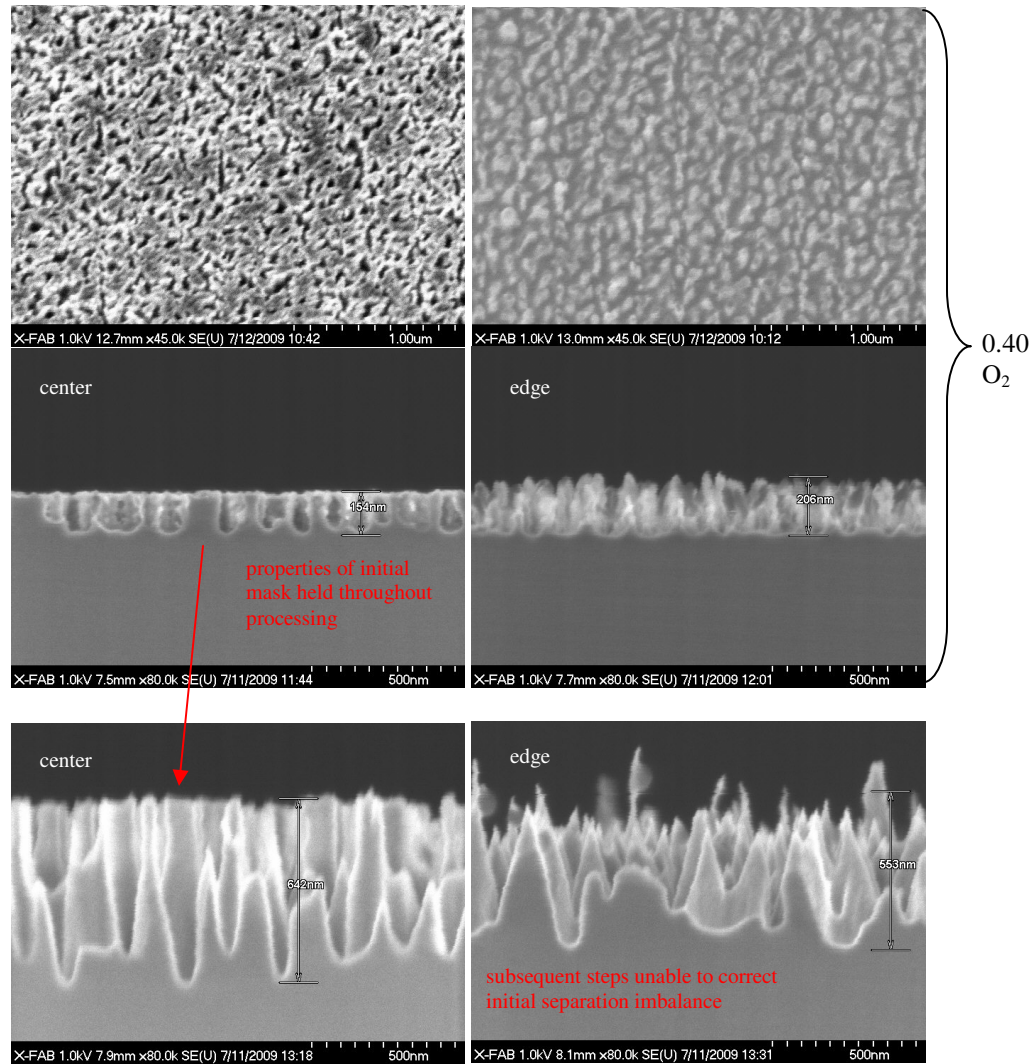


Figure 3.20: Beginning separation of center nano-structures. Form and density of initial structures carrier over throughout all subsequent etching steps. Normalized %O₂ value cited.

It suffices to say that although the edge pyramids of figure 3.20 seem nice, the current and more pressing target is to create a uniform etch mask, which could then be etched with a series of (other, different) steps to create pleasant looking pyramids but decoupling initial masking and structure profile, indeed as we are seeing in our dry etching this is a very strong relationship.

Enough with poor structures! The very positive results of figure 3.19, the optimized structures for the nano-loading effect to improve processability, are ready to be tested. The joke could be made that either now there is both a uniform temperature across the wafer and uniform species density above the wafer or the non-uniformities in wafer temperature match the non-uniformities of the plasma above the wafer. Regardless the observed result is a positive one. Testing is done by taking two wafers, etching one wafer first with the set of uniform nano-structures, and then etching both wafers with the exact recipe from figure 3.13 above (as seen, a recipe by itself yielding very jagged structures). The comparison of the SEM pictures in figure

3.13 with those in figure 3.22 is akin to a comparison of night and day. The contrast between the structures is more than stark. The unevenness of the sample without nano-loading is considerable, however with nano-loading any discrepancies in consistency are undetectable.

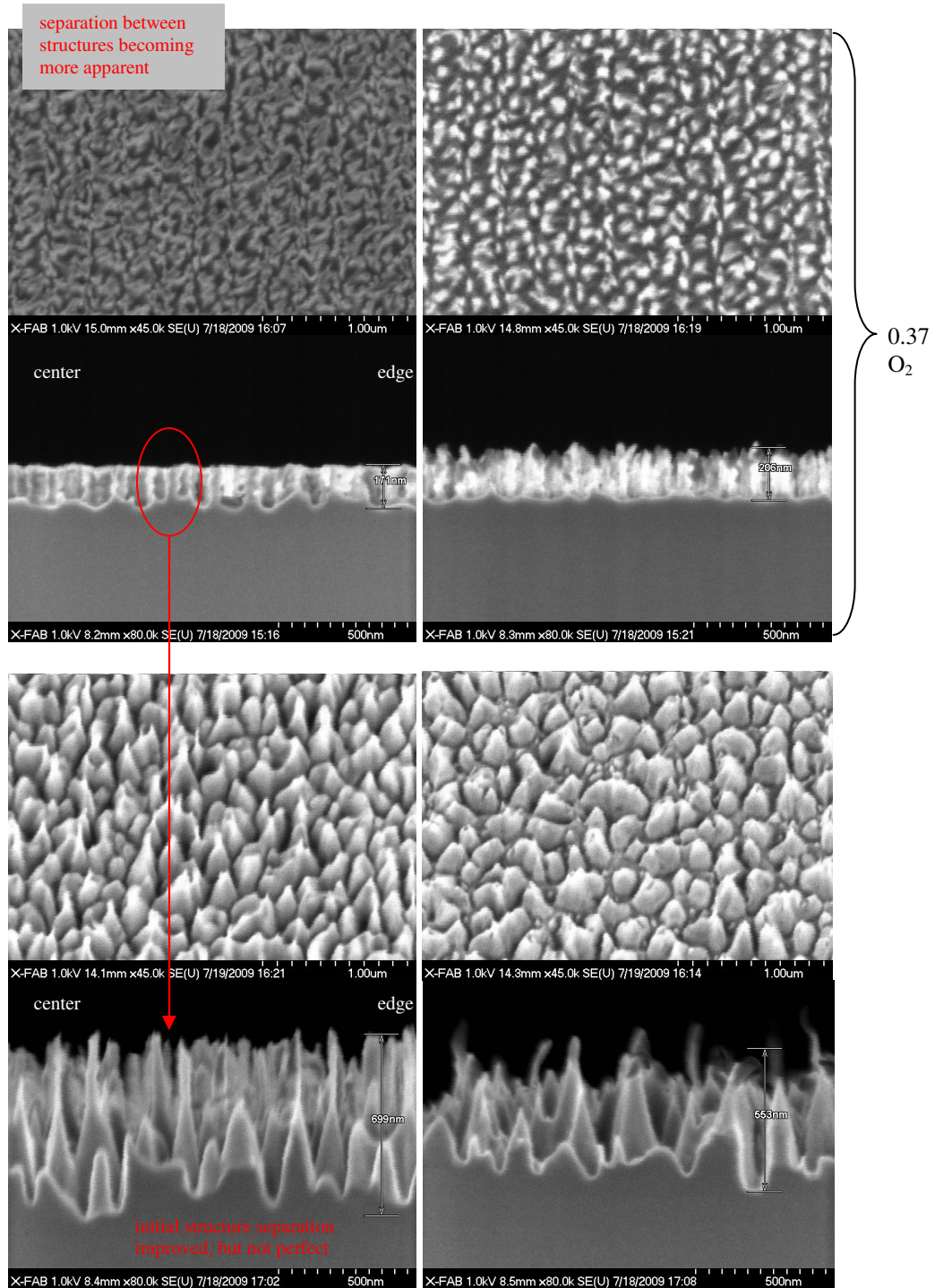


Figure 3.21: Etching of the not quite regular (but improved) initial nano-structures. Again form and density of initial structures carrier over throughout all subsequent etching steps.

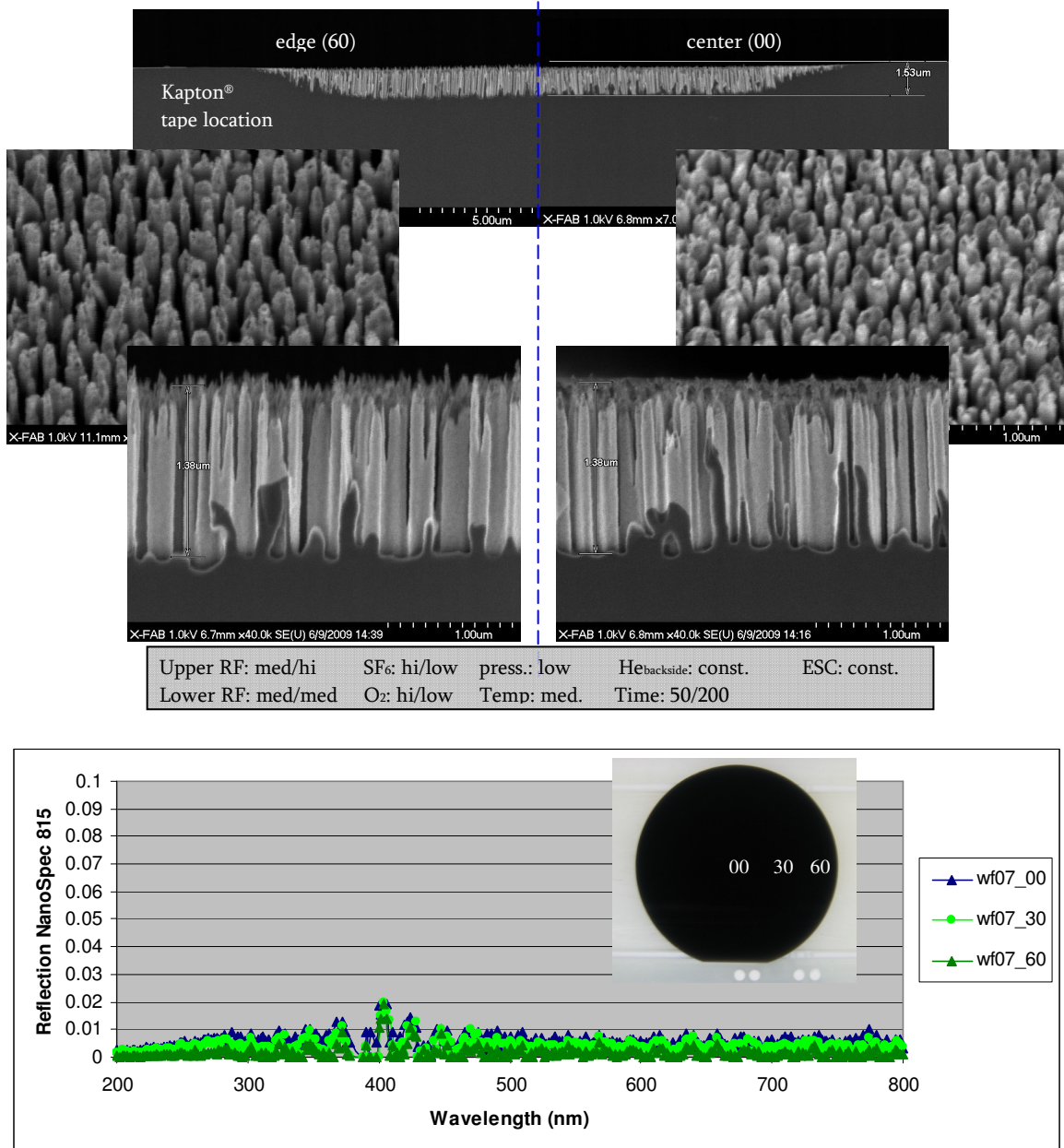


Figure 3.22: Wafer etched with the poor recipe from fig 3.13 however, as a separate initial masking step, receiving first the optimized nano-structures of fig 3.19. The results are an extreme improvement having uniform structure separation and density (albeit still too anisotropic). For the NanoSpec reflection the surface is completely uniform and black.

Without creating a distribution of nano-structures to be used as an etch mask the process becomes heavily restricted by chamber geometry, power coupling regimes and the type of gas delivery system. Furthermore this means that when the need or desire to influence structure shape increases the user is often left modifying structures only on a very small portion of the wafer. More importantly, as noted earlier, there are many operating conditions under which no structures

are seen, however it may be desirable to apply these steps to a surface already structured. This removes the need to use only processes that provide an initial masking.

Before moving on to such multi-step recipes influencing the surface topography further, an observed limitation should be discussed of the simple pillar structures created thus far. Simple differential growth of such needles may be described then by the equation:

$$\frac{dy}{dt} = ry - w \quad (3.13)$$

Where y is the height of the structure, r is the etch rate at the base of the structure (rate of growth) and w the etch rate at the apex (rate of loss). Here it should be pointed that w is taken and not wy . The rate at which silicon is etched away (stolen) from the top of a structure is being considered continuous, it is not some certain fraction of the whole needle. Plotting dy/dt against y and y against t shows the insights and limitations of such a differential equation, see figure 2.23. The critical point, the solution of the equation set equal to zero, is $y = w/r$. This is equivalent to the situation in which the rate of growth is equal to the rate of loss. The etching would simply carry on never altering whatever the initial starting height was. Looking at the graph of dy/dx it is seen that for values less than w/r the function is negative and for values greater than w/r the function is positive and equivalent to the situation depicted by the various isoclines above and below w/r .

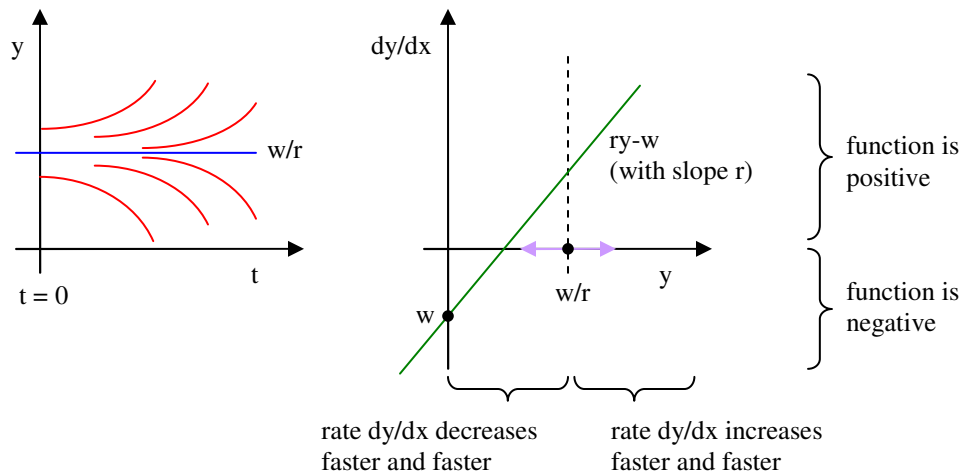


Figure 3.23: Initial height y with equation set equal to zero, $y = w/r$, is equivalent to the situation in which the rate of growth is equal to the rate of loss. The limitations here are that for starting heights greater than w/r the needles do not, as seen in figure 3.24, exhibit an infinite height.

A point beginning above w/r will increase in height through time and a point beginning below w/r will decrease in time. In math terms this critical point is said to be unstable (on both sides of the line w/r the solutions are moving away). The limitations here are that heights less than zero are nonsense and that the needles do not, as seen in figure 3.24, exhibit an infinite height potential. At a maximum height (or time, depending on how one considers the etching) the tops of the needles begin to fall over. The etching can be continued past this point however the structures will not show a net increase in size.

Naturally having unlimited growth with w is nonsense. A logistical growth scenario is needed to account for the ceiling on the maximum height.

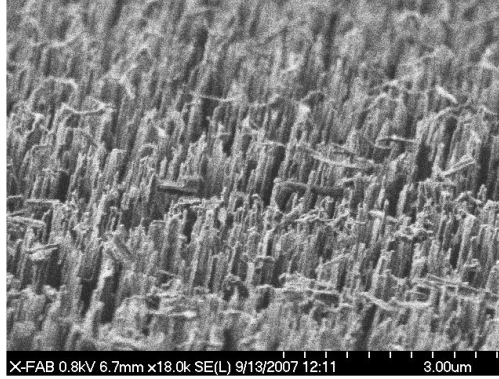


Figure 3.24: Maximum observed needle size reached as needle tips begin to break and fall over

For declining growth the simplest choice for a new w is:

$$w = a - by \quad (3.14)$$

And substituting w back into equation 2.13 gives,

$$\frac{dy}{dt} = ay - by^2 \quad (3.15)$$

The critical points of this new equation are zero and a/b . Figure 2.25 depicts graphs of both y against t and dy/dx against y .

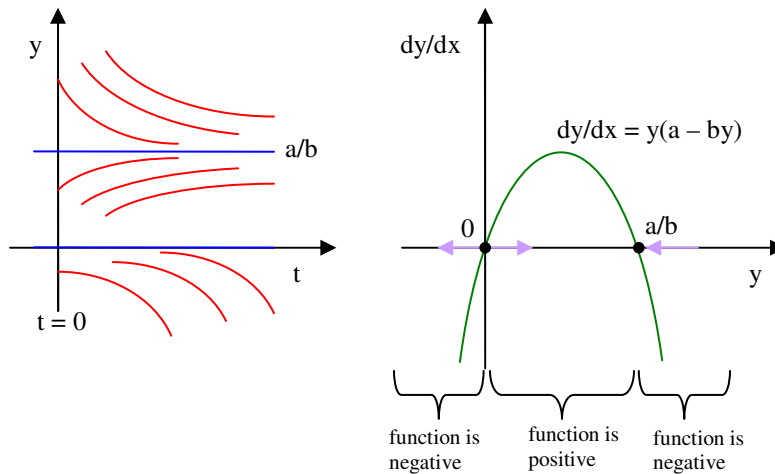


Figure 3.25: Logistical structure growth including a ceiling on height. Now for initial heights above zero or above a/b both converge to the height a/b . Figure 3.42 will re-address this issue.

A point beginning anywhere between a/b and zero will increase in height through time approaching but never reaching a/b . Points beginning below zero and above a/b will decrease in time to negative infinity and the line a/b respectively, however neither of these two cases correspond to any real situation. In math terms the critical point a/b is said to be stable (on both sides the solutions approach the line). The observed a/b for single step pillars was approximately

2 μm even after substantially long etching times, the introduction of multiple steps and pyramid like geometries creates a new sets of complications (for example spacing between pyramids limits the maximum height of a pyramid) and for the sake of maintaining a realistic scope will not be mathematically treated here.

3.2.2 Texturing silicon into nano-structures, part II: Pillars to pyramids

Until now only uniformity has been considered and it's clear that although this property has been drastically improved all the structures discussed share a very anisotropic shape. The addition of CF_4 fluorocarbons provides a second regime of passivation with an altogether different etch rate for added flexibility in texturing structures. Additionally the utilization of more than two etch steps is considered. Figure X shows two subsequent steps performed after the initial micro-masking with $\text{SF}_6/\text{O}_2/\text{CF}_4$ plasma. These structures are no longer the long anisotropic pillars and are after the third overall etch step have begun to resemble pyramids. Table 1 shows the difference between the recipes and from it we see the only difference is gas composition.

Table 1

wf	upperRF	lowerRF	pressure	SF_6	CF_4	O_2	Temp.	time (s)
A	const.	const.	low/med/hi	hi/med/low	0/0/0	low/med/hi	med.	40/100/100
B	const.	const.	low/med/hi	hi-x/med-y/low-z	x/y/z	low/med/hi	med.	40/100/100

In the three steps, each separated by a slash, a fixed sccm amount of SF_6 was replaced with the same sccm amount of CF_4 maintaining the same overall total gas flow. The resultant change in geometry from nano needles to nano pyramids is extraordinary.

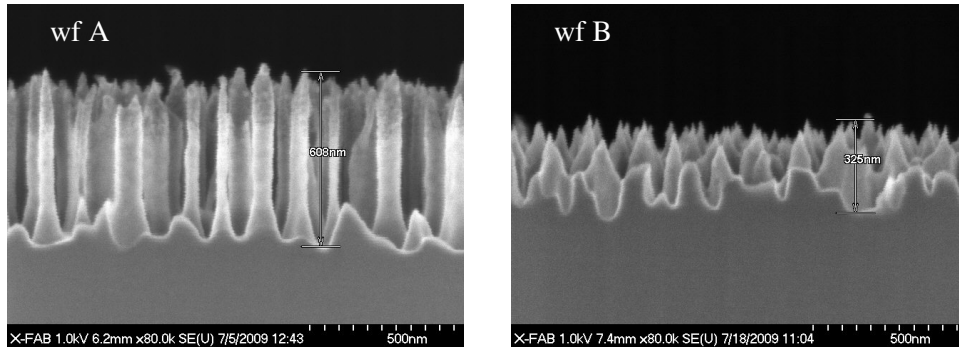


Figure 3.26: Nano structures a) needles without CF_4 and b) pyramids with CF_4 . Removing a fixed amount of SF_6 and replacing it with CF_4 . However significant amounts of carbon were never detected on the surface. Possibly linked to the presence of heavy carbon in the lower sheath altering ion trajectory as discussed in the introduction or simply playing a catalyzing role within the plasma.

Subsequent EDX analysis taken from the side of an individual pyramid shows no definite carbon concentration. Obviously the presence of multiple types of passivants is not the complete reason or the geometrical. It has been demonstrated with PROES methods (discussed in the introduction) that species of carbon do not respond to the time varying RF sheath (in fact sitting directly over the wafer, not riding on a “sheath”), this is attributed normally to the carbons weight. It could very well be that the presence of such heavy carbon in the sheath acts to deflect incoming ions and alter the trajectory creating the more pyramid appearance. It is also obvious that replacing a

fixed amount of SF_6 with CF_4 will lower the overall etch rate in each step as fluorine radical yield of CF_4 is less than that of SF_6 .

As already discussed figure 3.19 above shows the surface after the nano-loading, the first step of the process. Figure 3.28 shows the surface after the second and third step. After step three, in comparison to the wafer surface after step 2, smoother side walls, sharper points, and an overall reduction of the anisotropic trench profile of structures are observed.

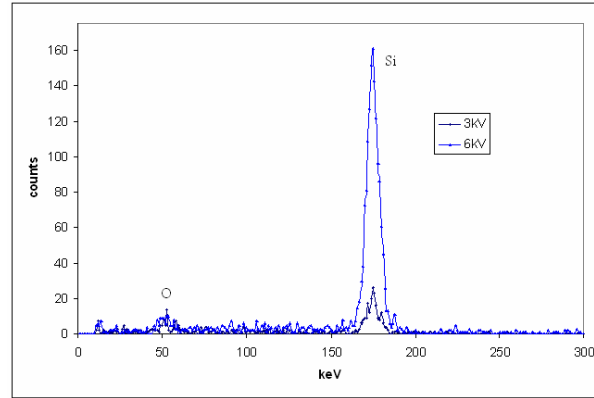


Figure 3.27: EDX of structures after etching shows no clear presence of carbon.

The upper and lower RF generators were held constant making the gas ratio of significance to note. In each step the overall flow is held (approximately) constant while the ration of SF_6 (etchant) to O_2 (passivant) is altered, exactly following the graph in figure 3.12 of section 1. The utility of steps rich in passivating chemistry of that normally provide no nano-structures are applied to modify the form of the existing surface topology.

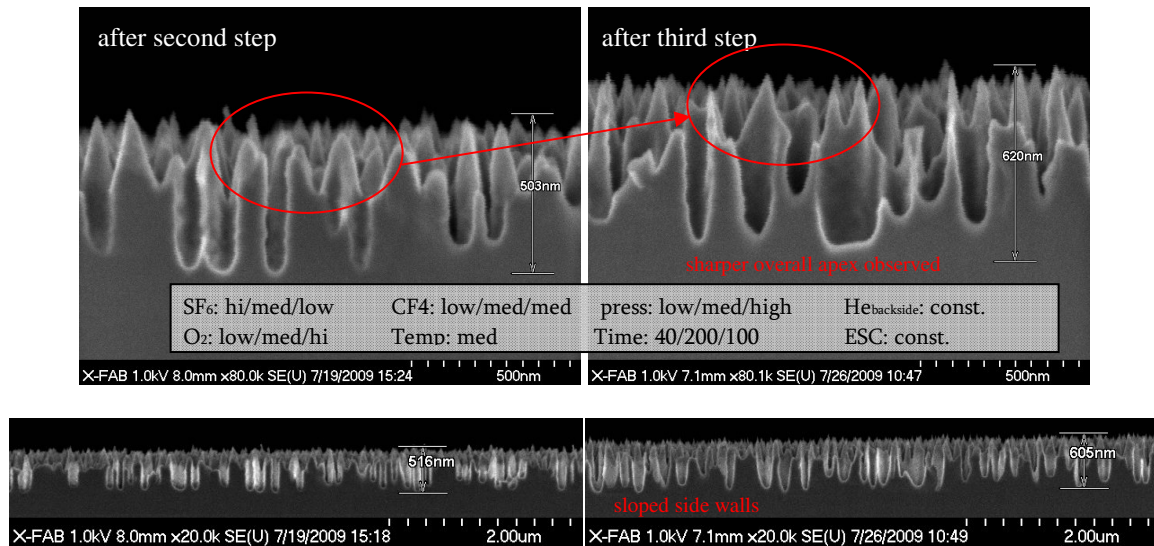


Figure 3.28: Structures after second and third steps. Upper and lower power held constant. After the second step we see a distinct pointing of the surface as the nano-mask is chiseled away and the silicon substrate around the nano-loading pillars is further etched. It should be stressed that alone steps two and three do not yield structures but here their rich passivant / light etchant mixture, especially seen after step 3, provide a very delicate surface texturing.

Figure 3.29 shows the extraordinary success of maintaining structure consistency of the wafer after the three steps. Simply increasing the etch time of the second step increases the overall pyramid height with the 200 s time creating nearly the optimal form. Times beyond this began to degrade the quality of the apex of individual structures. The full optical implications of this are treated in chapter three however a slight increase in height to the 800 nm or 900 nm mark would be desirable. Attempts to separate the second step, the step most generally responsible for structure height, into 2 separate steps were largely unsuccessful with the results shown in figure 3.30. Again the upper and lower RF generators were held constant while the rate of decrease in the quantity of SF_6 was slowed. Essentially a decrease in the rate of SF_6 decrease yielding a slightly higher etch rate, albeit too high as shown by figure 3.30.

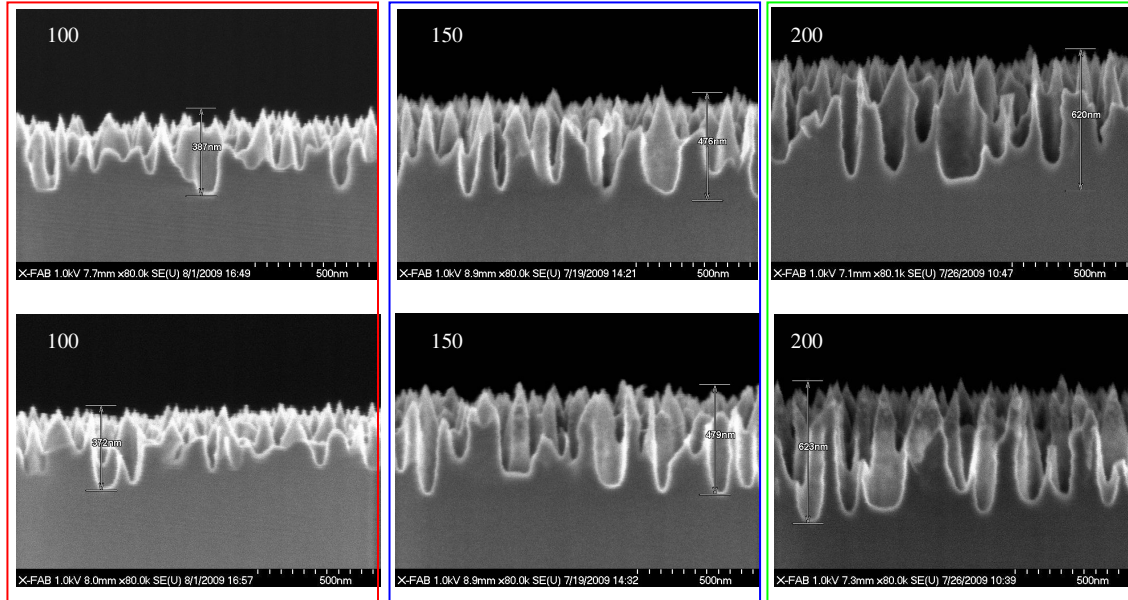


Figure 3.29: Three-step structures with a very pyramid-like form. With increased etch times during step 2 of 100, 150 and 200 we see a distinct increase in pyramid size. The uniformity between center and edge in all three cases is excellent.

The increased etch rate has cleared away more silicon material leaving posts on top of cones. The 20 degrees from top down view shows very well the now cleared away areas of silicon and the scattered array of thin posts while the side view shows that there is no significant increase in perpendicular etch, only a lateral chemical etching do to the greater abundance of fluorine radicals.

In this section structure form has been related to chemistry and the steps and thought processes involved in making decisions to add or subtract certain steps and alter certain process parameters have been demonstrated on blank wafer. The following section will treat the problems encountered when decreasing the area of exposed silicon. As it will be seen, altering the ratio of silicon and introducing hard etch masks will vary the etch results however where it is possible ideas and lessons introduced in this section will be referenced in the next section. The numerical parameter values may alter but the methods employed are not significantly different.

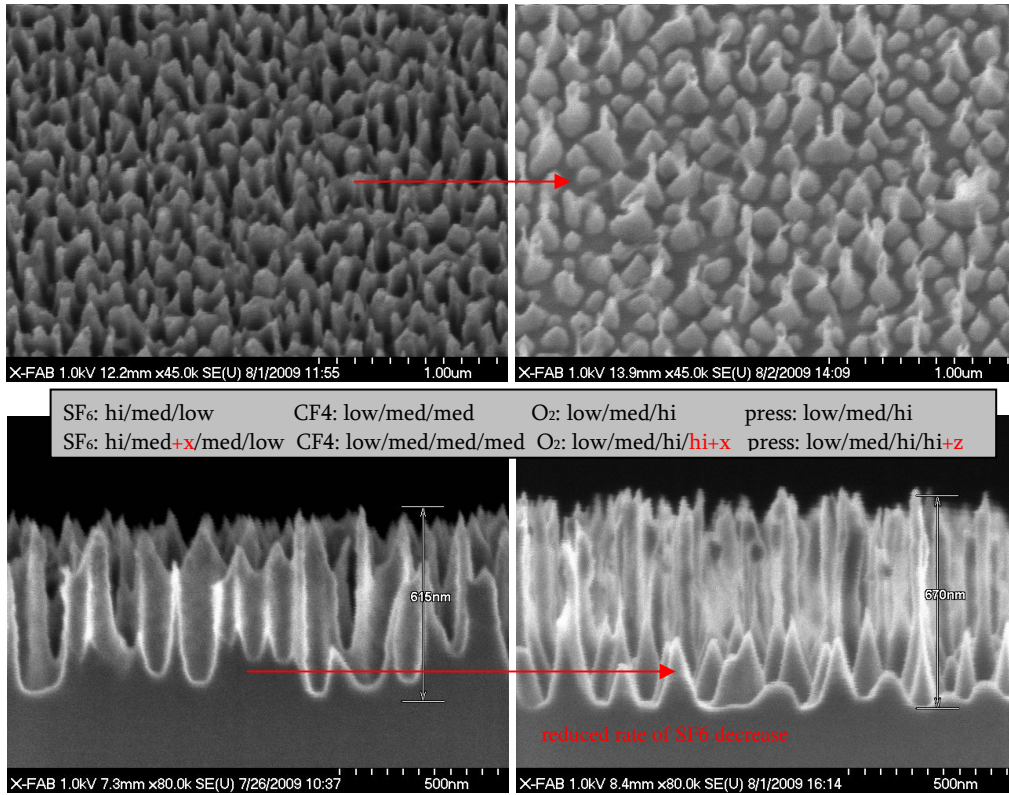


Figure 3.30: A reduction in the rate of SF₆ decrease (edge sites). Time steps for the left hand side are 40/100/100 and for the right hand side 40/100/100/100. X and Z are incremental constants highlighting the relatively small process changes which yield drastic structural results.

3.2.3 Experimental support for the theory of our nano-masking

The material of the previous section dealt with texturing more pyramidal shapes out of small sets of initial nano-structures of varying density and separation. This section would like to bring together the assumptions of figure 3.12 with several further experimental measurements and give a slightly more thorough account of the origin of the initial nano-structures. Figure 3.31 shows a comparison of AFM measurements taken on an unetched surface (left) and a surface having received only the first 10 s of an initial pillar etch (right). A magnitude scan (top row) of the surface measuring the absolute z (direction perpendicular to the surface) deflection of the tip shows very little difference between an unetched sample and a sample etched for 10 s. However the phase map difference is staggering. The oscillations of the cantilever of the AFM system are driven by a sinusoidal signal generator. This signal is saved and compared to the signal recorded by the photodetector. Deflection of the AFM tip caused by topographical changes is registered by a modification in amplitude of the detected signal when compared to the driving signal. Phase imaging however compares the phase lag of the cantilever oscillations relative to the original driving signal. The phase lag corresponds very well to variations in surface material, thus when the overall surface topography is nominally flat changes in attractive force between cantilever tip and surface are registered in the phase lag [X2.2]. This is pictured in figure 3.32. The claim that we are viewing oxide and silicon in the nominally flat surface (and not some other exotic contaminant) can be further supported by the results in figure 3.33. Auger emission spectroscopy (AES), a process measuring not more than the first 10 nm nanometers of a sample, shows the chemical composition of the surface being no more than variations of silicon and oxygen. For

clarity it is shown in figure 3.34 that the process of Auger emission here is slightly different than the process discussed in the beginning of this chapter for a surface in the vicinity of an ion.

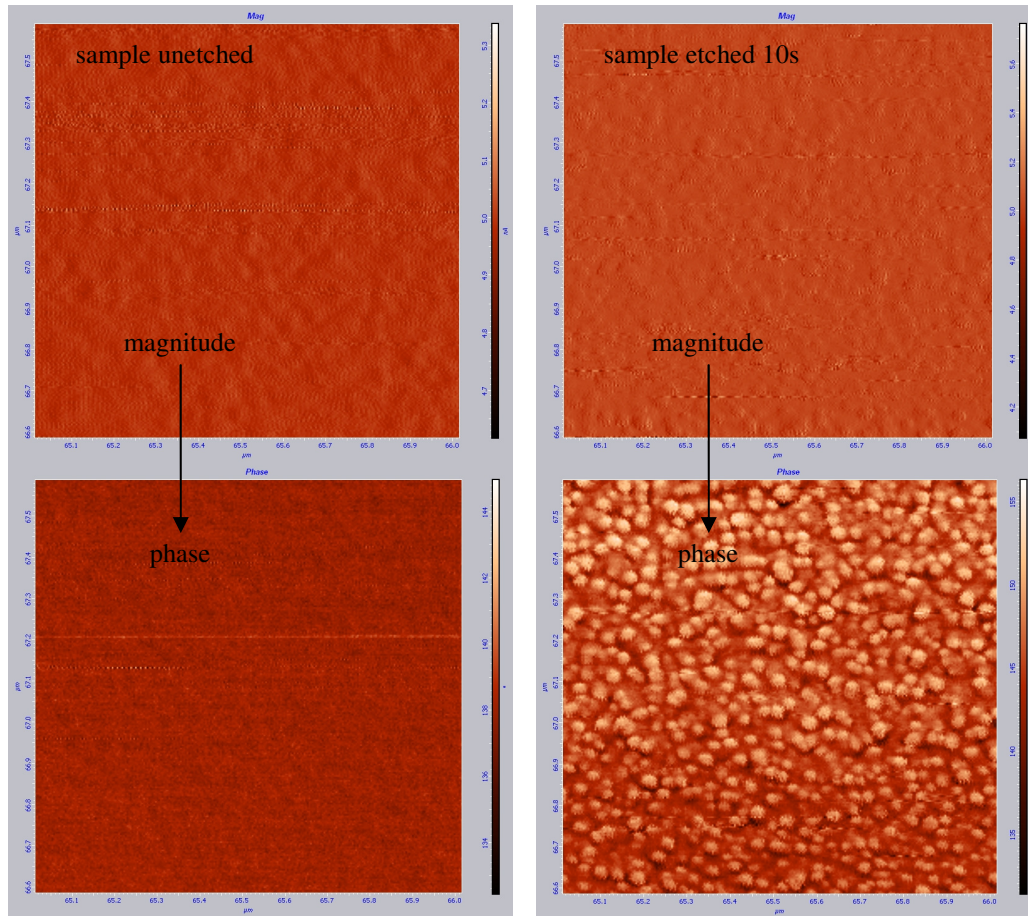


Figure 3.31: Atomic force microscopy (AFM), 1 μm x 1 μm sample. The difference between magnitude and phase pictures is stunning. Overall surface roughness is on the order of 1nm. Combined with the AES measurements of the next section, we are likely seeing a distribution of oxide on silicon.

This time we are bombarding the surface of the sample with an energetic primary beam of electrons (figure 3.34a) which causes the emission of electrons from an atom of the sample. When an electron is removed from a core level of an atom, leaving a vacancy, an electron from a higher energy level may fall into the vacancy (figure 3.34b), resulting in a release of energy. This occurs because an atom missing a core electron is unstable. The core hole can be filled by an outer shell electron, which loses energy moving to the lower level equal to the difference in orbital energies. Although sometimes this energy is released in the form of an emitted photon, the transition energy can be coupled to a second outer shell electron which will be emitted from the atom if the transferred energy is greater than the orbital binding energy. This second ejected electron is called an Auger electron figure (3.34c). The kinetic energy of the Auger electron is $E = E_A - E_B - E_C$ = the energy of the released Auger electron, where $E_A - E_B$ = the energy of the photon and E_C = the energy necessary to release the Auger electron [X2.2].

Although the main concern of this dissertation is to now transfer of nano-structures to the surfaces of diodes to be electrical evaluated, figure 3.35 presents how this “origin of needles” work could be taken further.

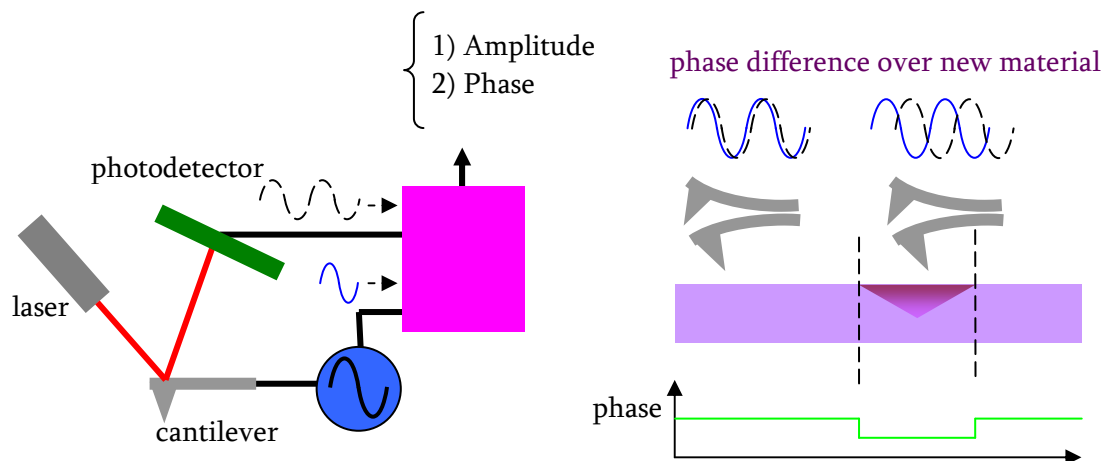


Figure 3.32: AFM functionality. The cantilever is driven by a sinusoid which is also saved and compared to the sinusoid recorded by the photodetector. Surface topography is imaged by comparing the differences in input and output sinusoidal amplitude. In phase imaging, the phase lag of the cantilever oscillation, relative to the signal sent to the cantilever is compared. The phase lag is very sensitive to variations in material properties (namely differences in the attractive force between cantilever tip and surface) [X2.2].

A more densely concentrated cluster of oxide “islands” formed by the passivant oxygen radicals, shown in blue, during the etch process will likely lead to the formation of structures by slightly inhibiting the ability of the etchants, shown in red to attack the silicon surface. This is exactly as we speculated in the beginning of this chapter, although at that time we drew only one “island”.

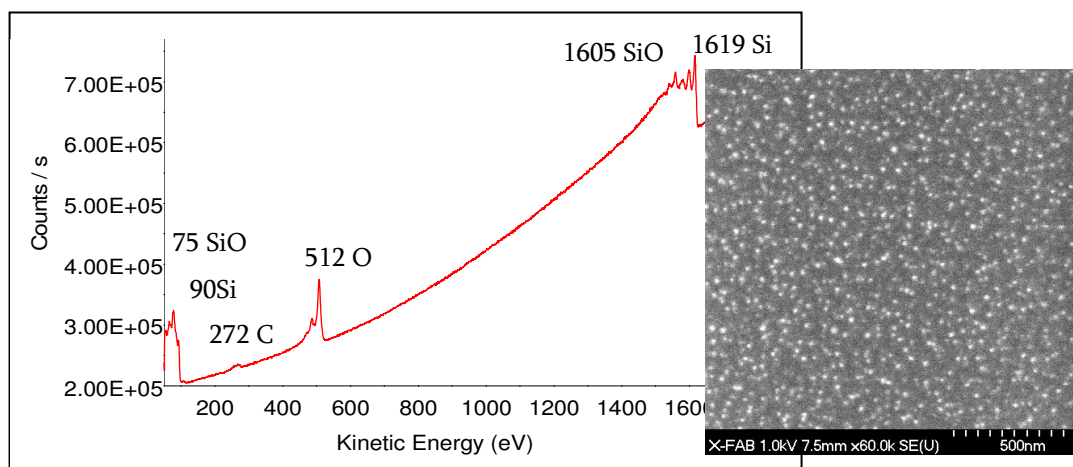


Figure 3.33: Auger electron spectroscopy (AES) of the sample from figure 3.31. The peaks correspond to various forms of silicon and oxygen. Nothing exotic was found. An SEM of the sample surface is shown for reference. Indeed bright white flecks are clearly visible on the surface and are very likely the same grains seen in the phase picture (it should be noted that focusing the SEM for this picture was quite a nightmare!)

A more thorough surface analysis program could be implemented by perhaps taking several different recipes, with corresponding SEM results showing variation in final pillar size and

separation, and etching further samples with each recipe for only 5 s, 10 s, 15 s, and 20 s. SEM and AFM comparison of samples combined with AES (and other surface characterization methods).

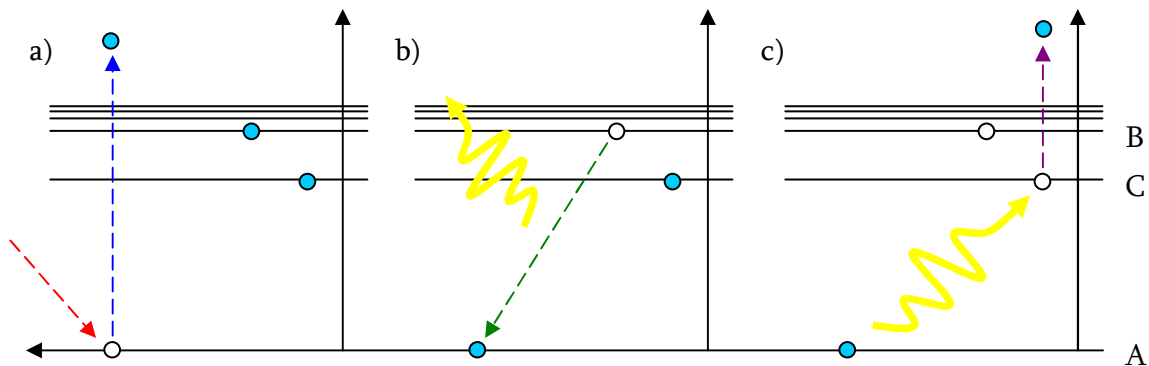


Figure 3.34: Auger emission spectroscopy. a) An electron from the stimulus beam causes the emission of an electron from a sample atom. b) When an electron is removed from a core level a vacancy is left behind and an electron from a higher energy level falls down releasing energy. c) This energy can be transferred to another electron ejecting it from the surface. This second ejected electron is called an Auger electron. This Auger electron's energy corresponds to the unique energy level transitions within the surface atom, which can then be identified [X2.2].

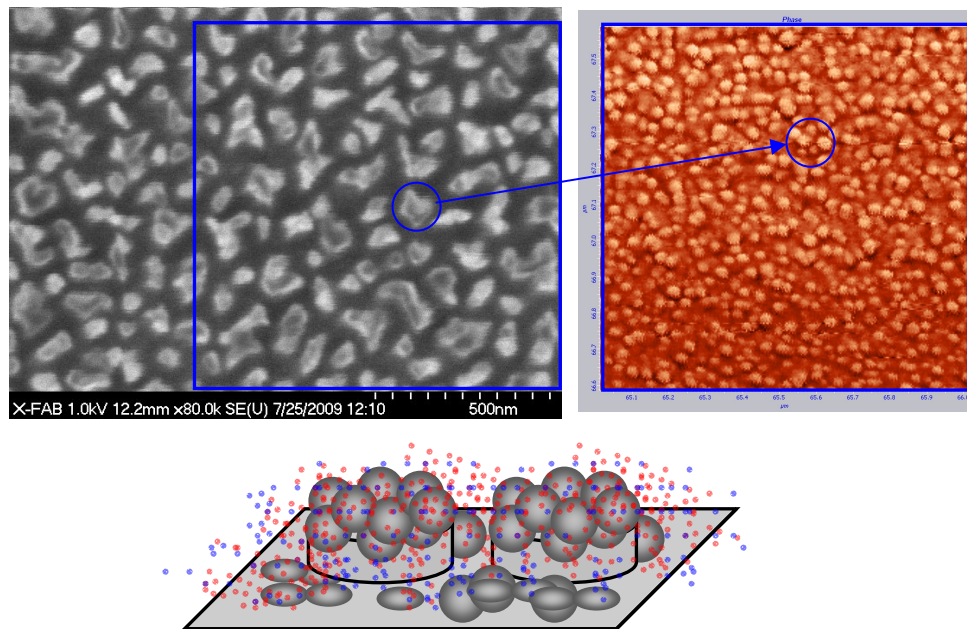


Figure 3.35: A slightly denser cluster of oxide “islands” during the etch process likely leads to the structure (pillar) formation. A passivant (oxygen) is shown in blue and an etchant (fluorine) is shown in red.

It would be very nice to see such experiments combined with characterization of the various densities of plasma radicals directly over the wafer surface. Thus species type and concentration could be linked (directly!) to resultant surface structure. One must not strain too hard to imagine a situation in which tool parameters could be modified based on species content generated above

the expanding plasma sheath, with the user being guided by a set of fixed morphological relationships between species and surface nano-structuring which dictate density and separation.

3.2.4 Modifying the percentage of open area

The three absolute priorities when reducing the size of silicon open area exposed to plasma for etching is first to realize the macro-masking material (photoresist, nitride, oxide, metal etc) will be partly etched and join the set of reactions within the bulk plasma discharge, and thus influence macroscopic parameters of the discharge (chamber bias voltage for example). Secondly, by shear proximity that some of the etched mask will diffuse to and join reactions at the wafer surface and influence the micro-(nano in this case) scopic process properties. Several trials with nitride and photoresist masks showed less than optimal results and for the sake of maintaining a realistic project scope an oxide mask was implemented and its chemical influence was subsequently ignored (obviously some negligible amount of the etched oxide mask will enter the plasma and could be described as a second oxygen source). Thirdly, the etch rate overall will drastically increase as the ratio of available fluorine radical etchants to silicon surface increases. With this third point in mind (and although the ideal form reached on unmasked silicon wafers was a multi-step process) it is a logical to assume the ideal starting point be a reduction of plasma ionization by lowering the power (less fluorine radical etchants) and a slight move further into the passivation gas ration region on the curve of silicon etch rate (increasing O_2 content with respect to SF_6) to reduce the etch rate shown previously in figure 3.12. Several trials are shown in figure 3.36. Of the three the sample approaching the etch rate of the first step of the ideal unmasked wafers (5.5 nm/s) does appear to have the best set of preliminary structures.

A more general distribution of ten trials is depicted in figure 3.37. Several important conclusions can be extracted from this figure. The two high power upper RF samples (1 and 3) exhibit rather opposite behavior about the ideal 5.5nm line.

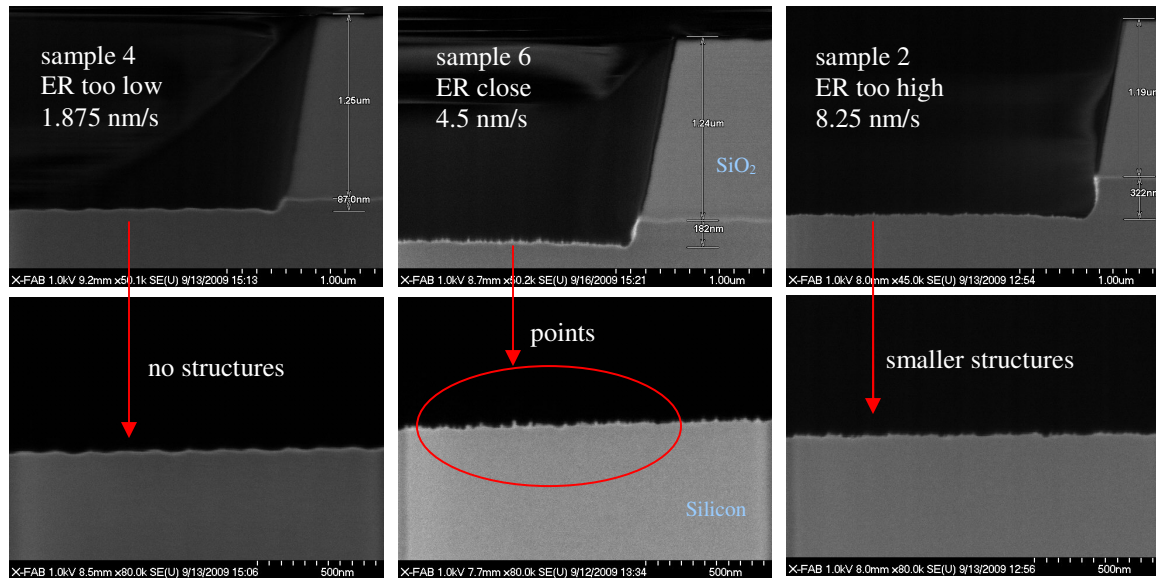
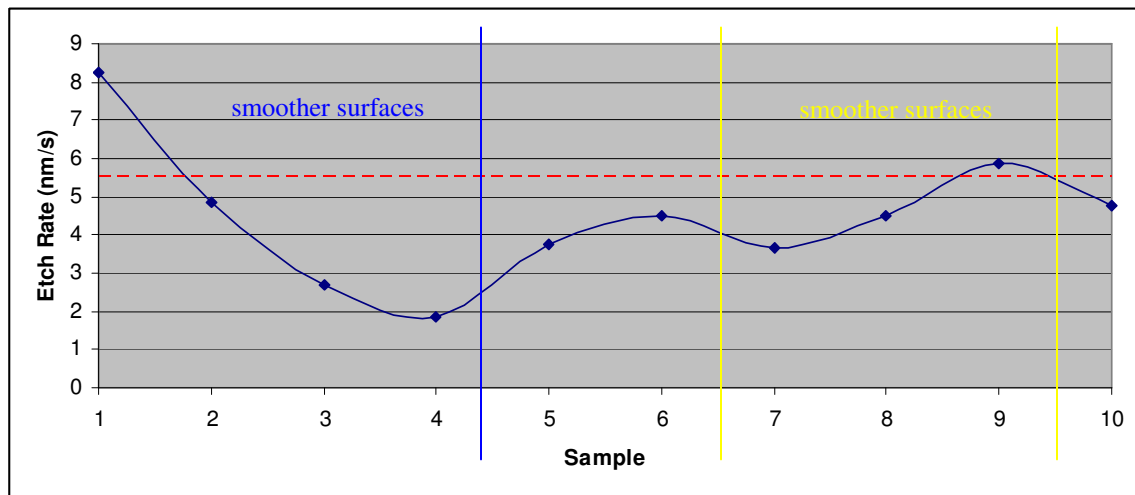


Figure 3.36: Finding the correct masked recipe by first approaching the correct etch rate from unmasked wafers (5.5 nm/s) as an initial guideline. Clearly the 4.5 nm/s sample is the closest and the SEM pictures show a correspondingly greater structure presence.

Exactly as mentioned earlier the influence of gas ratio is here the deciding factor. For sample 3 in spite of a large power coupled into the plasma the etch rate bottoms out because of the general

lack of fluorine content in the gas mixture. The next observation to make is that the samples 1 through 4 have smoother surfaces and that this is very likely due to the large lower RF voltage. A greater number of incident ions are bombarding the surface with enough energy to remove any oxide passivants. Samples 7 through 9 also have smoother surfaces however in the case of sample 7 and 8 it is likely too little lower RF power resulting in no physical ion etching of surface passivants and for sample 9 too much SF_6 for the given amount of exposed silicon (somewhat confirming the initial assumption that the etchant to passivant gas ratio should from the beginning be moved slightly to the passivant side from the blank wafer recipe). This leaves only samples 5, 6 and 10 of which samples 5 and 10 while exhibiting uniform macroscopic etch rates have a less than robust distribution of nano-structures from center to edge.

It is furthermore clear that once present the form and density of structures on such a small scale are no longer exclusively an activity of SEM evaluation. As a result of the etching needed to open the widows in the oxide mask itself it is clear that the silicon surface being etched this time from the beginning is not contaminant free and certainly not as smooth as the blank silicon wafer. Certainly the tiny point like structures observed thus far are quite different from the pillars created in the first step of the blank wafer recipe to be used for nano-loading.



sample	upperRF	lowerRF	Rfbias	press.	SF ₆	O ₂	CF ₄	Temp.	time	Etch rate
1	med	hi	0.94	med	med	med	low	low	40	8.25
2	low	hi	0.97	med	med	med	low	low	40	4.875
3	med	hi	0.89	low	low	hi	med	low	40	2.7
4	low	hi	1	med	low	hi	med	low	40	1.875
5	low	med	0.86	med	med	med	low	low	40	3.75
6	low	low	0.71	low	med	med	low	low	40	4.5
7	low	very low	0.56	low	med	med	low	low	40	3.675
8	low	very low	0.56	low	hi	med	low	low	40	4.5
9	low	low	0.69	low	hi	med	low	low	40	5.85
10	low	low	0.71	low	med	med	hi	low	40	4.75

Figure 3.37: Graph of initial trials to identify masked wafer process conditions for structures. Values are unitless and normalized to themselves or shown in the hi, med, low form of the last section, although are not the same hi, med, low (they are from the beginning a factor lower). See text for detailed analysis.

Reducing the upper RF power significantly below the low value begins to move outside the tools ability to maintain a plasma, let alone ignite it. The ratio of SF_6 to O_2 must be preserved to ensure the correct chemistry, there discrepancies in nano-scale etch rates must be present for structures to form, thus reducing the percentage of SF_6 further to less than 30 percent has already proved to create only smooth silicon surface. Reducing the pressure to below the low value is an unavailable option for the turbo pump given the already low gas flow values and therefore reducing the overall amount of gas while maintaining the overall ration (for example low SF_6 and low O_2) is also not an option. Indeed this last idea brings the process back into the small initial (hypersensitive) process window of figure 3.15 from the previous section. The process could be here abandoned or as proposed in the last section a user could simple work smarter than the tool and perform some set of surface preparation steps to allow for the creation structures through a set of loading steps. Just such a method is successfully implemented here.

As can be seen from the enlarged 500 nm scale photos of figure 3.36, differences in morphological properties between samples 2 and 6 are, with the exception of macroscopic etch rate, not easily seen while differences in center to edge uniformity of a single wafer can be nearly impossible to evaluate.

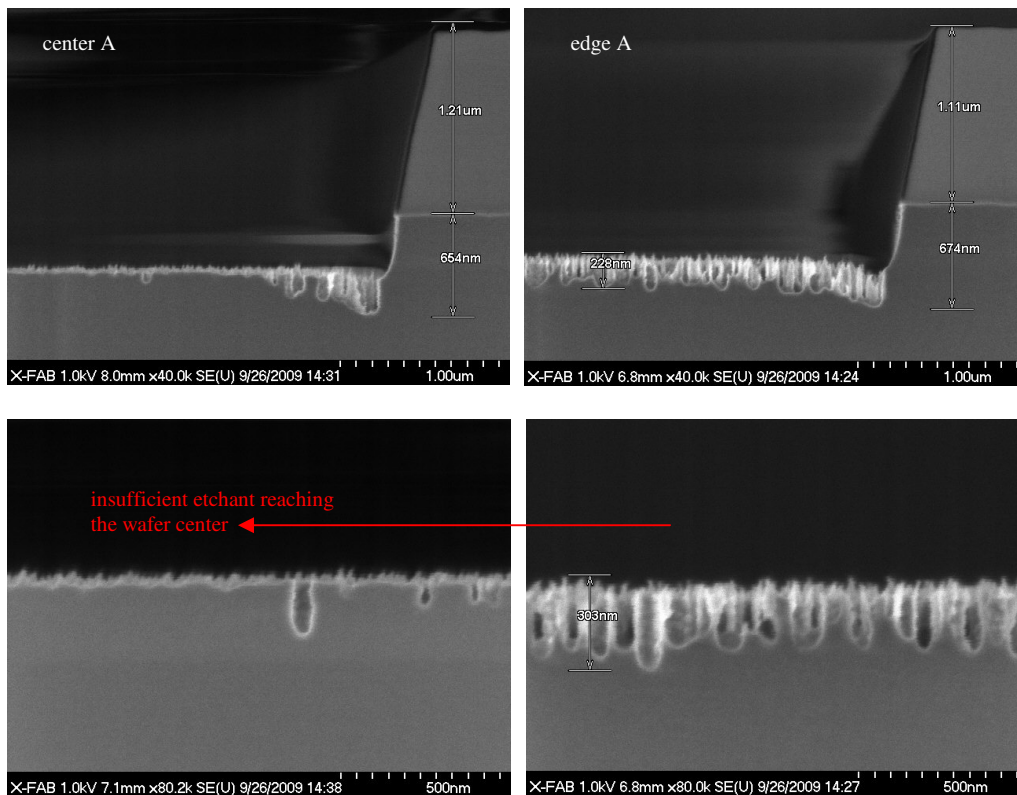


Figure 3.38: Dependence of preliminary nano-masking transfer (Sample A). Etch gas appears to not be reaching the center of the wafer.

Now imagine a situation in which ten samples are present all with relatively similar macroscopic etch rates and all having visible surface structures similar to the ‘points’ sample (indeed a situation encountered here!). Without the need of integrating further some times very time consuming analysis tools (such as AFM) a very effective solution applied in this pursuit was to add a second etch step to ensure homogenous structure distribution.

The results of etching the points of the ‘points’ wafer, a sample with an apparently very good center to edge uniformity, are shown in figure 3.38. The SEM photos show the surface after the two step process, unfortunately they resembling the undesirable uniformity of figure 3.18. This stresses two points. The first is the importance of the theoretical notions presented in section 3.1, namely that the morphological properties of structures with scales measured in hundreds of nano-meters (or their existence whatsoever) are most certainly dependant on the initial behavior of sub-nanometer phenomena, as seen above while the second step performed alone yields no nano-structures. The second point has not yet been addressed, searching for a suitable second step with dependences and non-uniformities compatible enough such that the combination of step one and two yields a uniform set of structures.

Table 2: Recipe steps separated by “/”

sample	upperRF	lowerRF	Rfbias	press.	SF ₆	O ₂	CF ₄	temp.	time
A	low/very hi	low/hi	.71/.78	low/low	med/low	med/hi	low/med	low	40/100
B	low/low low/very	low/low	.71/.72	low/med	med/low	med/hi	low/med	low	40/200
C	low/very hi	low/hi	.71/.79	low/med	med/low+	med/hi	low/low	low	40/100
D	low/very hi	low/hi	.71/.78	low/med	med/low	med/hi	low/med	low	40/200
E	low/very hi	low/hi	.71/.78	low/med	med/med	med/hi	low/med	low	40/100

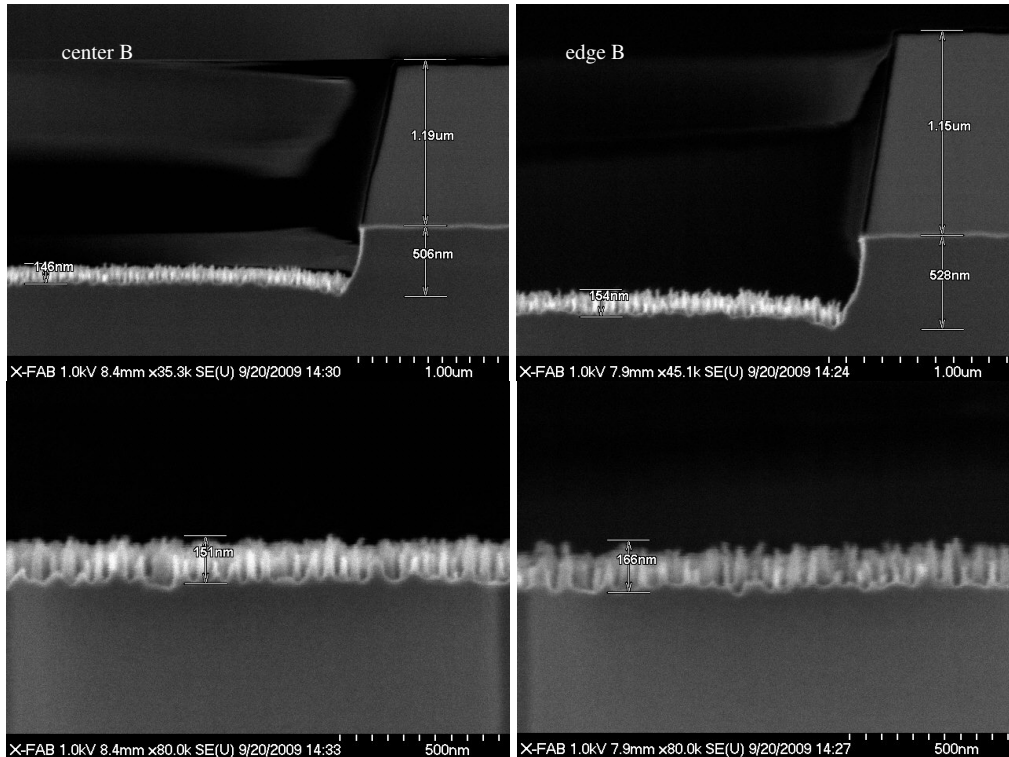


Figure 3.39: Sample B. Not an unreasonable amount of consumed silicon. Etching of structures appears relatively uniform.

The sensitivity of the very small open area of silicon requires a more delicate choice of subsequent steps. A set of nano-points as those shown in figure 3.36 cannot hope to provide a loading effect at for example tool parameter maximum values when the upper RF is set to 700 W, the bottom RF to 300 W, and the SF₆ gas flow to 200 sccm (an aggressive isotropic blanket silicon strip recipe).

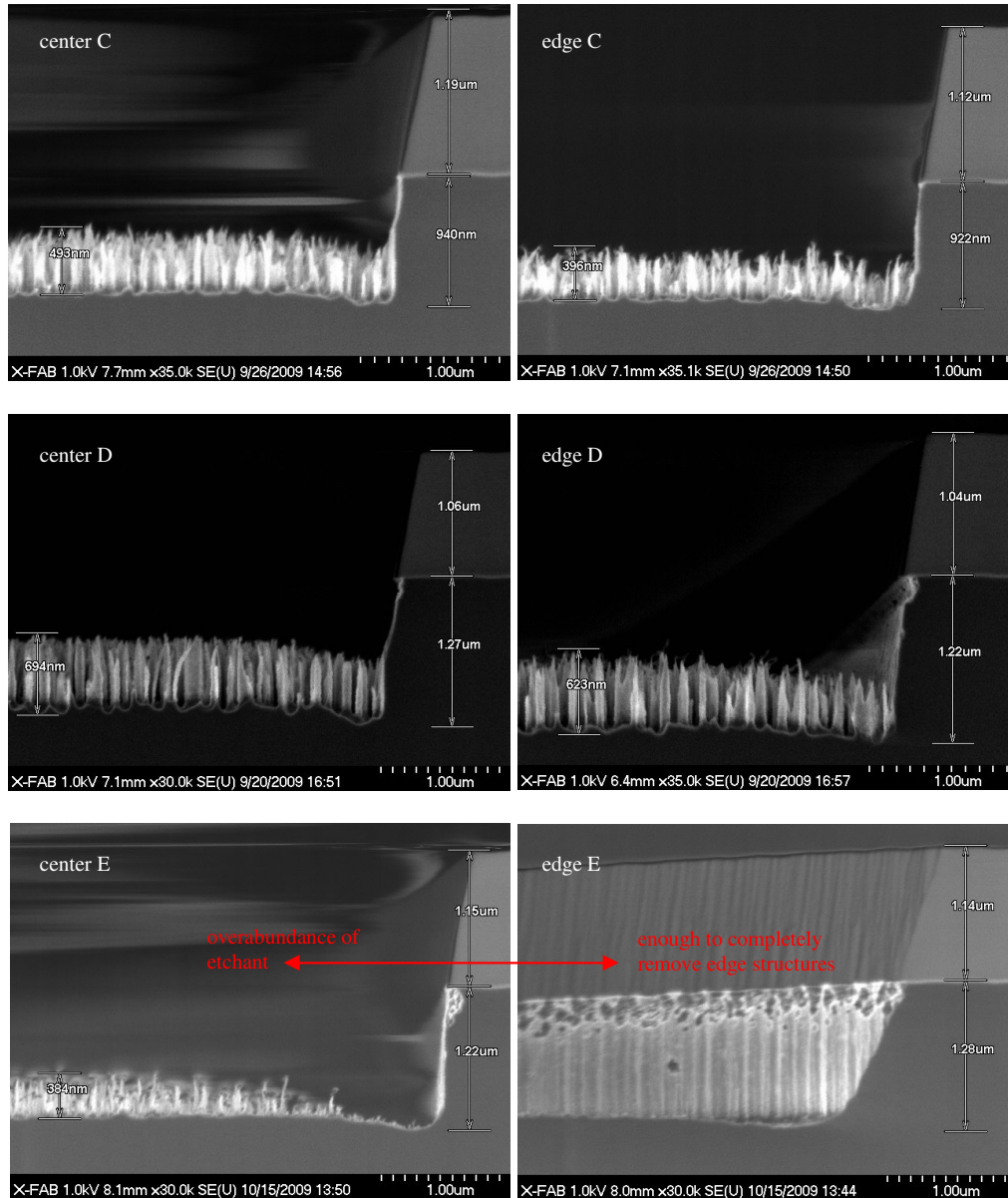


Figure 3.40: Sample C (top) Sample D (middle) Sample E (bottom). All samples show far too much consumed silicon.

Proper recipe selection must be done more wisely. Thus, seeing as the macroscopic etch rate target has been reached but without a definite set of pillars, the subsequent short term goal is now attempting to locate the ideal structures of figure 3.19 from figure 3.38 above, exactly as performed in the beginning with the blank samples, however out of the combination of two steps.

Table 2 provides parameter information for 5 two step samples the combination of which in all cases leads to a set of pillars to be used for texturing pyramids.

Figure 3.39 shows sample B and figure 3.40 shows samples C and D with structures half the size and twice the size of the ideal structures of the previous section. The overall consumed silicon for sample B is quite reasonable and for sample C it is not completely out of the question. The consumed silicon of Sample D and E is beginning to boarder on too much. These are still just attempts to generate a nano-loading effect and already more than 1.2 μm of silicon is gone. The greater abundance of SF_6 present in the processing of sample E is not drastically changing the perpendicular etch rate in comparison to sample D, but an increase in what could rightly be called the lateral etch rate is observed. The point structures of step 1 are mostly cleared away at the wafer edge and barely surviving at the wafer center. Signs of this are available in the pictures of sample D upon closer inspection. The tips of the structures appear to be etching more quickly at the edge in comparison to the center. Interestingly the samples E and A present reverse situations, E has structures present at the wafer center and not at the edge while A has structures present at the wafer edge and not at the center but for reversed reasons. The increase in pressure (decrease in the rate of out-pumping) and the increase in SF_6 of E provided a significant amount of SF_6 for reaction, and as can be seen the perpendicular etch rate is now more than double however the quantity of SF_6 at the wafer edge is too large while for A the quantity of SF_6 at the wafer center is too small.

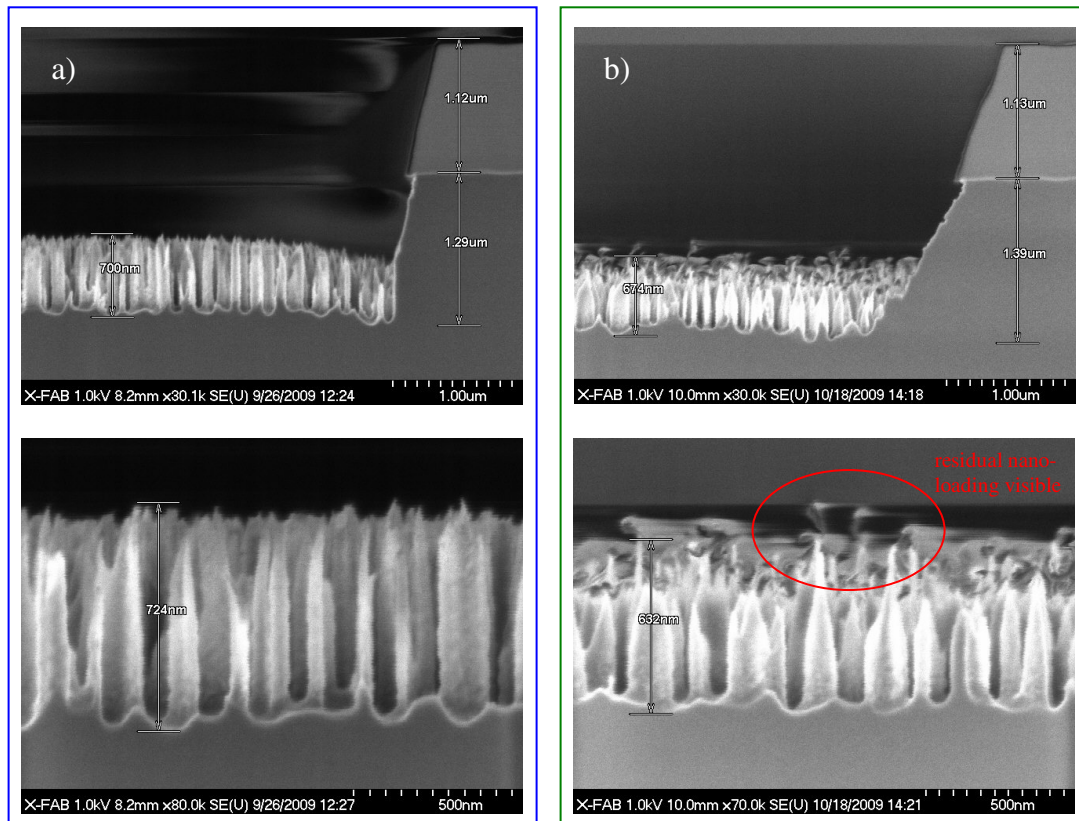


Figure 3.41: The center sites of two different masked wafers. Part a) shows a sample having been etched with a very small set of nano-loaders (from sample B), while part b) shows the result of etching the long structures (from sample C).

Of the structures it is sample B which seems quite ideal to move forward with, however upon etching it is very interestingly seen that the longer structures of sample C are the most conducive to further process. Figure 3.41 presents the center sites of two wafers and provides a relatively good example of what is meant. Part a) shows a sample having been etched with a very small set of nano-loaders (from sample B), while part b) shows the result of etching the long structures (from sample C).

Several other attempts were made to etch masked samples using structures with smaller nano-loading effect structures, but met with no success comparable to that of using the longer structures for nano-loading. The difference is quite striking. The structures of figure 3.41 part b) are clearly the correct wide based structures with observable taper ready to be further smoothed into pyramids exactly as was done on the blank wafers. The smaller nano-loaders, shown after etching in part a), are still definitely influencing the surface however it seems as though the strong pillar anisotropy is the result of the loading effect of the previous step simply setting the starting points for the next set of pillars. This is indeed one of the stated goals of nano-loading, however the second goal of introducing other chemistries and processing parameters not normally associated with initial needle formation but nonetheless desired to influence structure shape (significantly reduced SF_6 content in this) does not seem as effective here with the smaller nano-loading structures. Before continuing further an explanation needs to be offered.

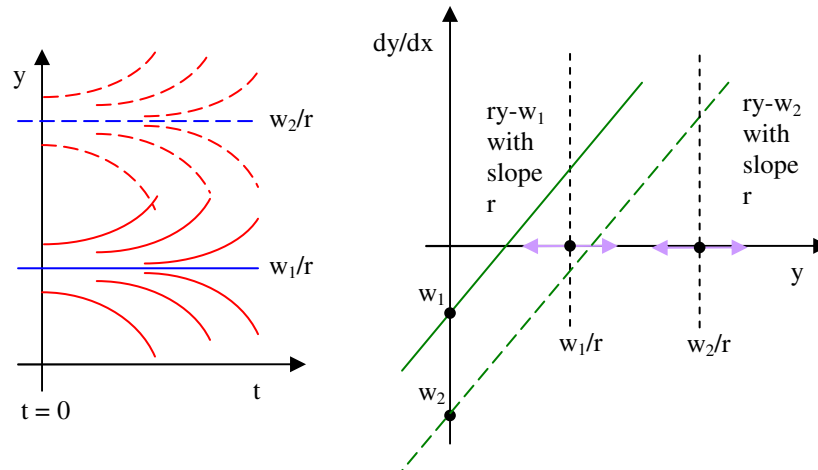


Figure 3.42: Differential etch rate of nano-loading effect. See text for complete explanation.

The paramount logical question is why are such long needles necessary for nano-loading on the masked wafers? There are two very plausible answers to this question. The first hypothesis is that although attempts have been made to reduce the etch rate to accommodate the smaller open area of the masked wafer, the process conditions are still not exactly the same as the blank wafer. The overall etch rate remains slightly more aggressive than desired and as such the nano-loading effect must be accommodated for the speed at which the nano-mask of pillars is etched away, in comparison to the blank wafers

Considered in section 3.2 was the differential etch rate, where y was the height of the structure, r was the etch rate at the base of the structure (rate of growth) and w the etch rate at the apex (rate of loss). In figure 3.42 the rate of loss w_1 is the rate of silicon loss from atop the nano-pillars used on blank wafers and the rate of loss w_2 is the rate of silicon loss from atop the nano-pillars used on masked wafers. Because of the similarities of the widths and heights of the initial structures on both masked and unmasked wafers the consumption at the base of the pillars r (the increase in height) will be considered constant in both cases. Indeed the whole idea is to introduce

a fixed r at the base of the pillars. Clearly with a greater value of the etch rate w , the consumption of the nano-loading mask, a greater initial height y is necessary

For the second answer it is entirely possible that silicon itself places a role as a catalyst. On a blank wafer it is obvious that at any given instant during the processes of section 3.2 there is an abundance of SiF_4 traveling just above the wafer surface being evacuated by the turbo pump. If the situation pictured in figure 3.44 is anything close to true, a reduction of open area can only be compensated by an increase in the size of structures for the nano-loading effect in an attempt to generate a small increase in consumed silicon directly above the location on the wafer being structured. A slightly confusing way of stating this would be that the etching of the needles influences the etching of the needles.

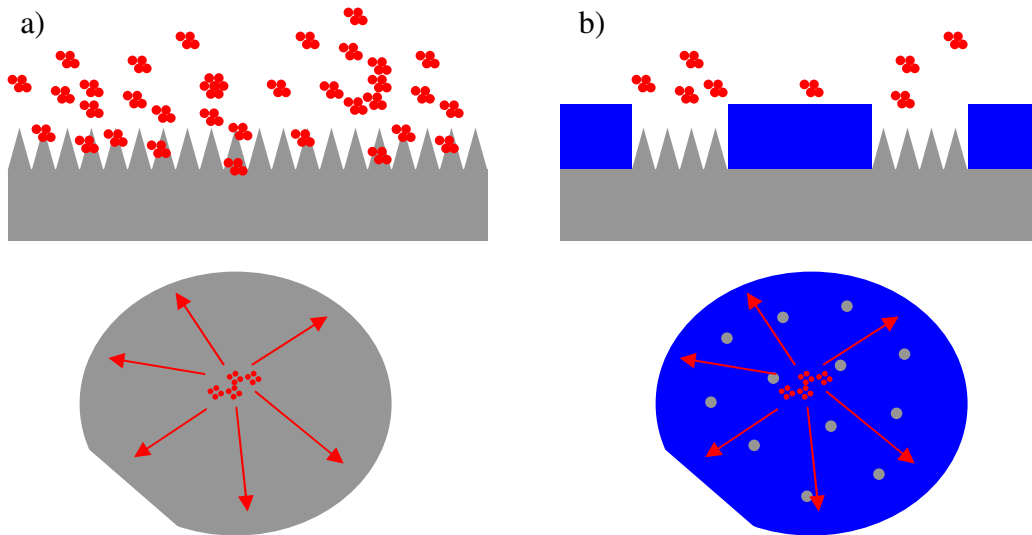


Figure 3.44: Consumed silicon as a catalyst on a) blank wafers and b) masked wafers. On a blank wafer it is obvious that at any given instant during the processes there is an abundance of SiF_4 traveling just above the wafer surface.

The initial etched silicon products play a role in subsequent steps of structuring. At the beginning of this section it was stated that the etch rate would drastically increase as the amount of available fluorine reactant to silicon surface had drastically increased. This current hypothesis can be understood as the exact reverse effect. The amount of silicon reactant (the consumed silicon etch product in the plasma) compared to silicon surface is drastically decreased. As silicon is etched the products are pumped away. Every point on the blank wafer acts as a source injecting silicon product into the plasma, thus each point on the wafer is (theoretically) influenced by all other points on the wafer (with silicon diffusing all over the place).

For the masked wafer however the number of point sources injecting silicon product into the plasma is greatly reduced and the diffusion of the very limited amount of etched product is not guaranteed (not likely at all) to traverse a path which crosses silicon surface, while for the blank wafer all paths to the turbo pump traverse silicon surface. This second situation could be studied using the PROES methods presented in introduction. The individual steps of the optimized structures for the 1% open area of silicon are shown in figure 3.46. Clearly it is possible to view the transition step-by-step from nothing to pyramids. Also seen, along with the author himself, is the silvery appearance of the unetched diode windows before etching and the black appearance of the most certainly etched diode windows after the complete process.

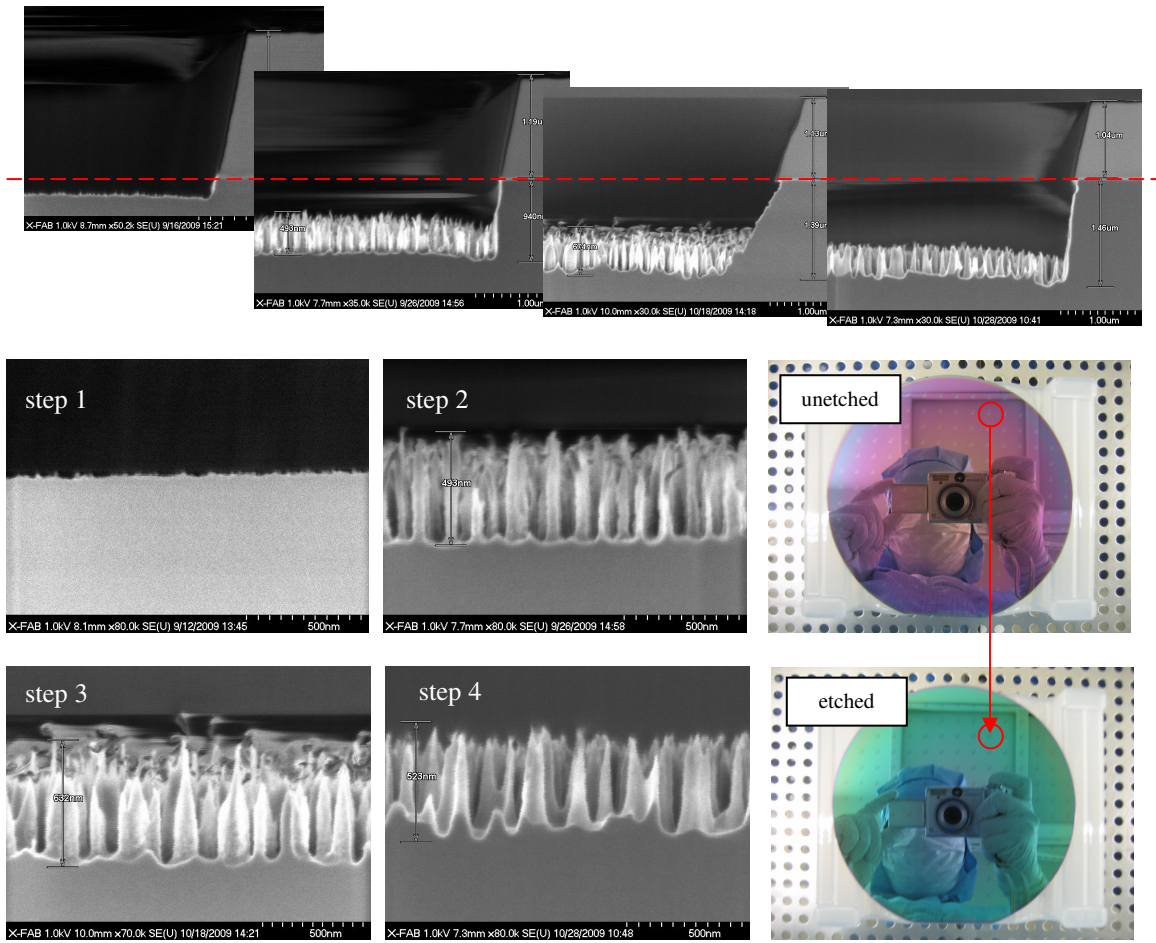


Figure 3.46: Steps of the optimal masked wafer (center sites shown). The top panel follows the overall consumption of silicon. The SEM photos of the lower left show the resulting morphology after each of the four steps. The digital pictures show the wafer unetched (silver diode surfaces) and the wafer after etching covered in small black photodiode surfaces.

Figure 3.47 shows the final structures. Ideal “black” behavior is exhibited until 500 nm at which point the reflection begins to climb steadily upwards, although not in any serious way as even at 800 nm the value of reflection is less than 4 percent. The quick answer is that these structures are not quite long enough to provide a continuous transition from air to silicon for wavelengths longer than 500 nm. A detailed explanation is left until the optical properties section. The shown amount of consumed silicon does not breach the stated maximum of 1.5 μm. The form indeed appears to be very pyramid like and as will be seen in the optical properties section is very friendly towards processes of covering. Of particular importance is to notice the uniformity, center to edge is indistinguishable.

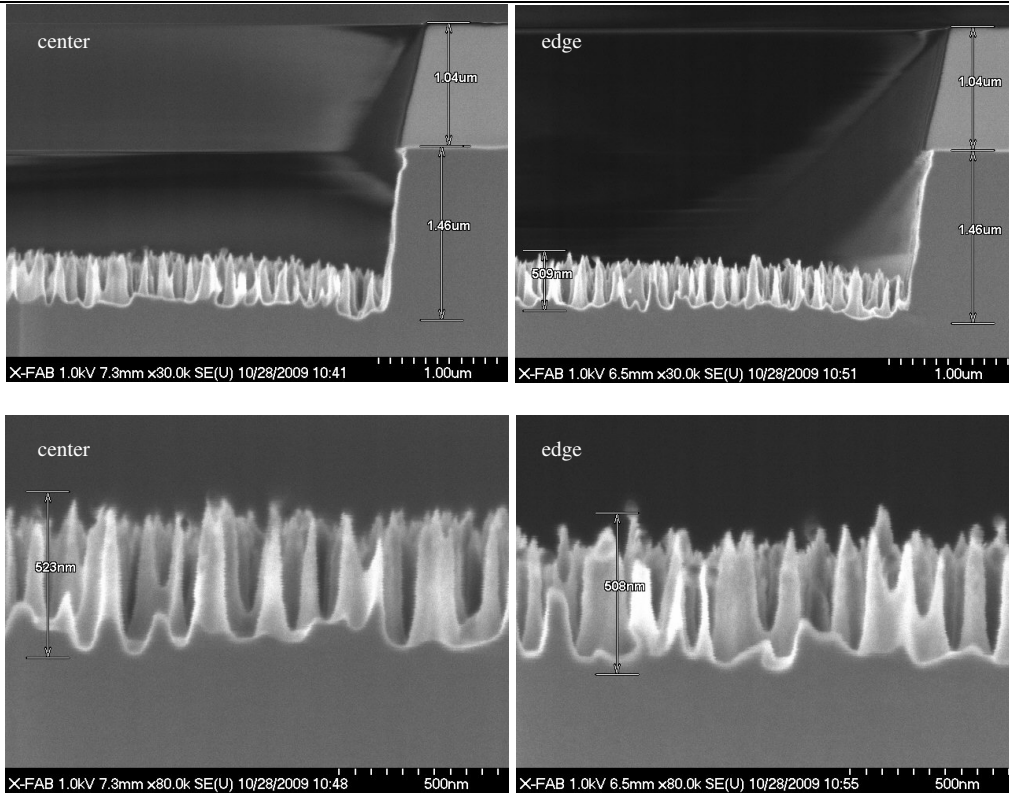


Figure 3.47: Uniformity of the optimal masked wafer looks excellent. Consumed silicon is fine and the structure shape shows a very nice taper from point to base.

3.3 Concluding comments on morphology

The activities and processes presented here have accomplished all stated morphological goals from the beginning discussion of this section. Known plasma chemistries for nano-structuring silicon were implemented, however through additional nano-scale loading effects and further sculpting steps sets of extremely uniform pyramid shaped structures, ideal for applications in reflection suppression, have been created.

4. Optical Properties

"Neither would it carry any imputation of falsehood to our simple ideas, if by the different structure of our organs, it were so ordered, that the same object should produce in several men's minds different ideas at the same time; e.g. if the idea, that a violet produced in one man's mind by his eyes, were the same that a marigold produces in another man's, and vice versa. For since this could never be known: because one man's mind could not pass into another man's body, to perceive, what appearances were produced by those organs; neither the ideas hereby, nor the names, would be at all confounded, or any falsehood be in either. For all things, that had the texture of a violet, producing constantly the idea, which he called blue, and those that had the texture of a marigold, producing constantly the idea, which he as constantly called yellow, whatever those appearances were in his mind; he would be able as regularly to distinguish things for his use by those appearances, and understand, and signify those distinctions, marked by the names blue and yellow, as if the appearances, or ideas in his mind, received from those two flowers, were exactly the same, with the ideas in other men's minds".

- John Locke, Essay Concerning Human Understanding

It is understood that in this passage Locke uses "idea" to mean something like "mental image", rather than its modern meaning and furthermore that he uses "texture" to mean something like "the physical property that gives rise to the experienced color". In modern philosophy this is summarized by the concept of qualia and a well-known example is that of the inverted spectrum depicted in figure 4.01.

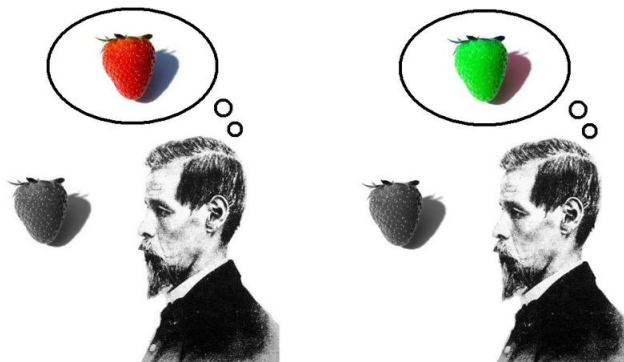


Figure 4.01: The inverted spectrum and the qualia of a strawberry. Two different perceptions of qualia are attributed to the object shown by the two differing visualized mental images. A second and altogether more interesting interpretation of the figure is that neither perception correctly represents the thing-in-itself!

Qualia is a term used in philosophy to describe the subjective quality of conscious experience. Examples of qualia are degrees of pain, taste, and the color of a strawberry as pictured. The figure is also very interestingly an excellent example of Immanuel Kant's thing-in-itself, an object considered transcendentally apart from all the conditions under which a subject can gain knowledge of it. Hence the thing-in-itself is, by definition, unknowable. Or perhaps for this discussion let it be only partially unknowable as this section will not undertake the project of

measuring the unknowable, rather it will undertake a thorough treatment of the partially unknowable subjective quality of, “That wafer is (not) black (enough)”, a concept only superficially introduced in section 2. Section 3.1 reviews reflection from a surface, section 3.2 introduces and implements the Finite-difference time-domain (FDTD) method for computational electrodynamics and to calculate anti-reflective properties of nano-structures. Section 3.3 catalogues the optical properties of some actual physical structures created on wafers and shows the relationship between measurement data and FDTD simulation. Section 3.3 also identifies key weakness is a “moth-eye” oversimplification to explain the antireflective nature of self-organized (meaning randomly organized or better yet, self-UNorganized) structures.

4.1. Reflection from a surface

In this first section we want to briefly remind the reader about the reflection of waves at a surface from the point of view of Maxwell’s equations. We also do not want to assume that the index of refraction for our material is real, meaning that there is an imaginary component and thus absorption in the material. Figure 4.02 shows the propagation vectors k , k' , and k'' for the incident, transmitted, reflected waves respectively. We choose the xy plane to designate the surface and the xz plane to designate the plane-of-incidence with a unit normal vector u as shown.

The two basic kinematic properties of the system are firstly that the angle of incidence equals the angle of reflection and secondly that the system can be described by the principles of Snell’s law: $\sin(\theta_i)/\sin(\theta_t)=n'/n$

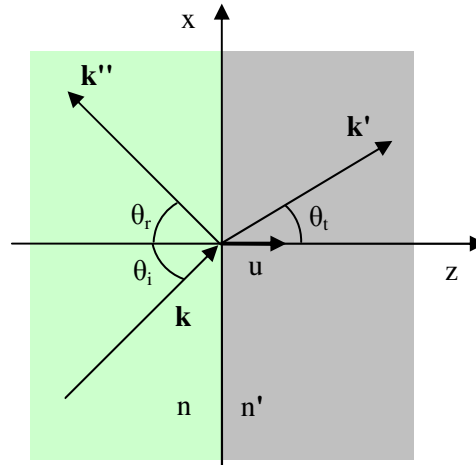


Figure 4.02: The propagation vectors k , k' , and k'' for the incident, transmitted, and reflected waves respectively. The xy plane designates the surface and the xz plane designates the plane-of-incidence [X3.5].

The dynamic properties of the system, the intensities of the reflected wave and refracted wave and the changes in phase and polarization, depend on the electromagnetic fields and their boundary conditions. Both materials in the figure possess a unique permittivity and permeability, yielding indices of refraction $n^2 = \mu\epsilon/\mu_0\epsilon_0$ and $n'^2 = \mu'\epsilon'/\mu_0\epsilon_0$.

Given an incident plane wave with frequency ω , the equations for the incident, transmitted, and reflected waves are [X3.6]

$$E = E_0 e^{i(k \cdot x - \omega t)}$$

$$B = \sqrt{\mu \epsilon} \frac{k \times E}{k} \quad (4.01)$$

$$E' = E'_0 e^{i(k' \cdot x - \omega t)}$$

$$B' = \sqrt{\mu' \epsilon'} \frac{k' \times E'}{k'} \quad (4.02)$$

$$E'' = E''_0 e^{i(k'' \cdot x - \omega t)}$$

$$B'' = \sqrt{\mu'' \epsilon''} \frac{k'' \times E''}{k''} \quad (4.03)$$

The magnitudes of the propagation vectors are

$$|k| = |k''| = k = \omega \sqrt{\mu \epsilon}$$

$$|k'| = k' = \omega \sqrt{\mu' \epsilon'} \quad (4.04)$$

Spatial and time variation of the electromagnetic fields must be the same at our boundary (z=0), meaning

$$(k \cdot x)_{z=0} = (k' \cdot x)_{z=0} = (k'' \cdot x)_{z=0} \quad (4.05)$$

Since the three wave vectors must be in the same plane

$$k \sin(\theta_i) = k' \sin(\theta_t) = k'' \sin(\theta_r) \quad (4.06)$$

And as we have seen $k''=k$, it follows that $\theta_i=\theta_r$ which was our first kinematic property and

$$\frac{\sin(\theta_i)}{\sin(\theta_t)} = \frac{k'}{k} = \sqrt{\frac{\mu' \epsilon'}{\mu \epsilon}} = \frac{n'}{n}$$

which was our second kinematic property, Snell's law.

Now applying Gauss' law and Stokes theorem to a Gaussian surface (a box) straddling the interface between the two materials we can derive the electromagnetic boundary conditions [a very thorough treatment of this derivation can be found in X3.10]

$$\begin{aligned} [\epsilon(E_0 + E_0'') - \epsilon' E_0'] \cdot u &= 0 \\ [k \times E_0 + k'' \times E_0'' - k' \times E_0'] \cdot u &= 0 \\ (E_0 + E_0'' - E_0') \times u &= 0 \end{aligned} \quad (4.07)$$

$$\left[\frac{1}{\mu} (k \times E_0 + k'' \times E_0'') - \frac{1}{\mu'} (k' \times E_0') \right] \times u = 0$$

To apply these boundary conditions it is convenient to separate the incident light into two cases, one in which the electric field is perpendicular to the plane-of-incidence and one in which it is parallel, shown in figure 3.03 a) and b) respectively. The general case of (arbitrary) elliptical polarization can be created from a linear combination of the two.

We consider first the case in which the E field is perpendicular to the plane of incidence. As we can see, this is another way of saying that the E field is parallel to the interface between the two materials and from equation 3.07 it is clear that in this case the first boundary conditions will not help us. However the from the third and forth boundary condition [X3.6]

$$E_0 + E_0'' - E_0' = 0$$

$$\sqrt{\frac{\epsilon}{\mu}}(E_0 - E_0'')\cos(\theta_i) - \sqrt{\frac{\epsilon'}{\mu'}}(E_0')\cos(\theta_t) = 0 \quad (4.08)$$

Solving for the amplitudes of the transmitted and reflected waves

$$\frac{E_0'}{E_0} = \frac{2n \cdot \cos(\theta_i)}{n \cdot \cos(\theta_i) + \frac{\mu}{\mu'} \sqrt{n'^2 - n^2 \sin^2(\theta_i)}}$$

$$\frac{E_0''}{E_0} = \frac{n \cdot \cos(\theta_i) - \frac{\mu}{\mu'} \sqrt{n'^2 - n^2 \sin^2(\theta_i)}}{n \cdot \cos(\theta_i) + \frac{\mu}{\mu'} \sqrt{n'^2 - n^2 \sin^2(\theta_i)}} \quad (4.09)$$

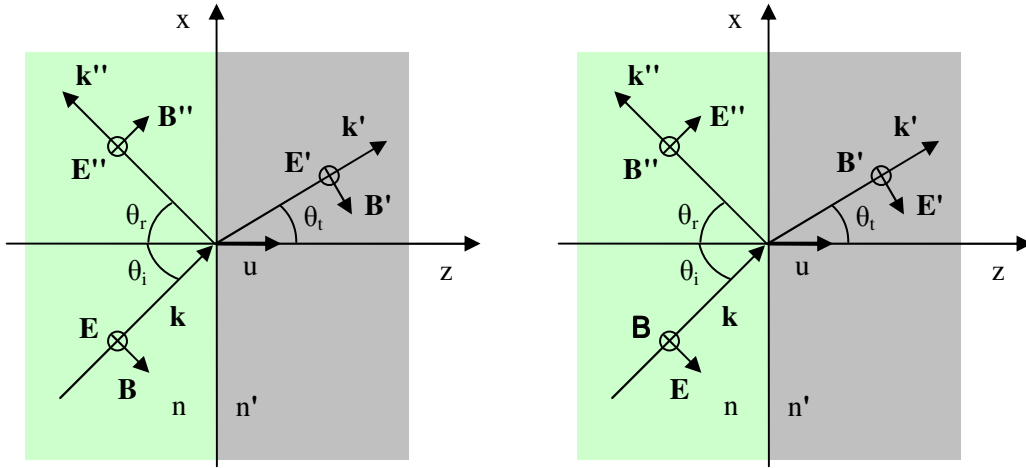


Figure 4.03: Reflection and refraction with polarization a) perpendicular (E is directed away from the reader) and b) parallel to the plane-of-incidence [X3.6].

Now we consider the case in which the E field is parallel to the plane-of-incidence. Again using the boundary conditions from equation 3.07

$$\begin{aligned} (E_0 - E_0'')\cos(\theta_i) - (E_0')\cos(\theta_i) &= 0 \\ \sqrt{\frac{\epsilon}{\mu}}(E_0 + E_0'') - \sqrt{\frac{\epsilon'}{\mu'}}(E_0') &= 0 \end{aligned} \quad (4.10)$$

Again solving for the amplitudes of the transmitted and reflected waves

$$\begin{aligned} \frac{E_0'}{E_0} &= \frac{2nn' \cdot \cos(\theta_i)}{\frac{\mu}{\mu'}n'^2 \cdot \cos(\theta_i) + n\sqrt{n'^2 - n^2 \sin^2(\theta_i)}} \\ \frac{E_0''}{E_0} &= \frac{\frac{\mu}{\mu'}n'^2 \cdot \cos(\theta_i) - n\sqrt{n'^2 - n^2 \sin^2(\theta_i)}}{\frac{\mu}{\mu'}n'^2 \cdot \cos(\theta_i) + n\sqrt{n'^2 - n^2 \sin^2(\theta_i)}} \end{aligned} \quad (4.11)$$

Equations 4.09 and 4.11 and called the Fresnel equations and are valid for reflections at a smooth interface (a sharp transition between both materials). Unfortunately for us if the index of refraction varies between the two materials over a few wavelengths these equations are no longer valid [X3.5]. This is chiefly because the derivation process used to obtain the boundary conditions in 4.07 is no longer valid.

One possible solution to this problem is to solve the equations computationally. The system (both materials and the interface) can be mapped to a grid with each grid cell containing the relevant permittivity and permeability. The cell size should be small enough to accurately model any geometric features of the interface and small enough comparatively to the incident wavelength to accurately capture any transitions between permittivity and permeability. In the next section such a method is introduced and implemented.

4.2. Modeling the electromagnetic field with FDTD

FDTD is a grid-based modeling method in computational electrodynamics used to solve differential equations in the time-domain [X3.11]. Discretized time-dependent Maxwell's equations are solved using central-difference approximations for the partial derivatives of time and space. To solve the ensuing difference equations a leapfrog method is implemented, this means in time and space the vector components of the electric field are solved at a given instant, and directly following in the same space but at the next instant of time the vector components of the magnetic field are solved. The process is repeated in every subsequent cell in the simulation space, to attain the sought transient or steady-state electromagnetic field behavior. Section 3.1.1 describes the discretization process and the leapfrog method in more detail. Section 3.1.2 shows the simulation results and translates the results with the vocabulary of the previous section.

4.2.1 The discretization process and the leapfrog method

The standard way to discretize differential equations is to replace derivatives by finite differences, for example second derivatives by second-order differences [X3.3]. To do so we introduce the Yee notation (the standard notation for FDTD) of space points and functions of space and time as follows [X3.11]. Given an arbitrary space point (i, j, k)

$$(i, j, k) = (i\Delta x, j\Delta y, k\Delta z) \quad (4.12)$$

The quantities Δx , Δy , and Δz are clearly the increments in the lattice space of the xyz coordinate system and i , j and k are arbitrary integers. Subsequently one can represent any space and time function u solved at this discrete space point in the grid (and adding a discrete time point!) by the function notation

$$u(i\Delta x, j\Delta y, k\Delta z, n\Delta t) = u_{i,j,k}^n \quad (4.13)$$

where Δt is the increments in the (lattice) time, taken to be uniform through the interval of evaluation, and n is another integer. Essentially the current value of the time-step is denoted in the super-script and the current value of the space-step is denoted in the sub-script.

For example with this notation the first (partial) derivative in the x direction for the point (i, j, k) using central difference expressions and evaluating the space derivative at a fixed time $n\Delta t$ (time “stands still” as we differentiate) appears as

$$\frac{\partial}{\partial x} u_{i,j,k}^n = \frac{u_{i+\Delta x/2,j,k}^n - u_{i-\Delta x/2,j,k}^n}{\Delta x} \quad (4.14)$$

The time-step remains at n in the superscript of u and for the desired space point $(i\Delta x, j\Delta y, k\Delta z)$ the difference is taking between the space-steps on each side of (i, j, k) at points $(i+\Delta x/2, j, k)$ and $(i-\Delta x/2, j, k)$. A similar expression is derived when evaluating the time component with the same central difference expression holding the space point (i, j, k) constant.

$$\frac{\partial}{\partial t} u_{i,j,k}^n = \frac{u_{i,j,k}^{n+\Delta t/2} - u_{i,j,k}^{n-\Delta t/2}}{\Delta t} \quad (3.15)$$

Here the space point subscript (i, j, k) remains the same and the time difference for $n\Delta t$ is taken between the time-steps on each side of n at points $(n+\Delta t/2)$ and $(n-\Delta t/2)$. Applying this notation to describe a one-dimensional propagating wave with central differences and taking the step size for time and space to be the value of 1 yields

$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} \quad (4.16)$$

$$\frac{\partial^2 u}{\partial t^2} = \frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{(\Delta t)^2} \quad (4.17)$$

Now substituting the two central difference equations into a one-dimensional scalar wave equation yields

$$\frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{(\Delta t)^2} = c^2 \left(\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} \right) \quad (4.18)$$

Rearranging terms, the solution for the most current u at a given grid point i is

$$u_i^{n+1} = (c\Delta t)^2 \left[\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} \right] + 2u_i^n + u_i^{n-1} \quad (4.19)$$

As all values of the equation appearing on the right hand side are known, this is said to be a fully explicit second-order accurate solution for u_i^{n+1} . Each value was acquired at n and $n-1$, the earlier time steps, and subsequently stored in the memory of the computer. The complete set of solutions for u_i^{n+1} is attained by simply repeating equation 3.14 over all space points, no simultaneous computation is necessary. The numerical FDTD solution of the scalar wave equation is then completed by repeating the process for u_i^{n+2} and so on across the whole area of simulation.

However we would like to find the solutions for two very specific differential equations, namely a time-varying electric field and time-varying magnetic field. We have already introduced Yee notion, and we can now introduce his very unique grid to map Maxwell's curl equations to a Cartesian coordinate system shown in figure 3.04.

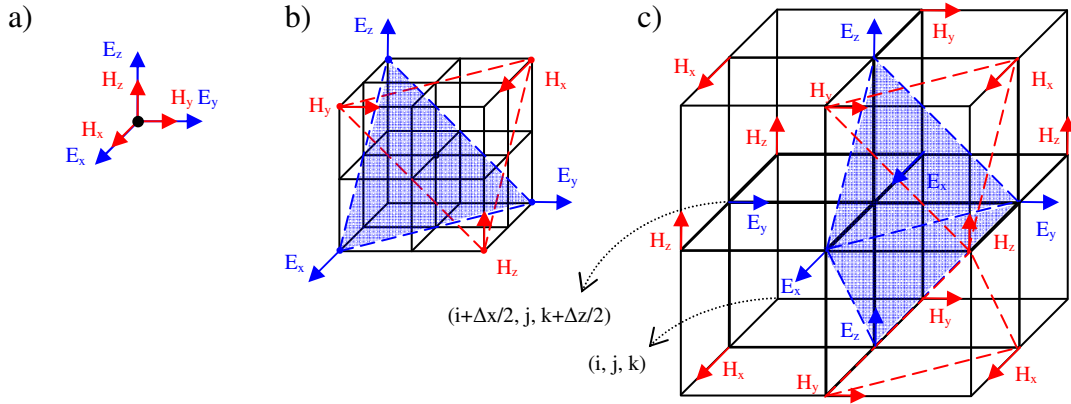


Figure 4.04: a) A point in space showing all six electromagnetic vector components [X3.12, 13]. b) The same point with the modified Yee representation separating all six vectors to specific corners of a cube. c) The finished Yee cell comprised of seven mirror images of the first cube. Notice that in the center of each the face of the cube is an E vector component surrounded by the exact four H components needed to derive this E from Maxwell's curl equations.

As we reviewed in section 4.1 there are normally six vector components (three E components and three H components) at ever point in space there shown in figure 3.04 a). Yee's first step was to separate the six components to specific corners of a cube as shown in 3.04 b). This cube can then be thought of as having "mirror images" of itself taken across its faces and further "mirror images of the images" taken across the faces of the images, essentially filling out the quadrants of figure 4.04 c). As will now be shown this leads to an extremely convenient geometry for discretizing and solving electromagnetic fields.

The two curl equations can be expanded by explicitly showing the vector components of the curl operator yielding a set of six coupled scalar equations, one of which is, in this case, the y component of the electric field is

$$\frac{\partial E_y}{\partial t} = \frac{1}{\epsilon} \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \right) \quad (4.20)$$

Let us calculate this for the E_y grid point in the center of the left most face from figure 3.04 c). However first notice that the two exact H components required (H_x and H_z) surround the desired E_y component!

Thus first denoting the space point at the desired E_y in figure 3.03 as $(i+\Delta x/2, j, k+\Delta z/2)$, then substituting a central difference centered at space point $(i+\Delta x/2, j, k+\Delta z/2)$ and at the time step n .

$$\frac{\partial}{\partial t} E_y \Big|_{i+\Delta x/2, j, k+\Delta z/2}^n = \frac{E_y \Big|_{i+\Delta x/2, j, k+\Delta z/2}^{n+\Delta t/2} - E_y \Big|_{i+\Delta x/2, j, k+\Delta z/2}^{n-\Delta t/2}}{\Delta t} =$$

$$\frac{H_x \Big|_{i+\Delta x/2, j, k+\Delta z}^n - H_x \Big|_{i+\Delta x/2, j, k}^n}{\Delta z} - \frac{H_z \Big|_{i+\Delta x, j, k+\Delta z/2}^n - H_z \Big|_{i, j+\Delta y, k}^n}{\Delta x} \quad (4.21)$$

And we now have a perfectly tailored set of central differences in space (using magnetic field vectors) to describe the variation in time of the electric field vector. The reader can see that simply rearranging the terms in 3.21 will give us a value for the electric field at a future time step using information from present and past time steps. This term in equation 3.21 between the two equal signs (E_y at the future time step minus E_y at the past time step) is called a leapfrog and is depicted in figure 3.05

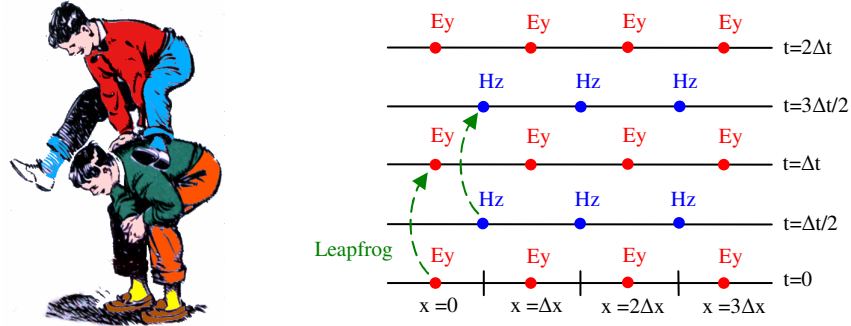


Figure 4.05: For a one-dimensional wave propagating in the x direction “Leapfrog” (named after the children’s game) is used for the time derivatives, while central differences are used for the space derivatives. The components are interwoven at time steps of $\Delta t/2$ with the electric field solved at one time step and the magnetic field solved at the next.

Leapfrog calculates the field values at interleaved time points so that field values 'leapfrog' over one another (with the process taking its name from the children’s game). Leapfrog is called a second order method, meaning that it outperforms other methods such as Euler which are only first order and, unlike Euler, Leapfrog is stable for oscillatory motion

4.2.2 FDTD simulation results of optical properties

The three most significant structures to evaluate with a set of FDTD simulations are the flat silicon surface, the nano-pillar (rectangle with sharp corners), and the nano-pyramid (the distance between the two upper most corners of the rectangle is reduced to zero creating a single apex). These three cases set an excellent set of boundaries on which our investigation can be established. The present application requires the final structures to be covered with an oxide creating an Air/SiO₂/Silicon interface stack, thus trivially it can be noted that all three cases share

the same set of indices of refraction with the only difference being a simple modification to the interface geometry. However it is exactly this simple modification of interface geometry which creates massively non-trivial differences in optical properties. In figure 4.06 the covered structures are illuminated by a 650 nm plane wave at an angle of incidence normal to the flat surface. One immediately notices the electric field E_x color scale and its strong demarcation of the simulation space into regions of reflected intensity, all three beginning with the air/oxide interface and ending with the bulk silicon interface. For the pyramid and pillar a third interface, an oxide/effective-medium interface, is observed.

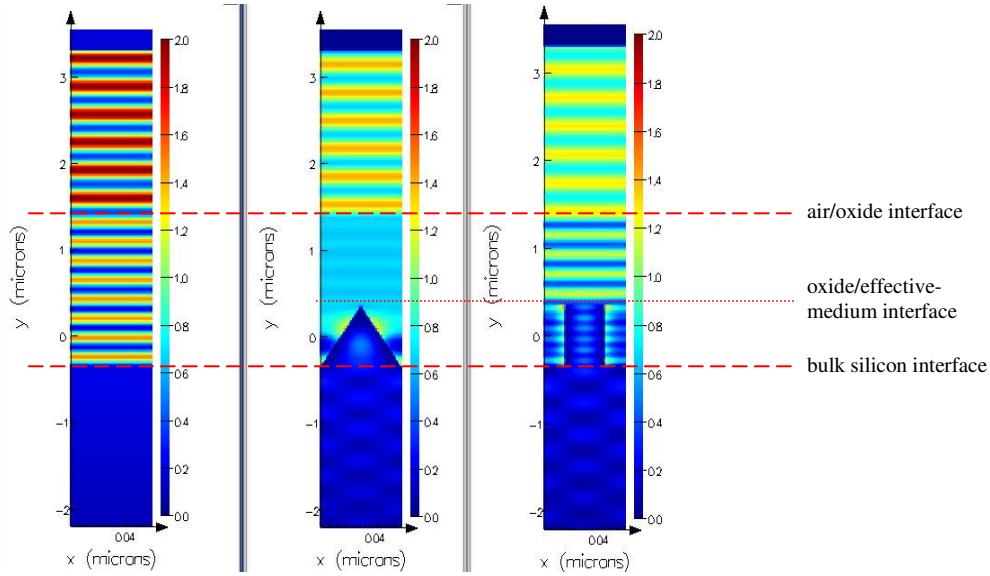


Figure 4.06: Plane wave (650 nm) incident on various surface structures covered with SiO_2 at an angle of incidence normal to the flat surface. The color scale of the E_x field intensity component demarcates the simulation into distinct regions separated by an air/oxide interface, oxide/effective-medium interface, and the bulk silicon interface.

The color pictures are indeed pleasant to look at however it is difficult not to notice the apparent checker board electric field pattern under the pyramid and also appearing (slightly) more intense under the pillar. The following graphs, depicted in figure 4.07, of electric field intensity E_x dependant on the y simulation axis through the direct center of each structure will better illuminate (pun intended!) the phenomenon. Beginning from the right hand side of each graph the incident plane wave travels through the air medium (refractive index $n = 1$) with $\lambda_0 = 650$ nm. It should be mentioned that the substantial increase in the number of data points centered around the line $y = 0$ is the result of an added grid refinement (the number of Yee cells is increased by a factor of 5) to assure an accurate result around the sharp corners of the nano-structures.

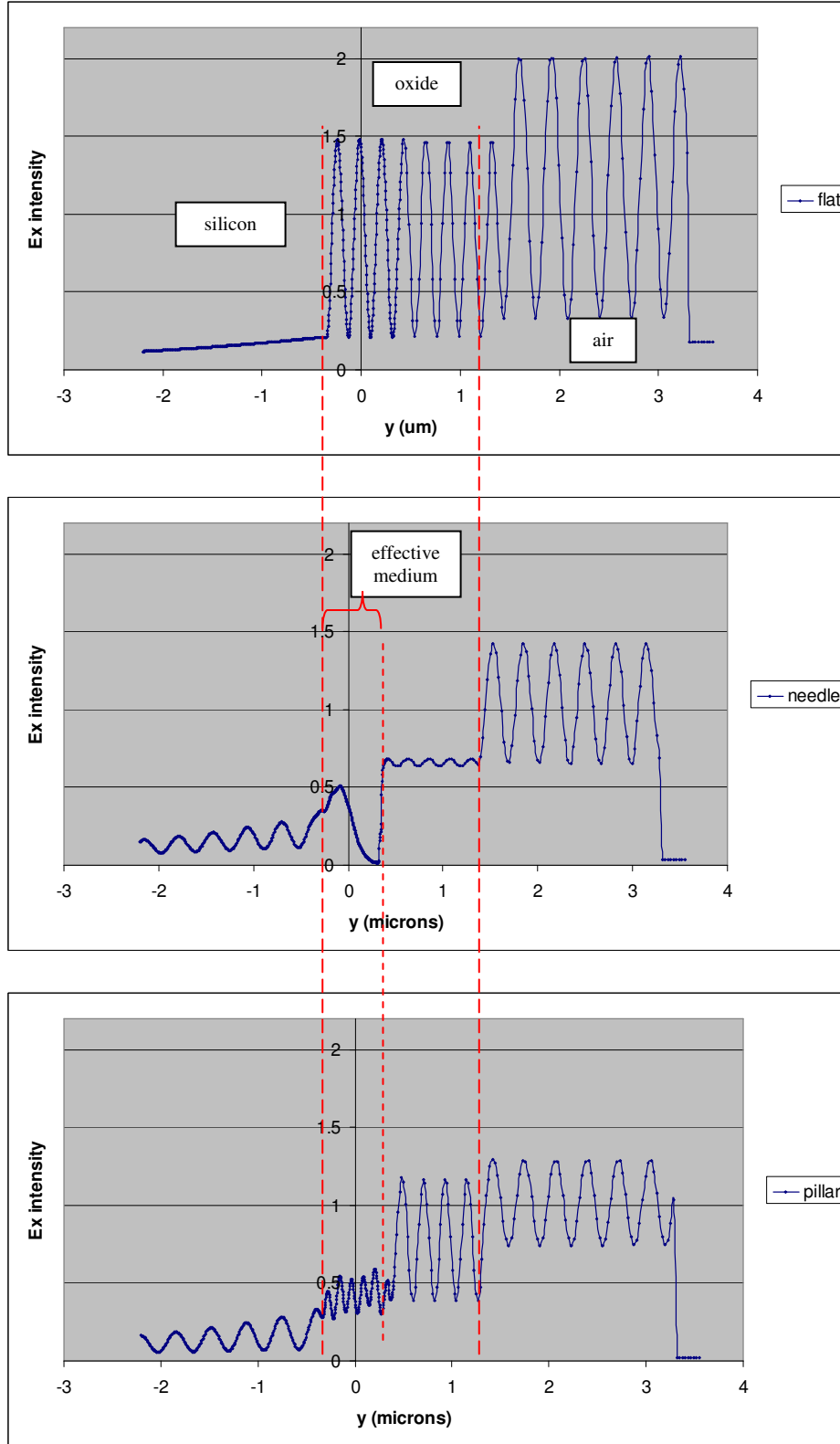


Figure 4.07: Ex intensity vs. y direction taken through the simulation center for a) the flat Si surface b) the needle and c) the pillar. The reflected Ex is the same for both nano-structures?

The grid refinement was preserved in the flat surface simulation to maintain similarity. The initial amplitude of the incident wave is 1 and thus no measured value will be greater than 2, as alterations in visible phase and amplitude (size of observed oscillations) are simply the result of the superposition of the incident wave (traveling in the $-y$ direction) and the numerous phases and amplitudes of the various reflected waves (traveling in the $+y$ direction). As the plane wave enters the SiO_2 medium the waves speed is reduced and travels with a reduced wavelength $\lambda_{\text{SiO}_2} = 650 \text{ nm} / n_{\text{SiO}_2}$ (thus maintaining its original frequency). Now the interesting properties begin. The Ex component of the flat surface simulation, as expected, contains three distinct regions, the pillar simulation contains four distinct regions, and the pyramid simulation contains $3 + \dots$? The effective-medium of the pillar/ SiO_2 mixture functions visibly like a single new layer which yielding an observed wavelength of $\lambda_{\text{effective}} = 650 \text{ nm} / n_{\text{effective}}$ within it, something like an index of refraction exactly between SiO_2 and silicon. The effective-medium of the pyramid/ SiO_2 mixture is more complicated, it appears to have one enormous half oscillation through its center, as though the plane wave has now experienced an increase in wavelength upon traveling into a denser medium! Stepping through the simulation pictured in figure 3.08 will add some clarity to the observed phenomena.

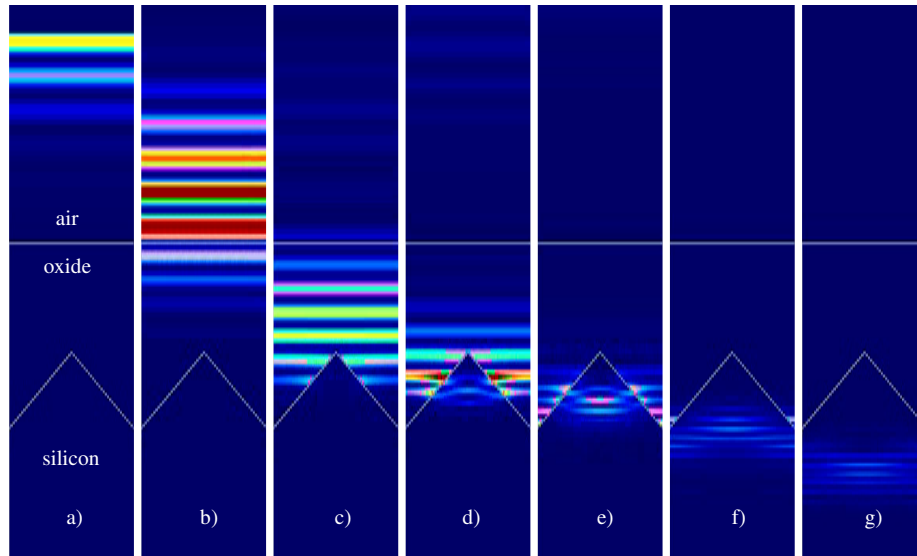


Figure 4.08: Stepping through the pyramid simulation (650nm light) in a) through g). Most interesting is the clear bending of the electromagnetic field in d) and c). Revisiting figure X above, there are clearly no visible standing wave fronts within the pyramid, only a single bright spot in the center.

Starting in part a) the incident pulse is seen approaching the oxide interface, in part b) and c) passing through the oxide and reflecting back a small percentage of the incident light. Part d) and e) show the essential phenomenon. The individual wave fronts are bent as the wave penetrates deeper into the pyramid. Seen on the left and right hand sides of the pyramid are the portions of the wave front still traveling with velocity corresponding to a $\lambda_{\text{effective}}$ in SiO_2 while within the pyramid the wave front now travels much more slowly a with velocity corresponding to $\lambda_{\text{effective}}$ in silicon. This leads to something akin to a traffic jam of wave fronts in the center of the structure and is the explanation for the lack of any visible standing waves within the pyramid in figure 3.06 above, only a single bright spot is seen in the structure's center.

This brings the discussion back to the checker board electric field pattern under the pillar and pyramid in figure 4.05, also appearing in figure 4.06 as the standing wave oscillations of

upsettingly large wavelength within the bulk silicon. Part f) and g) of figure 4.07 show wave fronts entering the bulk silicon with the correct spacing however with some oscillation between dark and light regions of intensity. The phase delay created by the aforementioned traffic jam is certainly the main factor in creating the pattern however an additional factor can be observed by taking the far-field transformation of the Ex field below the substrate surface (this is done in section 4.2.2.4). The nano-structures exhibit properties normally attributed to diffraction gratings, the components of which are too small to be visible in the step-by-step analysis of figure 4.08. However to treat this phenomenon it will first be necessary to discuss some general properties of reflection, and add an explanation for the nearly exact antireflective performance of such drastically different geometries (pillar and pyramid).

3.2.2.1 The effective index of refraction for nano-nanostructures

We begin with the total reflection from the structures in figure 4.09. These results are a combination of specular reflection, mirror-like reflection in which light from a single incoming direction is reflected into a single outgoing direction ($\theta_i = \theta_r$), and diffuse reflection, a process arising from an uneven or granular surface such that incident light is reflected at a number of angles. Although the surface is most certainly textured, it will become clear that it is exceedingly misleading to use the word diffuse with such (well-space) nano-structures.

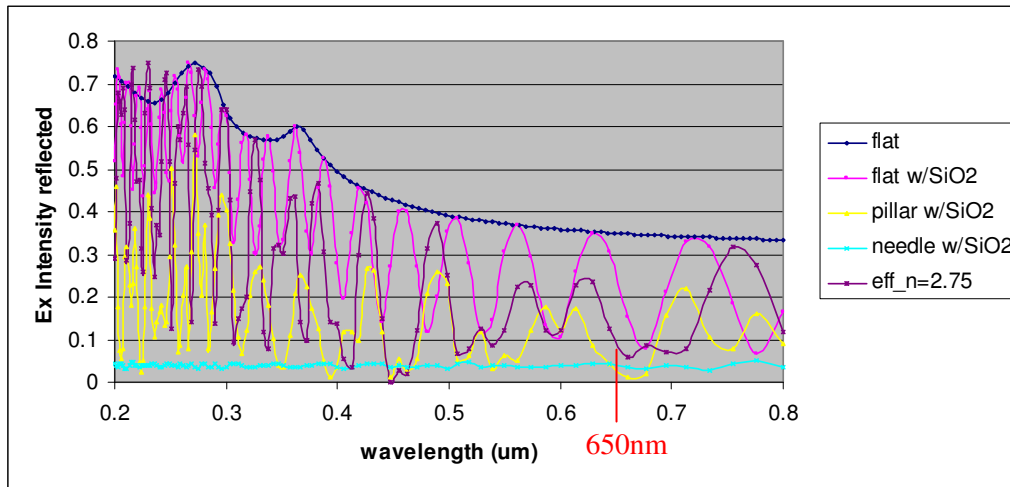


Figure 4.09: Reflection from covered structures. The pillars covered with oxide behave very much like a stack of two separate layers, with the pillar “layer” having an index of refraction, $n = 2.7$ (minus a portion of light absorbed within the “real” silicon pillar). The index $n = 2.7$ is taken as the value directly half way between the standard assumption of $n = 1.5$ for SiO_2 and $n = 4$ for silicon. This stack behavior for the pillar, as expected, leads to quarter-wave transformer properties which bring reflection to zero for several wavelengths. We also see the reason for the near identical performance of structures from the previous section at 650nm (the values overlap).

Notice how the oxide covered pillars behave very much like a simple two layer stack and exhibit properties of a quarter-wave transformer for some wavelengths, bringing reflection to zero. One such minimum in reflection is at approximately 675 nm and with it we see why the antireflective performance of the pillars and pyramids in the previous section was nearly identical at 650 nm (the curves overlap at 650 nm). This stack behavior was in fact already seen in the behavior of the Ex intensity through the covered pillar in figure 3.07 where the wavelength of oscillations within the pillar/oxide medium was reduced to a value which one would expect for a medium with an index between silicon and SiO_2 , implying that the oxide surrounded pillar mixture was just

another “layer” with a fixed index of refraction. An additional simulation depicted in figure 4.08 is exactly that, an isotropic medium with a fixed index of $n = 2.75$ used in place of the pillar, and although the two lines do not completely overlap many similarities clearly exist.

Subsequently one may be able to describe the behavior of the pyramids with an increasing number of layers having indices of refraction between SiO_2 and silicon. A multi-layer stack with fixed values of n is shown in figure 4.10 to explain the antireflective behavior of the pyramids. Each simulation is essentially a 700 nm thick gradient index built from an increasing number of layers where each layer’s index is greater than the previous layer and meant to represent an oxide covered pyramid. It is worth mentioning that such “theoretical” layers with a static index of refraction do not include frequency dependant absorption and the effects of nano-structures on transmission of light will be discussed later. Thus the stack behavior follows silicon pyramid behavior only as far as the index of silicon remains constant ($\approx 450 \text{ nm}$ to 800 nm) in the region interest.

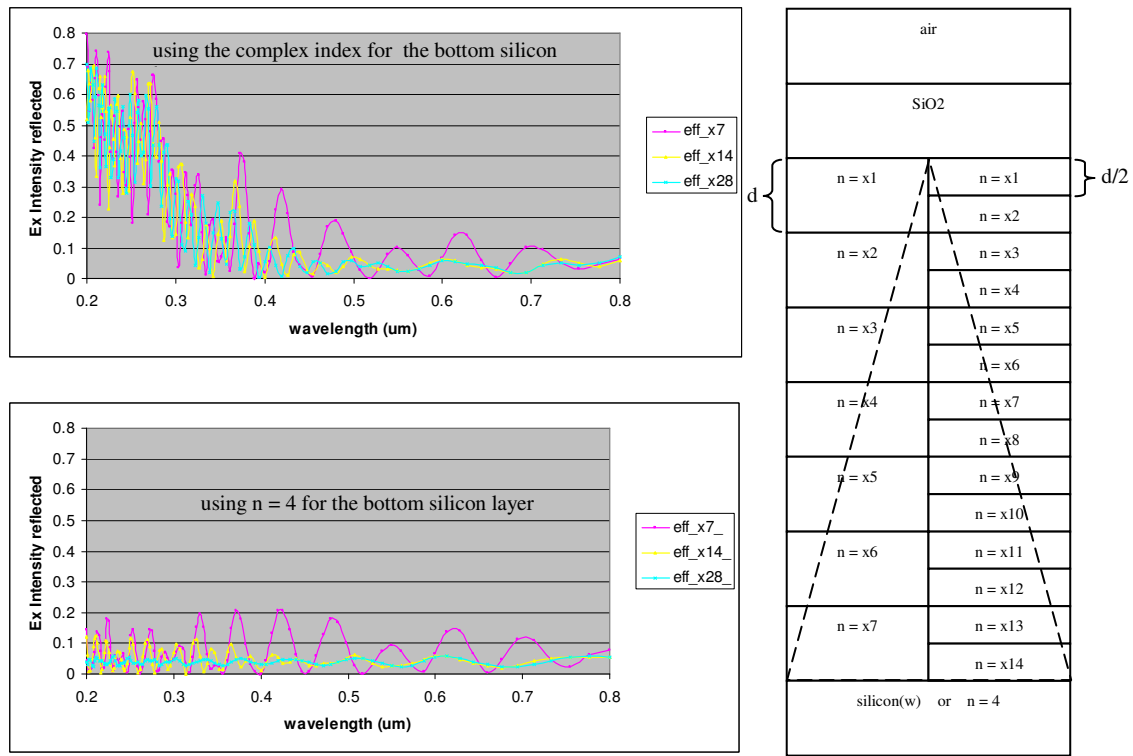


Figure 4.10: Limitations of a multi-layer stack model with fixed values to describe the antireflective behavior of the pyramids. A 700 nm thick gradient index created out of a number of layers with increasing indices of refraction (representative of the oxide covered pyramid) is simulated. The stack behavior follows silicon pyramid behavior only as far as the index of silicon remains constant ($\approx 450 \text{ nm}$ to 800 nm) in the region interest.

Naturally for the 7 layer case the thickness of each layer is 100 nm and the observed oscillations are simply the result of the superposition of reflections from the 7 interfaces (the jumps in index are still quite great). Notice how, as soon as the layer thickness falls well beneath the wavelength of the smallest wavelength in the region of interest (the jumps between index are smaller), the reflected intensity equals only the part reflected from the SiO_2 for the $n = 4$ substrate, the layers are matched for this case. However the index of refraction for silicon is no longer constant for wavelengths below 450 nm, it is very frequency dependant in this part of the region of interest in

for example contrast to that of the SiO_2 covering. This is shown in figure 4.11. A multi-stack model with fixed values for the indices of refraction of the individual layers thus, regardless of the number of layers, simulates the behavior of nanostructures very poorly as it cannot take into account the frequency dependant behavior of silicon.

A natural question to ask is what causes the index of refraction to be so varied in the region of interest for silicon and not for SiO_2 . Speaking generally, materials that are seen as transparent have characteristic frequencies which lay outside the spectrum of interest. The natural frequency of SiO_2 for example, a substance encountered everyday and always experienced as transparent, exists far below the visible spectrum where it becomes quite opaque in the ultraviolet. A second example is when one considers why Gallium Arsenide (GaAs) is a commonly used material for lenses in the infrared (2 μm to 30 μm). In this region GaAs is extremely transparent in spite of the fact that when one is holding the lens it appears completely opaque (GaAs is certainly not transparent in the visible spectrum). To account for this more thoroughly it is essential to incorporate the atomic nature of matter and its inherent frequency dependant properties into an explanation.

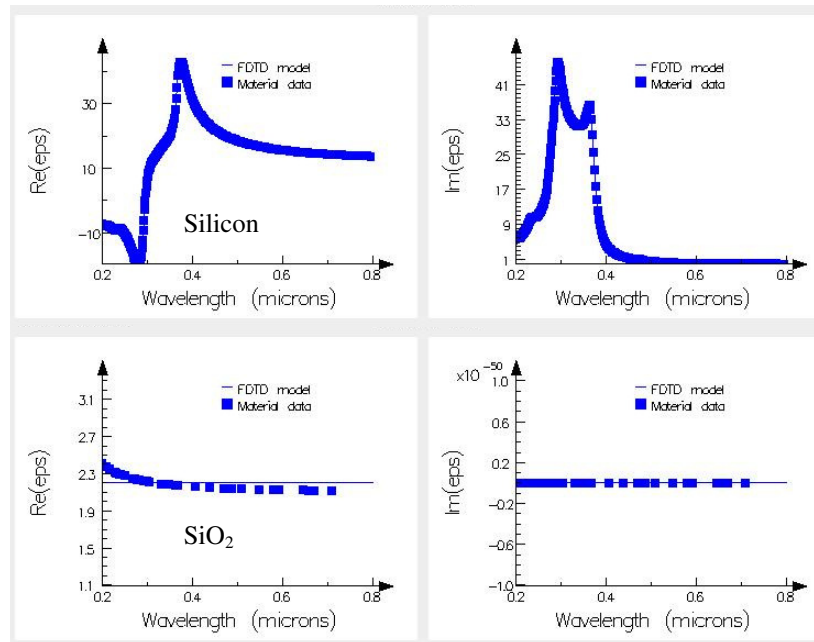


Figure 4.11: Complex index of refraction for silicon and SiO_2 [X3.9]. Two resonance peaks are seen between 200 nm and 400 nm for silicon. This presents a problem for utilizing silicon as a detector in this region.

When a material experiences an applied electric field, for example from incident light, a temporary change in the distribution of charge within the material is created. More specifically for the individual atoms the field applied alters the position of the negatively charged electron cloud with respect to the positive nucleus. This is pictured in figure 4.12 where it can be quite accurately represented as a mechanical system of spring compression and expansion. We will briefly remind ourselves of what takes place within this system as the result is not only necessary to interpret our simulation results but also to understand physical measurement data in the next section. With systems of this kind we expect to find an equilibrium position and a subsequent restoring force working against any disturbance (the electric field in this case) which returns the system to equilibrium. The form of the restoring force should be that of Hooke's law, $F = -k_E x$, where x is the displacement and k_E in this case could be thought of as an "electrical" spring

constant. If momentarily displaced the system will oscillate at its natural frequency $\omega_0 = (k_E/m_e)^{1/2}$, where m_e is the mass of the displaced electrons. Rearranging terms gives $k_E = m_e(\omega_0)^2$. However this is a driven system, as mentioned there is an electric field displacing the electron cloud with frequency ω . The force experienced by an electron in an electric field is $F = q_e E$, where q_e is the charge and E is the (in this case time varying) electric field strength. Thus the total force is the sum of the driving and restoring force and from Newton's Second Law these are set equal to the mass times the acceleration [X3.1]

$$F = q_e E(t) - k_E x = q_e E_0 \cos(\omega t) - m_e \omega_0^2 x = m_e \frac{d^2 x}{dt^2} \quad (4.22)$$

This is a second order linear differential equation and we can expect that the electrons will oscillate at the same frequency, ω , as the incident light. Thus the solution should look something like

$$x(t) = x_0 \cos(\omega t) \quad (4.23)$$

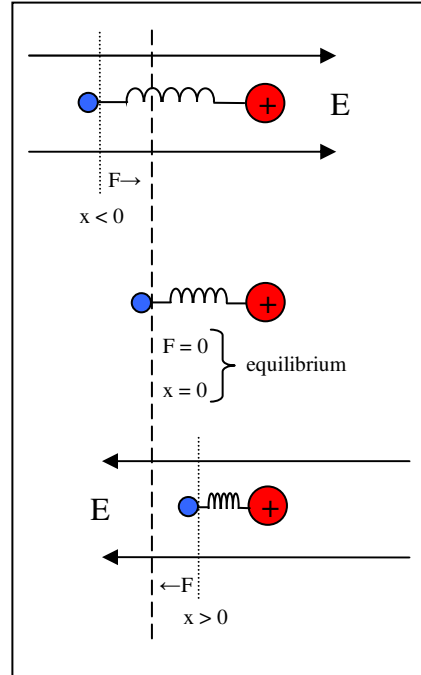


Figure 4.12: Deformation of an atom's electron cloud in the presence of an electric field. The behavior of the electrons can be visualized as the mechanical compression or expansion of a spring and its subsequent restoring force working against the electric field [X3.1].

Substituting this into equation 4.22

$$q_e E_0 \cos(\omega t) - m_e \omega_0^2 (x_0 \cos(\omega t)) = m_e \frac{d^2}{dt^2} (x_0 \cos(\omega t))$$

$$q_e E_0 \cos(\omega t) - m_e \omega_0^2 (x_0 \cos(\omega t)) = -m_e \omega^2 x_0 \cos(\omega t)$$

$$q_e E_0 \cos(\omega t) = m_e x_0 \cos(\omega t) (\omega_0^2 - \omega^2)$$

and solving for x_0

$$x_0 = \frac{q_e / m_e}{\omega_0^2 - \omega^2} E_0 \quad (4.24)$$

Thus, relatively speaking, the displacement between the electron cloud and nucleus in the spring system is

$$x(t) = \frac{q_e / m_e}{\omega_0^2 - \omega^2} E_0 \cos(\omega t) = \frac{q_e / m_e}{\omega_0^2 - \omega^2} E(t) \quad (4.25)$$

For an incident wave with frequency $\omega < \omega_0$ the direction of the electric field is changing slowly enough for the oscillator to follow the changes (for example 800 nm incident light in our region of interest as regards silicon). When the $\omega > \omega_0$ the direction of the electric field is changing multiple times during a single period of the natural frequency and the oscillator will remain relatively unaffected.

This picture is not yet complete as this is only considering the displacement of one electron. The dipole moment for a general array of N elements of charge is [X3.5]

$$P = \sum_{i=1}^N q_i x_i \quad (4.26)$$

As the charge of each array element (electron) is the same and we consider the displacement of each electron to be the same (although still time varying) the polarization density of the dipole moment may be simplified to

$$P(t) = q_e x(t) N \quad (4.27)$$

Substituting equation 3.26 above for $x(t)$

$$P(t) = N \frac{q_e^2 / m_e}{\omega_0^2 - \omega^2} E(t) \quad (4.28)$$

For a homogeneous linear dielectric medium (no free charges) this polarization is aligned with and proportional to the field $E(t)$. Remembering that ϵ_0 is the vacuum permittivity, χ is the electric susceptibility of the medium ($\epsilon_r - 1 = \chi$) and that ϵ_r is the relative static permittivity.

$$P(t) = \chi \epsilon_0 E(t) = (\epsilon_r - 1) \epsilon_0 E(t) = (\epsilon_r \epsilon_0 - \epsilon_0) E(t) = (\epsilon - \epsilon_0) E(t) \quad (4.29)$$

where permittivity ϵ equals $\epsilon_r \epsilon_0$. Now combining this result with equation 3.29 we get

$$\begin{aligned} N \frac{q_e^2 / m_e}{\omega_0^2 - \omega^2} E(t) &= (\epsilon - \epsilon_0) E(t) \\ \epsilon &= N \frac{q_e^2 / m_e}{\omega_0^2 - \omega^2} + \epsilon_0 \end{aligned} \quad (4.30)$$

Now taking the relationship $n^2 = (\epsilon u_r/\epsilon_0)$, and in this application ignoring u_r , the permeability (degree of magnetization a material obtains in response to an applied electric field).

$$n^2(\omega) = \frac{\epsilon}{\epsilon_0} = \frac{N \frac{q_e^2}{m_e} + \epsilon_0}{\omega_0^2 - \omega^2} = 1 + \frac{Nq_e^2}{\epsilon_0 m_e} \left(\frac{1}{\omega_0^2 - \omega^2} \right) \quad (4.31)$$

This relationship for the frequency dependence of the index of refraction is, for the necessary understanding of the work presented here, close to correct. Unfortunately we see that at resonance, when $\omega_{0j} = \omega$, the function we have derived for the index of refraction is discontinuous, a result that is certainly not in agreeance with measured data. Without giving a complete mathematical treatment this is due to a missing damping term in the denominator, $i\gamma\omega$.

$$n^2(\omega) = 1 + \frac{Nq_e^2}{\epsilon_0 m_e} \left(\frac{1}{\omega_0^2 - \omega^2 - i\gamma_j\omega} \right) \quad (4.32)$$

The inclusion of a damping term, $i\gamma_j\omega$, in equation 4.33 is ascribed to dissipative frictional forces on the forced oscillations from the close proximity of atoms in a material. Electromagnetic energy is lost thermally in the form of random atomic motion.

The $n(\omega)$ of equation 3.33 is now a relatively accurate picture of the frequency dependent index of refraction. As ω increases, approaching ω_{0j} , the value $((\omega_{0j})^2 - (\omega)^2)$ begins to decrease and leads to a greater value of n . With ω approaching a natural frequency the amplitude of the system's oscillations will greatly increase (resonate) accompanied by damping and strong absorption of energy from the incident wave (the damping term is now the dominant term in the denominator), creating an absorption band. All substances possess an absorption band somewhere in the electromagnetic spectrum. In comparison notice that when $(\omega_{0j})^2 \gg (\omega)^2$ a relatively constant index of refraction over the desired spectrum is observed. Thus when a wave's frequency is significantly different from the natural frequency oscillations are small and little energy is lost. At resonance the oscillations are large and the electromagnetic field does an increased amount of work on the charges, leaving the material quite opaque in this frequency region. One final note is that because the spectrum of interest for this work is small in comparison to the complete electromagnetic spectrum, we have treated material as having only one resonance frequency. In reality there are several natural frequencies for any given material, thought of perhaps as each atom having several oscillatory modes, each with a unique frequency and oscillator strength. A complete treatment would involve quantum mechanical transition probabilities and is outside the scope of this present work.

A typical graph of the frequency dependant index of refraction is shown in figure 4.12, which should look suspiciously like the silicon index (real part). The region surrounding the central resonance frequency ω_0 (where the slope of the $n(\omega)$ curve is negative) is referred to as the absorption band, as wavelengths (frequencies) corresponding to ω_0 are absorbed quite strongly. Where the slope of the curve is positive, on the left and right side of the band, $n(\omega)$ increases with frequency. It is in precisely these regions of absorption, frequently called regions of resonant absorption, that a material will simultaneously experience an increasing in reflection at the center frequency ω_0 [X3.5]. This is observed in both simulation results and in the measurement data of physical structures. Clearly only in those regions of the electromagnetic spectrum where anomalous dispersion takes place is the imaginary part of the refractive index appreciable, as a

positive imaginary component of the refractive index implies that the wave is absorbed as it propagates through the medium.

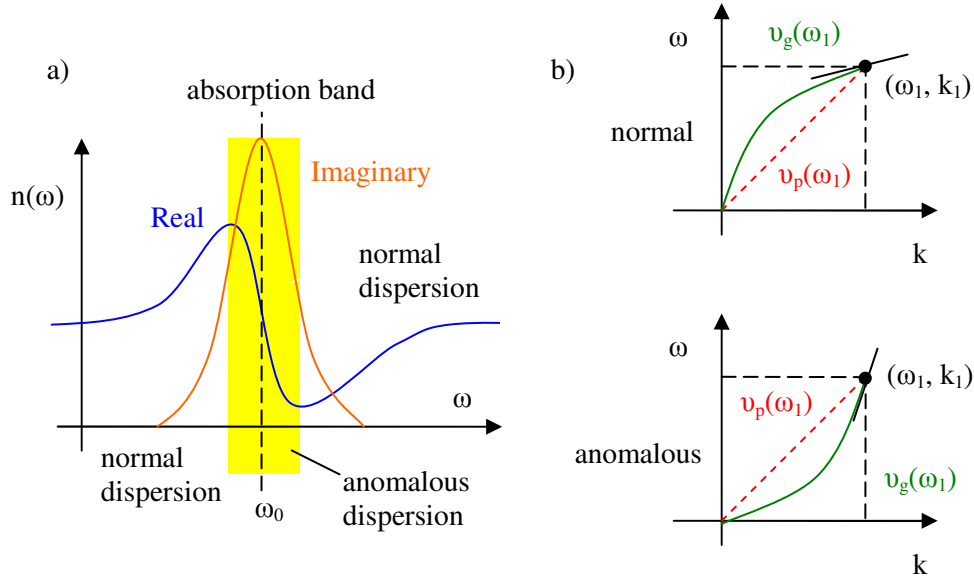


Figure 4.13: a) Anomalous dispersion at a resonance frequency. Notice that normal dispersion occurs as the frequency and index increase (the index has a positive slope). Anomalous dispersion occurs in the absorption band where the slope of the index is negative. b) Both normal and anomalous dispersion relations are shown [X3.1]. Although absorption is much higher in the absorption band it is precisely in this region that a greater reflection is observed (this will be seen in our physical measurement data!).

Figure 4.13 b) makes takes point one step further. The quantity v_g is the group velocity and is frequently denoted as the velocity with which wave energy is transmitted. If true, then group velocity is synonymous with the general velocity of the waveform. However in anomalous dispersion the index of refraction's rate of change (with respect to the frequency) experiences a sign change and the group velocity can exceed the speed of light ($v_g > c$). Group velocity is in this case no longer an indicator of velocity of the waveform. A further discussion of dispersion appears in section 4.2.2.3, as it seriously complicates non-normal incidence with an FDTD simulation.

4.2.2.2 Periodicity of nano-structures and diffraction gratings

We are now ready to continue with the previously mentioned diffraction grating properties of the nano-structures and evaluate the diffuse components of reflection for various geometries. Figure 3.14 is meant to remind the reader of the relation between (slit) diffraction and the far-field projection. Illuminated slits at normal incidence emit light which is in phase, but the optical path length traveled by light from each slit may be, at a point of observation, considerably different. This yields additional interference in the flux-density distribution which then appears as a single slit diffraction pattern envelope modulating a quickly changing double slit interference system. The single slit envelope drops from a maximum at the center and the intensity meets its first zero on both sides of the main central peak at $\sin(\theta_1) = \pm\lambda/b$. The centers of the quickly changing (modulated) frequency appear at values of $\pm\lambda/a$.

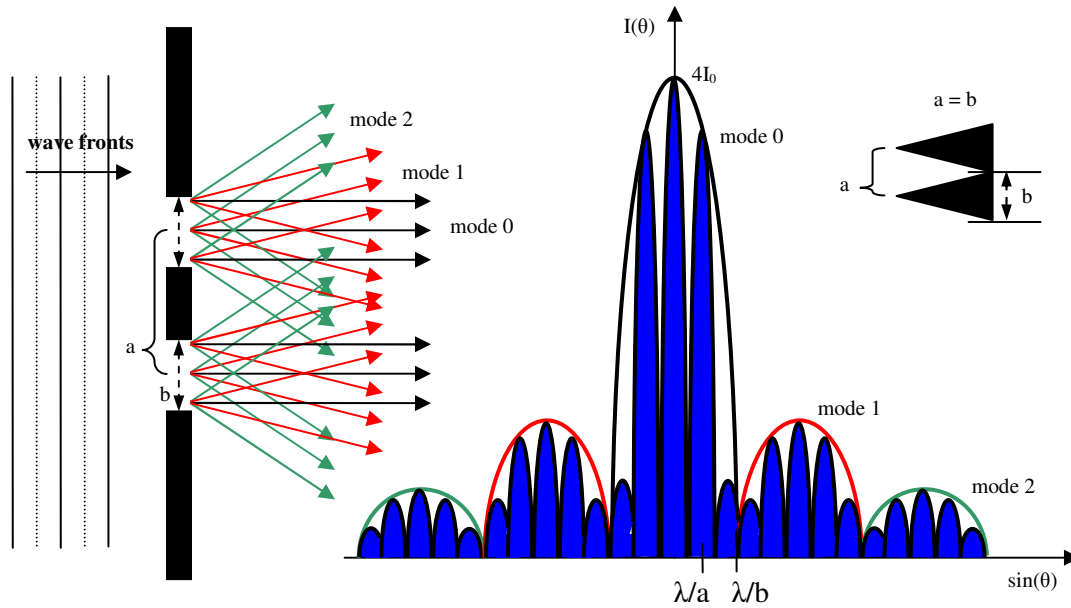


Figure 4.14: The double slit. Two slits of width b and center-to-center separation a . Either slit alone would generate the overall diffraction pattern seen, however combined. The flux-density distribution is a single slit diffraction pattern envelope modulating a quickly changing double slit interference system [X3.1].

The angles are given by the grating equation, $\sin^{-1}(\theta) = m\lambda$, however for the pyramids $a = b$ and this is why the first mode (the zero mode) gets smaller with pyramid width. Shown in figure 4.15 is the far-field intensity for 200 nm, 400 nm, and 800 nm incident light reflected from an array of a) 100 pyramids and b) 100 pillars. Shown for clarity of comparison are dashed red lines helping illustrate, as expected, the significant disparity between pyramids and pillars with regard to the amount of light reflected at zero degrees (the specular component) and other angles of reflection (the diffuse component). As expected the pyramid structure outperforms the pillar in reflection suppression. However more interesting is that the ± 1 modes are only noticed in the far-field for an incident wavelength of 200 nm. This may have some small influence on creating the checker board pattern in figure 4.06, although the cause of this is nearly completely the phase shift of the wavefront due to differing velocities in silicon and oxide. Figure 4.16 further highlights what is happening. The pyramid width is doubled to 400 nm and now shown are the far-field intensities for incident wavelengths of 200 nm, 300 nm and 400 nm. Notice that the 400 nm incident light on

a 400 nm wide structure (400 nm grid spacing) matches exactly to the 200 nm incident light on a 200 nm wide structure (200 nm grid spacing). Components at ± 30 degrees are seen and can be verified simple with the grating equation. At wavelengths greater than the width of the pyramid no modal components are possible. Subsequently, a first design rule can be stated. Structures with widths less than 200 nm have no modal components for incident light above the 200 nm wavelength, the lower limit of interest for this research. Reflected intensity from sets of structures with differing widths is shown below in figure 4.17 along with a now a very apparent lower size width limit.

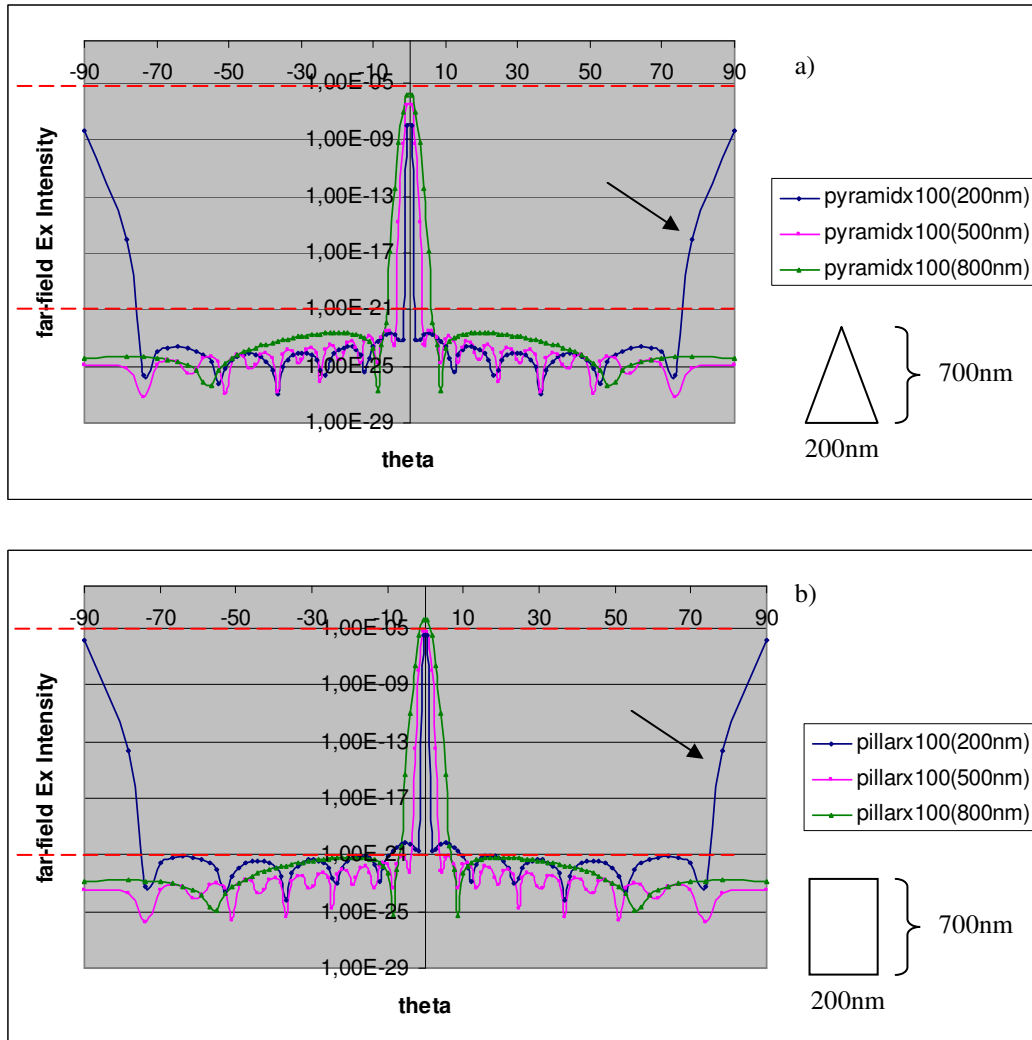


Figure 4.15: Logarithmic scale differences in far-field intensity for 100 element arrays (without SiO_2 covering) of a) pyramids and b) pillars. The dashed red lines help illustrate the extreme difference in the amount of light reflected in the specular component at zero degrees and what could be called the diffuse component, other angles of reflection, between pyramids and pillars. Clearly the pyramid structure reduces reflection most significantly. The ± 1 modes are only noticed in the far-field for an incident wavelength of 200 nm.

Only a slight increase in reflectivity is seen for larger (red) wavelengths at a width of 100 nm further reduction of the width to 50 nm brings the reflected intensity to a completely unacceptable

level for this application. The 50 nm wide pyramids begin to approach the behavior of a silicon surface, only in the higher frequencies where the wavelength to peak spacing is still fitting is reflection repressed, however as if the spacing were further decreased this would the silicon behavior would begin to infringe on the anti-reflective behavior of even the higher frequencies. Geometrically one can easily imagine a surface of pyramids with an ever shrinking width. As the width (distance between peaks) decreases to zero the surface begins to resemble blank silicon. Figure 3.18 sets a limit for the minimum structure height at above 500 nm.

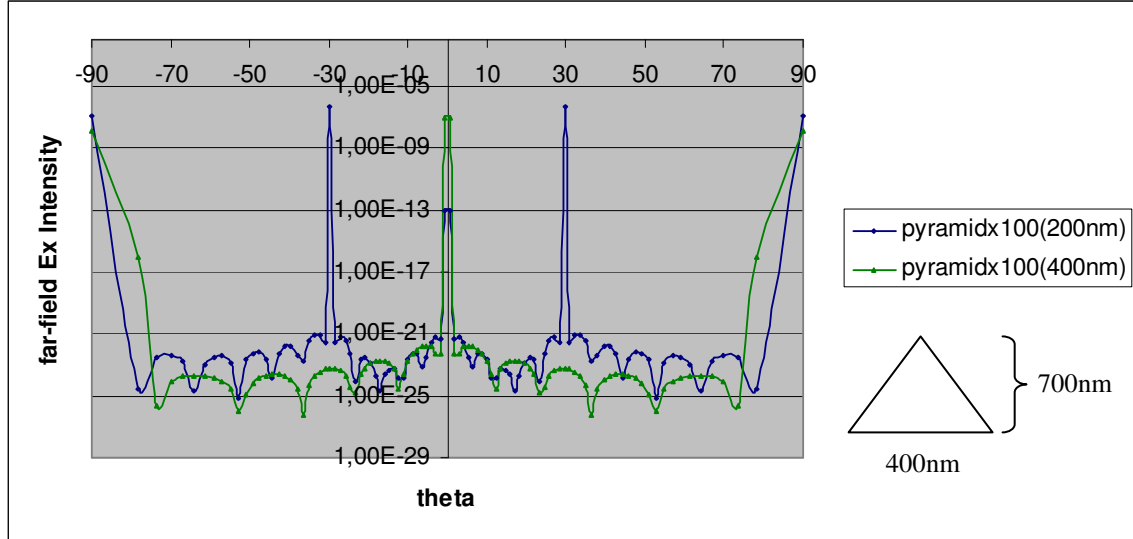


Figure 4.16: Logarithmic scale far-field intensities for incident wavelengths of 200 nm, 300 nm and 400nm illuminating an array of pyramids 400 nm wide (twice the width of pyramids in figure 4.15). Notice that the 400nm incident light on a 400 nm wide structure (400 nm grid spacing!) matches exactly to the 200nm incident light on a 200 nm wide structure (200 nm grid spacing!).

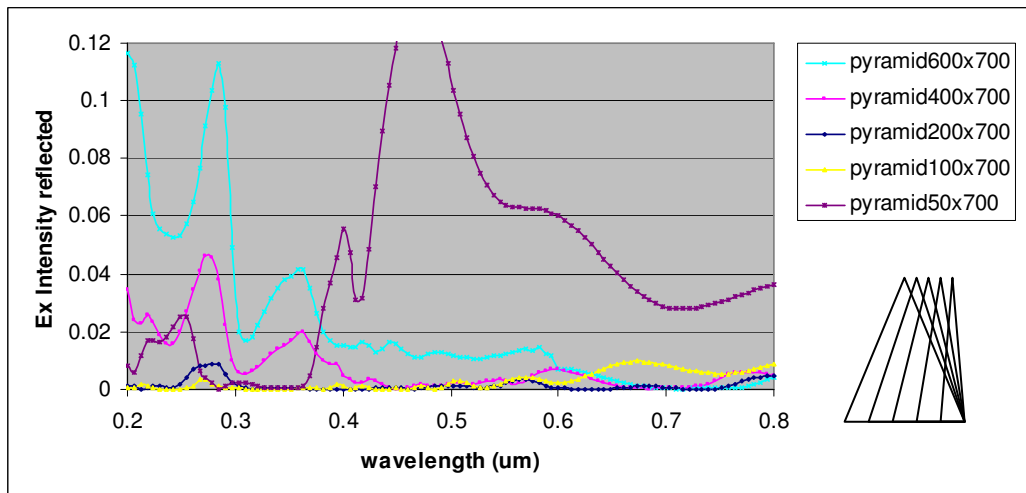


Figure 4.17: Reflected intensity for various widths (changing the grating constant!) of pyramids without the SiO_2 covering. The optimal results for the given wavelength range (200 and 800 nm) are between 100 and 200 nm.

One major point of this chapter and a consequence of this diffraction grating investigation for the pyramid nano-structures is, what would the modal components look like for two adjacent pyramid structures that slightly (or significantly) overlapped? The “real” answer (more complicated than a simple diffraction grating) will be deferred to section 4.3, however figure 4.19 nicely introduces our problem.

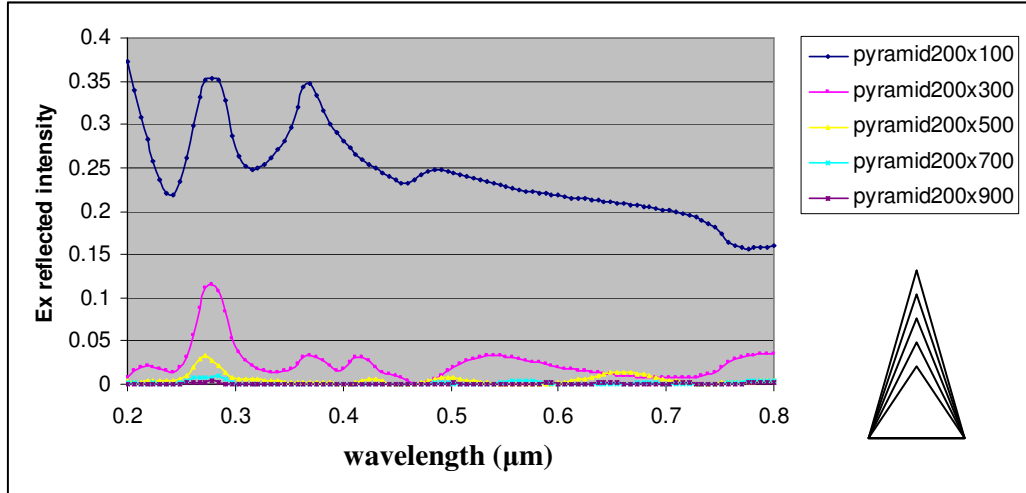


Figure 4.18: Reflected intensity with pyramid height. The optimal results for the given wavelength range (200 and 800 nm) are above 500 nm.

It should be very apparent that $x_1 + dx$ will not be constant over the wafer. If it were to be a fixed value the structure would, with periodic boundaries, function, as a periodic grating with several different periodicities. In reality The width of pyramids will vary up to a maximum width and down to a minimum width creating a smearing effect in the reflected Ex intensity at various modal angles. This means there will not be no definite diffraction orders. To answer this problem any further will require us to first make some observations with simulation boundary conditions and then introduce 3D structures. The detailed picture we receive at the end of this will be worth our work.

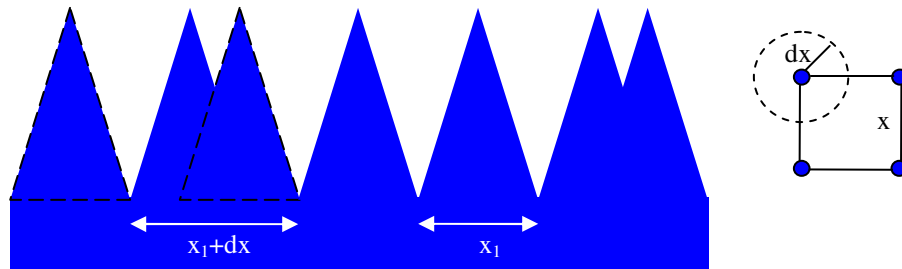


Figure 4.19: Overlapped gratings. Additional modal (diffuse) components are added to the reflected intensity when separation between structures is not fully achieved. Essentially two adjacent unseparated pyramids (as shown) create a greater “effective” structure width. This particular structure when simulated with periodic boundaries will remain a diffraction grating (albeit more complicated) only if dx is a fixed value.

4.2.2.3 Boundary conditions and non-normal incidence in 2D structure simulation

Returning now to the absence of oscillations within the substrate of the flat silicon sample from figure 4.07 will correctly introduce the notion of boundaries in the simulation space, a non-trivial problem and important here as we wish to now change the angle of incident light. The nonappearance of oscillations is the result of Perfectly Matched Layer (PML) upper and lower boundary conditions. The computational simulation grid always requires truncation when a PDE is solved with numerical methods. Such a truncation when performed in an inelegant way can lead to the introduction of considerable artifacts into the simulation. In some situations truncation is not necessary for example the left and right boundary conditions of the three structures (flat, needle, pillar) simulated thus far have been periodic. This is possible because the structures themselves represent only one small part of a large periodic structure. This drastically reduces the necessary size of the simulation space as, shown in figure 4.20, behavior of structures very distant from the simulation space are interpreted as really existing and incorporated into the overall result (making for example diffuse components coming back from very large reflected angles available with the need of only one small structure/simulation-space).

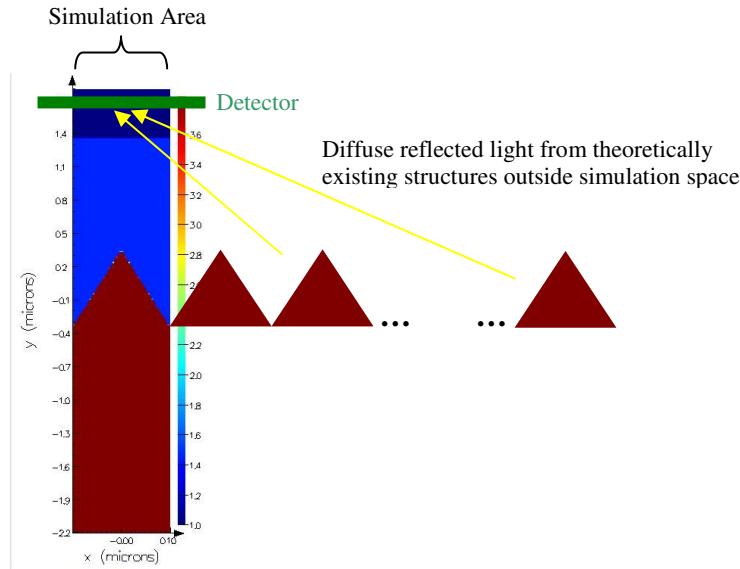


Figure 4.20: Left and right side periodic boundary conditions. The simulation space is essentially extended to infinity.

Thus problem of boundary conditions can be exemplified by the flat oxide covered silicon surface now shown in figure 4.21 with an upper PML boundary and a lower ideal metal (hard wall reflection) boundary. It should be stressed that the PML is not really a boundary condition rather it is an absorbing boundary LAYER [X3.19]. When a wave enters the layer it decays with an exponential attenuation and this means that any reflected waves returning from inside the PML do so exponentially smaller and are thus quite negligible. A following point is that any transition between two media of differing indices gives rise to a reflection at the interface (before entering the PML in this case), that is unless the PML itself tailored to the interface (meaning there are silicon PMLs, Air PMLs, SiO₂ PMLs, etc). A PML can be thought of as truncating the simulation space with a medium that is exponentially absorptive.

Thus the first of two points for the inclusion of non-normal incidence light into the simulation space is shown in figure 4.22, simply taking a cross section of the simulation space

perpendicular to the surface (along the y axis) as has been done is no longer possible. The observed wavelengths will be distorted by the superposition of the incident and reflected wave. Secondly and more importantly a new boundary condition, called a Bloch boundary, must be introduced. One needs to ensure that the field that exits from one side of the FDTD simulation immediately enters from the opposite side. A comparison of using periodic and Bloch boundary conditions is depicted in figure 4.23 and 4.24 using 650 nm light at various angles incident on an oxide covered silicon surface.

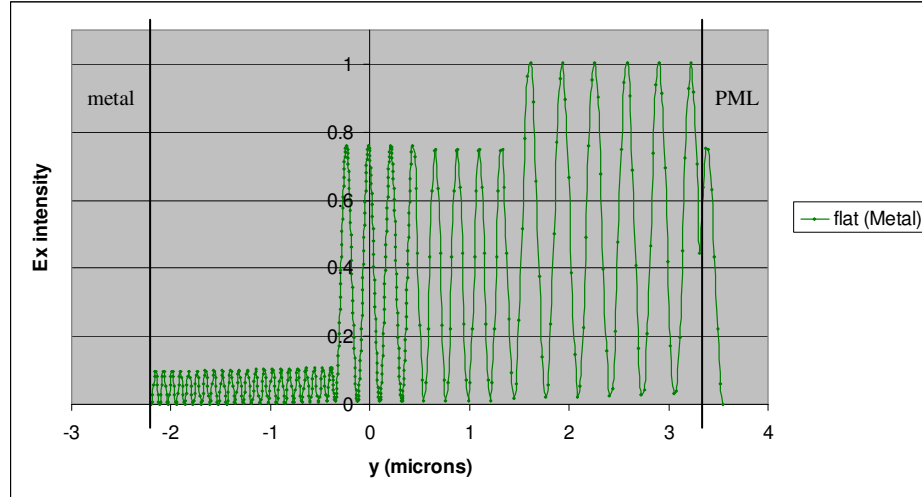


Figure 4.21: Metal (hard wall reflection) boundaries for oxide covered silicon. The correct wave length is now visible within the silicon.

The phase difference between a1 and b1 is zero when the angle of incidence is perpendicular to the surface thus the overlap in results is to be completely expected. The oscillations begin to appear in the results using periodic boundary conditions around 30 degrees (for smaller angles the phase difference remains negligible). As the angle of incidence further increase so does the phase difference, eventually leading to completely nonsensical results. Notice how the Bloch boundary correctly simulates the decrease in reflected intensity for increasing angles of incidence.

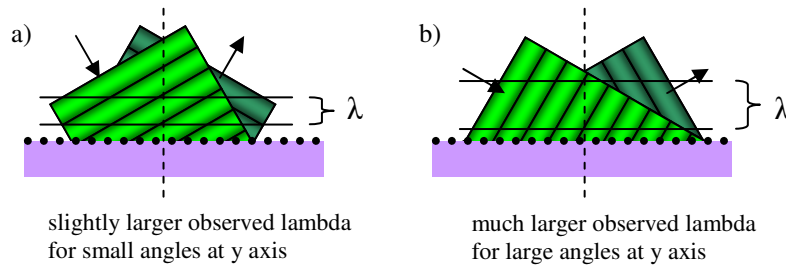


Figure 4.22: The increase in observed wavelength as a result of the superposition of incident and reflected waves at the axis of measurement a) small for small angles and b) large for large angles

The reader is reminded that the amplitude is lower for greater angles of incidence up to the Brewster angle. According to the Fresnel equations (from section 3.1) for light which is polarized in the plane of incidence (p-polarized), as we have here, or light polarized perpendicularly to the plane of incidence (s-polarized). For a pair of refractive indices, n_1 and n_2 , the value of reflection (R_p) decreases to zero as the angle of incidence for p-polarized light increases. The

point at which all incident light is refracted (zero reflection) is called the Brewster's angle and is approximately 56 degrees for an air glass interface.

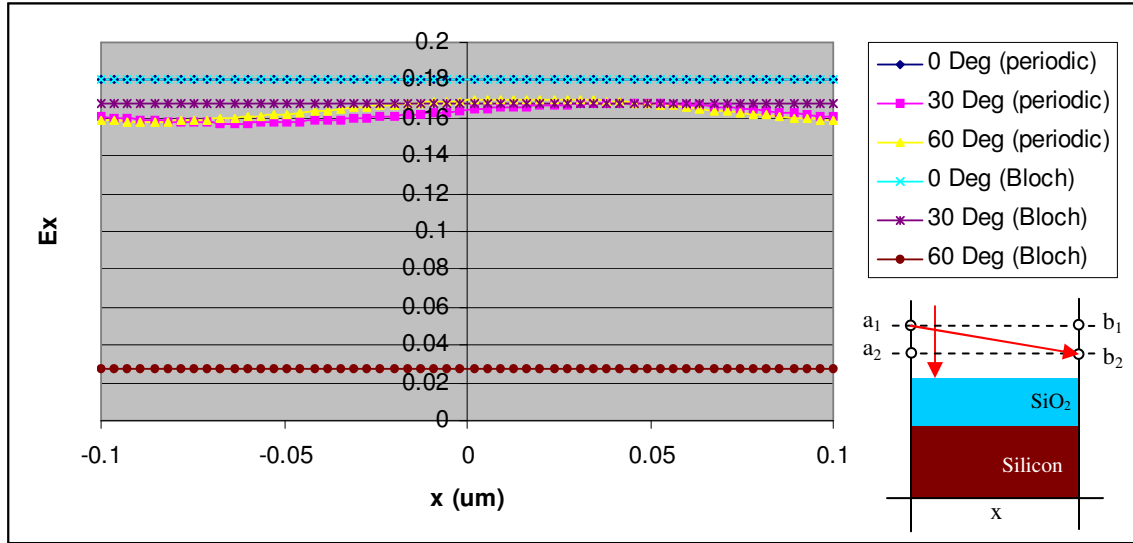


Figure 4.23: Total reflection of 650 nm incident light at various angles from an oxide covered silicon surface using periodic and Bloch boundary conditions. The overlap at perpendicular incidence is to be expected as the phase difference between a_1 and b_1 is zero, however notice the oscillations in the results using periodic boundary conditions for angles other than zero.

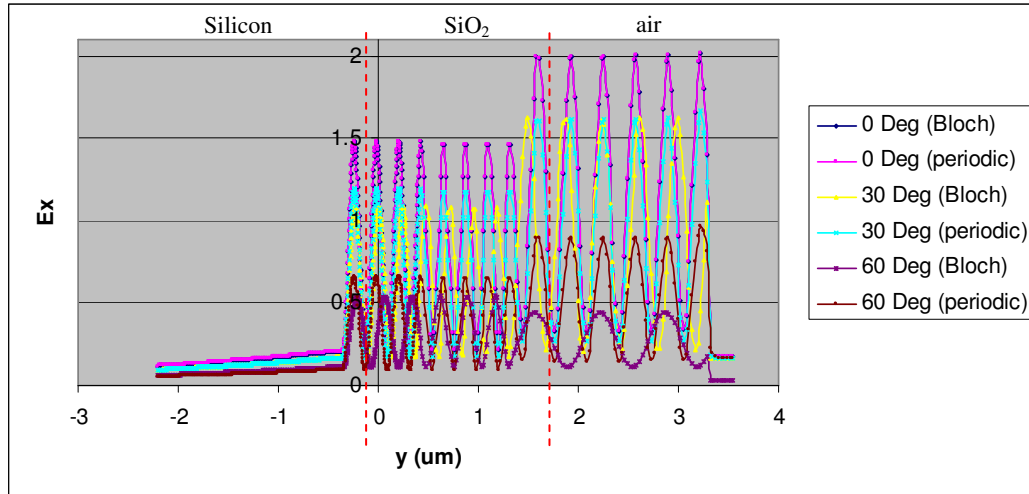


Figure 4.24: Bloch and periodic boundary conditions for the simulation y axis. For zero degree incidence there is no difference observed, as expected the plane wave travels through the simulation in both cases with identical phase and amplitude. For 30 degree incidence a slightly greater wavelength is seen as

It should be mentioned that this only is true for indices of refraction which are real numbers. Typically for light absorbing materials such as metals and semiconductors, cases in which n is a complex number, R_p does not completely reach zero. In many applications light is unpolarized, being a mix of s - and p -polarizations, and in such a case the reflection coefficient is simply $R =$

$(R_s + R_p)/2$. Regardless, this provides a concise explanation as to why the amount of observed reflected light decreases with increasing angle of incidence. However, Brewster's angle is in general not applicable to the nano-structured surfaces, meaning the effective medium created by the pyramids. This is shown in figure 3.24 is the reflection from 400 nm wide 700 nm tall pyramids. This will combine everything we have examined in the two last two sections with non-normal simulation incidence.

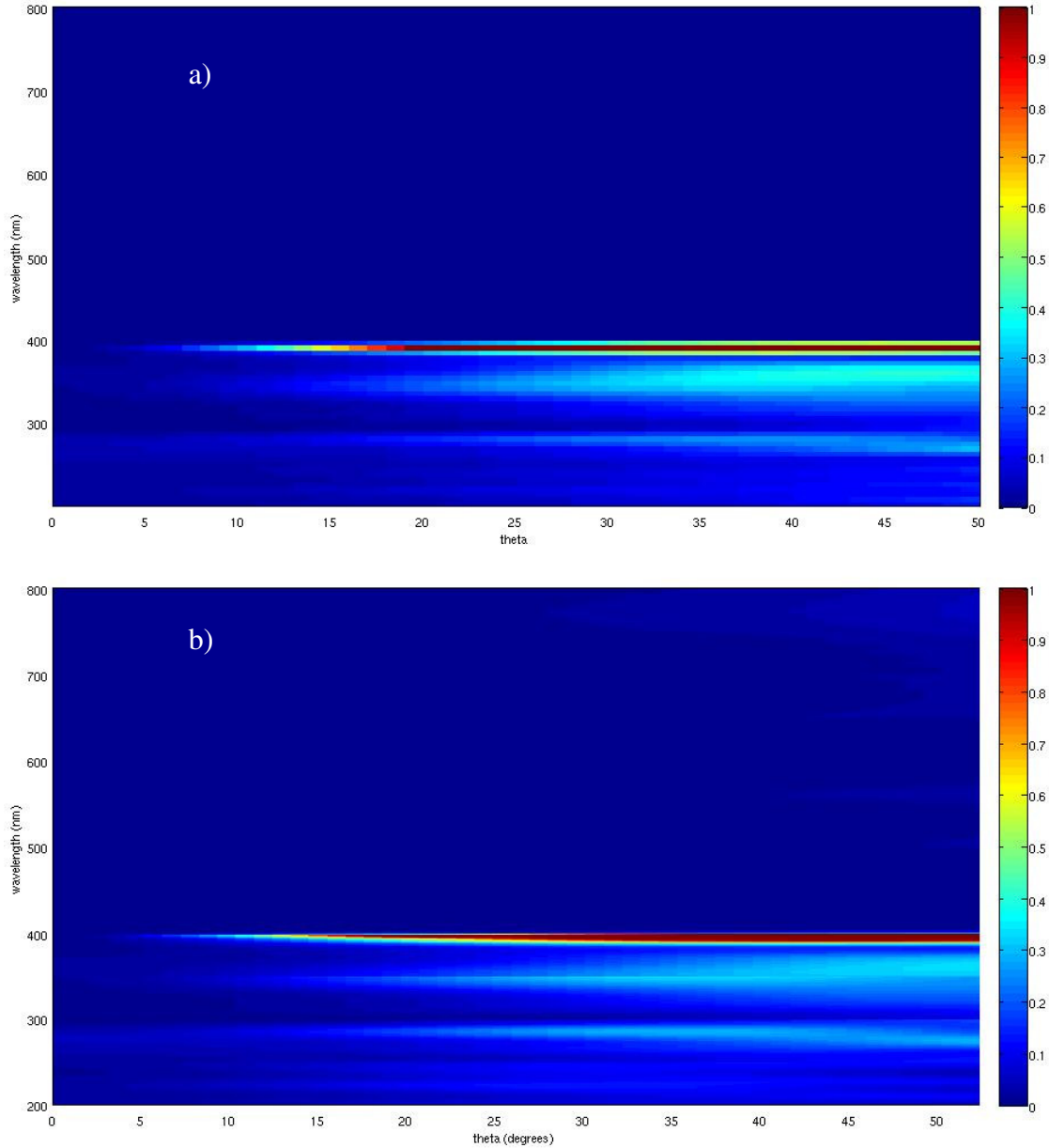


Figure 4.25: Reflection from 400 nm wide pyramids (700 nm tall), a) the full spectrum in a single simulation angle by angle and b) the full spectrum divided into twelve 50 nm wide bandwidths and simulated angle by angle. In both cases the predicted peaks at approximately 250 nm and 350nm are visible followed by a massive reflection at 400 nm corresponding exactly to the periodicity of the structures, seen in figure 3.15 where the additional angle at 30 degrees is added.

Part b), as will be discussed in the next figure, is slightly smoother and reveals some detail in the larger wavelengths.

In a) a single simulation is performed angle by angle (0 to 50 degrees) with the full spectrum (200 nm to 800 nm) and in b) the full spectrum is again simulated angle by angle however this time it is divided into twelve 50 nm wide bandwidths and simulated. As will be discussed in figure 3.25, part b) is smoother (clearly around the spike at 400 nm) and reveals additional detail in the longer wavelengths.

The graphs can be separated into two main regions, above 400 nm and below. In the lower region the predicted peaks at approximately 250 nm and 350 nm are visible (as a result of the resonant absorption regions in silicon's index of refraction, discussed in detail earlier) followed by a massive reflection at 400 nm corresponding exactly to the periodicity of the structures, seen in figure 3.15 where the additional angle at 30 degrees is added.

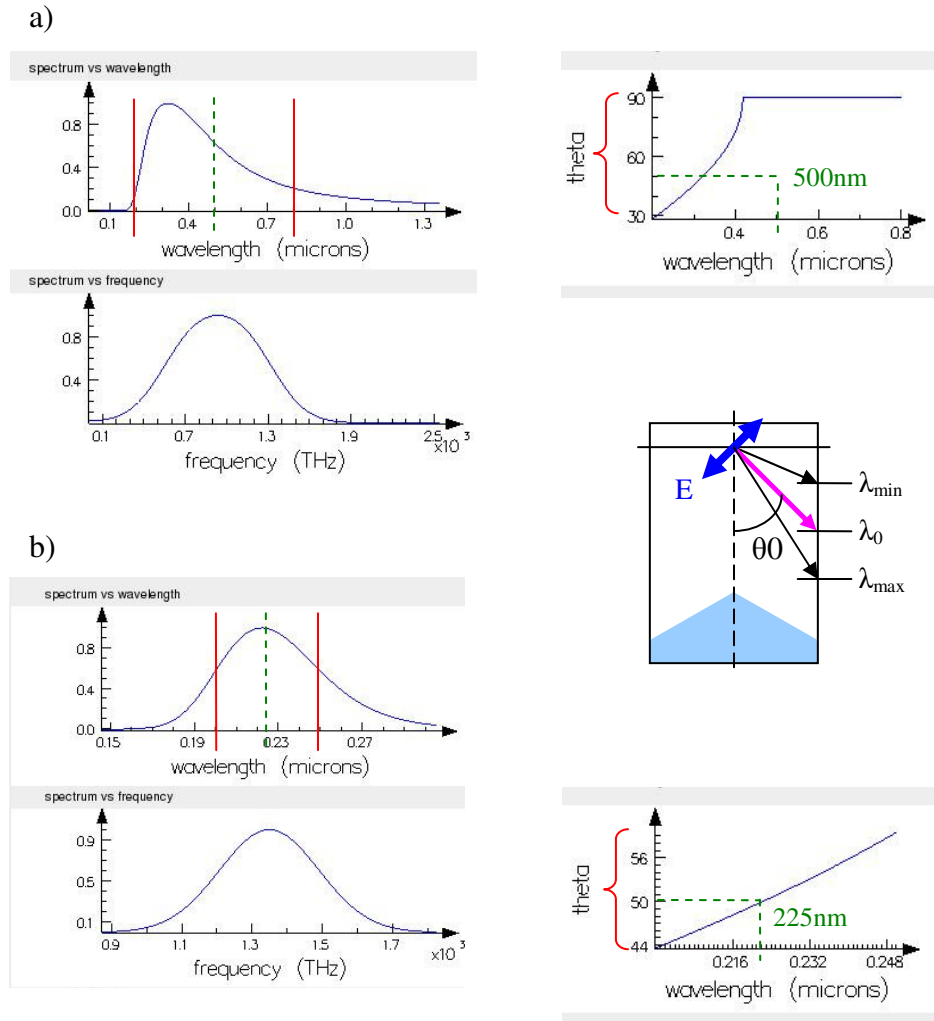


Figure 4.26: Dispersion in the broadband injection angle. In plane components of the wave vector are shown at the center wavelength λ_0 which has the correct θ , and the smallest and largest wavelength in the incident wave, λ_{\min} and λ_{\max} . In a) the full spectrum is shown and in b) a 50nm bandwidth is shown. Part b) clearly offers calculable k vectors [X3.7].

In the region above 400 nm there is only an effective medium, there is no periodicity whatsoever. In part b), the more accurate simulation, very slight reflection values begin to be observed as the more extreme angles are simulated, greater than 30 degrees. This is understandable as at greater and greater angles the pyramids form less and less of a gradient index (incidence perpendicular to the pyramid's point is optimal – hence Brewster's angle is not applicable).

The explanation for the higher quality of part b) from figure 4.25 is shown in figure 4.26. The full bandwidth wave from 200 nm to 800 nm all at once yields a less than desirable result as the dispersion of k is quite terrible. Theoretically a person could simulate one wavelength at a time with FDTD sweep through the desired angles. However the time needed for such a simulation is totally unacceptable. Assuming 1nm steps between 200 nm and 800 nm, assuming 1 degree steps between 0 and 60 degrees, and an approximate simulation time of 3 hours per step:

$$(600 \text{ wavelengths}) (61 \text{ degrees}) (10 \text{ min}) = 366000 \text{ min, approximately 254 days}$$

This is a very very long wait indeed. A much smarter way to proceed is to separate the spectrum into only 6 center wavelengths and simulate

$$(6 \text{ wavelengths}) (61 \text{ degrees}) (10 \text{ min}) = 3660 \text{ min, approximately 2.5 days}$$

This is a much more acceptable simulation time and the limited amount of error introduced can be removed post simulation by some simple data manipulation. The center wavelength of each bandwidth (λ_0) will have the correct angle of incidence and the correct theta for the non-center wavelengths can be calculated from [X3.7]

$$\theta_{new} = \sin^{-1} \left(\frac{\lambda_i}{\lambda_0} \sin(\theta_0) \right) \quad (4.33)$$

Notice in x-axis in figure 4.25 b) compared to a). The axis extends past 50 degrees as some longer wavelengths were actually calculated at angles greater than 50 due to this dispersion in the simulation space. In terms of space, if a signal is broadband (containing all relevant wavelengths) there will be a correspondingly large number of sinusoidal components present and the pulse will thus have a wide range of k values. As discussed all components of the wave travel with differing phase velocities in a dispersive medium (any medium other than a vacuum). The shape of signal will subsequently change as it travels through the medium. This separation of waves dependant on k vector is not limited to just angle of injection. There are very serious implications to where light will be absorbed within the nano-structures based on frequency. The next section will address such problems in detail.

4.2.2.4 Visualization in 3D to observe absorption within nano-structures

Up to this point we have limited our discussion to 2D simulations however the inclusion of 3D structures provides an extremely useful way to visualize incident electromagnetic fields on structured surfaces. The 3D forms used are shown in figure 4.27 (without SiO₂ for geometric clarity). The blue arrow perpendicular to the normal incidence indicates the direction of the incident E field.

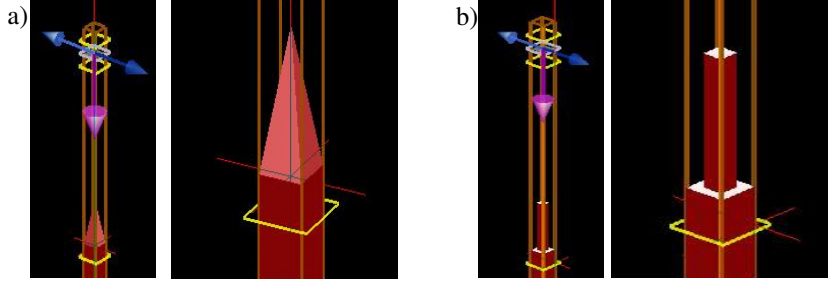


Figure 3.27: a) 3D pyramid and b) 3D pillar layout (SiO₂ covering not shown)

A logical first question is whether or not a difference in reflected intensity is visible when expanding the simulation to an additional dimension. Given the transverse nature of light, such a difference should most certainly not be found, however demonstrating this explicitly will give us the chance to further push the investigation of the diffraction grating properties of needles and pillars to their limit. Firstly however, as seen and discussed earlier, FDTD is not merely a set of sources and monitors it is a complete simulation space. Figure 3.28 shows the newly observable 3D properties of the simulation space, namely the ability to view “slices” of the space. For this study plane (linearly polarized) waves were used for which the direction of the time varying field is fixed. The figure is oriented to the coordinate system seen in the previous 3D simulation results. The electric field is parallel to the x-axis (the substrate surface), where from the differential form of Maxwell’s equations we see

$$\begin{aligned} \frac{\partial \bar{E}_z}{\partial y} - \frac{\partial \bar{E}_y}{\partial z} &= -\frac{\partial \bar{B}_x}{\partial t} \\ \frac{\partial \bar{E}_x}{\partial z} - \frac{\partial \bar{E}_z}{\partial x} &= -\frac{\partial \bar{B}_y}{\partial t} \\ \frac{\partial \bar{E}_y}{\partial x} - \frac{\partial \bar{E}_x}{\partial y} &= -\frac{\partial \bar{B}_z}{\partial t} \end{aligned} \quad (4.34)$$

As we see the time dependant magnetic field can only have a component in the y-direction. Taking a harmonic electric field (wave) of the form $E_x(z, t) = E_{0x} \cos[\omega(t - z/c)]$, the associated magnetic field is then

$$B_y(z, t) = -\int \frac{\partial E_x}{\partial z} dt = \frac{1}{c} E_{0x} \cos\left[\omega\left(t - \frac{z}{c}\right)\right] \quad (4.35)$$

The direction of propagation of this wave is in the negative z-direction (the cross-product of the E and B fields) and having the same time-dependence both fields remain in phase at all space locations. Obviously this agrees with the intensity cross sections of figure 3.28.

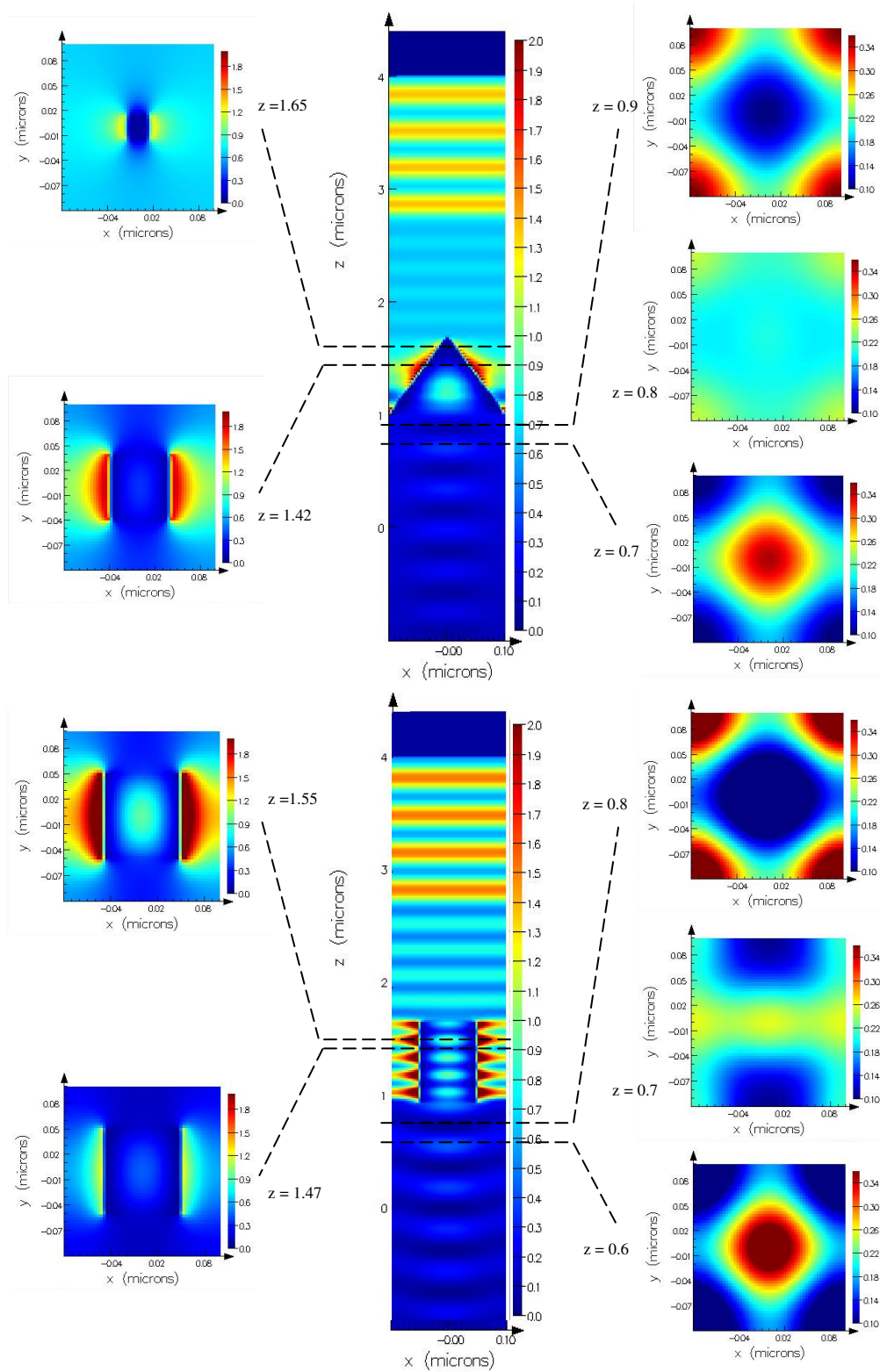


Figure 4.28: E_x intensity within the structures and beneath in the substrate, illuminated with 650 nm. Shown to more clearly highlight the modulated pattern, the reader should be aware of the reduced intensity scale used in the silicon substrate pictures on the right-hand side.

If each cross section were rotated 90 degrees, this would be the pattern of the magnetic field (differing only in intensity).

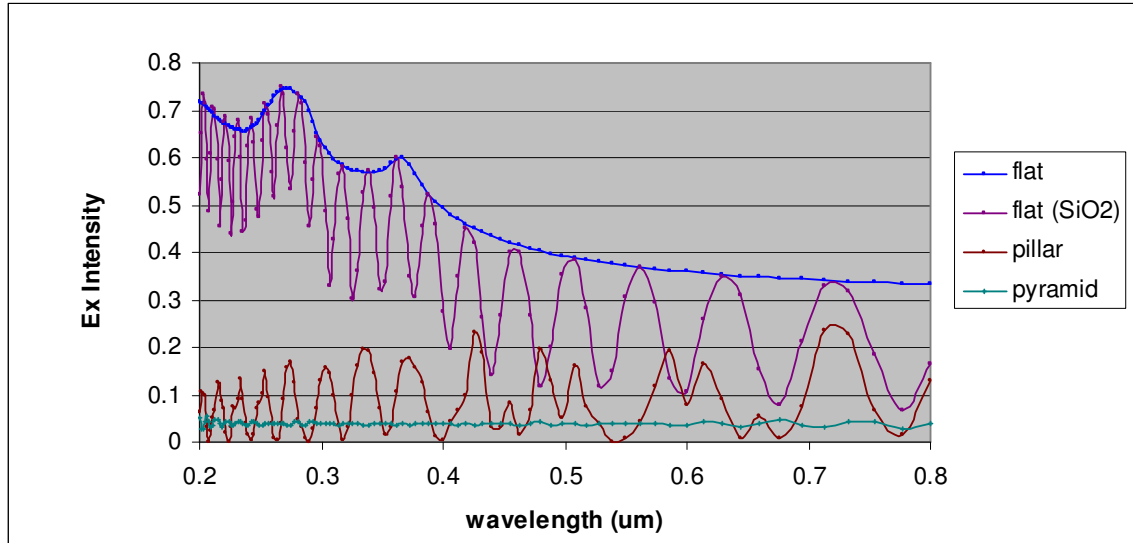


Figure 4.29: Reflected intensity from 3D structures. Results are the same as for 2D structures with the exception of the pillars (discussed in the next figure).

Shown in figure 4.29 is a graph of reflected intensity versus wavelength. Clearly there is only one significant difference when compared to the 2D variant of this graph. This difference being between the reflected values of two-dimensional structures and three-dimensional structures are reflected intensity values from the pillars. This is due to the ratio of silicon to SiO_2 , for 2D pillars the ratio is 1/2 silicon to 1/2 SiO_2 , however for 3D pillars the ratio is 1/3 silicon 2/3 SiO_2 , seen in figure 4.30.

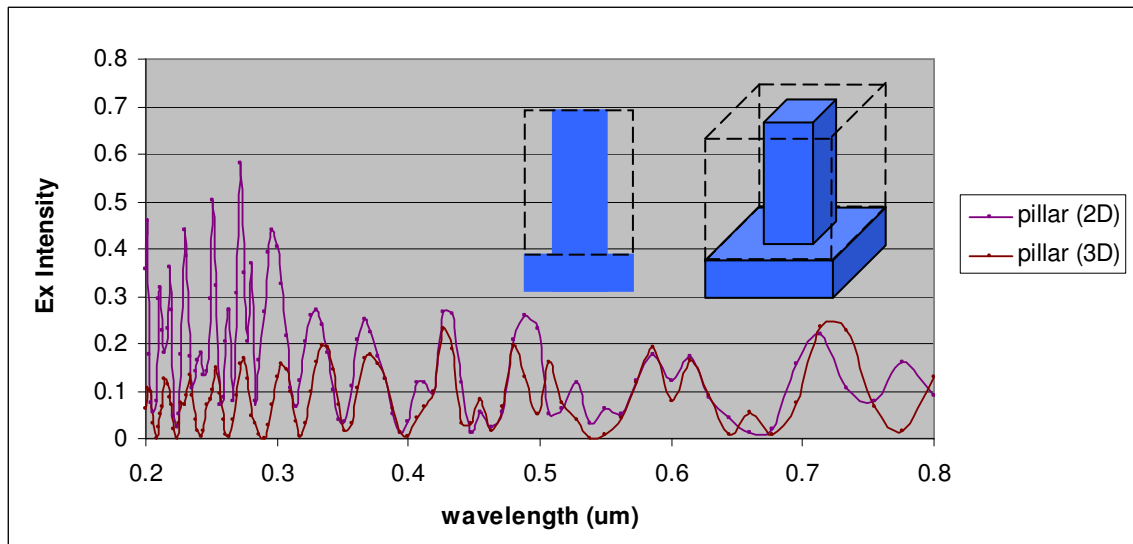


Figure 4.30: Two-dimensional vs. three-dimensional pillars. The “effective” index (the ratio of silicon to SiO_2) is not constant. For pillars the ratio is 1/3 Silicon 2/3 SiO_2 .

The properties of the gradual introduction of silicon through the pyramid shape are apparent. However it can be made abundantly clear just where the light has gone, as this will dramatically influence the electrical performance of devices in the following chapter. Figure 3.31 shows cross sections of 3D pyramids in the xz direction taken from the center of the structure. The wavelength of incident light is increased by 100 nm starting in part a) at 200 nm to part g) at 800 nm a somewhat startling result is the location of absorption in the wavelength range less than 400 nm, light is absorbed very early in the substrate, directly on the pyramid surface in fact.

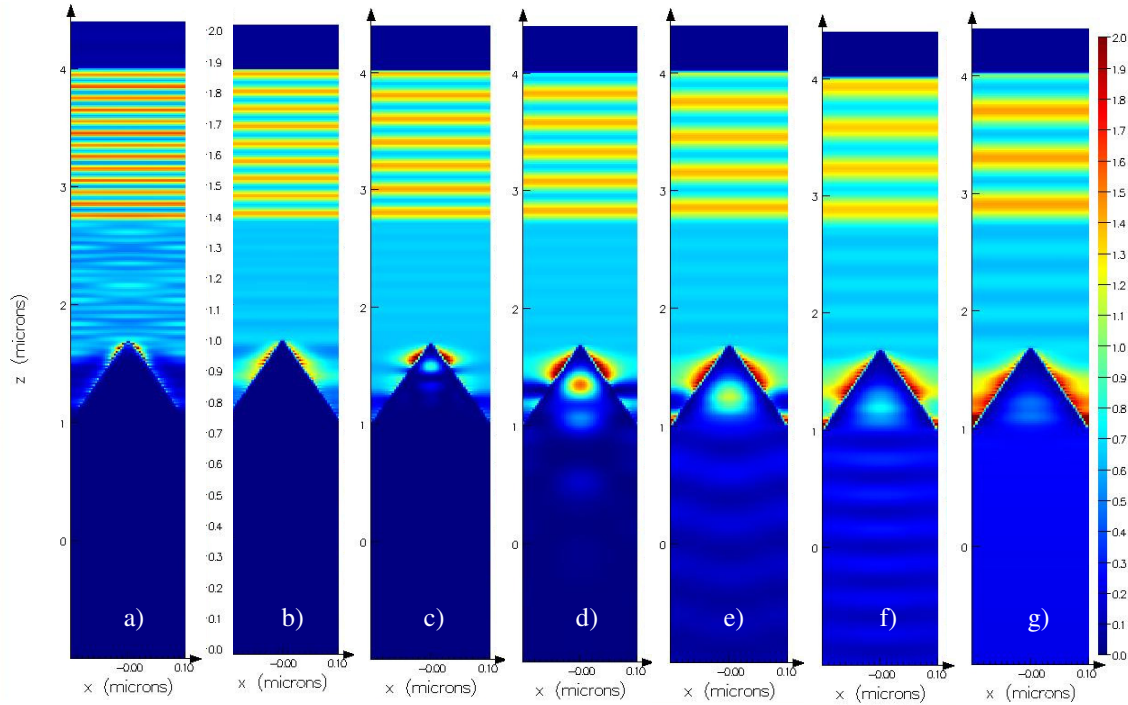


Figure 4.31: FDTD Simulation space showing E_x Intensity through the center of a 3D pyramid structure using incident light of wavelength a) 200nm b) 300nm c) 400nm d) 500nm e) 600nm f) 700nm g) 800nm. Of utmost importance is to notice where blue light (less than 400nm) is absorbed directly on the pyramid surface. The grid spacing is 3nm x 3nm x 3nm.

This will be a key explanation for subsequent integration activities to increase carrier lifetime and to understand final spectral sensitivity measurements in the next chapter. For wavelengths greater than this range but still less than 800 nm a hot spot is seen in the center of the structures, moving closer to the silicon surface with increasing wavelength. This behavior can be further demonstrated by showing the residual wavelength dependant intensity at various depths under the nano-structures. Figure 4.32 a) shows residual intensity under the surface of bare silicon and oxide covered silicon at depths of 20 nm, 200 nm and 1 μ m. These curves can be clearly derived from the reflected intensity curves of figure 4.29. The intensity minimums, where the SiO_2 oscillation meets the flat silicon curve, correspond exactly to the locations in which the oscillation curve of the reflected intensity from SiO_2 touches the curve of the reflected intensity from silicon (for example at approximately 725 nm). Part b) shows residual intensity under the surface of oxide covered pillars and oxide covered pyramids at depths of 20 nm, 200 nm and 1 μ m. The pillars out perform the pyramids at wavelengths where the quarter-wavelength behavior is observed. However it is difficult to evaluate the intensity of transmitted light beneath several structures in comparison to that of the intensity beneath a flat surface as absorption will clearly

take place within the structure. The consequences of this will be discussed in the chapter on electrical properties, but the process follows from the band gap of silicon, depicted in figure 4.33.

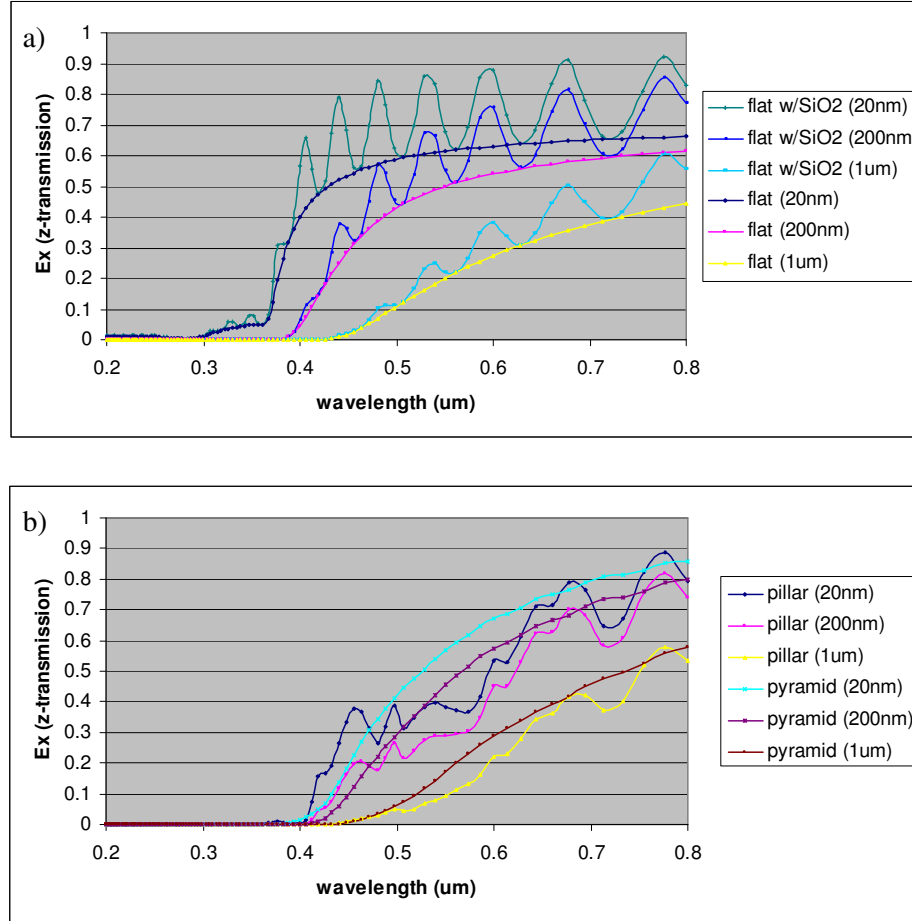


Figure 4.32: Transmitted intensity at various depths beneath a) covered and uncovered silicon and b) 3D pyramids and 3D pillars. Again notice how the oxide covered pillars behave exactly like a quarter-wave transformer for some wavelengths bringing reflection to zero. As a result pillars experience a slight increase in transmission at these exact wavelengths in comparison to the pyramids. Overall however the pyramids exhibit a smooth behavior across all wavelengths (no visible oscillations in transmitted signal).

We assume a photon beam of intensity I_0 (photons/cm² s) and containing photons of only wavelength λ , directed at a sample of thickness ℓ , as depicted below in figure 4.33. At any distance x in the sample the intensity of the beam can be calculated by the probability of absorption in any incremental distance dx . Therefore the degradation of beam intensity is proportional to the remaining intensity at a distance x from the sample surface:

$$-\frac{dI(x)}{dx} = \alpha I(x) \quad (4.36)$$

The coefficient α (cm⁻¹) is called a material's absorption coefficient and varies with wavelength however the function is clearly centered at the E_g band gap value of the material. This leads to the

statement that photons with energies greater than the band gap are readily absorbed while photons with energies less than the band gap are transmitted (any absorption is negligible). Solving equation 3.38 and selecting x to be the complete sample length:

$$I_t = I_0 e^{-\alpha \ell} \quad (4.37)$$

This equation describes the relationship between incident and transmitted light and is commonly referred to as the Beer-Lambert law. However, regardless of depth, in some instances it is possible to make some observations and simple absolute comparisons, for example a comparison of 20 nm beneath the bare and oxide covered silicon shown at approximately 725 nm makes clear that the intensity of transmitted light beneath a pyramid is much higher than that of the silicon.

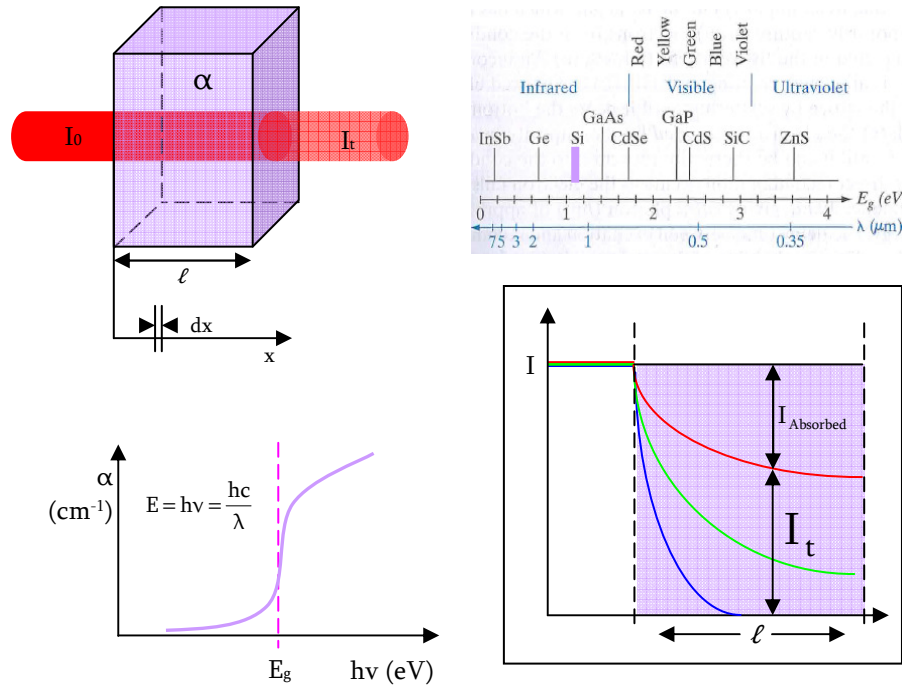


Figure 4.33: Optical absorption [X4.1]. (Top left) A beam of intensity I_0 (photons/cm² s) containing photons of only wavelength λ is directed at a sample of thickness ℓ . At any distance x in the sample the intensity of the beam can be calculated knowing the material's absorption coefficient α (cm⁻¹), (bottom left) a function which varies with wavelength and is centered at the band gap E_g of the material. (Top right) Photons with energies greater than the band gap are readily absorbed while photons with energies less than the band gap are transmitted. (Bottom right) Shorter wavelengths will then be absorbed earlier in the material (for example blue, green and red as shown).

There is one important detail to note. Careful observation of figure 4.31a) leads one directly to the limitations of the current grid size. In figure 4.34 the space is grid size of the simulation space is made twice as small and a noticeable improvement in picture resolution is observed. This is to be expected, as at the 4 nm grid spacing a 200 nm wavelength has less than (remember the wavelength is reduced in silicon!) 50 cells per wavelength, whereas at 2 nm grid spacing there are twice as many.

We conclude this section with some comments on 3D structure shape by considering the differences in the far-field projection of pyramids with a square base. Figure 4.35 shows the far-field projection of this pyramid for multiple array sizes.

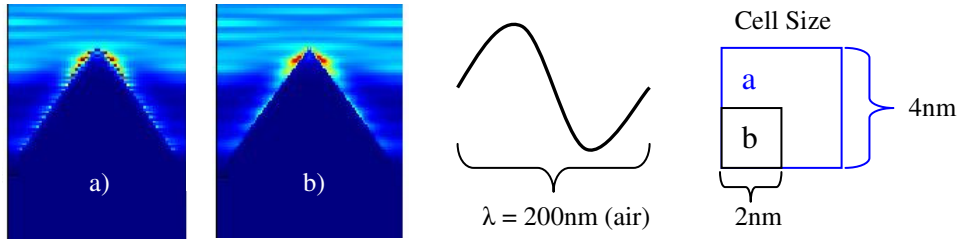


Figure 4.34: Increasing cells per wavelength. A significant difference is seen in image resolution. a) 4nm x 4nm x 4nm Cell size and b) 2 nm x 2 nm x 2 nm Cell size

Far-field projections of the intensity beneath the (uniformly illuminated arrays of) pyramids support the 2D diffraction grating behavior, but do so more colorfully.

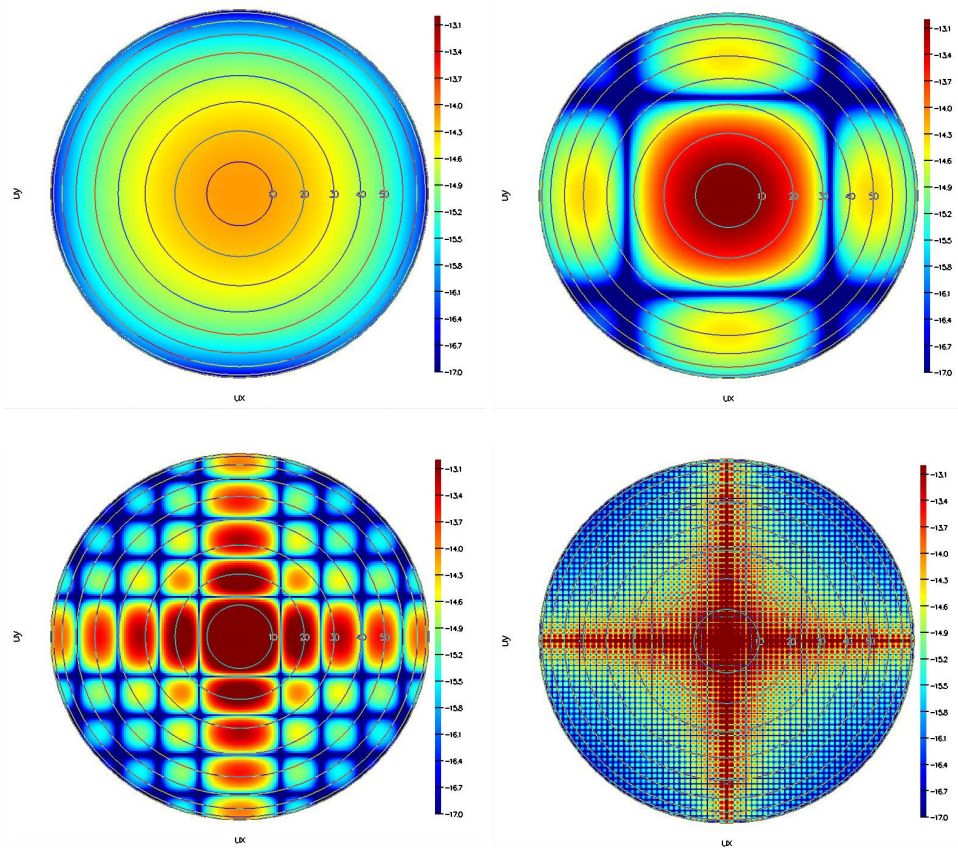


Figure 4.35: Logarithmic far-field projections of transmitted diffraction pattern from 20 nm underneath the surface of the pyramid a) Single structure, b) 2x2 array, c) 5x5 array and d) 500x500 array. Incident light is 650 nm. [X3.1]

The very colorful far-field pictures are indeed nice to look at, however for scientific clarity let us consider then the same monochromatic plane wave traveling in the direction and incidence on the square aperture in figure 3.36. We are interested in calculating a far-field flux density. The total disturbance arriving at P, a point on a screen a distance x from the aperture, is (from [X3.1])

$$\tilde{E} = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \iint_{\text{Aperture}} e^{ik(Yy + Zz)/R} dS, \quad \text{where } dS \text{ is the differential area element } dydz \quad (4.38)$$

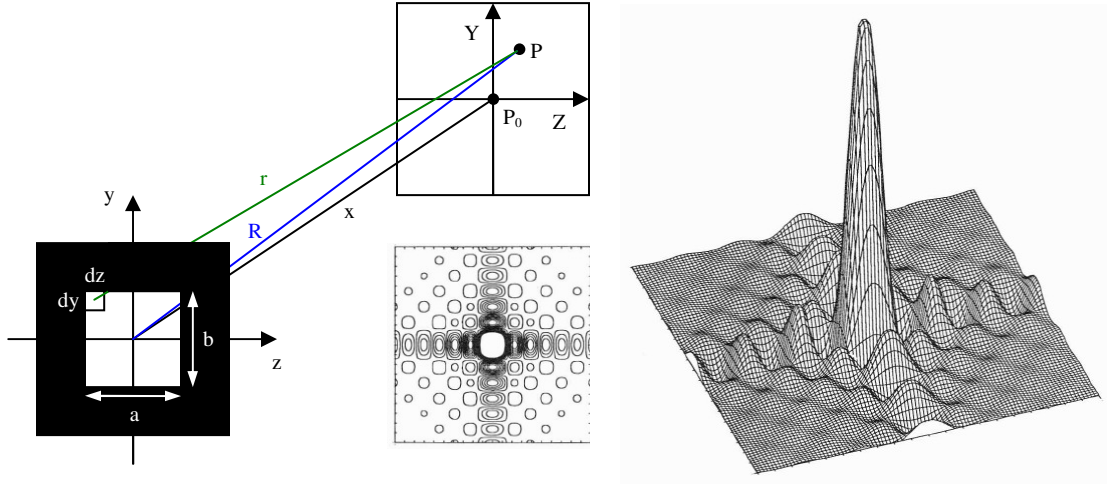


Figure 4.36: The far field projection of a rectangular aperture yields results equivalent to the square based pyramid and pillar [X3.20].

Equation 3.38 can be more specifically written as

$$\tilde{E} = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \int_{-b/2}^{+b/2} e^{ikYy/R} dy \int_{-a/2}^{+a/2} e^{ikZz/R} dz \quad (4.39)$$

solving the first integral yields

$$\int_{-b/2}^{+b/2} e^{ikYy/R} dy = b \left(\frac{e^{i\beta} - e^{-i\beta}}{2i\beta} \right) = b \left(\frac{\sin \beta}{\beta} \right), \quad \text{where } \beta = (kY)/(2R) \quad (4.40)$$

solving the second integral yields

$$\int_{-a/2}^{+a/2} e^{ikZz/R} dz = a \left(\frac{e^{i\alpha} - e^{-i\alpha}}{2i\alpha} \right) = a \left(\frac{\sin \alpha}{\alpha} \right), \quad \text{where } \alpha = (kZ)/(2R) \quad (4.41)$$

and substituting 3.40 and 3.41 back into 3.39

$$\tilde{E} = \frac{A \mathcal{E}_A e^{i(\omega t - kR)}}{R} \left(\frac{\sin \beta}{\beta} \right) \left(\frac{\sin \alpha}{\alpha} \right), \quad \text{where } A = ab, \text{ the area of the aperture!} \quad (4.42)$$

Since the irradiance (the intensity) is proportional to the square of the amplitude of the electric field, taking the real part of 4.42 and then squaring

$$I(Y, Z) = I(0) \left(\frac{\sin \beta}{\beta} \right)^2 \left(\frac{\sin \alpha}{\alpha} \right)^2 \quad (4.43)$$

Thus, $I(0)$ is the intensity at the center point P_0 , where Y and Z both equal zero. The instances in which β or α is equal to a nonzero integer multiples of π or in which Y and Z are equal to nonzero integer multiples of $\lambda R/b$ and $\lambda R/a$, respectively, yield $I(Y, Z) = 0$ and create the rectangular grid of nodal lines, as shown.

The increase in the number of structures in figure 4.35 is equivalent to the results obtained from a double slit diffraction grating. The contributions from the two slits will overlap at all locations on the plane of observance and although possibly equal in amplitude a non-trivial phase difference may be present. Each slit alone creates an equivalent flux-density distribution and as the number of slit elements is increased the results of all the slits are superimposed and the assorted parts combine to give a multi-wave interference system modulated by the single slit diffraction. In fact any array of either slit apertures or obstructions which produce a periodic change in the incident wave's amplitude and or phase is a diffraction grating, such as an array of pyramids or pillars. The multiple-slit experiment is a transmission amplitude grating as the incident wave front while traversing the grating experiences an array of opaque and transparent regions. If the material of the grating system were total transparent there would be a trivial modulation in wave amplitude, however the variations in optical thickness over the system would create phase modulation and is thus named a transmission phase grating. In the case of an array of nano- pyramids or pillars on a silicon surface the system performance exists in an area between both types of transmission gratings, modulating both amplitude and phase of the transmitted wave.

We now leave the idealized nano-structure grating system behind and begin to focus on the problem presented to us earlier in figure 4.19, self-(un)aligned sets of nano-structures, structures without easily definable densities and heights. This will eliminate any reduction of black silicon to a simplified "Moth's eye" picture. For the reader not familiar with this effect it can be said briefly that the surface of a moth's eye is covered with natural nano-structures, repeating patterns of hexagonal structures, approximately 200 nm high with a 300 nm center to center spacing, and the fabrication of structures meant to mimic this "moth's eye" takes place normally through some form of ion milling or lithography [X3.21, 22]. Both fabrication processes have nothing to do self-organization. Notice we just used the words "pattern" and "repeating" (words used to describe gratings with an incredibly rigorous order). Reflection is reduced as we have seen earlier in this chapter when structures are smaller than the wavelengths of interest the refractive index of the structures can form a gradient between the air and the medium, effectively removing the air-substrate interface [X3.2].

4.3 Simulations compared to actual physical structures

The best point to begin a discussion of optical properties is with those measured on structures that most definitely do not fit the desired application, but demonstrate very well in which direction a process can be modified in order to create structures with the correct set of properties. Figure 4.37 shows rather curious results and introduces the rather limited ability of the human eye to gauge the actual optical response of a set of nano-structures.

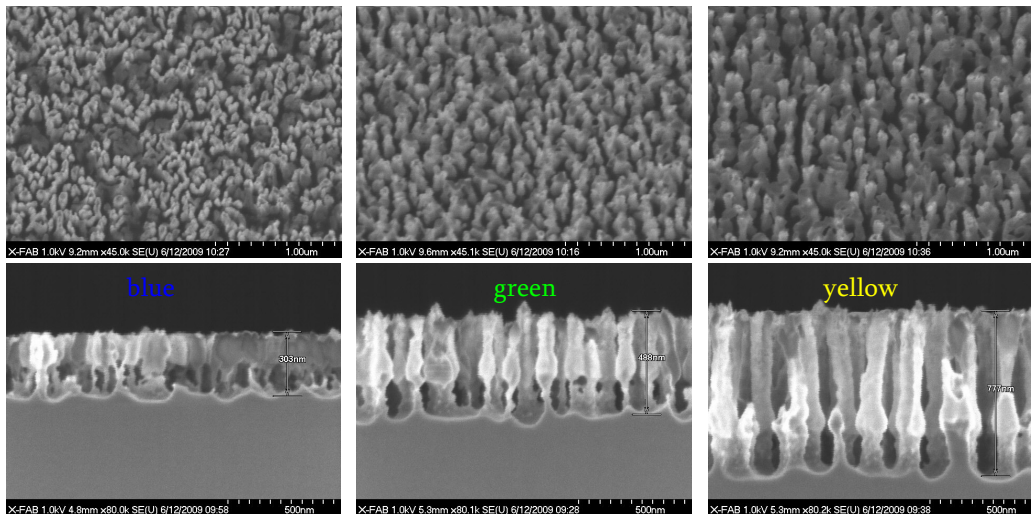
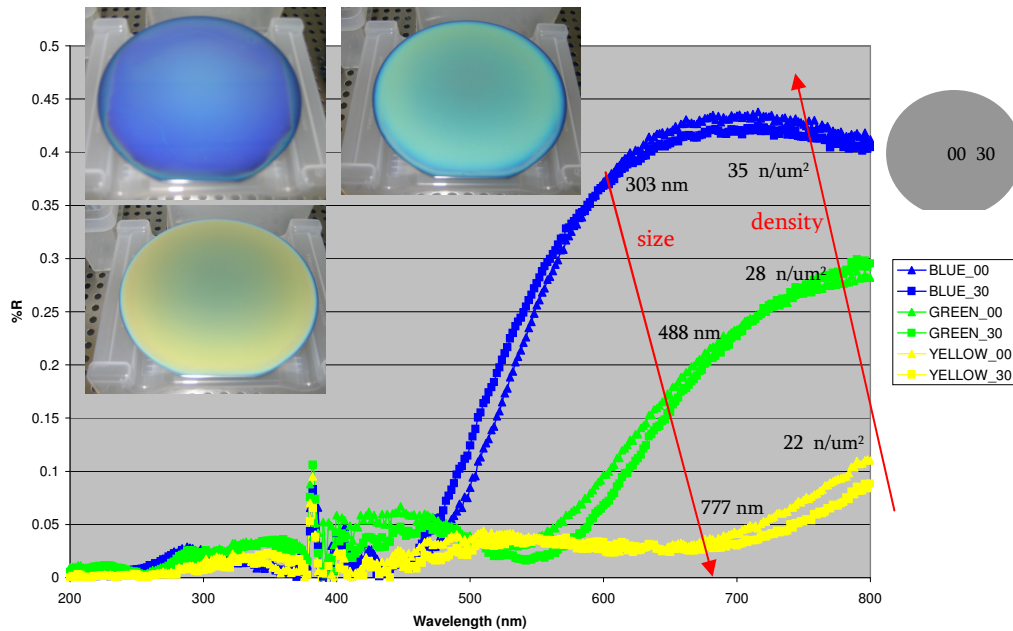


Figure 4.37: Black silicon... it's not just black anymore...

The digital photographs showing the distinctly blue, green and yellow wafers from above would certainly lead a casual observer to speculate that the reflected light content from each wafer represented only wavelengths in the corresponding color spectrum of visible light. This would be somewhat akin to the effect of a pigment, material which as a result of wavelength-selective absorption changes the color of reflected or transmitted light. An observer would most

naturally conclude (without having access to SEM pictures of nano-structures) that the explanation involves deposited layers of some substance or another, as a great many materials selectively absorb certain wavelengths of light. Indeed scratching the surface of these wafers, with a car key for example, reveals the reflective silicon surface beneath. Our observer would certainly conclude their explanation of a “coating” to be correct, much the same way scratching the door of their neighbor’s automobile with the same car key would reveal the silver metal surface beneath.

Actually it is entirely likely that we are discussing two problems for the observer, an apparent pigment coating on the wafer and the data collected by the photo-pigments of the observer’s eye (which could very well also yield the initially suggested incorrect curves). The three types of color photoreceptors (called cone cells in vertebrates) are located in the retina of the eye, each having different absorption spectra. Variation in perceived color is dependant upon the proteins found in the cone photoreceptor, each having a different response curve leading to the selective absorption of certain wavelengths of light. Pictured in figure 4.38, longer wavelengths of visible light, peaking in the yellow region (564–580 nm), are designated by L cones. Medium-wavelengths, peaking in green (534–545 nm), are detected by M cones. Finally shorter wavelengths of light, in the blue-violet region (420–440 nm), are collected by the S cones.

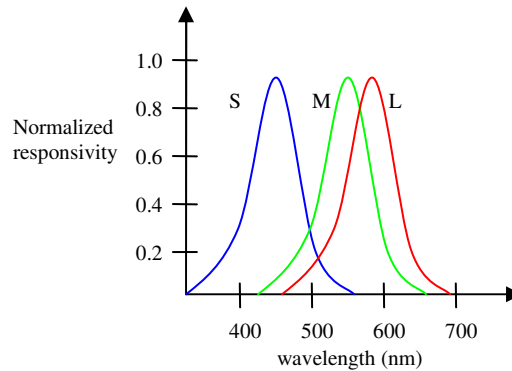


Figure 4.38: Normalized responsivity spectra of human cones, the photoreceptor cells in the retina of the eye. S, M, and L types, spectral sensitivity. [X3.23]

After optical stimulation the brain receives a set of signals from the three cone types and subsequently discriminates between the protein attenuated signals resulting in a conscious perception of color qualia. Although this process is well outside the topic of this dissertation, it is clear that the process is again grounded in the quantum chemical properties of pigments and not completely representative of a reflected spectrum from a structured surface. However we are most fortunate enough to have corresponding SEM pictures (it’s not a pigment coating) and a photodetector capable of objectively measuring wavelength content (no attenuated protein signals). Thus, the effect is the result of a (not-always-so) simple geometric modification of the silicon surface.

The shape of the structures in figure 3.37 could quite justifiably be referred to as swords, with a definite blade, hilt and handle region (most certainly for the yellow structures). As seen, with increasing structure size (height) wavelengths of greater and greater length are incorporated into the antireflective behavior as measured by direct reflection, not quite matching our observer’s initial assumption that the reflected content was composed of only the corresponding visible spectrum wavelengths around for the perceived color. An inverse relationship between density and antireflective behavior seems (could seem) to be visible, however figure 3.39 demonstrates the rather difficult task of counting the structures. This is an especially bothersome

problem when structures do not exhibit a distinct separation from one another. From above SEM results show a maze of wall-like structures tracing around the wafer surface. For fun readers are encouraged to follow an experiment carried out by the author. Encourage office colleagues to count the number of structures in a $1\text{ }\mu\text{m}^2$ area to ensure a relatively stable consensus for the density value. The results are guaranteed to be as inconsistent as they were here. Perhaps there is a more intelligent way to characterize the surface?

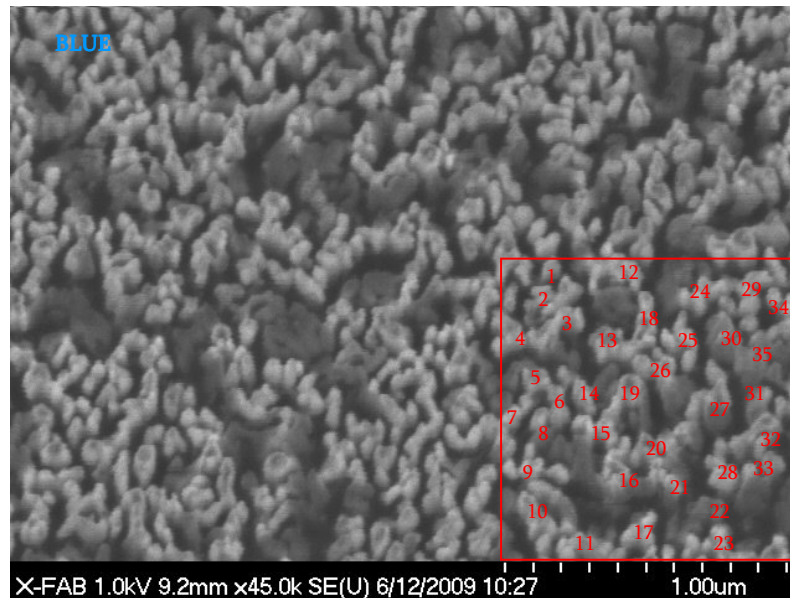


Figure 4.39: Complexity of accurately counting nano-structures. What counts as a nano-structure, particularly for cases in which a “wall” is observed? A more applicable method of calculating density is presented in the next section. The result of “walls” versus individual structures on reflection is addressed at the end of this chapter.

Figure 4.40 are the best overall results attained in the morphology chapter. Pictures from the side and at a 20 degree angle are shown for clarity (differences are noticeable) but we immediately see a very important difference between center and edge location top-down SEM pictures and corresponding reflection values. Separation between structures on the wafer edge is unmistakably better, leading to a more cobweb pattern. The wafer edge is consistently darker (blue = specular, green = diffuse, black = combined). Notice now figure 3.41, the same recipe however with a reduced second step etch time which reduces the overall structures size. Again we see the same differences in structure separation to reflection values. This can be moved an additional step further when we see the results of figure 3.42. This wafer was etch be a completely different process in a drastically different machine, but we see very stark similarities between it and figure 4.40. In fact based on the incredible antireflective performance of the structures in figure 4.42 we can discern the areas in which our structures need to be improved. Notice the near perfect cobweb pattern formed in the top-down SEM pictures compared to the best results obtained here. Structure separation appears to be absolutely key. Also very fine, extremely sharp points seem to considerably reduce the reflection due to the two resonant peaks of silicon.

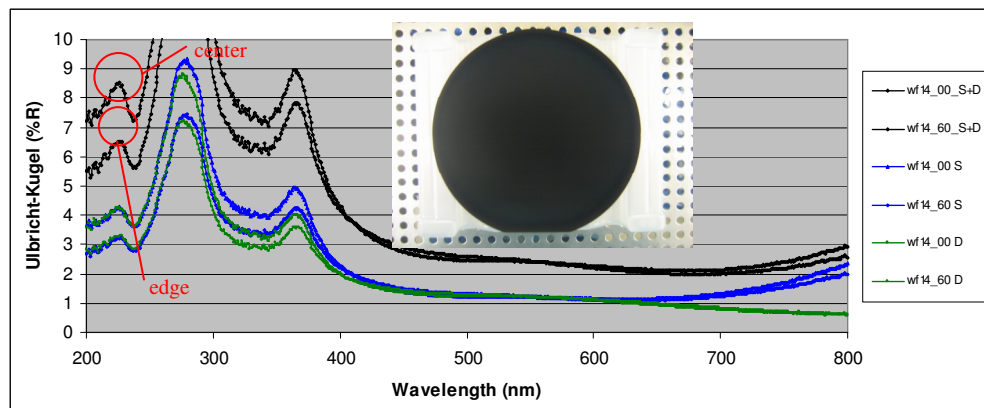
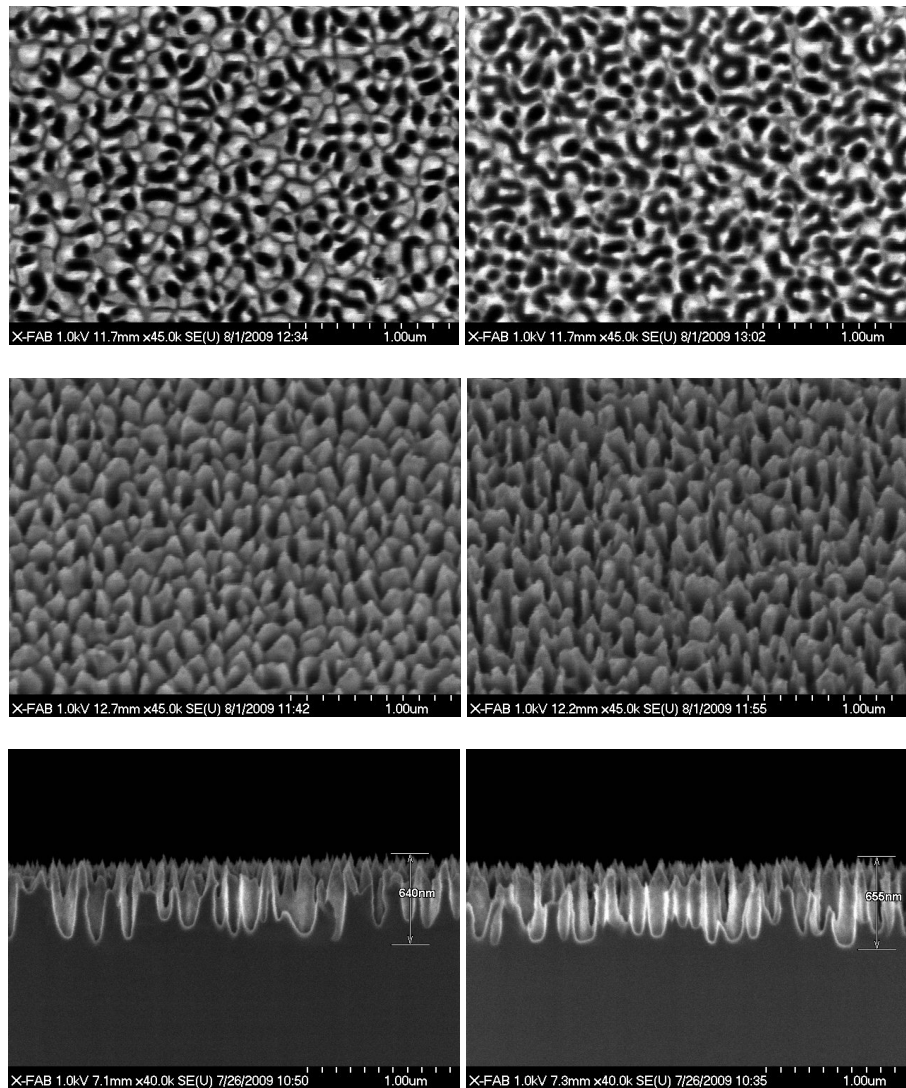


Figure 4.40: Integrating (Ulbricht) sphere results from the optimal nano-structured wafer from the morphology chapter (left = center, right =edge).

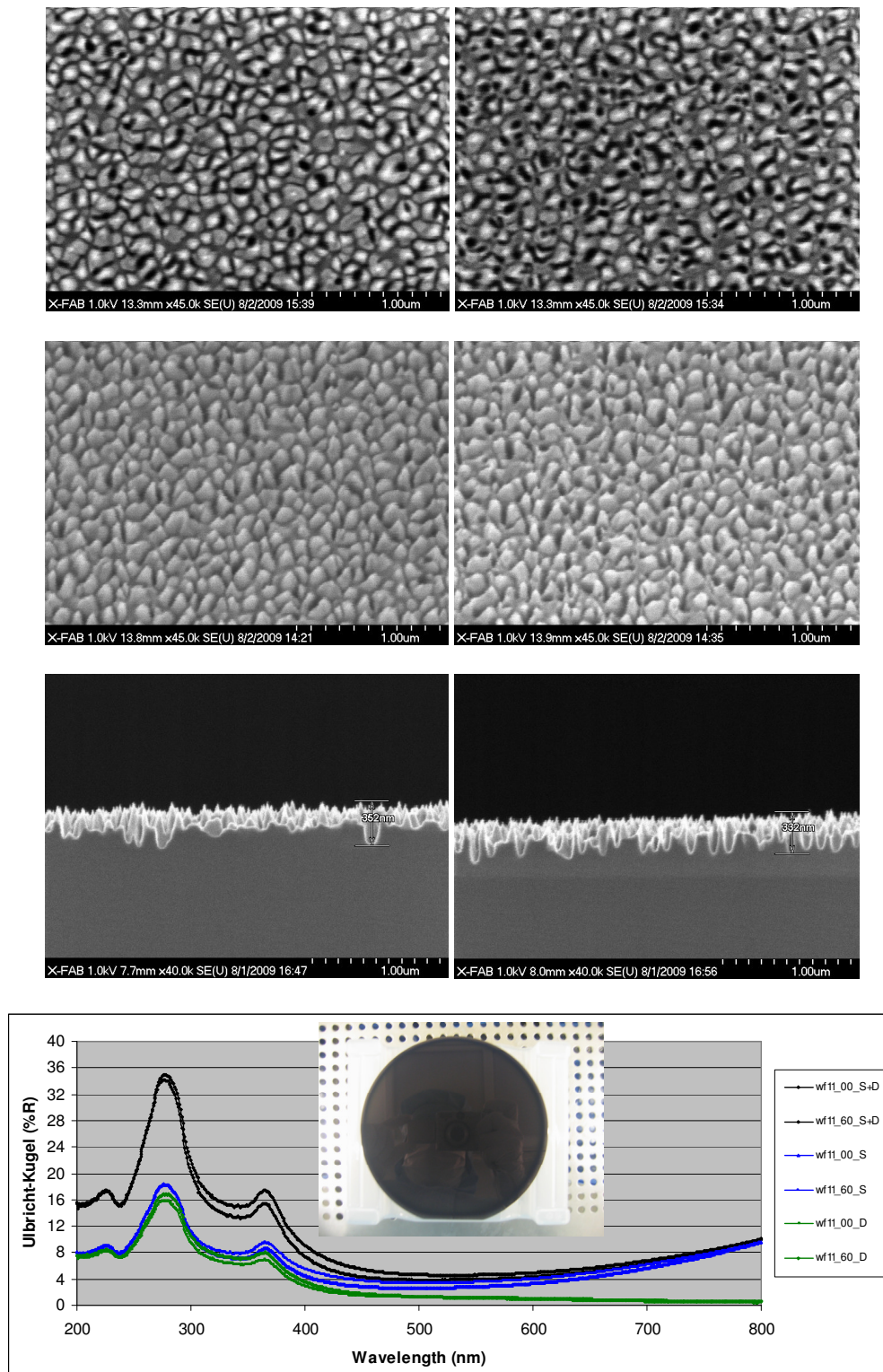


Figure 4.41: Integrating (Ulbricht) sphere results from a nano-structured wafer etched as above however with a reduced time. The reduction in antireflection is clear (left = center, right =edge).

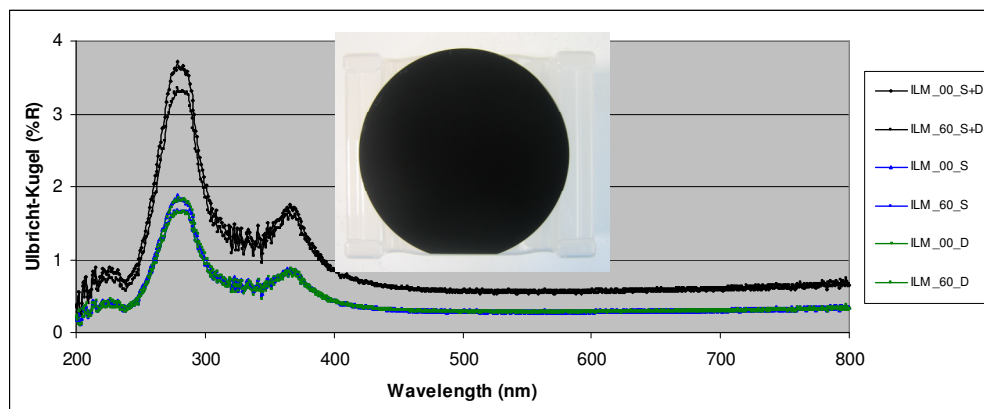
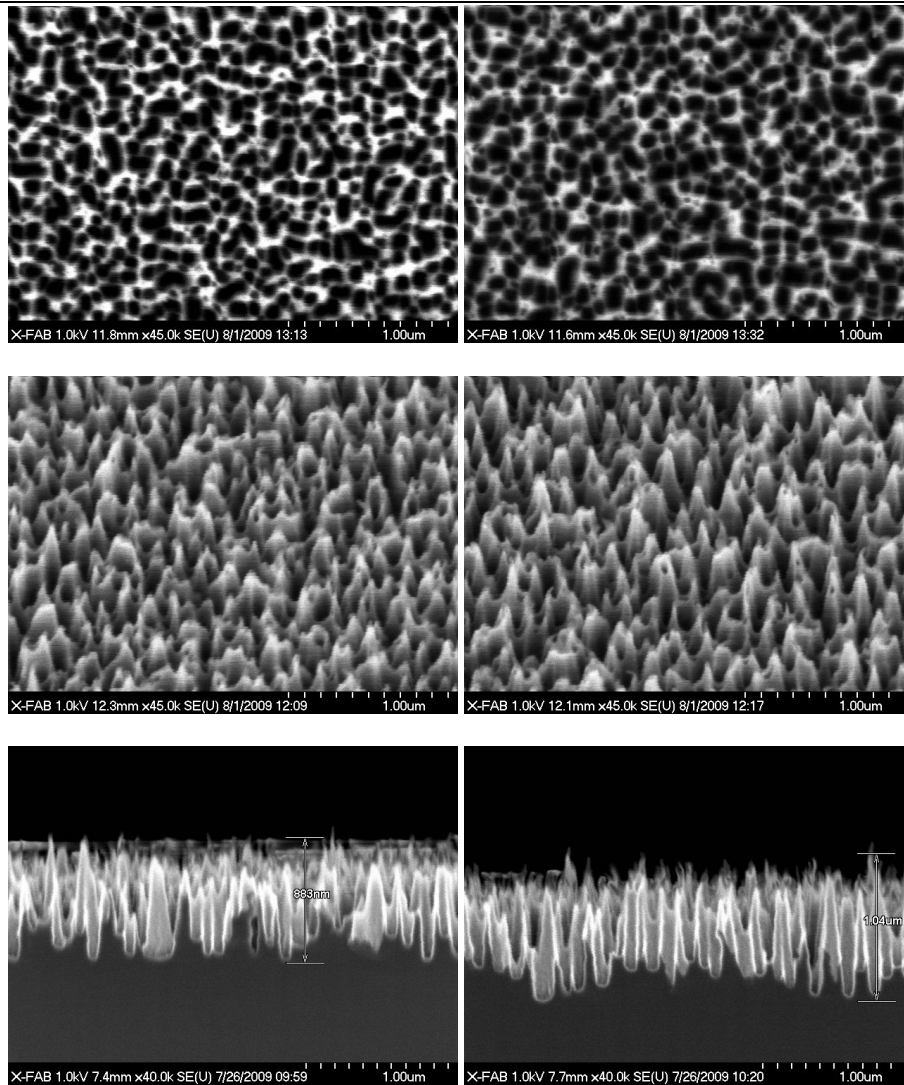


Figure 4.42: Integrating (Ulbricht) sphere results from a very antireflective wafer, although not etched (Thank you very much to Michel Fischer of TU-Ilmenau for this sample!!) (left = center, right = edge).

4.3.1 The lack of periodicity in self-organized nano-structures

Correlation is a mathematical tool used to find (sometimes less than obvious) repeating patterns, such as the presence of a periodic signal buried under noise. The degree of similarity between two waveforms is determined by a time-lag (or space-lag as we will see here) applied to one of the two signals with respect to the other. In signal processing this is often referred to as a sliding dot- or inner-product. Frequently a person using a correlation is seeking a short known signal trait within a considerably longer duration signal. The term autocorrelation is used when a person, instead of correlating between two different signals, correlates a given signal to itself, again with a time-lag (or space-lag) applied to one of the two signals [X3.8].

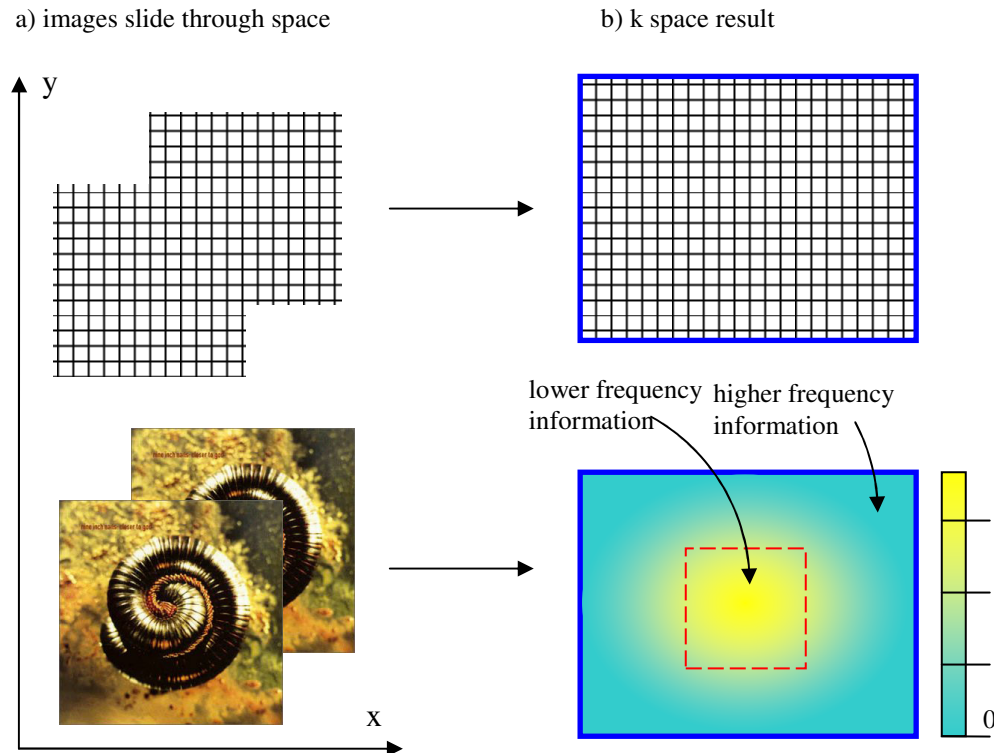


Figure 4.43: Picturing the relationship between k-space and an image with autocorrelation. In part a) the top image (a perfect grid) and the bottom image (a color picture) are shown overlapped by a copy. It is this overlap that we are interested in evaluating. Values are assigned to each xy position (based on color scale for example). The images begin in the same xy position (the value of the color scale is the same at each xy), if the copy is shifted only a small amount with respect to the original image it is likely that the color scale value at a given xy coordinate is still similar. Only the very finest lines (highest space frequencies) will no longer overlap. As the copy is shifted further and further even the larger space frequencies cease to correlate at a given xy coordinate. This process yields the k space frequency plot in b) with low frequencies towards the center and higher frequencies toward the edge. The grid clearly has a periodic k-space, as its copy is shifted through space there is periodically a perfect overlap [X3.8].

Figure 4.43 makes the autocorrelation in space more tenable. The relationship between k-space and the corresponding image is shown. Part a) has an image of a perfectly periodic grid (top) and a non-periodic color picture (bottom) with a copy of the same image overlapped. We are interested in evaluating the quality of this overlap. Each xy coordinate in the space is assigned a value, possible based on the color map but this could be any user defined scale. The target is to

create a topographical map of the image, a function of x and y (the position) and z (the value of the function at the xy point). Both original image and copy begin at the same xy position. If the copy is shifted only very slightly with regard to the original image, the z value of the original image and the copy at a given xy coordinate is likely still the same value (or at least very similar). This may not be true for the very highest space frequencies (the very finest lines of detail), as they will no longer overlap. Shifting the copy further will result in the larger space frequencies (the larger blocks of solid color) ceasing to correlate at a given xy coordinate.

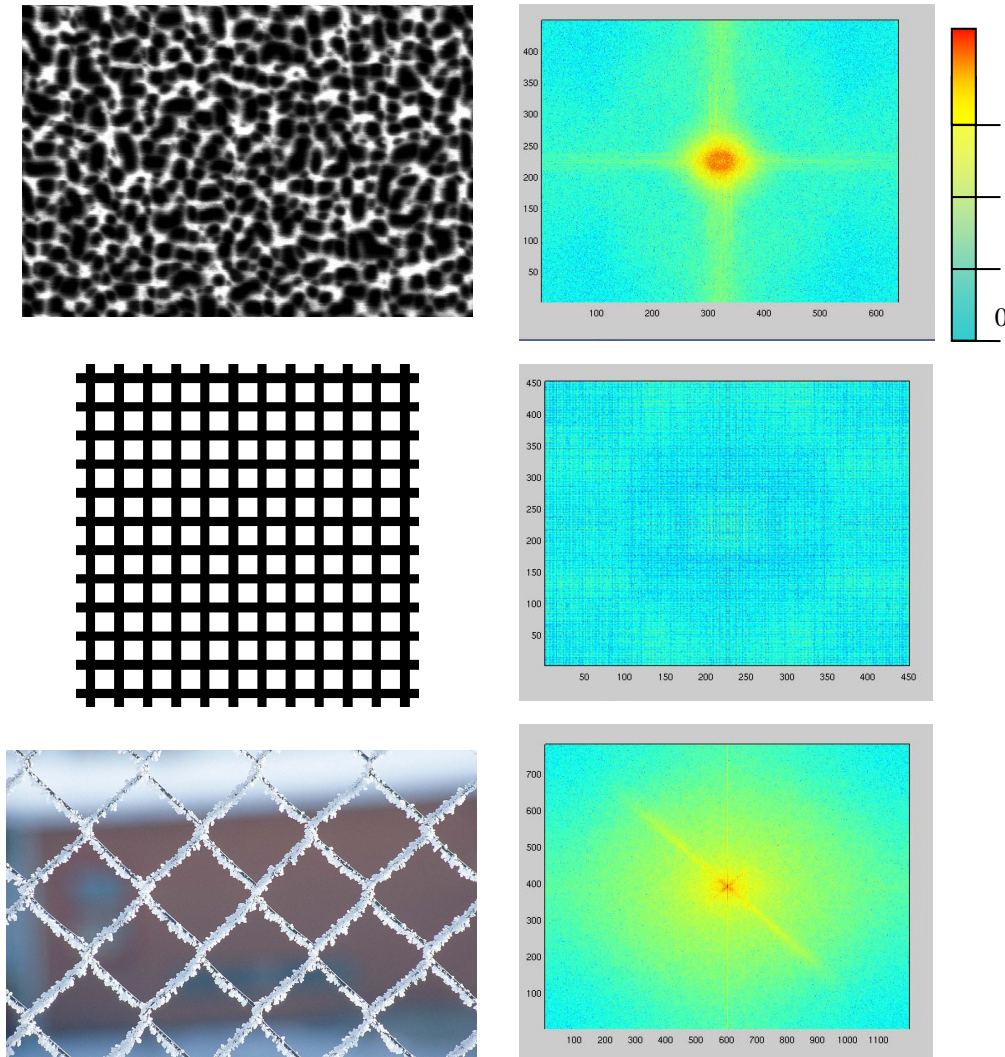


Figure 4.44: Evaluation of periodicity by autocorrelation. The pictures on the left-hand side are shown next to their k -space equivalents on the right-hand side. Auto-correlations were calculated with Matlab. The nano-structures have an extremely small correlation length, meaning the area encompassing the majority of the Gaussian distribution is concentrated in the direct center of the graph (red center surrounded by the yellow ring). The periodic grid not surprisingly creates a periodic k -space. The image of the fence proves to be quite periodic. The Gaussian tapers off well into the higher frequencies (see the large yellow area of correlation).

This process yields the k space frequency plot in b) with low frequencies towards the center and higher frequencies toward the edge. For example if the k -space results of the color picture were

transformed back to space, but the information outside the red box was excluded, the result would be the color picture minus its high-frequency components. The picture would appear blurry without and sharp lines. Excluding the information inside the red box would have the opposite effect. The resulting space image would retain its sharp lines (its outline). The grid clearly has a periodic k-space, as its copy is shifted through space there is periodically a perfect overlap. One can imagine that a solid color image (constant value) would have a solid color (constant value) k-space.

We now apply this method to evaluate the periodicity of our nano-structures. Figure 4.44 shows the autocorrelation results of several images: (from top to bottom) very antireflective nano-structures, an ideal grid, an image of a fence (possibly periodic). As discussed, the grid, due to its periodicity, creates a k-space that is also a periodic grid. The fence image creates a very Gaussian k-space retaining a great deal of correlation out into higher frequencies. The etched nano-structure sample shows nothing in common with the grating, indeed there is no ordered periodicity about the surface distribution of nano-structures. Their formation and distribution across the wafer surface are a stochastic processes. Although not done in detail here, this method can also be used to calculate average island (nano-structure) density by selecting a certain bandwidth in the center of the Gaussian distribution [X3.25]

This method can be used to create models of more randomly organized (meaning less correlated) surfaces. Manipulating a frequency domain picture to be closer to that of the one in figure 4.44 for the nano-structures (a very narrow, noisy Gauss) and transforming it back to the space domain can create a random surface distribution. A natural question is, how does this impact the simulations of the previous sections? The answer is given in figure 4.44. Seen in this figure are a various 3D surfaces created in the FDTD simulation space, in a) we see the ordered sets of pyramids, in b) a disordered array of pyramids and in c) a very needle-like surface constructed from the inverse transform of a frequency domain Gaussian distribution with a very narrow sigma multiplied by white noise.

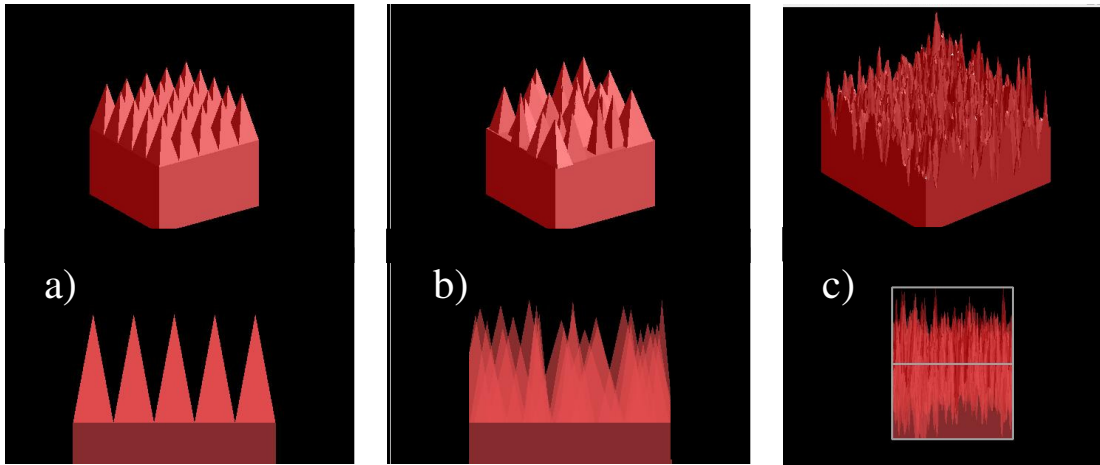


Figure 4.44: a) an ordered 5x5 array of pyramids with height = 700 nm, x base-width = 200 nm, y base-width = 200 nm, and spacing between tips = 200 nm, b) a disordered surface of pyramids (all distances the same as in part a) however with a random offset for each parameter between ± 100 nm. c) A Gaussian distribution multiplied by white noise creating a random surface.

The random surface matches the real black silicon surface because of the similarity in autocorrelation. On an array of pyramids if we move any given distance from a starting point we will always know the value of the function because the surface repeats with a fixed interval. If however we move from a starting point on the black silicon surface we know that for very small

distances the value of the function (the height, meaning surface topography) will be close to where we started, the surface is in our case never represented by a periodic function in frequency space. The slope may be quite large but it is always a smooth transition. However past a certain distance it is totally unknown what the value of the function will be. The sharp transition can be implemented by decreasing the sigma of the Gaussian distribution. Noise is necessary because it must be an unknown move up or down in function value.

A comparison between integrating sphere measurements, an ordered array of pyramids with the same height and a random surface with the same height is shown in figure 4.45. Clearly the ordered array of structures creates the optimal anti-reflective coating. The random surface not only appears to look like the real surface, its antireflective behaviour is also very similar.

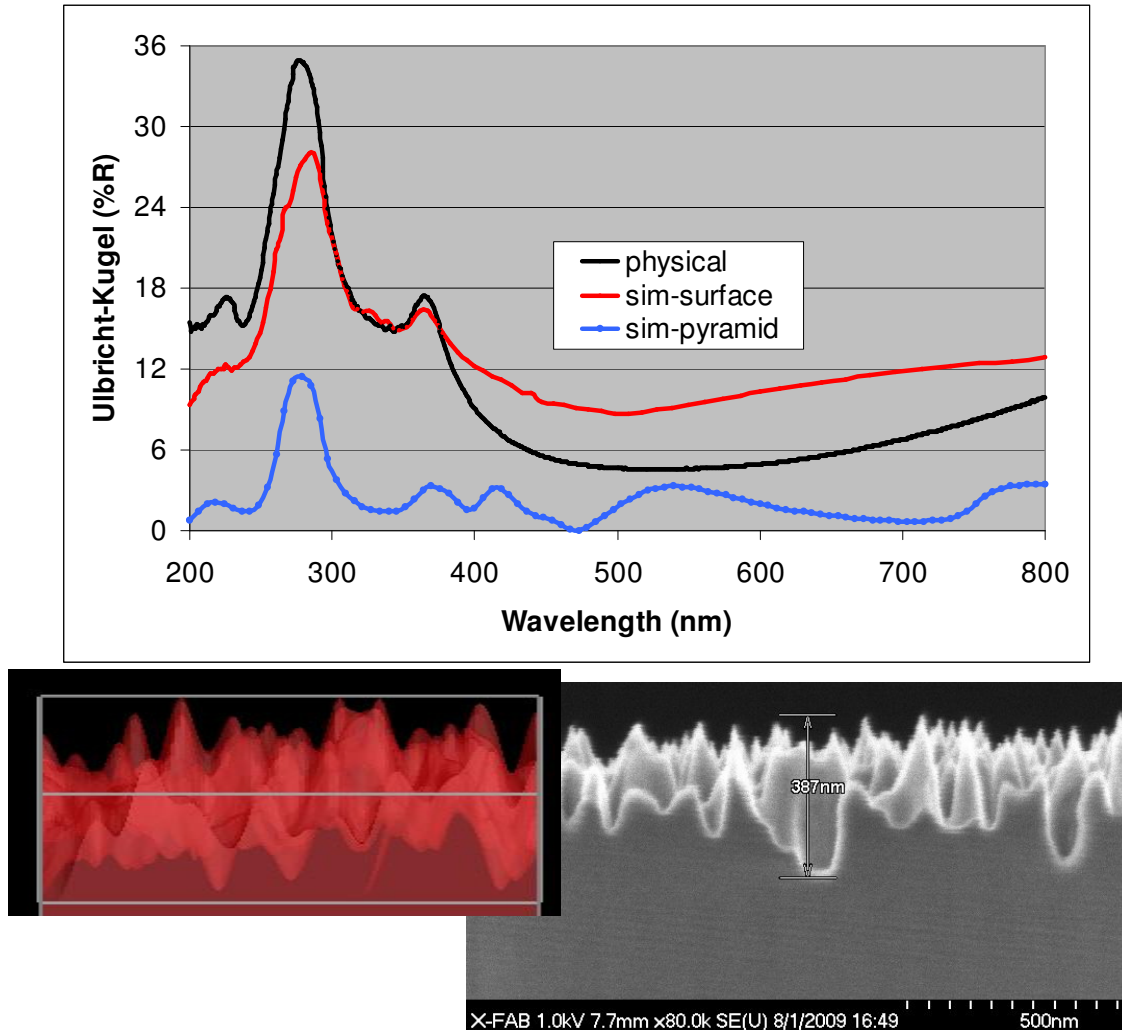


Figure 4.45: Comparison of integrating sphere measurement (in black) to an ordered array of pyramids (in blue) all with the same height and to a “random surface” (in red) with the same maximum height as the structures in the SEM (approximately 400 nm). The simulated random surface in red matches the physical measurement in black more closely. Disparities between the red and blue curves are likely due to the definition of maximum height (setting the highest possible peak value) and average height (setting the number of peaks reaching the maximum height). More peaks allowed to reach the maximum height appear to reduce reflection.

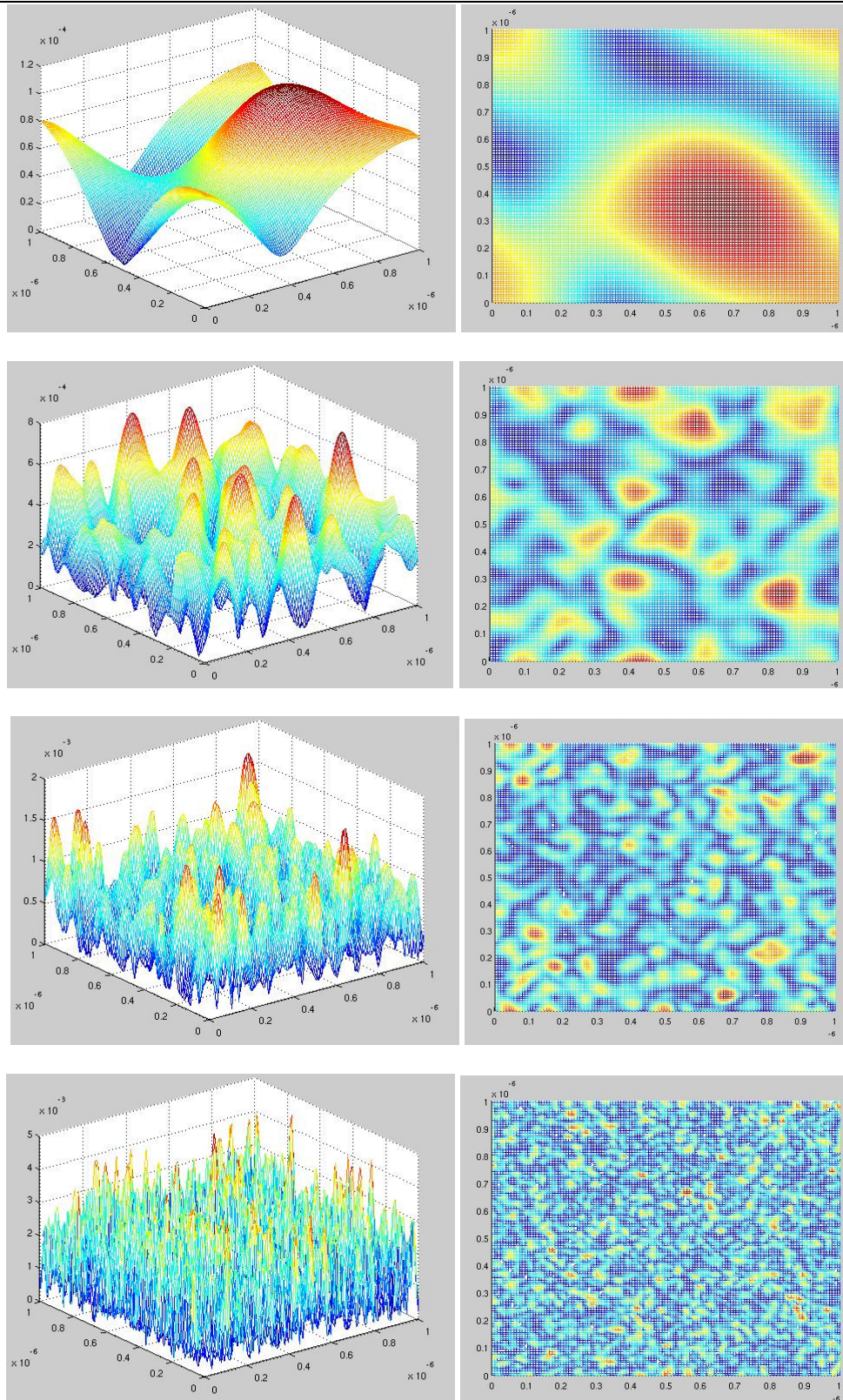


Figure 4.46: Surfaces created in Matlab with increasing sigma (from top to bottom increasing sigma)

Variations in the sigma parameter are shown in figure 4.46 depicting a surface with a very slow transition (wide sigma) to an extremely narrow sigma (a case so narrow that it is not worth considering the existence of the Gauss!). Different functions could be implemented for different structures. A square wave for slightly more anisotropic structures, a triangle wave would yield results similar to Gauss, and very interesting results could be implemented with the so called Mexican Hat function.

4.4 Concluding comments on optical properties

This chapter has introduced and implemented Finite-difference time-domain (FDTD) simulations to model the electromagnetic interaction of light with nano-structures on approximately the same scale as the spectrum of interest. Background on the discretization process of Maxwell's equations and the Leapfrog technique for time-domain problems was shown using the Yee approach. The FDTD results provided insight into the nature of the effective index of refraction for nano-structures and yielded a detailed discussion of the derivation of the index of refraction including resonance peaks and the corresponding absorption around such a peak.

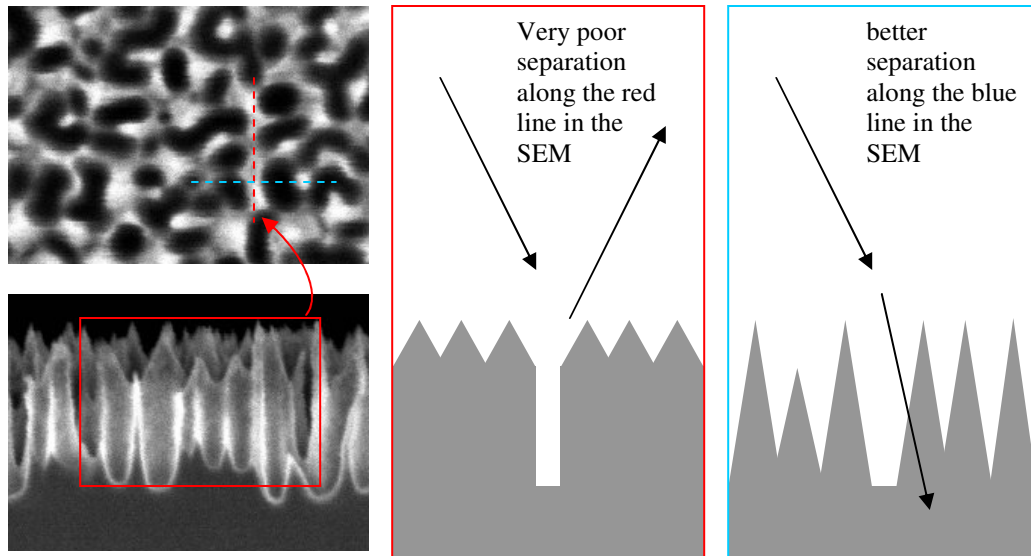


Figure 4.47: Necessity to achieve structure separation. Electric field components of figure 4.03 when incident in the blue plane (shown on the SEM with a blue dashed line) experience a more “gradient-like” index, and have a smaller reflected electric field component. Electric field components incident in the red plane (shown on the SEM with a red dashed line) experience something similar to Fresnel reflection from a flat surface. All physical reflection measurements were performed with an unpolarized laser (equal intensity in the blue and red plane, and all planes in between). We can then assume that the reflected intensity measured is a combination of both Fresnel-like, reflection from electric field components incident in planes containing “wall” structures (like the red plane), and insignificant reflected intensity, reflection from electric field components incident in planes of a more gradient-index nature (like the blue plane).

Furthermore the difference in functionality between the two extremes of nano-structure shape (pillar and pyramid) yielded the interesting results that the pillar can be represented quite accurately as a medium with a single index of refraction value (a geometric ratio between the amount of “pillar” and the amount of “medium” surrounding the pillar). The pyramid was shown to be a true gradient index.

The periodicity of nano-structures was investigated and demonstrated clear diffraction grating properties, which did not prove to be a useful way to evaluate the measured specular and diffuse components of real physical structures. The problem being namely that overlapping pyramids (for example unseparated through the etch process) disturb the conditions of periodicity (diffraction into component angles) corresponding to the structures overall width. As the ideal width of a needle was set at between 100 nm and 200 nm for the light spectrum of interest (200 nm to 800 nm) any periodic structure with widths over 200 nm will begin diffracting light into non-specular angles. Furthermore the dependence of structure size height width on reflected intensity was calculated with heights over 500 nm yielding negligible reflection.

Boundary conditions for the FDTD were discussed in the process of simulating non-normal incidence light on 2D structures. Dispersion in the injection angle for a simulation created difficulties in simulating broadband spectrums in a single simulation, as only the center wavelength would have the correct theta, while the smallest and largest wavelength in the incident wave would experience the most extreme error in incident theta. Simulating each wavelength, one after the other, was not possible as it yields incalculable computational times. The solution provided broke the desired spectrum into several smaller spectrums and provided satisfactory results while limiting the amount of dispersive error.

Simulations in 3D gave excellent insight into the form of electromagnetic waves within the nano-structures, particularly by providing the opportunity to view cross-sections of individual structures at various depths (and also very colorful far-field results). Transmitted intensity beneath structures provided the final relationship between antireflective properties and captured light. The chance was also taken to discuss material absorption coefficients as it varies with wavelength and the band gap value the material, leading to the statement that photons with energies greater than the band gap are readily absorbed while photons with energies less than the band gap are transmitted.

The explanatory strength of simulations was best correlated to the optical properties of physical structures (Ulbricht-Kugel measurements of specular and diffuse components – not a grating!), when a truly random surface was simulated (a surface with a very limited correlation length). An immense amount of work could still be done in the area of simulation of real structures but for the sake of brevity we will now move on to the electrical characterization and integration of black silicon diodes.

5. Electrical properties

Up to this point the nano-structuring of silicon has been discussed in detail and the resulting optical properties have been related to the morphological process utilized. We discussed the goals of this project as not only being the creation and characterization of nano-structures exhibiting ideal antireflective properties, but also their integration into real photodiodes, within the framework of a real process. In semiconductor manufacturing this is fittingly called process integration and will be discussed in section 5.1. The main sources of possible device damage encountered in this project will be presented in 5.1.1 (lattice damage and dangling bonds on the wafer surface) and the implemented solutions with their evaluation will be shown in 5.1.2. Section 5.2, appropriately titled Diodes, will provide an explanation of illuminated diode functionality in section 5.2.1 and end with the much anticipated spectral sensitivity measurements of diodes in section 5.2.2.

5.1 Process integration

The process integration engineer considers the interactions between different process operations from the outset, rather than optimizing them separately. The role that one fabrication step plays on influencing subsequent steps, or possibly the results of previous steps.

For the results discussed thus far, plasma etching has played the important role. However for the creation of really functioning diodes, even in the most simplified picture, one requires n- and p-wells (doping and high temperature drive-in processes), contacts and interconnections (metal deposition and etching), isolation from other devices (more wells, field oxide), and each of these steps requires some form of lithography. Essentially the creation of a single diode takes place through multiple steps, in numerous machines, processed by many different engineers, technicians, and operators. The possibility for failure is certainly present and the methods used to determine where and which steps created a problem can be quite complicated as often a mistake is not known until the electrical properties of the device under investigation are measured.

Clearly the topic of process integration is considerably large and consequently not every process step and extraneous circumstance will be evaluated. The focus of this section will then be on what we know we damaged, the diode surface through doping and plasma etching, and on what we can do to repair it.

5.1.1 Defects through etching and doping and subsequent annealing

The first of the two primary defects considered here is damage to the substrate lattice, and although most often (rightly) attributed to doping (n- and p- well creation) it is most certainly not limited to doping. During a plasma etching process the substrate is bombarded with highly energetic ions, which not only sputter atoms from the surface, but will also implant themselves in the substrate. Temperature annealing of substrates after standard implantation processes normally must take place at a sufficiently low-enough temperature to ensure that the desired doping profile of impurity ions does not experience too great a migration. As will be explained, ions of the same type and energy (for example accelerating across a fixed sheath width) do not in general penetrate to exactly the same depth (even a standard doping profile has a tolerance range). Furthermore it has been verified that incident ions can penetrate quite deeply into a substrate when the angle of incidence is parallel to a lower-index crystallographic direction [X4.6]. For example as the sheath becomes crowded with diffusing gas and etch products, subsequent collisions alter the incoming trajectory of ions, which then travel along different incoming paths meeting different crystal

orientations. One can imagine that this latter effect would be compounded as various nanostructures begin to appear on the surface exposing such orientations to further non-perpendicularly incident ions.

Once in the crystal ions can travel down the open spaces between atomic planes, essentially channels with sides comprised of atomic planes, with crystallographic damage occurring as the energetic ions collide with the nuclei and electrons in the substrate, losing energy and eventually coming to rest as an interstitial, pictured in 5.01. These collisions with lattice nuclei may dislocate the lattice atom itself leaving behind a vacancy and creating an interstitial out of the former lattice atom [X4.10]. Furthermore As these collisions are random, both in number and in amount of energy transferred, ions of the same type do not have the same range, the distance traveled into the substrate. Indeed, when considering implanted incident ions, we can easily take as a starting point some general features of standard ion implantation processes (obviously noting that for normal implant processes the energies of incident ions, and thus implantation depths are, much much greater).

Regardless of whether the target substrate is amorphous or crystalline, the implantation range is dependant on the density, mass, and atomic number of the target's atoms, and on its temperature during processing. Clearly this means that the range is also strongly dependant on the energy, mass and atomic number of the incident ions. If the target is crystalline the aforementioned crystallographic orientation also influences the range. It is not a simple problem to calculate a definite range when searching for something such as a "depth of substrate damage" parameter during a process because of the extremely wide difference in exactly these values just mentioned which one normally considers.

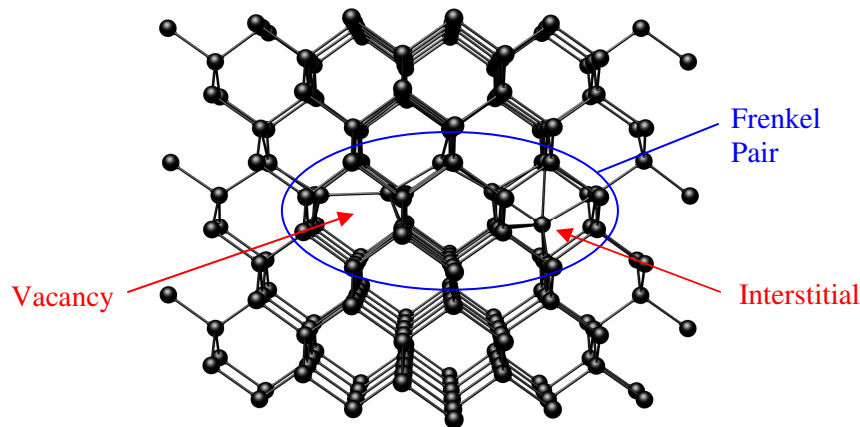


Figure 5.01: Vacancies and interstitials in silicon, when in close proximity to one another these two defects are called a Frenkel pair. Such defects create magnificent opportunities for electron trapping and recombination. [X4.10]

For example the mass of projectiles striking the surface varies from 5.5×10^{-4} AMU for electrons to hundreds of AMU for the heaviest atoms, and as the surface becomes structured new paths for these projectiles are opened (or closed) [X4.7]. It is customary for simulations to assume that for the high energy small particles most energy loss occurs as a result of interactions with electrons bound in the target [X4.8]. Here, the quantum mechanical interactions are of primary importance as they fix the amount of energy transferred to the bound electrons. In contrast quantum mechanical effects are ignored for heavy projectiles with lower energies and in the case of amorphous substrates it is possible to make some very good estimates [X4.9]. For crystal targets this is not so well defined, resulting from the channels created by atomic planes mentioned above,

as in such targets the incident projectiles can have a much smaller energy loss with distance and likewise a larger range in comparison to amorphous targets.

The second type of defect, actually it takes some time getting used to calling it a defect, is the surface itself. One can imagine this quite well when bearing in mind that an atom missing a neighbor to whom it would normally be bound will possess an unbonded valence electron. This is called a dangling band and can be expected quite normally then at places where a material terminates for example at a surface interface, such as silicon-air or silicon-oxide. This dissertation uses silicon but one could consider any tetrahedral semiconductors, the elements of column IV, the III-V and II-VI compounds. Such materials have a structure in which each atom has four neighbours surrounding it in such a way as to create a polyhedron composed of four triangular faces with vertices at each of these neighbours (hence the name tetrahedral). The number of valence electrons per atom is four and the arrangement of bonds, each bond having two electrons, is shown pictorially in figure 5.02 a). The arrangement has an extremely covalent quality, indeed a property displayed by these substrates.

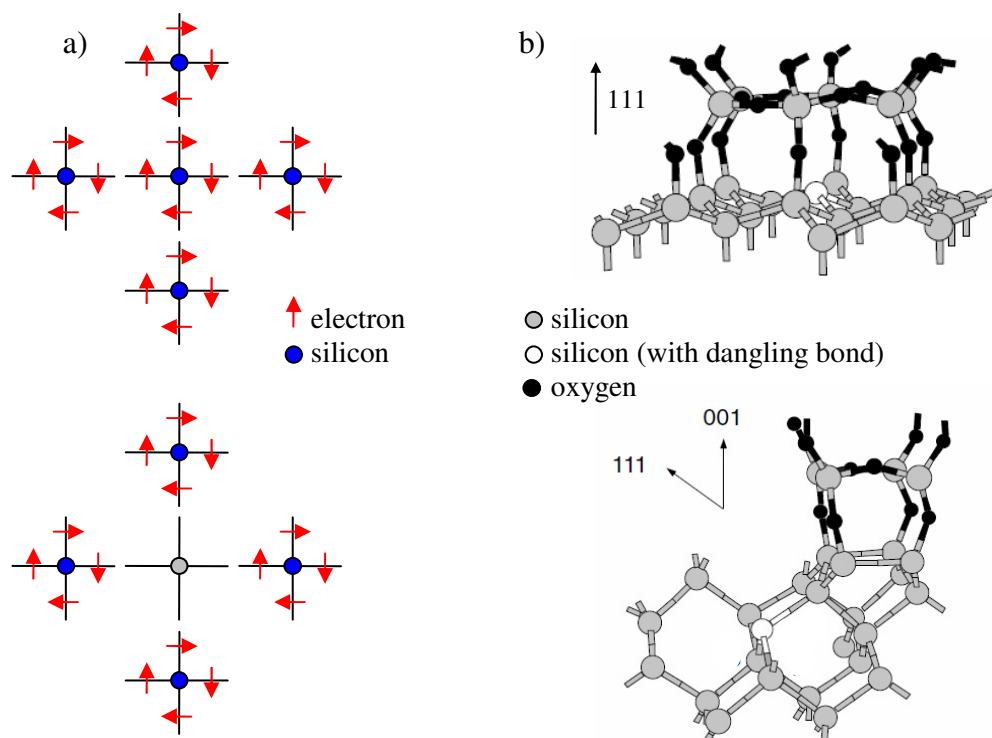


Figure 5.02: a) Simple Lewis structure of the dangling bond [11] b) 3D schematic of the dangling bond at the Si-SiO₂ interface for several crystallographic orientations [X4.13]

This relatively simple picture of the dangling bond is justified when one considers the molecular description. Removal of one silicon atom creates a situation where the four sp³ hybrids which, pointing towards the vacancy, have lost their corresponding bond partner for normal formation of bonding and anti-bonding states. These are now silicon “dangling orbitals” with energy equal to the sp³ energy. With a vacancy there are four electrons to allocate across the four “dangling bond states”, essentially one electron per dangling bond, making palatable the more simplified Lewis structure. But under what circumstance would our silicon be missing a neighbour? Moving to figure 5.02 b) we see the structure of the dangling bond at the oxide silicon interface for several

directions. Not simply a consequence of the realistic non-zero roughness of the silicon substrate (although this will certainly deform the Si-SiO₂ interface), well characterized voids appear at the interface between the two media due to the different chemical potentials which silicon and oxygen atoms possess leaving unsaturated valencies (our dangling bonds) [X4.12, 13]. In CMOS technology these are typically referred to as a form of point defect, with their presence degrading MOS electrical performance [X4.2]. In the figure we see the point defect, the vacancy in the silicon.

As the diodes used in this work are formed by implanted wells in a crystalline silicon substrate subsequently receiving a black silicon surface texturing process through plasma etching, we can certainly expect a great many dislocations, vacancies, and interstitials within the substrate and a great number of dangling bonds (or at the very least a less than ideal result) covering the surface of the etched nano-structures. The question is why and how do these etched and implanted conditions concern our illuminated silicon photodiode?

Light incident on the surface of the diode creates excess carriers (their generation process will be discussed in detail in section 5.2). Radiative recombination is the process by which an electron (for example generated by our incident light) undergoes a transition to the ground state, moving from the conduction band to combine with a hole in the valence band, releasing the excess energy as a photon. If the electron moves directly from the bottom of the conduction band to the hole near the top of the valence band, this process is called a direct transition and is only possible in a direct band gap semiconductor. Such a move is not possible in an indirect band gap semiconductor because the conservation of crystal momentum (an interesting topic not discussed here in detail, but the interested reader ought to see, “The Feynman Lectures on Physics: Vol. 3”) would be violated.

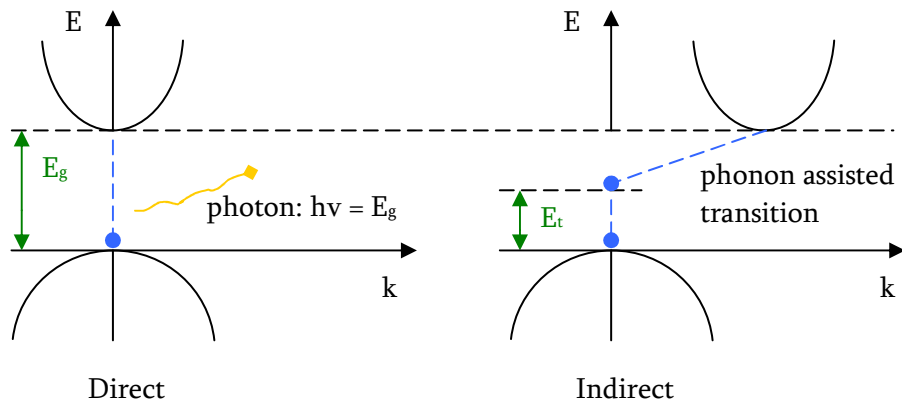


Figure 5.03: a) Direct and b) indirect transitions in semiconductors. Silicon has an indirect band gap meaning that excited carriers require a phonon to transition (recombine) back to their ground state. Unfortunately point defects can “phonon assist” excess light generated carriers back to their ground state. [X4.1]

Here, with no direct move across the band gap, the process requires the absorption or emission of a phonon, a unit of calculation used in quantum mechanics with momentum equal to the difference between the electron and hole momentum, in essence “facilitating”, or better said, “accounting for”, the transition. The participation of a phonon makes the transition far less likely to occur per unit time, and means that radiative recombination in indirect band gap materials is far slower than in direct materials, with both processes shown in figure 5.03. One can imagine quite easily then why a light-emitting diode is almost exclusively made out of direct band gap materials.

This sounds very good. The recombination process in our indirect band gap silicon can proceed only (comparatively) slowly through a phonon assisted process. Unfortunately a move across the band gap can also be facilitated by a crystallographic defect, which performs essentially the same role as the phonon. Essentially because radiative recombination is slower by comparative standards in indirect materials, it will account for only a small portion of recombinations in comparison to those resulting from point defects. If however the excited carriers were to be kept from the point defects, they would eventually recombine through normal radiative methods (photon emission, heat generation as phonons in the lattice, energy transfer to a third particle).

Although defect recombination has been discussed there is a second more general problem, namely that a carrier, once generated, must make its way out of the nano-structure (before recombining!) to contribute to current in any significant way. Mobility is the quantity associated with how easily a carrier can travel through a particular material. Denoted by the symbol μ , it is the ratio of carrier velocity in the field direction to the magnitude of the electric field and expressed in centimetres squared per volt-second ($\text{cm}^2/(\text{Vs})$). In a given semiconductor the main cause of carrier mobility reduction is scattering, normally appearing in two forms, lattice scattering and impurity scattering [X4.14]. Lattice scattering is in some sense unavoidable as it is the scattering of a carrier by the normal thermal motion of the lattice. The frequency of this phenomenon then increases as expected with temperature with a corresponding reduction in mobility. Interestingly, the above mentioned phonons are related to the (collective) vibrations of the lattice atoms, thus lattice scattering is sometimes called phonon scattering.

As device operating temperature is not a topic investigated in this present work, impurity scattering is the source of mobility reduction which most interests us here. These impurities from the doping process (although our implanted ions from the etch process could just as easily scatter electron and hole carriers) are typically ionized (charged).

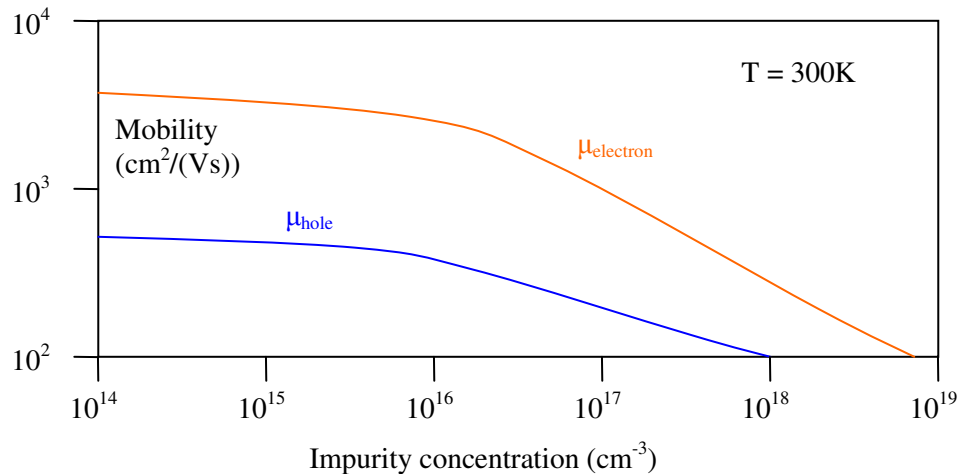


Figure 5.04: Variation of mobility with doping impurity concentration for silicon [X4.1]. The main cause of carrier mobility reduction is scattering and the chance that a carrier will collide with an impurity and scatter is increased with greater substrate doping.

An electron or hole approaching the charged impurity will scatter (deflect) due to Coulombic forces between itself and the impurity. The amount of deflection is related to the velocity of the carrier and the distance to the impurity. Understandably the chance that a carrier will collide with an impurity is increased the more greatly a substrate is doped (considered within a given time, reducing the mean free time between collisions) [X4.15]. This effect will be more pronounced at

lower temperatures (lower carrier velocities) as a slower moving carrier is likely to be scattered to a greater degree through a collision with a charged ion than is a carrier with a much greater momentum. This mobility quantity is seen for both electrons and holes in silicon dependant on impurity concentration in figure 5.04. Thus doping, although necessary to create the diode, has a two-fold negative effect: defect generation (resulting directly in recombination) and reduction of mobility (indirectly leading to recombination by decrease the mobility of carriers and thus their ability to escape the nano-structures).

A summary of this section can be seen in figure 5.05. One can very easily imagine photocurrent, excess carriers generated by incident light, traversing (or attempting to traverse) a needle etched out of this doped crystalline silicon. There is a certain inability to predict the distribution profiles of the implanted ions and damage within the needle accurately, and it is unknown which crystalline sites the implanted ions occupy, unknown damage centers and dislocations that have been introduced during the implantation and etching process and a degree of uncertainty as to the extent these defects can be removed by appropriate procedures. These point defects generate an assortment of additional energy levels within the silicon Fermi band diagram, as seen in figure 5.05 a), ready to “assist” our carriers back to their ground state (as a phonon would) resulting in the a loss of generated carriers, and thus a shortening of the overall lifetime of carriers.

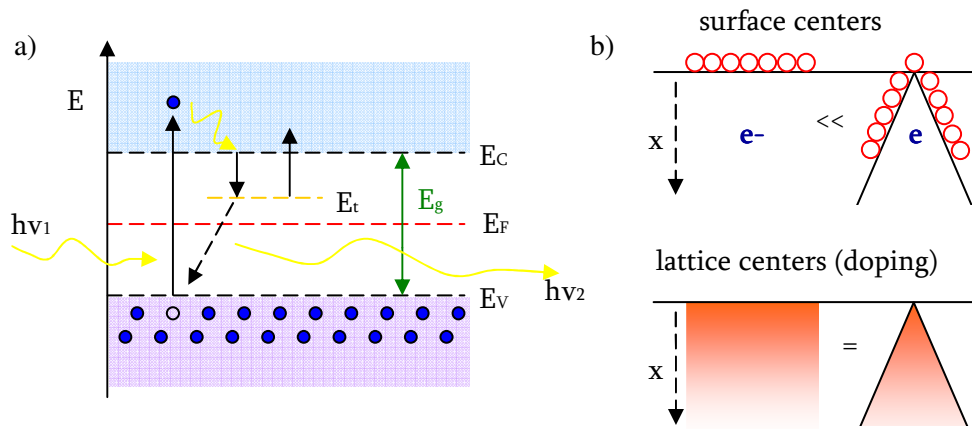


Figure 5.05: a) Photoelectric excitation and recombination. E_t is a point defect appearing as a new energy level in the Fermi band diagram of silicon which can, as shown, either trap the carrier and release it or assist an indirect band gap transition sending the carrier back to its ground state. b) Point defects arise at the surface, in the form of vacancies (dangling bonds) and as one can imagine the sheer amount of surface area is increased in the case of the nanostructure (greatly increasing the number of surface recombination opportunities). Additional point defects and mobility reduction are the result of the actual doping process. For needles and an unetched substrate the quantities will be (very close to) equal in the x direction, assuming the point of the needle corresponds with the exact original surface of the diode.

This section has introduced the mechanisms of damage resulting in carrier loss. The next section will attempt to characterize the actual amount of damage done to the substrate through our diode creation process (including the nano-structuring) and the effectiveness of integration to remove the defects.

5.1.2 Process integration results by microwave photoconductive decay (μ PCD)

For a given temperature and in the absence of light, silicon (or any semiconductor for that matter) exhibits a given concentration of holes and electrons determined by the doping level. As we have discussed, light incident on the silicon surface will generate excess carriers disturbing the thermal equilibrium. In such circumstances the conductivity of the silicon is increased [16], however if the external light source is terminated, the excess carriers in the silicon recombine, returning to the thermal equilibrium value through the mechanisms we have discussed in the previous section. Interestingly for greater excess carrier concentrations the (instantaneous) rate of recombination is also greater leading to a recombination process with exponential decay. Plotting the conductivity as a function of time would yield an exponentially decaying curve with time constant equal to the carrier lifetime [X4.17]. This process is called photoconductivity decay.

In this work recombination lifetime measurements on unmasked wafers were performed using a WT-85 lifetime scanner from Semilab. A standard wafer cassette is positioned on the tool's loading stage and the robot arm handles automatic loading and unloading of wafers controlled via the software. The measurement process used inside the tool has a slightly modified name, Microwave Photoconductivity Decay (μ PCD), which is a fast non-contact means for measuring the carrier lifetime of silicon wafers and in keeping with the theme of using standard industrially implemented equipment is also a standard measurement technique in silicon device fabrication to detect harmful contamination which causes carrier lifetime degradation after implantation processes.

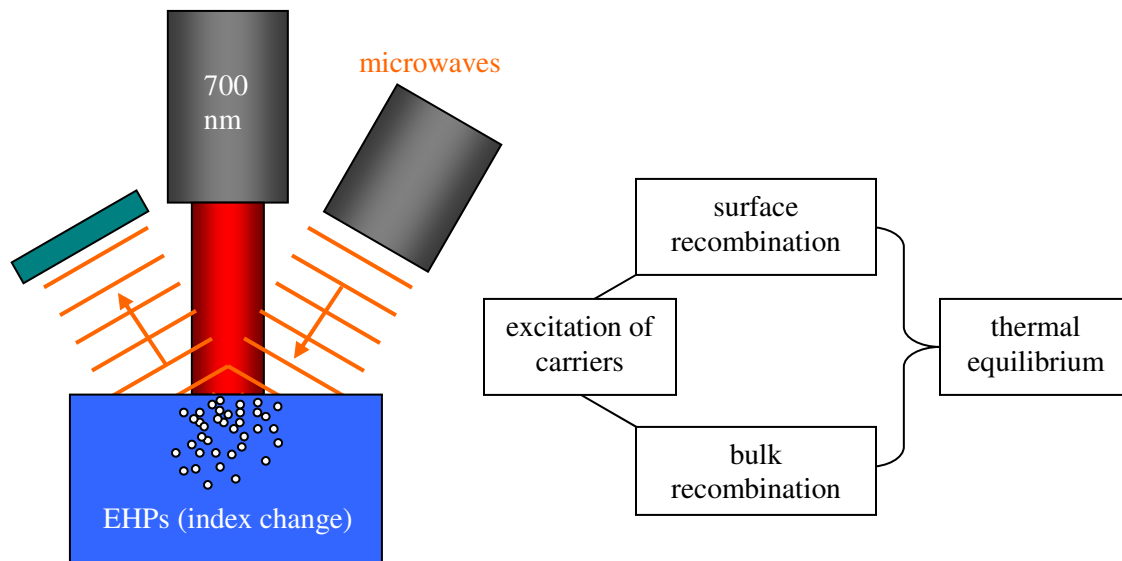


Figure 5.06: The method of μ PCD. When the laser light source is removed, the excess carriers recombine to the original equilibrium level. The rate of recombination is dependent on the concentration of excess carriers, the conductivity decays exponentially to the equilibrium value. The time constant of this exponential decay is the carrier recombination lifetime of the semiconductor. [X4.16, 17, 18]

In practice a 700 nm laser pulse generates excess carriers beneath the surface of the wafer, increasing the conductivity of the sample and correspondingly the index of refraction. Subsequently a microwave signal reflecting off the wafer surface experiences an increased intensity at its detector. The detector output becomes a signal correlated to the concentration of excess carriers in the wafer. Figure 5.06 illustrates the system schematically with a corresponding

block diagram depicting an important concept, namely that after carrier excitation the recombination process proceeds through two mechanisms, bulk defects and surface defects [X4.18].

The morphological properties chapter devoted itself to the creation of the darkest (most antireflective) structures possible with figure 3.29 (and figure 4.40) representing approximately the best results while maintaining a decent uniformity. This recipe and several variations were evaluated with μ PCD before and after the black silicon etch (BSE) process to determine carrier lifetime values. Additional measurements were taken between subsequent process integration steps after BSE to help evaluate the effectiveness of each integration step in an attempt to maximize carrier lifetimes of structured wafers. Figure 5.07 shows two wafers before the BSE process with recombination lifetimes at approximately 16 μ s. Samples after BSE show no meaningful change in lifetime, certainly not the dramatic change that one would initially expect given the extreme plasma etched surface. The explanation is that, for this tool, both wafers before and after BSE, are equally bad. The blank wafers have a set resistivity (doping) from the ingot growth process and although their surface appears mirror like, no activities have been taken to ensure their electrical performance. For the machine, both samples appear beneath the noise floor. However, as we see, thermal oxidation steps demonstrate the sensitivity of samples after geometric modification of the silicon surface with a BSE process.

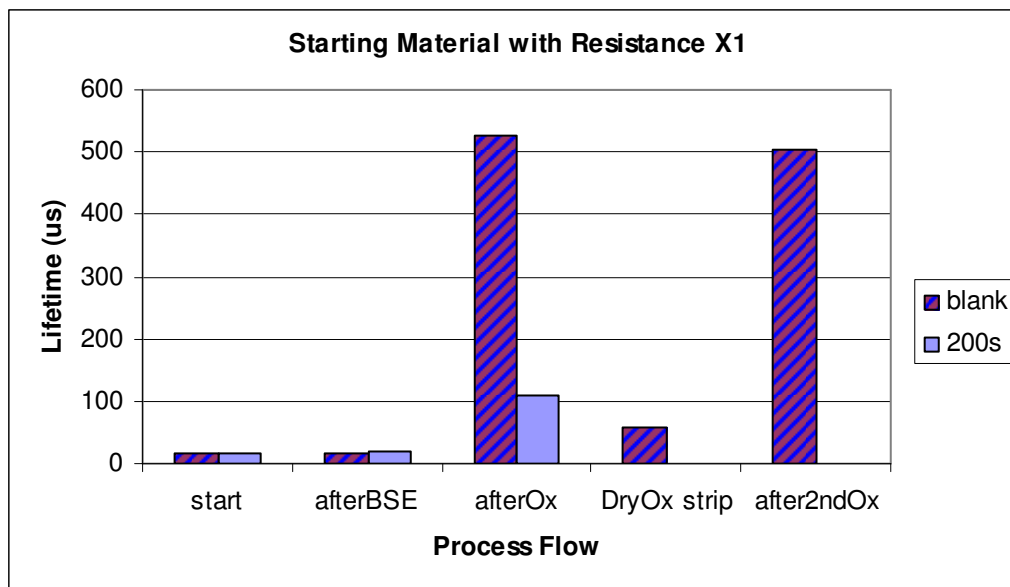
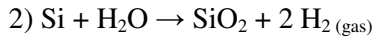
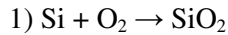


Figure 5.07: Evaluation of carrier lifetimes for two wafers, one structured with black silicon (200 s – corresponding to figure 4.40) and an unetched wafer. Notice the incredible jump in carrier lifetime after the surface is thermally oxidized. This oxidation accentuates the (previously undetectable) degradation of carrier lifetime induced by the plasma etching process. The thermal oxide for the blank wafer was plasma stripped using SF_6 (predictably reducing lifetimes) and the process was repeated. The slightly higher values after plasma stripping are certainly due to an incomplete oxide strip and the slightly lower values after the second thermal oxidation are most certainly due to having a substrate that has now received some plasma damage.

The oxidation process highlights the carrier degradation of the surface as a result of the plasma etching process with the blank wafer reaching a value over 500 μ s and the etched wafer reaching a value of over 100 μ s, but the positive aspect of the thermal oxidation can be treated as follows. Thermal oxidation processes are carried out at temperatures as high as 1200 $^{\circ}\text{C}$ using

either molecular oxygen or water vapour as the oxidant [X4.19]. Logically these two types of thermal oxide are referred to as either 1) dry or 2) wet oxide respectively [X1.5].



When growing thick oxides, such as field oxide, the need for a higher growth rate may be exchanged for overall oxide quality and a wet oxidation process may be used [X4.2]. For our situation the dry oxidation is preferred (indeed it was the process implemented) as the faster wet growth process is, for defect reduction at a nano-structured surface meant to conduct current, less desirable because of an increased number of dangling bonds at the oxide-silicon interface creating, as we have discussed, opportunities for leaking excess carriers [X4.19]. Interestingly, although not done here, it is possible to add a certain (small) percentage of hydrochloric acid to the furnace gas mixture in an attempt to remove metal ion contaminants in the oxide by way of chlorine [X4.19].

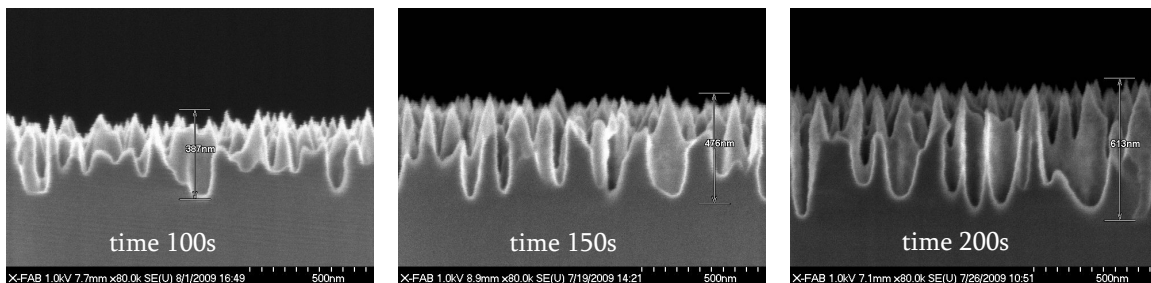
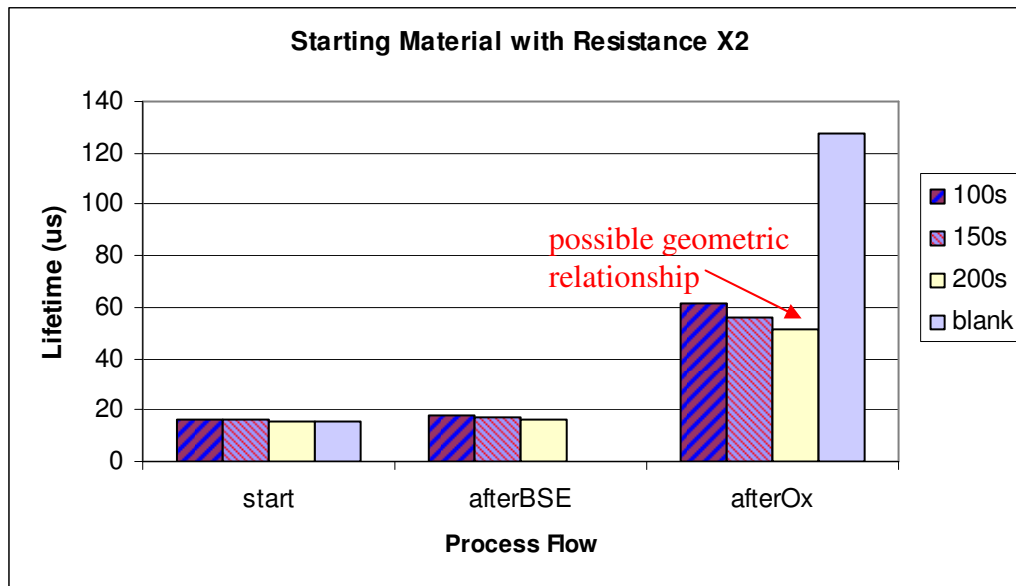


Figure 5.08: The geometric tendency is (possibly) visible, an increase in structure size corresponds to an increase in surface area and thus a decrease in lifetime; however bulk quality plays a much stronger roll in determining lifetimes. For example compare the lifetime of the 200 s sample (approximately 53 us) to the lifetime of the 200 s sample in figure 4.07 (approximately 110 us). These two samples received the exact same black silicon recipe.

In addition to dangling bond reduction, a second feature of the thermal oxide as regards increasing carrier lifetime is that thermal oxide is not like a deposited layer on top of the silicon.

This process consumes silicon from the substrate integrating it into the grown SiO₂ layer, in effect growing upwards out of the substrate and downwards into the substrate. Statistically speaking 46% of the oxide thickness is beneath the original silicon surface, and 54% is above it [X4.2]. At the beginning of this chapter the ways in which plasma damages a silicon surface was discussed. This simultaneous growth upwards and growth downwards phenomenon consuming surface silicon has most certainly eliminated a great deal of tattered and dislocated crystalline silicon at the surface. Thus it is less a question of the generation of new and unique (“dangling”) interfaces states during etching/implantation and likely more a question of a thin layer of plasma damaged silicon, this includes containments on the wafer surface, nano scale and sub-nano scale jagged edges, implanted ions resulting in dislocations, vacancies, and interstitials. The thermal (gate) oxidation process will not only create a high quality (very well-characterized) cap over the silicon, a cap used in standard CMOS transistors (the gate oxide!) trusted to allow a decent amount of current flow with a minimum amount of carrier loss beneath it, but also the source of the silicon in the SiO₂, the consumption of the actual battered surface silicon of the substrate with implanted ions and contaminants shifted out into the SiO₂ layer away from the surface. The reader is reminded that SiO₂ grows from the bottom up, oxygen diffuse to the silicon interface and subsequently reacts with a silicon atom which extends the oxide into the silicon. This process incorporates substrate contaminants into the oxide as the silicon atoms around the contaminant become part of the oxide.

We will ignore the fact that oxide gates of transistors in the most standard processing arrangement are created on a wafer parallel to the surface, meaning dangling bonds of only one crystallographic direction, for example Si(100), are considered and characterized in a device. As we saw in the previous section different crystallographic directions will form different Si-SiO₂ interfaces with slightly different dangling bond structures. The needles will self evidently then represent most normal crystallographic directions along with all exotic variations in between. Certainly one could consider an increase (or decrease) in surface carrier recombination at a different crystallographic direction based on a less compressive (or a more compressive) passivation of the silicon surface in this direction.

Figure 5.08 introduces possibly a geometric dependence of lifetime. This is not difficult to imagine. Larger structures increase the overall surface area and thus increase the amount of surface available for recombination which subsequently lowers lifetimes. An argument could be made that a longer etch time increases the number of ions to the surface and thus the damage to the silicon, but a longer etch time also removes more and more layers of (damaged) surface silicon, as the SEM pictures reveal the structures are increasing in size due to the removal of more and more silicon (yes, but not at the needle tips...). It's a tough call. What figure 5.08 really emphasizes is the role of starting material quality on lifetime. One need only compare the blank wafer (starting material X1) of figure 4.07 with the blank wafer (starting material X2) here in 4.08. A difference of approximately 400 us is observed! This is also observable with etched structures. Compare the lifetime of the 200 s sample (approximately 53 us) to the lifetime of the 200s sample figure 5.07 (approximately 110 us). These two samples received the exact same black silicon recipe. After cleaning and thermal oxidation we see that the jump between structured and unstructured samples is not a factor of 6 as with the starting material X1 but only a factor of 2. Clearly the (poor) bulk quality in this case is playing a more decisive role in setting overall attainable carrier lifetimes.

Samples from the X2 starting material received two further processes after thermal oxidation (results shown in figure 4.09), firstly an oxide covering made of an initial Tetraethyl orthosilicate (TEOS) oxide Si(OC₂H₅)₄ layer and final boron/phosphorous doped TEOS (BP TEOS) oxide layer, and then secondly a planarization step (called a “Reflow”) [X1.5]. The TEOS/BPTEOS we have talked about from the very beginning, although we have not used these terms. The goal of the project is to have oxide covered nano-structures and this TEOS/BPTEOS is

exactly this oxide covering. At higher temperatures (>600 C) Tetraethyl orthosilicate forms oxide.



This process takes place in a plasma-enhanced chemical vapor deposition (PECVD) reactor, essentially a very similar plasma chamber to those we have discussed for etching, however with a greater focus on diffusion of species to the surface (species not intending to chemically attack the substrate, such as fluorine does to silicon) and no ion hammering of the substrate. The background plasma must also necessarily then be neutral and is normally an oxygen or oxygen-argon plasma mixture. After deposition neither TEOS nor BPTEOS is particularly planar, however due to the much more porous nature of BPTEOS it is quite easily planarized with an elevated temperature reflow process. This combination of oxides (in this order, and not forgetting our very high quality thermally grown gate oxide at the silicon interface) is optimal for both maintaining electrical performance, having the BPTEOS layer first could very well lead to having boron and phosphorous dopants interacting at the needle surface with generated carriers, and higher planarization, having TEOS second would require a high temperature reflow, which could further influence device doping profiles [X1.5].

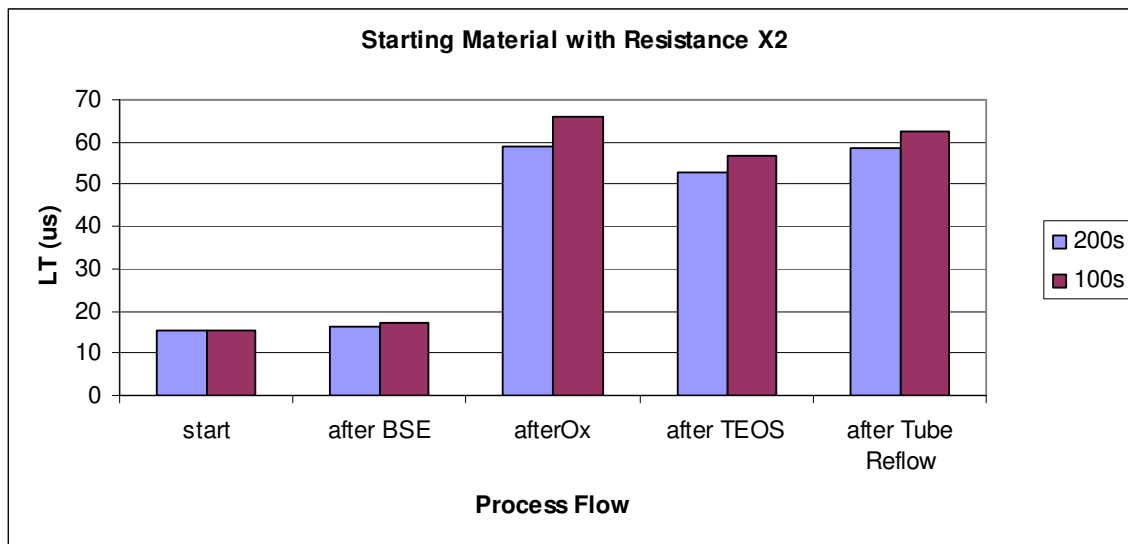


Figure 5.09: Lifetime values after two further processes following thermal oxidation for the X2 starting material are seen. Although for samples with differing reflections the microwave decay should yield a comparatively accurate time constant, the decrease in lifetime after TEOS could be the result of a decrease in transmitted laser signal to the sample resulting in a slightly modified time constant or simple surface damage during the PECVD process.

This last point about temperature is often referred to as the “thermal budget” of a process and is the primary advantage of the PECVD process in comparison to a standard CVD process in which the required energy for the chemical reaction of tetraethyl orthosilicate into oxide is provided by heating the whole reaction chamber (and the wafer!) like a furnace process. Instead the PECVD process lets the plasma do the heating, leaving the wafer’s thermal budget relatively untouched, certainly important if wafers have already been through an aluminum metallization (creating metal lines and interconnections).

Lifetimes after oxide covering are still very high but have come down slightly. Why? Although for samples with differing reflections, samples which absorb a greater or lesser percentage of the 700 nm carrier stimulus laser, the microwave decay should yield a comparatively accurate time constant, here the slight drop in carrier lifetime after the TEOS step has likely very little to do with carrier degradation in the substrate or at the oxide-silicon interface and likely much more to do with a decrease in transmitted laser signal to the sample resulting in a slightly modified time constant. Criticism of this position could certainly be made as it is quite apparent from our above description of the PECVD process that there are inevitably some number of high energy electrons and ions incident on the wafer surface during deposition and that our thin thermal oxide can in no way guarantee device integrity during the beginning of the deposition process. The surface has received some unavoidable amount of damage.

In summary we have seen that lifetime is a property of not only the bulk silicon substrate but also the condition of the silicon surface. To measure the lifetime accurately it is important to generate a concentration of excess carriers as uniformly as possible throughout the wafer. This was realized with a 700 nm stimulus laser wavelength, which is long enough to penetrate many microns into the silicon substrate (as we know from FDTD simulations of the previous chapter shorter wavelengths can be completely absorbed in a matter of nano-meters). After carrier generation there are two parallel processes of recombination, the excess carriers recombining in the bulk according to (as we have demonstrated) the intrinsic quality of the bulk substrate and excess carriers diffusion to and recombination at the surface. Our processes have little influence on the quality of the bulk (this is set by the bulk material vendor), but as we have shown, there are a great number of integrative activities that can be performed to significantly improve the surface conditions and correspondingly the lifetime of nano-structured silicon surfaces.

5.2 Diodes

The section will implement nano-structuring of the diode surface with the optimal integration activities of the last section (both wet chemical cleaning steps, thermal oxidation, furnace planarization) on actual diode surfaces. Preceding a discussion of the spectral sensitivity of these diodes is a brief review of diode properties and the functional area of interest for this work is provided is provided.

5.2.1 Illuminated PN junctions

Before illuminating our junction let us take a very brief moment to remind the reader of the several diode concepts (and the vocabulary that will be used in this dissertation). Let us begin by considering two distinct regions of n- and p-type semiconductor brought together to form a device. This newly formed pn junction has initially a greater concentration of electrons in the n-type material and holes in the p-type. Because of these large carrier concentration gradients holes immediately diffuse towards the n-type material and electrons diffuse towards the p-type material, pictured below in figure 5.13. However these diffusing carriers leave behind uncompensated donor ions and therefore this resultant diffusion cannot go on indefinitely. Near the junction interface a positive space charge within the n side develops while a corresponding negative space charge develops within the p side. The resulting electric field points in the opposite direction to the diffusion of species and this creates a drift of n-type carriers back towards the n material and p-type carriers back towards the p material. The electric field causing the carrier drift builds to a point at which drift exactly balances diffusion so there is no net current flow to either the p or n side. We assume this field appears across a transition region W (also called the space charge region or depletion region) at the junction and the potential across W is called the contact potential or built-in potential V_0 . We further assume that the electric field outside of W is zero and that there is a constant potential V_n in the n-type side of W and a constant potential V_p in the p-type side such that $V_0 = V_n - V_p$. These assumptions are called the

depletion approximation and result in having very few carriers inside the transition region at any given time as the electric field sweeps them directly across. Since the dipole about the junction must have an equal number of charges on each side it can be that the transition region extends unequally into the n and p regions based on relative level of doping.

A very useful feature of such a junction is that current flows very easily across the junction from p to n when the p region has a positive external bias relative to the n side, called forward bias, and that nearly zero current flows when the p region has a negative external bias relative to the n side, called reverse bias. With regard to this bias it is assumed that the entire voltage drop of the externally applied voltage V_f appears across the transition region and this voltage is taken to be positive when the applied voltage to the p side is positive relative to the n side.

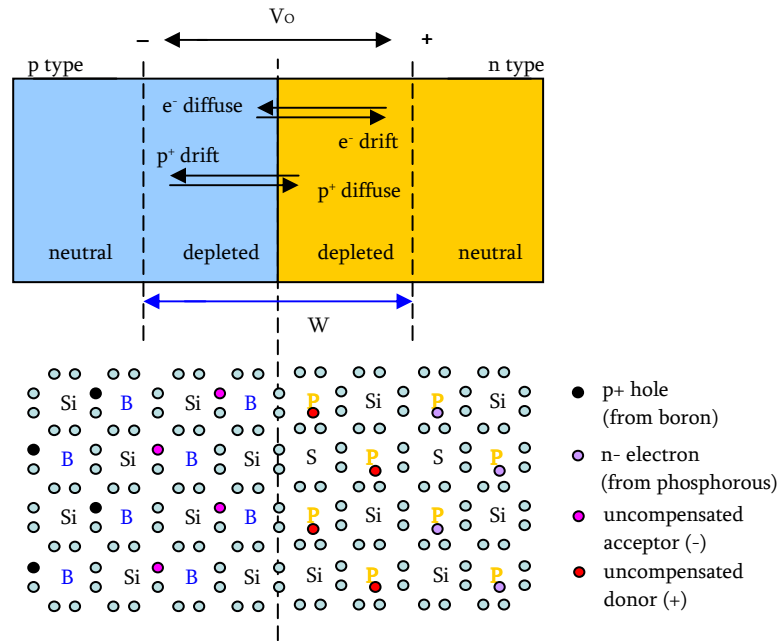


Figure 5.13: The pn junction [X4.1]. As a result of n and p majority carrier diffusion across the interface a depletion region of width W forms with a potential V_0 across it. This potential creates a drift of n carriers back towards the n material and p carriers back towards the p material. Without an additional external source this process reaches an equilibrium with drift exactly balancing diffusion.

As can be seen in figure 4.14 the electrostatic potential barrier at the junction interface is decreased from V_0 by the applied forward voltage V_f to the smaller value $V_0 - V_f$ [X4.1]. For a reverse bias the opposite takes place, the potential barrier is increased to $V_0 + V_f$. This is clear because the direction of the applied external field either reinforces or opposes the built-in equilibrium field. Since it's required that the correct number of charges, that is to say uncompensated donors and acceptors, become exposed for the given new total electric field value we also expect the width of W to decrease or increase with forward bias or negative bias respectively. With no applied bias the Fermi levels E_{FP} and E_{FN} are equal and the size of the electrostatic potential barrier on the Fermi-diagram is simply the electronic charge q times the magnitude of the potential barrier V_0 .

However with bias V_f we see that there is a separation of the Fermi energy bands directly correlating to the increase in the potential barrier at the junction. The diffusion current is then

majority carrier electrons from the n side overcoming the potential barrier to diffuse to the p side while holes surmount their barrier and diffuse to the n side. The drift current is made up of the minority carriers from the n side and p side and is interestingly quite insensitive to the size of the potential barrier. On the p side every minority carrier electron which reaches the transition region is swept down the barrier regardless of the barrier size (regardless of forward or reverse bias) and similarly every minority hole on the n side reaching the transition region is swept down to the p side. The key insight is that the supply of minority carriers is generated by thermal excitation of electron-hole pairs (EHPs) and those EHPs generated within a diffusion length L_n of the transition region or in the transition region itself compose the drift current (commonly called generation current). The total current across the junction is therefore always the sum of the drift and the diffusion current. At equilibrium this value is zero

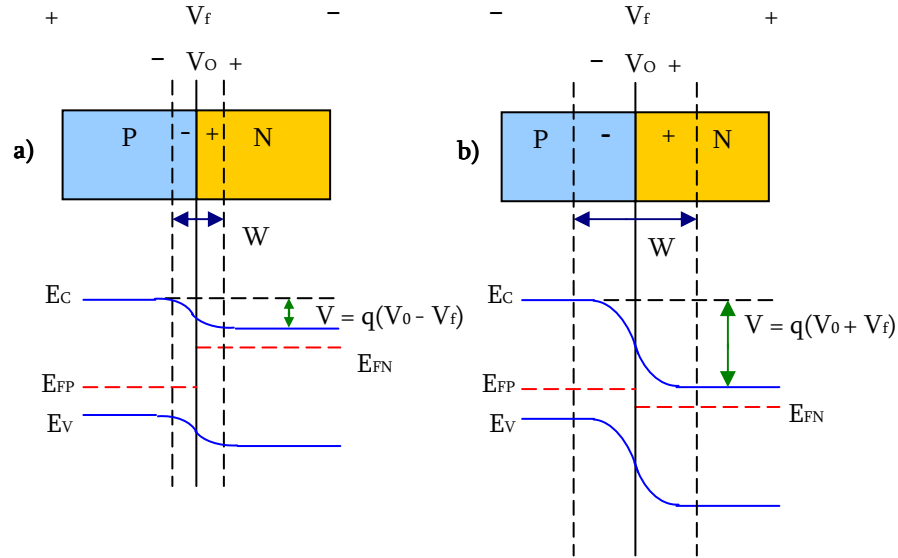


Figure 5.14: PN junction with a) forward and b) reverse bias [X4.1]. Diffusion current is majority carrier electrons from the n side overcoming the potential barrier to diffuse to the p side while holes surmount their barrier and diffuse to the n side. Drift current is minority carriers and is insensitive to the size of the potential barrier. Every minority carrier which reaches the depletion region is swept down the barrier regardless of the barrier size (regardless of forward or reverse bias). This supply of minority carriers is generated by thermal excitation.

$$I = I_{diff} - |I_{gen}| = 0 \quad \text{for } V_f = 0 \quad (5.01)$$

The generation current can be graphical seen in figure 5.15 in the complete I-V plot of a normal pn junction. Here the current I is taken to be positive from p to n under forward bias and primarily made of diffusion while under reverse bias diffusion is assumed to be negligible.

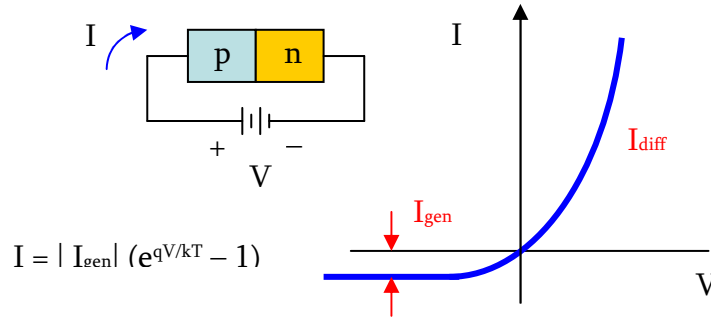


Figure 5.15: IV characteristics of a PN Junction [X4.1] The total current I under any bias (forward shown) is, just as in equilibrium, the diffusion current minus the absolute value of the generation current.

With a forward bias present the probability of a carrier diffusing across the junction is increased by $\exp(qV/kT)$ while in reverse bias the diffusion current is reduced by the same exponent (for negative V values the exponent approaches zero). During equilibrium the diffusion current is equal in magnitude to the generation current and so under bias I_{diff} is then always I_{gen} multiplied by $\exp(qV/kT)$. Therefore the total current I under any bias is, just as in equilibrium, the diffusion current minus the absolute value of the generation current:

$$I = |I_{gen}| e^{(qV/kT)} - |I_{gen}| = |I_{gen}| (e^{(qV/kT)} - 1) \quad (5.02)$$

Equation 5.02 is normally referred to as the diode equation and with it we have very quickly reviewed the topics needed to understand the illuminated junction, namely the current associated with the drift of minority carriers across the pn junction by thermal EHP generation either in the transition region W and then separated by the junction field or with a diffusion length of the junction.

Figure 5.16 introduces the effect on the pn junction I-V characteristics when the junction is illuminated with photons having $h\nu > E_g$ and a generation rate g_{op} (EHP/cm³s). It's clear that the incident photons will uniformly create EHPs on both sides of the junction however it is the EHP minority carriers (holes on the n side and electrons on the p side) which will readily drift (slide down) the potential barrier and create the new optical current I_{op} (flowing from n to p). Thus this new current can be combined with the total current I flowing from p to n and the diode equation becomes:

$$I = |I_{gen}| (e^{(qV/kT)} - 1) - I_{op} \quad (5.04)$$

This new diode equation can be simply understood as the standard equation reduced by an amount relative to the optical generation rate. However when V equals zero, the value at which standard equilibrium was considered, there is now clearly a net current flowing ($I \neq 0$). The characteristic curve crosses the y axis at a negative value in relation to g_{op} . The current I_{op} flowing at $V = 0$ has a corresponding optical voltage V_{op} and as a result the built-in junction potential V_0 will be modified by V_{op} . This is more elegantly demonstrated by revisiting the Fermi diagram of a junction at equilibrium and comparing it with itself while illuminated, both below in figure 5.17.

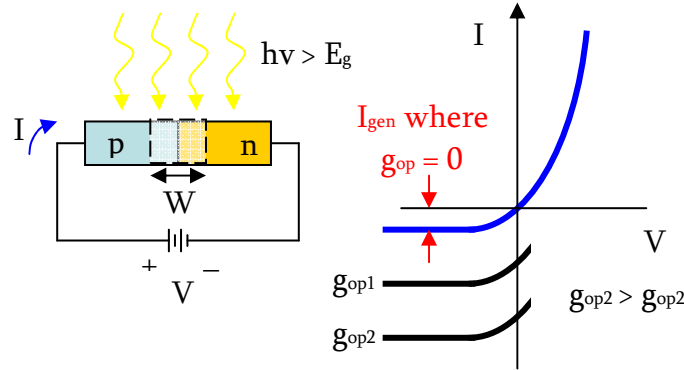


Figure 5.16: Optical generation of carriers in a pn junction [X4.1]. The new diode current can be understood as the standard diode current reduced by an amount relative to the optical generation rate.

The appearance of this forward voltage V_{op} across an illuminated junction is called the photovoltaic effect. Increasing the intensity of the light beam illuminating the surface increases the number of photons in the light beam, and thus increases the number of EHPs generated, but does not increase the energy that each EHP possesses. The energy of an individual electron and hole in a generated pair does not depend on the intensity (number of photons) of incoming light, but rather on the energy (actually frequency, $E = h\nu$) of the specific photon which generated the pair. It should be said that for a given frequency of incident light, although the rate at which EHPs are generated is directly proportional to the intensity, there exists a certain minimum frequency of light below which no EHPs can be generated regardless of intensity. This frequency is normally called the threshold frequency.

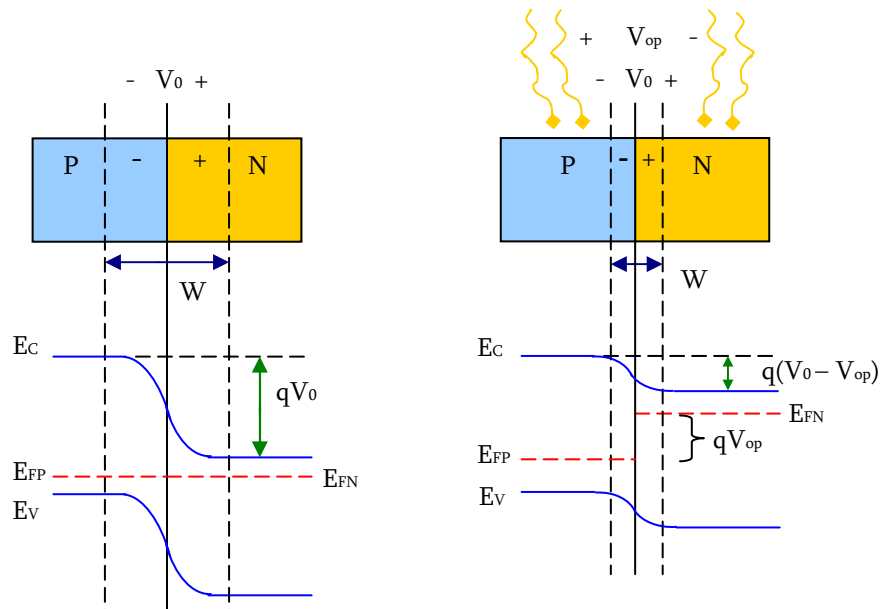


Figure 5.17: Effect of illumination on the equilibrium Fermi Level [X4.1] See text for discussion.

The junction can be operated in the first third or fourth quadrant of the characteristic I-V curve depending on the desired application. This is show below in figure 5.18. In the first

quadrant the illuminated junction delivers power to the load R_L and it should become clear that this will be the least interesting region for our desired (sensing) application.

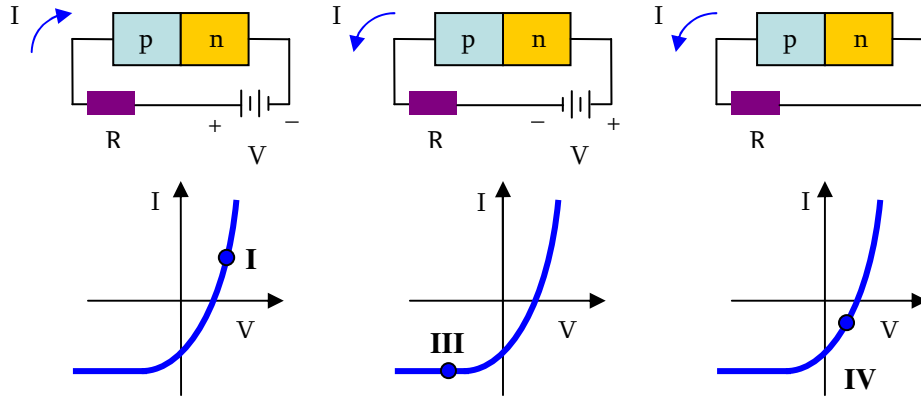


Figure 5.18: Operation of the illuminated junction [X4.1]. Region III is of most interest for the work presented here, when a current flows across the illuminated junction proportional to the optical generation rate.

In the third and fourth quadrants power is taken from the load. However in quadrant four the total current is negative and the junction voltage is positive but the voltage is restricted to values less than the built-in potential V_0 . The main operational interest of this dissertation is in the third quadrant when a current flows across the illuminated junction proportional to the optical generation rate. This offers a valuable way of measuring illumination levels or of converting time-varying optical signals into electrical signals.

5.2.3 Spectral sensitivity results

The illumination thus far has been considered evenly distributed across our PN junction diode however in an actual CMOS process made up of p- and n-wells implanted into a p- or n-type substrate wafer, the simplest method of diode construction is a single well implanted into a substrate of opposite carrier type. Figure 4.19 illustrates exactly the situation used in this dissertation, an n-well in a p-substrate wafer. With this geometry holes will be the minority carrier generated on the needle-side of the junction.

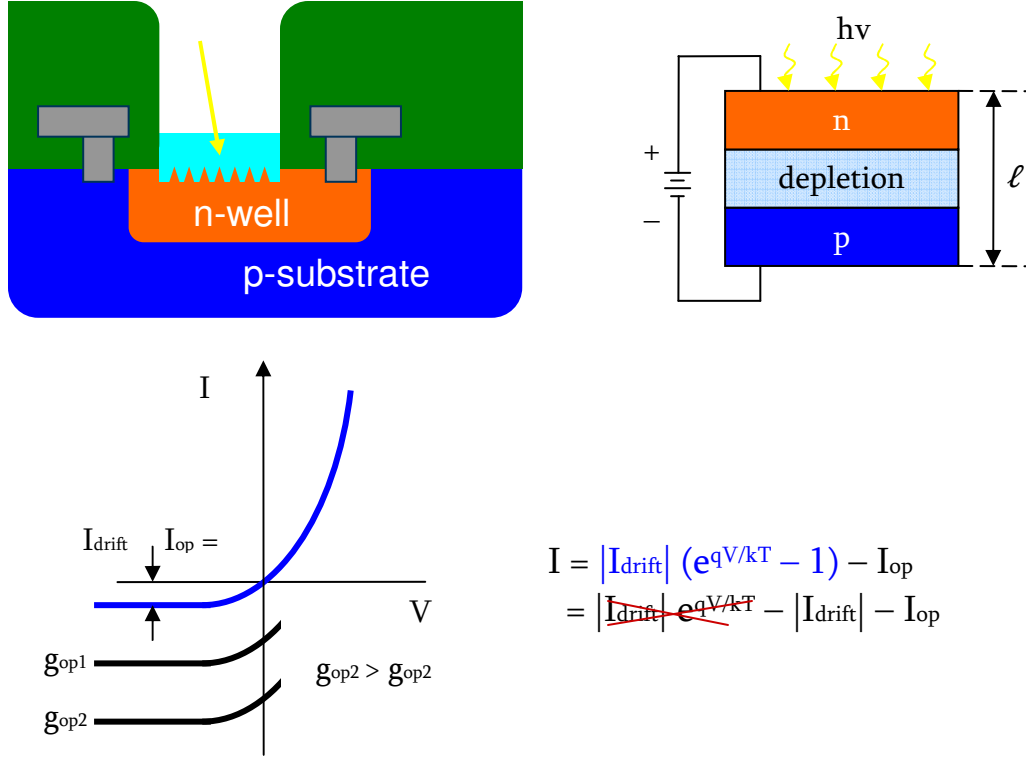


Figure 5.19: The CMOS diode layout used in this dissertation, an n-well implanted in a p-substrate wafer. Contact is made to the n- and p-areas through metal contacts from within the stack (all of the corresponding dielectric and metal layers of the given technology creating and interlinking various devices on the wafer).

The reader is reminded that L is the thickness from figure 4.33 in the optical properties chapter. Indeed figure 5.20 shows theoretical spectral sensitivities of diodes with thicknesses of 1um and 10um. A more simplified form of this graph was established in the introduction of this dissertation while discussing the overall target of the project. We are now ready to discuss this graph in greater detail beginning with the slated linear line in yellow, the ideal spectral sensitivity. It appears as though our device is not able to handle as efficiently waves of a decreasing wavelength in comparison to waves of an increasing wavelength. This is of course nonsense, we're just misinterpreting the science causing of the slope of the line.

We begin by asking what is the energy in a single photon of light at 400 nm (blue light)? and at 700 nm (red light)? Since $E = hc/\lambda = (6.626 \times 10^{-34})(3.0 \times 10^8)/(400 \times 10^{-9})$, the energy of the blue light is, $E_{blue} = 4.96 \times 10^{-19}$ J and the energy of the red light is $E_{red} = 2.83 \times 10^{-19}$ J. Evidently there is more energy in the blue light, but how many photons per joule exist for light at wavelength $\lambda = 400$ nm? and at 700 nm? $E = \text{Energy/photon}$, so to create 1 J of energy we will need N photons. $N \times E = 1$ joule, hence $N = 1/E = \lambda/hc$ and $N_{blue} = 2.01 \times 10^{18}$ photons and $N_{red} = 3.52 \times 10^{18}$ photons. Evidently there are more photons per joule of red light.

Thus again, revisiting figure 4.33 from the optical properties chapter we see that so long as the incident light has an energy E_g above the gap, the light will be absorbed. Light is however quantized in photons (no half photons allowed) and as such the smallest unit of energy absorbed and able to create an electron hole pair which separates contributing to the photocurrent corresponds to one photon. So although blue light as we have just calculated brings a whole lot of energy, it brings relatively few photons in comparison to red light. The energy of red light is less

than blue but still more than enough to be absorbed and create electron hole pairs. This accounts for the slope of the ideal sensitivity line.

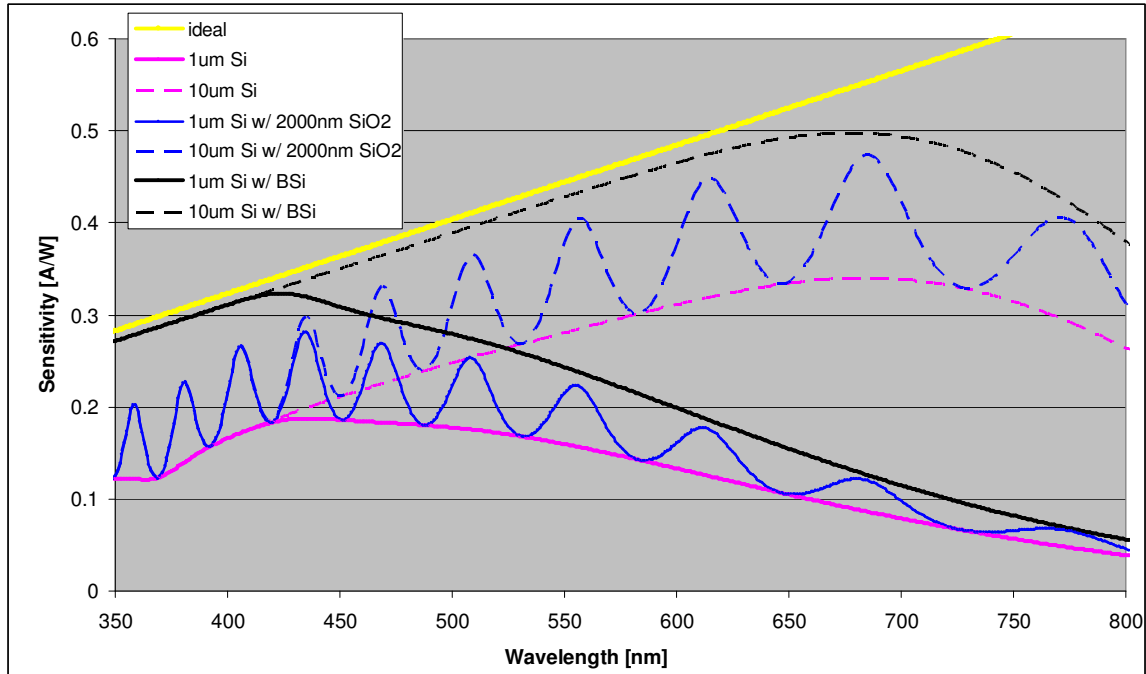


Figure 5.20: Ideal spectral sensitivity. All lines shown are in fact ideal, assuming every non-reflected photon generates an electron hole pair which contributes to the measured photo-current. [X4.4]

The difference between the 10 um photo diode and the 1um photodiode is seen as a worsening performance in the red region, but again considering the diagram in 4.33 of the last chapter, a thickness L of 1 um is simple not enough physical distance to absorb all the incident red light. One more fact about the ideal line, before continuing on to measurement of the black silicon photodiodes, is that this line is for a medium that absorbs photons all wavelengths of light. For silicon, a real medium, eventually the energy of the incident light will be less than the band gap E_g , and silicon will become transparent, absorbing nothing (approximately from infrared and on to all waves of greater wavelength).

Moving now to the spectral sensitivity of real diodes etched with black silicon on a wafer prepared with a chess board layout (for each black silicon diode there is a neighboring oxide covered diode next to it for baseline comparison). For a target in the red light region from the previous idealized cases we take approximately 0.5 A/W for wavelengths around 700 nm and we see that this target is satisfactorily reached. Considering blue light is more complicated. The idealizations presented earlier are for defect free diodes (no defect related recombination), however as we have discussed doping and etching are the primary causes of recombination, assuming the starting material is of suitable quality. The area in which blue light is absorbed (definitely revisiting figure 4.30 of the last chapter) is directly within the plasma etched (bombarded) needle, likely an area rampant with dislocations and interstitials. Furthermore, this needle is etched directly into an n-well with a doping intensity profile strongest at the wafer surface and dropping off in intensity with distance perpendicular to the wafer surface, again meaning the doping concentration of the needle (essentially the number of dislocations and

interstitials per unit volume) is greater and greater as one traverses from the base of the needle to its top. As we have seen light under 500 nm is nearly totally absorbed before the halfway point.

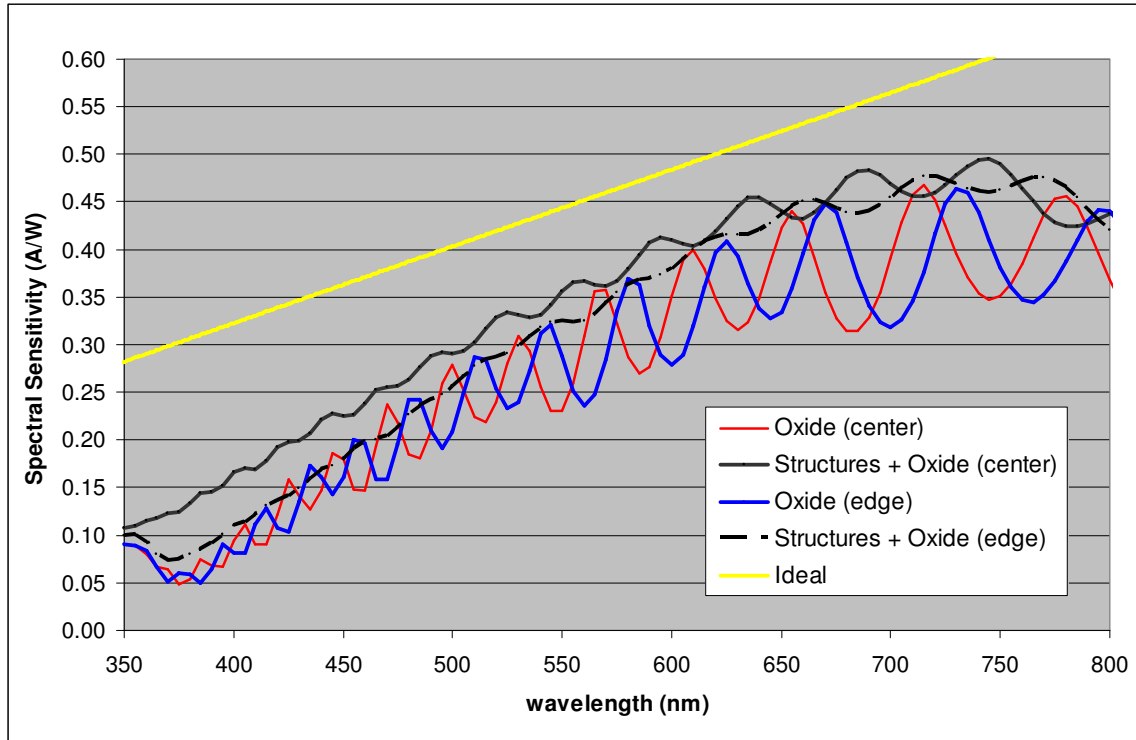


Figure 5.21: Spectral sensitivity of real diodes. This wafer is prepared in a chess board fashion (every black silicon diode has an oxide covered diode next to it for baseline comparison. The red target, around 0.5 A/W at wavelengths greater than 700 nm is reached. For blue light the center site has been very successful. Performance of the edge site is most certainly due to trapping and recombination resulting from remaining defects related to the nature of etching in the TCP. Small oscillations in the black silicon performance (center and edge) are the results of remaining oscillations in the reflection from the oxide covered diodes.

It is then clearly an absurdity to think that the black silicon system could out perform an antireflective coating at its tailored wavelength, for example from the figure in the introduction showing a 50 nm nitride layer with zero reflection at approximately 50 nm. We do not under any circumstance consider any significant amount of incident light as having been absorbed to any degree within the nitride coating. Here our plasma-battered antireflective coating is doing most of the absorbing and likely most of the recombining (as seen from red lights near target performance – red as we have seen is absorbed below the needle). The real comparison of merit is then to consider the difference in gained sensitivity between the ideal oxide covered silicon and the ideal needles, in figure 5.21 this is approximately (imagining an average line through the oxide oscillations) about 0.1 A/W. This is about the same between 375 nm and 425 nm. Our structures in this same region have accomplished about 0.07 A/W and this is a really incredible achievement. We can say with confidence that the center site has been very successful. The etching of structures has not only cleared away some of the more heavily doped (defect ridden) uppermost layer of the n-well (while increasing the diodes performance!), it has also flattened out the oscillatory response of a normal antireflective coatings behavior making it highly insensitive to the wavelength of incident light, meaning the diode performs very well at plus or minus 25nm

on either side of a given center wavelength (certainly not something the oxide covering can claim!). The poorer performance at the edge site is undoubtedly caused by trapping and recombination due to remaining defects associated with the very the nature of etching in the TCP.

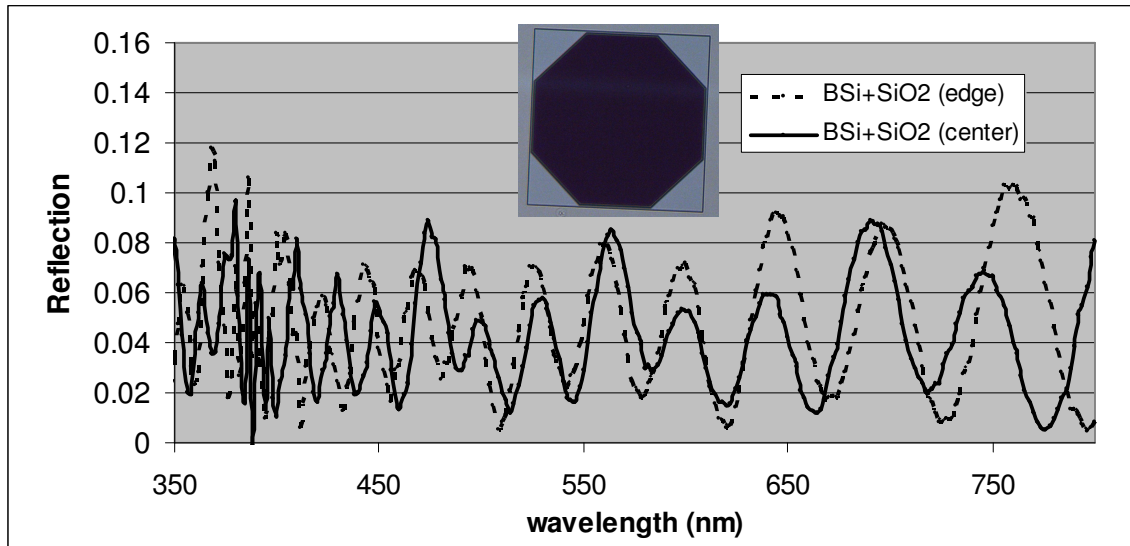


Figure 5.22: Reflection from oxide covered black silicon diodes. Oscillations in the measured spectral sensitivity are due to the oscillations present here around an average value of about 4%. This however does not explain the difference in center to edge performance of structures particularly between 200 nm and 500 nm.

This can be more analytically proven by looking at reflections from the oxide covered diodes before measurement as both center and edge perform with an average reflection of about 4% and oscillations of approximately the same size, see figure 5.22. Essentially the center and edge structures are equally “black”. Furthermore SEM photos of the etching showed that structures at center and edge were of (nearly exactly!) the same size and shape with an indiscernible difference in consumed silicon (thus not a question of more or less doping or more or less n-well). The small remaining oscillations in the black silicon spectral sensitivity performance (center and edge) are simply the results of remaining oscillations in the reflection from the oxide covered diodes. Furthermore both center and edge oxide-only reference diodes perform in exactly the same way (again center to edge black silicon diode differences cannot be n-well doping related).

Thus discrepancies in spectral sensitivity between center and edge sites, diodes on the edge performing below the results of diodes located at the wafer center, are (incredibly probable) due to more aggressive etching of the wafer edge in comparison to the wafer center. This corresponds exceedingly well to firstly all studies characterizing the etch tool itself shown in the introductory section describing the TCP, field strength from the upper power coil is coupled far more aggressively into the chamber at the wafer edge in comparison to the center, and secondly this is completely consistent with all etch rate and SEM investigations of structures at the wafer edge in comparison to the wafer center. Take figure 3.13 from the morphology chapter, this is an extreme example and we have painstakingly worked to have the same structures at center and edge by applying nano-loading effects to help distribute chemical (gas/radical) etching evenly across the wafer (masked wafers help this significantly too), but the point is that ion delivery (the number of ions) in this system is always greater at the wafer edge as plasma is more efficiently ionized at the wafer edge.

5.3 Concluding comments on electrical properties

Further process integration activities could very quickly answer the question of ion damaged silicon, for example an increase in the amount of thermal oxidation of the needles after etching, increasing the oxide thickness (eliminating additional atomic layers of damaged silicon). As discussed incident ions have a range into the wafer but due to the random nature of inter-atomic collisions experienced by an ion inside the silicon crystal it is difficult to say exactly how deep and where a single ion would penetrate. Since the plasma ionization (plasma density) is much greater at the wafer edge, as for example seen in this dissertation by increased etch rates at the wafer edge, it is not hard to conclude that the fraction of incident ions reaching at the edge is greater than the fraction of incident ions reacting the same depth at the center (Thus more defects, interstitials and dislocations, at the wafer edge in comparison to the center). In fact, given the results of studies characterizing the plasma etch tool used here, it would be very awkward to not see some difference between center and edge spectral sensitivity performance.

Before concluding one more thing to note is that it is most certainly not possible to attribute any one exact peak in the oscillations of the reflection data from the oxide covered black silicon diodes in figure 5.22 to any one exact peak in the spectral sensitivity of figure 5.21 (like in a simulation) as the reflection of the covered diodes is measured locally (the measurement spot is significantly smaller than the diode's area) and irregularities in oxide thickness to move the peaks and troughs of the oscillations around. This has been proven experimentally. However the size of the oscillations and their average value of approximately four percent stay the same regardless of where the measurement is taken.

6. Summary and Outlook

The activities presented in this dissertation have shown in detail the creation and implementation of black silicon as an antireflective coating for photodiodes. The plasma chamber and plasma processes were discussed from as fundamental a position as possible. Plasma chemistries known for nano-structuring silicon were implemented, however through additional nano-scale loading effects and further sculpting steps, sets of extremely uniform pyramid shaped structures which are ideal for applications in reflection suppression have been created. The nano-loading of surfaces, etching a set of highly uniform nano-pillars on the silicon surface as an initial step, eliminated to a strong degree the inherent non-uniformities of the chosen plasma tool. This was essential, as without the nano-loading step, only a small process window of extremely anisotropic nano-structures and a poor center-to-edge wafer uniformity was possible. The nano-loading mask was subsequently plasma etched with chemistries that did not by themselves exhibit any nano-structures, and yielded the possibility to create less-anisotropic nano-structure geometries with properties more desirable for reflection suppression. Due to a greatly increased silicon etch rate resulting from the decrease in exposed silicon, an extended nano-loading approach was necessary to successfully transfer plasma processes from blank wafers to masked wafers. Furthermore the actual appearance of nano-structures on the silicon substrate etched in the SF_6/O_2 plasma mixture was explained by generalizing the etch rate of silicon dependant on fluorine concentration (with a varying fluorine to oxygen gas ratio) into two regions, one of excess fluorine generation (weakly passivating) and one of excess fluorine consumption (strongly passivating). We have experimentally verified this explanation, as well as could it be with the tools available. The PROES method briefly discussed in chapter 2 could be implemented to give a more detailed analysis of plasma parameters and their effect on the resulting plasma (size/uniformity of the sheath above the wafer) and the type of resulting nano-structures observed.

FDTD simulations were used to better understand the relationship between nano-structure morphology and optical properties. It was demonstrated that the majority of light absorption in the spectrum from 200 nm to 800 nm takes place within the silicon nano-structure antireflective coating, in contrast to a standard anti-reflective coating. The problem of nano-structure periodicity was addressed, namely that overlapping nano-structures resulting from an incomplete plasma etching separation disturb the conditions of periodicity (diffraction into component angles) corresponding to the structures overall width. Initial simulations of nano-structures utilized single structures with periodic boundary conditions and demonstrated clear diffraction grating properties. These initial simulations did not prove to be a useful way of evaluating the measured specular and diffuse components of real physical structures. However they did prove to be more representative of an ideal antireflective coating. It was shown that incomplete structure separation disturbs the conditions of the ideal gradient index. The ideal width of a needle was set between 100 nm and 200 nm for the light spectrum of interest (200 nm to 800 nm) any periodic structure with widths over 200 nm will begin diffracting light into non-specular angles. Furthermore the dependence of structure size height width on reflected intensity was calculated with heights over 500 nm yielding negligible reflection. Simulated surfaces are best correlated to the optical properties of physical self-organized nano-structures (integrating sphere measurements of specular and diffuse components) when a random (non-periodic) surface is simulated. Further simulations could be performed to more exactly correlate the width of the Gaussian distribution's sigma (this distribution itself must not necessarily be Gaussian) in frequency space and amount of

noise introduced to physical measurements, although it is not clear how beneficial such theoretical parameters will be for the experimentalist working in the clean room.

Boundary conditions for the FDTD were discussed in the process of simulating non-normal incidence light on nano-structures. Dispersion which occurs in the injection angle for incident light within a simulation created difficulties in simulating broadband spectrums in a single simulation. In such situations only the center wavelength has the correct theta, while the smallest and largest wavelengths of the incident light experience an extreme error in incident theta. Incalculable computational times (measured in “years”) were encountered when attempting the most straight forward solution, simulating each wavelength, one after the other. The final solution implemented broke the desired spectrum into several smaller spectrums and provided satisfactory results while limiting the amount of dispersive error. An interesting result regarding boundary conditions discussed in this dissertation is that of placing a gradient-index in front of PMLs to possibly improve their angle-dependant limitations (discussed in Appendix A as it is not required reading for the implementation of nano-structures in photodiodes). When the mechanisms are more clearly understood, this could possibly be a very significant discovery. Simulations in 3D gave insight into the location of electromagnetic intensity within the nano-structures, particularly by providing the opportunity to view cross-sections of individual structures at various depths (and also very colorful far-field results). Transmitted intensity beneath structures demonstrated again the amount of absorption taking place with the structures themselves.

It was shown with microwave photoconductivity decay that nano-structures possessing excellent properties of reflection suppression do not always correlated to nano-structures possessing ideal properties of electrical conduction. This is a result of the plasma etching to modify surface topology drastically reducing carrier lifetimes by greatly increasing recombination through lattice damage and an overall increase in surface area. Regardless, with a relatively small amount of thermal oxidation of the nano-structures, the achieved photodiode efficiency in the red region (675 nm - 750 nm) met the 0.5 A/W target, while in the blue region (375 nm - 425 nm) an improvement of +0.07 A/W over oxide covered diodes was achieved for center cites. Additional process integration activities could be applied to further improve these results.

Resources

This section has been organized for the reader (as best as possible) by content, as follows:

- X1. Plasma**
- X2. Black silicon**
- X3. Optics**
- X4. Photodiodes and semiconductor processing**

X1. Plasma

Textbooks

- [1] Mitsuharu Konuma, “Plasma Techniques for Film Deposition”, Alpha Science International Ltd, 2006.
- [2] Lieberman, Lichtenberg, “Principles of Plasma Discharges and Materials Processing”, Wiley-Interscience, 2nd edition, 2005.
- [3] Bittencourt, “Fundamentals of Plasma Physics”, Springer, 3rd edition, 2004.
- [4] Chen, Chang, “Lecture Notes on Principles of Plasma Processing”, Springer, 1st edition, 2003.
- [5] Elwenspoek, Jansen, “Silicon Micromachining”, Cambridge University Press, 1998.

Articles

- [6] Vendor, Boswell, “Numerical Modeling of Low-Pressure RF Plasmas”, IEEE Transactions on Plasma Science, vol.18, no. 4, August 1990.
- [7] Surendra, Graves, “Electron Acoustic Waves in Capacitively Coupled, Low-pressure rf Glow Discharges”, Physical Review Letters, vol.66, no.11, March 1991.
- [8] Vendor, Boswell, “Electron sheath interaction in capacitive radio frequency plasmas”, J. Vac. Sci. Technol. A, vol.10, no.4, Jul/Aug 1992.
- [9] Turner, “Pressure Heating of Electrons in Capacitively Coupled rf Discharges”, Physical Review Letters, vol.75, no.7, August 1995.
- [10] Lieberman, Godyak, “From Fermi Acceleration to Collisionless Discharge Heating” (Invited Review) , IEEE Transactions on plasma science, vol.26, no.3, June 1998.
- [11] Brake, Pender, Fournier, “The Gaseous Electronic Conference (GEC) reference cell as a benchmark for understanding microelectronics processing plasmas”, Physics of Plasmas, vol.6, no.5, May 1999.
- [12] Kawamura, Lieberman, Lichtenberg, “Stochastic heating in single and dual frequency capacitive discharges”, Physics of Plasmas, vol.13, 2006.
- [13] Mussenbrock, Brinkmann, “Nonlinear electron resonance heating in capacitive radio frequency discharges”, Appl. Phys. Lett., vol.88, 2006.

[14] Mussenbrock, Brinkmann, “Nonlinear plasma dynamics in capacitive radio frequency discharges”, Plasma Sources Sci. Technol. vol.16, 2007.

LAM[®] related

[15] Schoenborn, Toda, “Application of a high density inductively coupled plasma reactor to polysilicon etching”, J. Vac. Sci. Technol. A, vol.11, no.4, Jul/Aug 1993.

[16] Hopwood, Guarnieri, Whitehair, Cuomo, “Electromagnetic fields in a radio-frequency induction plasma”, J. Vac. Sci. Technol. A, vol.11, 1993.

General ICP related

[17] Denda, Miyoshi, Komukai, Goto, Petrović, Makabe, “Functional separation in two frequency operation of an inductively coupled plasma, J. Appl. Phys. vol.95, no.3, Feb 2004.

[18] Kitajima, Takeo, Petrović, and Makabe, “Functional separation of biasing and sustaining voltages in two-frequency capacitively coupled plasma”, Appl. Phys. Lett., vol.77, no.4, July 2000.

[19] Boyle, Ellingboe, Turner, “Independent control of ion current and ion impact energy onto electrodes in dual frequency plasma devices”, J. Phys. D: Appl. Phys. 37, 2004.

[20] Boyle, Ellingboe, Turner, “Electrostatic modelling of dual frequency rf plasma discharges”, Plasma Sources Sci. Technol. 13, 2004.

[21] Ryan, Plumb, “Gas-phase reactions in plasmas of SF₆ with O₂ in He”, Plasma Chemistry and Plasma Processing, vol.8, no.3, September 1988.

[22] Wright, Chen, Federlin, Forbes, “Manufacturing issues of electrostatic chucks”, J. Vac. Sci. Technol. B, vol.13, issue 4, 1995.

[23] Daviel, Peccoud, Mondon, “Electrostatic Clamping Applied to Semiconductor Plasma Processing”, J. Electrochem. Soc., vol.140, issue 11, 1993.

[24] Klick, Bernt, “Microscopic approach to an equation for the heat flow between wafer and E-chuck”, J. Vac. Sci. Technol. B, vol.24, issue 6, 2006.

PROES related

Textbooks

[25] Grove, “Analytical Emission Spectroscopy”, Grove, vol.1, part 1, Decker, 1971.

Articles

[26] Tochikubo, Suzuki, Kakuta, Terazono, Makabe, “Study of the structure in rf glow discharges in SiH₄/H₂ by spatiotemporal optical emission spectroscopy: influence of negative ions”, J. Appl. Phys., vol.68, no.11, Dec 1990.

[27] Tochikubo, Makabe, Kakuta, Suzuki, “Study of the structure of radio frequency glow discharges in CH₄ and H₂ by spatiotemporal optical emission spectroscopy”, J. Appl. Phys., vol.71, no.5, Dec 1992.

-
- [28] Petrović, Tochikubo, Kakuta, Makabe, “Spatiotemporal optical emission spectroscopy of rf discharges in SF₆”, J. Appl. Phys., vol.73, no.5, March 1993.
- [29] Leroy, Stratil, Perrin, Jolly, Belenguer, “Spatiotemporal analysis of the double layer formation in hydrogen radio frequency discharges”, J. Phys. D: Appl. Phys. 28, 1995.
- [30] Mahony, Wazzan, Graham, “Sheath dynamics observed in a 13.56 MHz-driven plasma”, Appl. Phys. Lett., vol.71, no.5, August 1997.
- [31] Czarnetzki, Luggenhölscher, Döbele, “Space and time resolved electric field measurements in helium and hydrogen RF-discharges”, Plasma Sources Sci. Technol. 8 1999.
- [32] Gans, Schulz-von der Gathen, Döbele, “Spectroscopic measurements of phase-resolved electron energy distribution functions in RF-excited discharges”, Europhys. Lett., 66, 2004.
- [33] Dittmann, Drozdov, Krames, Meichsner, “Radio-frequency discharges in oxygen: II. Spatio-temporally resolved optical emission pattern”, J. Phys. D: Appl. Phys. 40, 2007.
- [34] Schulze, Heil, Luggenhölscher, Mussenbrock, Brinkmann, Czarnetzki, “Electron beams in asymmetric capacitively coupled radio frequency discharges at low pressures”, J. Phys. D: Appl. Phys. 41, 2008.
- [35] Gans, Schulze, O’Connell, Czarnetzki, Faulkner, Ellingboe, Turner, “Frequency coupling in dual frequency capacitively coupled radio-frequency plasmas”, Appl. Phys. Lett. 89, 2006.
- [36] Schulze1, Gans, O’Connell, Czarnetzki, Ellingboe, Turner, “Space and phase resolved plasma parameters in an industrial dual-frequency capacitively coupled radio-frequency discharge”, J. Phys. D: Appl. Phys. 40, 2007.
- [37] Schulze1, Donkó, Heil, Luggenhölscher, Mussenbrock, Brinkmann, Czarnetzki, “Electric field reversals in the sheath region of capacitively coupled radio frequency discharges at different pressures”, J. Phys. D: Appl. Phys. 41, 2008.

X2. Black Silicon

Textbooks

- [1] McQuarrie, Simon, “Physical Chemistry”, University Science Books, 1997.
- [2] Bubertt, Jenett, “Surface and Thin Film Analysis”, Wiley, 2003.
- [3] Biederman, “Plasma Polymer Films”, Imperial College Press, 2004.

Articles

- [4] Younkin, “Surface studies and microstructure fabrication using femtosecond laser pulses”, PhD Dissertation, Harvard University, Cambridge Massachusetts, August 2001.
- [5] Wu, “Femtosecond laser-gas-solid interactions”, PhD Dissertation, Harvard University, Cambridge Massachusetts, August 2000.

[6] Jansen et al, “The black silicon method: a universal method for determining the parameter setting of a fluorine-based reactive ion etcher in deep silicon trench etching with profile control”, J. Micromech. Microeng. 5, 1995.

[7] Jansen, Boer, Unnikrishnan, Louwerseand, Elwenspoek, “Topical Review: Black silicon method”, J. Micromech. Microeng. 19, 2009.

[8] Kremin, “Kombiniertes form- und kraftschlüssiges Fügeverfahren für Silizium”, Diplomarbeit, Ilmenau, 2005.

[9] Dussart, Mellhaoui, Tillocher, Lefauchaux, Volatier, Socquet-Clerc, Brault, Ranson, “Silicon columnar microstructures induced by an SF₆/O₂ plasma”, J. Phys. D: Appl. Phys. 38, 2005.

[10] Pereira, Pichon, Dussart, Cardinaud, Duluard, Oubensaid, Lefauchaux, Boufnichel, Ranson, “In situ x-ray photoelectron spectroscopy analysis of SiO_xF_y passivation layer obtained in a SF₆/O₂ cryoetching process”, Appl. Phys. Lett., 94, 2009.

X3. Optics

Textbooks

[1] Hecht, “Optics”, 4th ed., Addison Wesley, 2002.

[2] Gomez-Reino, Perez, Bao, “Gradient Index Optics”, Springer 2002.

[3] Strang, “Computational Science and Engineering”, Wellesley-Cambridge Press, 2007.

[4] Taflove, Hagness, “Computational Electrodynamics: The Finite-Difference Time-Domain Method”, 3rd Edition, Artech House Publishers, 2005.

[5] Feynman, “The Feynman Lectures on Physics: Vol.2”, Definitive Edition, Pearson Addison Wesley, 2006.

[6] Jackson, “Classical Electrodynamics”, 3rd edition, Jon Wiley & Sons, 1999.

[7] Raether, “Surface Plasmons on Smooth and Rough Surfaces and on Gratings”, Springer, 1988.

[8] Oppenheim, “Discrete-Time Signal Processing”, 3rd edition, Prentice Hall, 2009.

[9] FDTD Solutions: Reference Guide, Release 6.5, Lumerical Solutions, Inc., Vancouver BC, 2009.

Articles

[10] Yamamoto, “Surface Plasmon Resonance (SPR) Theory”, Tutorial, Kyoto University, 2008.
http://www.konan-u.ac.jp/~dcfm/applphys/web_material/spr_tutorial/sprtheory.html,
last retrieved July 17th 2010.

[11] Yee, “Numerical Solution of Initial Boundary Value Problems Involving Maxwell’s Equations in Isotropic Media”, IEEE AP Trans., vol.14, no.3, 1966.

-
- [12] Taflove, "Application of the Finite-Difference Time-Domain Method to Sinusoidal Steady-State Electromagnetic-Penetration Problems", IEEE EMC Trans., vol.22, no.23, 1980.
- [13] Umashankar, Taflove, "A Novel Method to Analyze Electromagnetic Scattering of Complex Objects", IEEE EMC Trans., vol.24, no.24, 1982.
- [14] Umashankar, Taflove, "Radar Cross Section of General Three-Dimensional Scatterers", IEEE EMC Trans. vol.25, no.4, 1983.
- [15] Mur, "Absorbing Boundary Conditions for the Finite-Difference Approximation of the Time-Domain Electromagnetic-Field Equations", IEEE EMC Trans., vol.23, no.4, 1981.
- [16] Umashankar, Taflove, Beker, Harfoush, Yee, "Detailed FD-TD Analysis of Electromagnetic Fields Penetrating Narrow Slots and Lapped Joints in Thick Conducting Screens", IEEE AP Trans., vol.36, no.2, 1988.
- [17] Luebbers, Hunsberger, Kunz, Standler, Schneider, "A Frequency Dependant Finite-Difference Time-Domain Formulation for Dispersive Materials", IEEE EMC Trans., vol.32, no.3, 1990.
- [18] Umashankar, Taflove, Jurgens, Moore, "Finite-Difference Time-Domain Modeling of Curved Surfaces", IEEE AP Trans., vol.40, no.4, 1992.
- [19] Johnson, "Notes on Perfectly Matched Layers (PMLs)", <http://www-math.mit.edu/~stevenj/18.369/pml.pdf>, last retrieved July 17th 2010.
- [20] Guasti, "Diffraction pattern of a circle/square aperture", JMO, vol.40, no.6, 1993.
- [21] Glaser, Ihring, Morgenroth, Seifert, Schröter, Baier, "High temperature resistant antireflective moth-eye structures for infrared radiation sensors", Microsystem Technologies, 11, 2005.
- [22] Brückner, Pradarutti, Stenzel, Steinkopf, Riehemann, Notni, Tünnermann, "Broadband antireflective surface-relief structure for THz optics", Optics Express, vol.15, no.3, 2007.
- [23] Rushton, "Densitometry of Pigments in Rods and Cones of Normal and Color Defective Subjects", Investigative Ophthalmology and Visual Science, 5, 1966.
- [25] Michely, Krug, "Islands, Mounds and Atoms: Patterns and Processes in Crystal Growth Far from Equilibrium", Springer, 1st edition, 2003.

X4. Photodiodes and semiconductor processing

Textbooks

- [1] Streetman, Banerjee, "Solid State Electronic Devices", 6th Edition, Prentice Hall, 2005.
- [2] Wolf, "Silicon Processing for VLSI", vol.1, Lattice Press, 1986.
- [3] Kern, "Handbook of Semiconductor Wafer Cleaning Technology", William Andrew publishing, 1994.

Articles

[4] Vörckel, “Integration of Black-Silicon to Photo Diode Flow”, taken with permission from internal presentation covering project targets (X-FAB Semiconductor Foundries AG, 2009)

[5] Thomas, “SIMS doping profile”, taken with permission from internal process specifications of the X-FAB XC06 0.6um technology (X-FAB Semiconductor Foundries AG)

[6] Gibbons, “Ion implantation in Semiconductors – Part I: Range Distribution Theory and Experiments”, IEEE, vol.56, no.3, 1968.

[7] Gibbons, “Ion implantation in Semiconductors – Part II: Damage Production and Annealing”, IEEE, vol.60, no.9, 1972.

[8] Robinson, Oen, “Computer studies of the slowing down of energetic atoms in crystals”, Phys. Rev., vol.132, 1963.

[9] Lindhard, Scharff, Schiott, “Range concepts and heavy ion ranges”, Mat. Fys. Medd. Dan. Vid. Selsk., vol. 33, 1963.

[10] Beck, Hatcher, Tsetseris, Caussanel, Schrimpf, Fleetwood, Pantelides, “Elemental Defect Processes in Radiation-Induced Displacement Damage in Si”, MURI Review, 2006.

[11] Lannoo, “The role of dangling bonds in the properties of surfaces and interfaces of Semiconductors”, Revue Phys. Appl. 25, 1990.

[12] Pacchioni, Skuja, Griscom, “Defects in SiO₂ and Related Dielectrics”, Springer, 1st edition, 2000.

[13] Stirling, Pasquarello, Charlier, Car, “Dangling bond defects at Si-SiO₂ interfaces: atomic structure of the Pb1 center”, Phys. Rev Lett., vol.85, no.13, 2000.

[14] Caughey, Thomas, “Carrier mobilities in silicon empirically related to doping and field”, Proc. IEEE, 55, 1967.

[15] Alamo, Swirhun, Swanson, “Measure and modeling minority carrier transport in heavily doped silicon”, Solid State Electronics, 28, 1985.

[16] Sanders, Kunst, “Characterization of Silicon Wafers by Transient Microwave Photoconductivity Measurements”, Solid-State Electron. 34, 1991.

[17] Gaubas, Kaniava, “Determination of Recombination Parameters in Silicon Wafers by Transient Microwave Absorption”, Rev. Sci. Instrum. 67, 1996.

[18] Masaya, Atsushi, Eisuke, Hiroyuki, Shingo, “Bulk carrier lifetime measurement by the microwave reflectance photoconductivity decay method with external surface electric field”, Appl. Phys. Lett. vol.80, no.23, 2002.

[19] Jaeger, “Introduction to Microelectronic Fabrication”, Thermal Oxidation of Silicon, Prentice Hall, 2001.

-
- [20] Chen, Liu, Dai, Weng, Jih, “Extended One-Dimensional Analysis to Effectively Derive Quantum Efficiency of Various CMOS Photodiodes”, IEEE Trans. on Electronic Devices, vol.54, no.10, 2007.
- [21] Pang, Rathman, Silversmith, Mountain, DeGraff, “Damage in Si by ion milling or reactive ion etching”, J. Appl. Phys. vol.54, no.6, 1983.
- [22] Misra, Heasell, “Electrical damage to silicon devices due to reactive ion etching”, Semicond. Sci. Technol. 5, 1990.
- [23] Yoo, Kim, Thamilselvan, Lakshminarayn, Kim, Lee, Yoo, Yi, “RIE texturing optimization for thin c-Si solar cells in SF₆/O₂ plasma”, J. Phys. D: Appl. Phys. 41, 2008.
- [24] Rentsch, Kohn, Bamberg, Roth, Peters, Lüdemann, Preu, “Isotropic plasma texturing of mc-Si for industrial solar cell fabrication”, IEEE, 2005.
- [25] Christofides, “Annealing kinetics of defects of ion-implanted and furnace-annealed silicon layers: thermodynamic approach”, Semicond. Sci. Technol. 7, 1992.
- [26] del Alamo, Swirhun, Swanson, “Simultaneous measurement of hole lifetime, hole mobility and bandgap narrowing in heavily doped n-type silicon”, IEDM Technical Digest 85, 1985.
- [27] Swirhun, Kwark, Swanson, “Measurement of electron lifetime, electron mobility and bandgap narrowing in heavily doped p-type silicon”, IEDM Technical Digest, 1986.

Appendix A: Improvement of PMLs by gradient-index

In figure A.01 we see on the left-hand side the real part of an oscillating wave e^{ikx} corresponding to the x -axis in the complex- x plane on the right-hand side [X3.19]. At $x < 5$ a deformity in the complex plane is introduced and the oscillating wave exponentially decays as we move further along x . This deformity is referred to as the “absorbing” region (the PML).

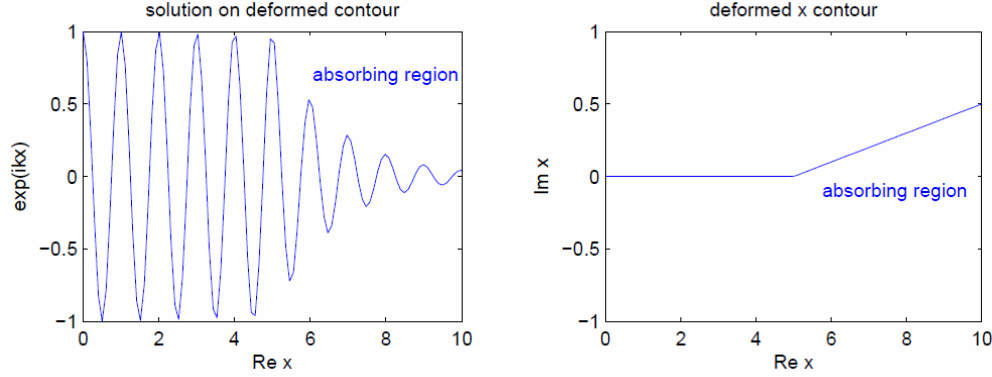


Figure A.01: An oscillating wave e^{ikx} corresponding to the x -axis in the complex plane. We see for values above 5 (the “absorbing” region or PML) the wave is exponentially damped [X3.19].

To avoid confusion we now denote the complex x by x_c and the real x by x_r . We have $x_c(x) = x_r + i f(x)$, where $f(x)$ is some function indicating how the imaginary axis has been deformed. In figure A.01 the function $f(x)$ was chosen to be zero for $x \leq 5$ and a positive constant for $x > 5$, which gave the exponential decay. There are no reflections from a PML because this attenuation is only an analytic continuation of the original solution from purely x_r to x_c . Furthermore this attenuation is independent of frequency, all wavelengths decay at the same rate, but unfortunately this same attenuation is dependent on the angle of the light.

The problem is that the attenuation is actually not dependant on x , it is dependant on k of the oscillating wave, or better said the kx component of the oscillating wave. This means that, regardless of PML thickness, waves sufficiently close to glancing incidence will have substantial reflections from the PML (as the very bottom of the PML is a perfect metal boundary - discussed in the optical properties chapter). The current rule of thumb is nothing over angles of 55 degrees [X3.19].

What we propose here is not to modify the PML itself, but rather to “bend” light out of the simulation space and into the PML. In figure A.02 a graph of 600nm TM light is shown incident from a medium with $n=1$ on a medium of $n=2$. The blue line shows the reflection calculated with the Fresnel equation,

$$R_{TM} = \frac{n_1 \sqrt{1 - \left(\frac{n_1}{n_2} \sin(\theta_i) \right)^2} - n_2 \cdot \cos(\theta_i)}{n_1 \sqrt{1 - \left(\frac{n_1}{n_2} \sin(\theta_i) \right)^2} + n_2 \cdot \cos(\theta_i)} \quad (\text{A.01})$$

The green line is the reflection yielded by the corresponding FDTD simulation. At approximately 70 degrees the simulation begins to give nonsense reflection values due to reflections from the upper and lower PMLs. Or better said, at approximately 70 degrees we see Fabry-Pérot resonance between the interface of the $n=1$ and the $n=2$ material and the perfect metal bottom of the PMLs. This situation is greatly improved in figure A.03. Here the Fresnel reflection is again shown in contrast to the FDTD result. However this time in front of the PMLs a finite number of thin layers are introduced, creating a gradient-index. This is exactly how we modeled the nano-structures in figure 4.10 with a multi-layer stack. This could be understood as light “bending” towards the higher index of refraction, towards the PML! This lines up the k component (k_x for example, assuming the axis of damping for the PML is x) of the incident light with the directional component of the PML.

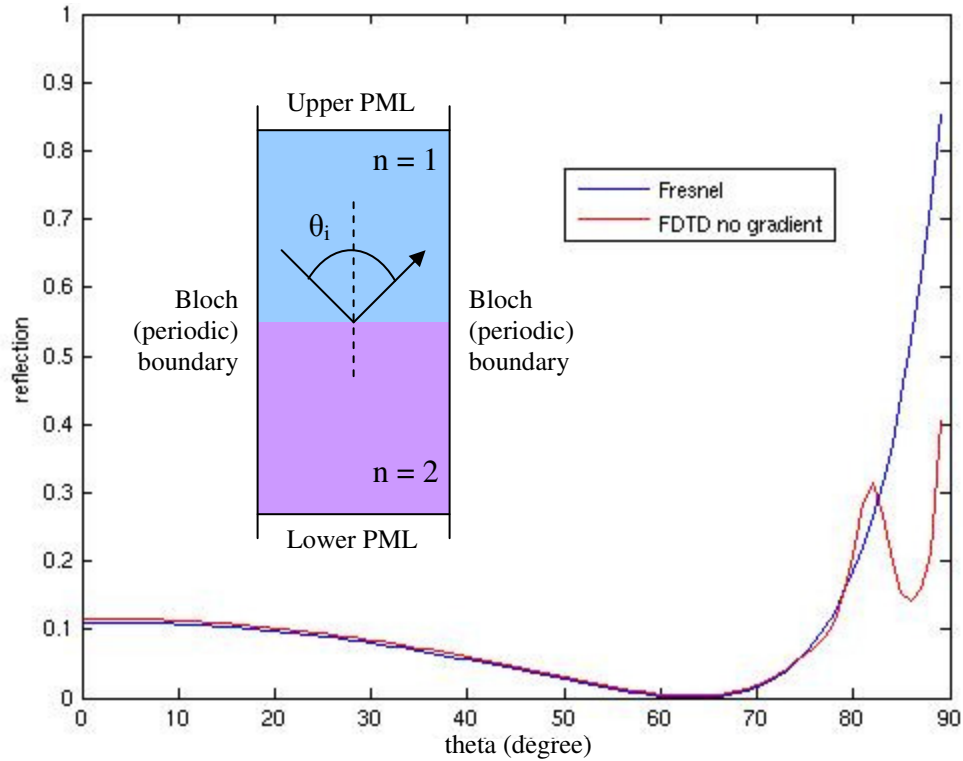


Figure A.02: Reflection dependant on angle of incidence between a material having $n=1$ and a material having $n=2$. The blue line is calculated with the Fresnel equation (TM light) and the green line is the FDTD simulation result. At approximately 70 degrees the simulation result begins to return nonsense.

We see a noticeable improvement in performance up to incident angles of 87 degrees, however the FDTD result no longer follows the Fresnel line as exactly for less extreme angles of incidence. An attempt at introducing a smoother transition is shown in figure A.04, where instead of a finite number of layer, a pyramid structure is introduced in front of the PML. For the silicon pyramid we see that the reflection again follows the Fresnel line and the PML appears to function up to 87 degrees. The pyramid of a lower refractive index, $n=1.1$ (closer to the $n=1$ medium) creates the very best situation, following the Fresnel line perfectly up to approximately 85 degrees. To the author’s knowledge this is seen (at the time of writing) nowhere in current literature is such a method discussed to improve the performance of PMLs. This novel set of first

experiments will be expanded upon in the future when the mechanisms are more clearly understood.

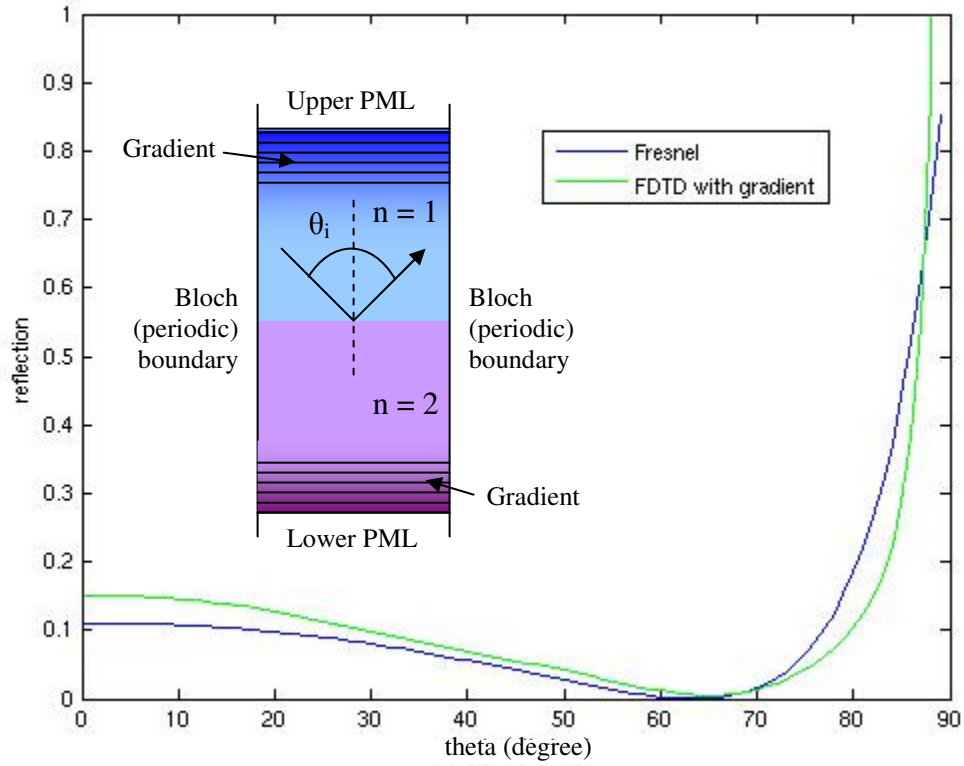


Figure A.03: Modified FDTD result using a multi-layer gradient-index in front of the PML. The gradient is created by a finite number of layers with each subsequent layer added in the direction of the PML having a slightly higher refractive index. The PML appears to function up to angles of 87 degrees, however the overall performance of the multi-layer gradient does not exactly follow the Fresnel result.

The next logical step would be to replace the finite number of layers and pyramids used in this work with an function of epsilon, different functions (linear, exponential) may lead to better results. Additionally it has been shown in this dissertation that the gradient properties of pyramids are best when light is perpendicularly incident on the structures. The addition of a direction-independent gradient would be beneficial.

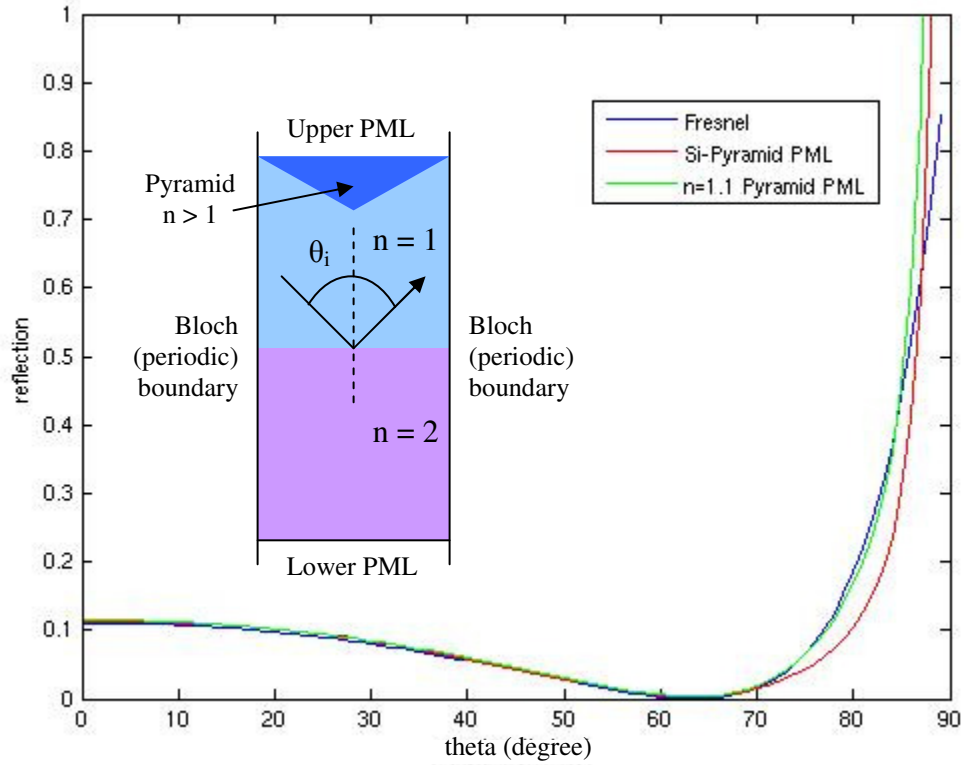


Figure A.04: Modified FDTD result using a pyramid gradient-index in front of the PML. The gradient is created by the addition of a pyramid structure, which should represent a more smooth index transition in comparison to the multi-layer method. The PML appears to function up to angles of 87 degrees for the silicon pyramid (Si-pyramid), however the overall performance of the pyramid with index $n=1.1$ exactly follows the Fresnel result up to 85 degrees.

Appendix B: Thesen zur Dissertation

„Black Silicon for Photodiodes: Experimentally Implemented and FDTD Simulated“

vorgelegt von A. Williamson

- The appearance of nano-structures on the silicon surface when processed in the SF_6/O_2 plasma mixture is explained by generalizing the etch rate of silicon dependant on fluorine concentration (with a varying fluorine to oxygen gas ratio) into two regions, one of excess fluorine generation (weakly passivating) and one of excess fluorine consumption (strongly passivating). This is experimentally verified.
- Beyond a certain silicon substrate temperature, nano-masking ceases to be possible.
- Inherent non-uniformities of the plasma processing chamber which yield a small process window of extremely anisotropic nano-structures and a poor center-to-edge wafer uniformity are eliminated by the introduction of a nano-loading step. The silicon wafer surface is first etched with a set of highly uniform nano-pillars.
- The nano-loading mask is subsequently plasma etched utilizing gas compositions that do not by themselves result in the appearance of nano-structures on the silicon wafer surface. This yields new less-anisotropic nano-structure geometries with properties more desirable for reflection suppression.
- Due to a greatly increased silicon etch rate resulting from the decrease in exposed silicon, an extended nano-loading approach is necessary to successfully transfer plasma processes from blank wafers to masked wafers.
- The majority of light absorption in the spectrum from 200 nm to 800 nm takes place within the silicon nano-structure antireflective coating, in contrast to a standard anti-reflective coating. This is demonstrated with FDTD.
- Simulations are best correlated to the optical properties of physical self-organized nano-structures (integrating sphere measurements of specular and diffuse components) when a non-periodic surface is simulated.
- Improvement of the angle-dependent performance of perfectly matched layers (PMLs). We demonstrate how to utilize PMLs with extreme non-normal incidence light by including gradient index layers in front of the absorbing boundary.
- Clear separation between nano-structure tips is experimentally demonstrated as the deciding factor to improve reflection suppression.
- Reflection suppression is not strongly correlated to electrical performance as plasma etching to modify surface topology drastically reduces carrier lifetimes (increase in recombination due lattice damage and to an overall increase in surface area).
- Photodiode efficiency in the red region (675 nm - 750 nm) meets the 0.5 A/W target, while in the blue region (375 nm - 425 nm) an improvement of +0.07 A/W over oxide covered diodes is achieved for center cites.

Appendix C: Erklärung gemäß Anlage 1 der Promotionsordnung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. Prof. Dr. Martin Hoffmann
2. Prof. Dr. Erich Runge
3. Dr. Andreas Voerckel
4. Dipl.-Ing. Chris Kremin
5. Dipl.-Ing. David Leipold
6. Dipl.-Ing. Michael Fischer

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalte der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch angesehen wird und den erfolglosen Abbruch des Promotionsverfahrens zu Folge hat.

Ilmenau, 15. Juli 2010

Adam Williasmon