Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia

Quantitative infrared analysis of the asymmetric stretching peak of SiO_2 films thermally grown

João Duarte Neves Cruz

Supervisors:
Dr. Antoine Goullet (IRESTE, Université de Nantes)
Prof. Dr. Rodrigo Martins (FCT, Universidade Nova de Lisboa)

Lisboa, March 1995

Abstract

In this work we make a quantitative infrared analysis of the asymmetric stretching peak of SiO_2 films thermally grown, and with thickenesses ranging between ≈ 16 nm and ≈ 527.5 nm.

For this effect, it was developed a software, FITIR, which is basically a nonlinear least-squares curve fitting program capable of executing two analysis methods: gaussian decomposition and dispersion analysis.

The results obtained by dispersion analysis are "biased" due to the fact that the studied spectra show higher values than the real ones. This problem has been partially solved, reducing the "bias" to, what we believe, acceptable values.

Gaussian decomposition strongly suggests that the asymmetric stretching peak is composed, not by one or two vibrational modes as suggested in literature, but by three, and that at least two of these modes show a longitudinal-optic − transverse-optic (LO−TO) frequency splitting. This implies that LO modes are only observed in oblique incidence spectra of radiation polarized with the electric field vector parallel to the plane of incidence.

The majority of the gaussian parameters show large changes on their values with film thickness, thus explaining the asymmetric stretching peak shift with film thickness.

When the dielectric function is modelled by a set of lorentzian oscillators with a gaussian distribution of frequencies, dispersion analysis of transmittance spectra corroborates quite reasonably the results obtained in gaussian decomposition, namely in the number of vibrational modes and their respective relative intensities. However, this model starts failing for reflectance spectra. With this method of analysis, some properties of the dipoles responsible for the vibrational peaks were determined, as well as the SiO_2 films complex dielectric function and complex refractive index.

The dispersion parameters, following the example of the gaussian parameters, also show a high dependence on film thickness.

There is some temptation to justify these parameters changes by assuming the existence of bond strains in the oxide, which in turn would also change with film thickness. However, if this is true, its effect is rather irregular since there are parameters that change, with film thickness, but not for all peaks.

Key words

SiO_2 films. Fourier-transform infrared spectroscopy. Nonlinear least-squares curve fit. Gaussian decomposition. Dispersion analysis. FITIR. Crystalline-like behaviour. Amorphous-like behaviour. LO−TO frequency splitting. Lorentzian oscillators with gaussian distribution of frequencies.

Resumo

Pretende-se com este trabalho analisar quantitativamente o pico de alongamento assimétrico de filmes de SiO_2 formados termicamente e com espessuras entre ≈16 nm e ≈527.5 nm, observado na espectroscopia de infravermelhos.

Desenvolveu-se, para este efeito, um programa de computador, FITIR, que baseia o seu funcionamento no ajuste não linear de curvas pelo método dos mínimos quadrados e que vai executar dois métodos de análise: decomposição com gaussianas e análise dispersiva.

Os resultados obtidos através da análise dispersiva estão “distorcidos” pelo facto dos espectros de infravermelhos estudados apresentarem valores superiores aos reais. Este problema foi parcialmente resolvido, reduzindo a “distorção” a valores, que julgo serem, aceitáveis.

Da decomposição com gaussianas observa-se uma forte evidência no sentido do pico de alongamento assimétrico ser formado, não por um ou dois modos vibracionais conforme sugerido na literatura, mas sim por três, e que pelo menos dois dos modos apresentam uma divisão em frequência Longitudinal-óptica − Transversal-óptica (LO−TO). Isto implica que os modos LO só são observados nos espectros de incidência oblíqua com radiação polarizada cujo vector campo eléctrico é paralelo ao plano de incidência.

A maioria dos parâmetros das gaussianas apresenta grandes variações com a espessura do filme de óxido, o que explica o deslocamento do pico de alongamento assimétrico com a referida espessura.

Quando a função dieléctrica é modelada por um conjunto de osciladores lorentzianos com uma distribuição gaussiana de frequências, a análise dispersiva dos espectros de transmitância corrobora bastante razoavelmente os resultados obtidos na decomposição com gaussianas, nomeadamente no número de modos vibracionais e respectivas intensidades relativas. No entanto, este modelo começa a falhar nos espectros de reflectância. Com este método de análise foi possível determinar algumas propriedades dos dipolos que originam os picos vibracionais bem como a função dieléctrica complexa e índice de refracção complexo dos filmes de SiO_2.

Os parâmetros da dispersão, seguindo o exemplo dos parâmetros das gaussianas, também apresentam grandes variações com a espessura do filme.

Poder-se-à tentar justificar esta variação dos parâmetros, obtidos com ambos os métodos de análise, assumindo a existência de “esforços” nas ligações químicas (“bond strains”) do óxido, que por sua vez também variariam com a espessura do filme. No entanto, a ser verdade, o seu efeito é bastante irregular dado que existem parâmetros que só apresentam variações com a espessura do filme em alguns picos.

Palavras chave

Filmes de SiO_2. Espectroscopia de infravermelhos com transformada de Fourier. Ajuste não linear de curvas pelo método dos mínimos quadrados. Decomposição com gaussianas. Análise dispersiva. FITIR. Comportamento do tipo cristallino. Comportamento do tipo amorfo. Divisão de frequência LO-TO. Osciladores lorentzianos com distribuição gaussiana de frequências.

Acknowledgements to

Dr. Antoine Goullet	for the fruitful discussions and enlightening tutorials throughout the execution of this project.
Prof. Dr. Rodrigo Martins	for his availability to review and correct this report.
All my friends	who helped me with their suggestions.
My family	for their support and care.

Table of Contents

INTRODUCTION **1**

1 - SILICON DIOXIDE FILMS **3**

1.1- INTRODUCTION 3
1.2 - THERMAL OXIDES 4
1.3 - DEPOSITED OXIDES 6
1.4- INFRARED 'SIGNATURE' 7
1.4.1- Vibrational energies of a diatomic molecule 8
1.4.1.1- Harmonic oscillator 8
1.4.1.2- The anharmonicity of molecular vibrations 9
1.4.2- Vibrational energies of a polyatomic molecule 10
1.4.3 - Vibrational modes in a-SiO_2 films 11

2 - THIN FILM OPTICS **19**

2.1- INTRODUCTION 19
2.2- WAVE PROPAGATION IN A LOSSY DIELECTRIC 19
2.3- REFLECTION AND TRANSMISSION OF LIGHT AT THE SURFACE OF A TRANSPARENT MEDIUM 21
2.4- REFLECTION AND TRANSMISSION OF LIGHT AT THE SURFACE OF AN ABSORBING MEDIUM 23
2.5- REFLECTION AND TRANSMISSION OF LIGHT BY A SINGLE FILM 23
2.6- N AND K AS FUNCTIONS OF ε_1 AND ε_2 27

3 - DIELECTRIC DISPERSION THEORY **29**

3.1- INTRODUCTION 29
3.2- DIELECTRIC FUNCTION FOR CRYSTALLINE SILICON DIOXIDE 29
3.3- DIELECTRIC FUNCTION FOR AMORPHOUS SILICON DIOXIDE 31

4 - FITIR , A NONLINEAR LEAST-SQUARES CURVE FITTING PROGRAM **33**

4.1- INTRODUCTION 33
4.2- CHI-SQUARE FITTING 34
4.3- MATHEMATICAL SOFTWARE 36
4.4- STATISTICAL TREATMENT 38
4.4.1- Goodness-of-fit 38
4.4.2 - Error estimates on the parameters 39
4.5 - DATA SAVING 39
4.6 - FITIR FLOW CHART 40

5 - MODELS EQUATIONS **43**

5.1- INTRODUCTION 43
5.2- MODEL FOR DISPERSION ANALYSIS 43
5.2.1- Dielectric function 44
5.2.1.1- Finding the integration limits 46
5.2.2- Transmittance equation 46
5.2.3- Reflectance equation 48
5.3- MODEL FOR GAUSSIAN DECOMPOSITION 52

6 - EXPERIMENTAL PROCEDURE **53**

7 - Spectra analysis 61
7.1- INTRODUCTION 61
7.2- GAUSSIAN DECOMPOSITION 61
7.2.1- Transmittance spectra....61
7.2.2- Reflectance spectra 65
7.3- DISPERSION ANALYSIS 72
7.3.1- Transmittance spectra....72
7.3.2- Reflectance spectra 78
7.4- MATHEMATICAL SOFTWARE PERFORMANCE 79

8- CONCLUSIONS 81

APPENDIXES 82

Introduction

The properties of amorphous SiO_2 (a-SiO_2) films on single crystal Si (c-Si) substrates are very important from a technological viewpoint, given their application in the electronics industry.

The structure of the oxide layer and the nature of the SiO_2 /Si interface have been extensively studied. However, they still generate considerable disagreement, namely concerning the existence of bond strains and the exact structural significance of the peaks occuring between $\approx 1000\ cm^{-1}$ and $\approx 1300\ cm^{-1}$. Although these features are generally accepted to arise mostly from asymmetric stretching vibrations of Si–O–Si bridging sequences, different interpretations still hold. Therefore, the study of this frequency region is very pertinent.

The present work is organized in the following manner. In Chapter 1, we make a brief introduction to silicon dioxide films, namely their applications, growing/deposition techniques, and infrared properties, where we enfasize the different interpretations and theories developed to try to explain the peaks structure between $\approx 1000\ cm^{-1}$ and $\approx 1300\ cm^{-1}$. Chapters 2 and 3 are devoted to the basic development of the equations on which dispersion analysis is based upon: Fresnel equations for wave propagation and dielectric dispersion theory. In Chapter 4, the software specially developed for this work, FITIR, is briefly described. Chapter 5 gives the more important steps taken to calculate the models equations for gaussian decomposition and dispersion analysis. In Chapter 6 the experimental procedure is described, and the obtained transmittance and reflectance spectra are plotted. These spectra are analysed in Chapter 7 and the results compared with the results of other authors.

1 - Silicon dioxide films

1.1- Introduction

Oxide films are essential in the fabrication of integrated circuits, as they [1]: provide isolation between devices; mask against impurities, dopant diffusion, and environmental effects; provide stability against mechanical damages, e.g., scratching; act as an insulating layer between conductors; act as the gate dielectric.

Silicon dioxide films, formed on a silicon surface, have the highest obtainable quality, in terms of electrical characteristics. In fact, for semiconductors other than silicon, the attempts to form oxide films with good electrical properties have been made without any significant success.

Silicon dioxide films have the further advantage of being easily etched off with buffered hydrofluoric acid (solution of 34.6 % NH_4F, 6.8 % HF, and 58.6 % H_2O by weight) to which the underlying silicon is impervious [2].

The superiority of these films in comparison to other oxide films justify the overwhelming use of silicon technology in the industry of integrated circuits.

Oxide layers with thicknesses from 2 to 1000 nm are of technical interest [3]. Their properties depend on their intended use.

Basically, silicon dioxide films may be divided in two groups: thermal oxides and deposited oxides. They complement each other. For instance, both gate and field oxide layers of a MOSFET (figure 1.1) are generally thermally grown because only this process is able to provide the highest quality oxides having the lowest interface trap densities [4]. On the other hand, the multilevel interconnection technology requires insulating films to be deposited below 400°C. For such low temperatures the best quality films are obtained by plasma enhanced chemical vapor deposition (PECVD) of silicon dioxide using tetraethoxysilane $\left[\mathrm{TEOS, Si(OC_2H_5)_4}\right]$ and oxygen as gas reactants [5].

Deposited oxides also complement the thermal oxides in the way that the formation of silicon dioxide films over materials other than silicon (which is the case of the blue layers of figure 1.1) may only be done by deposition techniques, since thermal oxides are formed by the oxidation of the silicon wafer (section 1.2).

Both types of oxides will be briefly described in the two following sections. Section 1.4 gives a brief overview about the infrared properties of these oxide films.

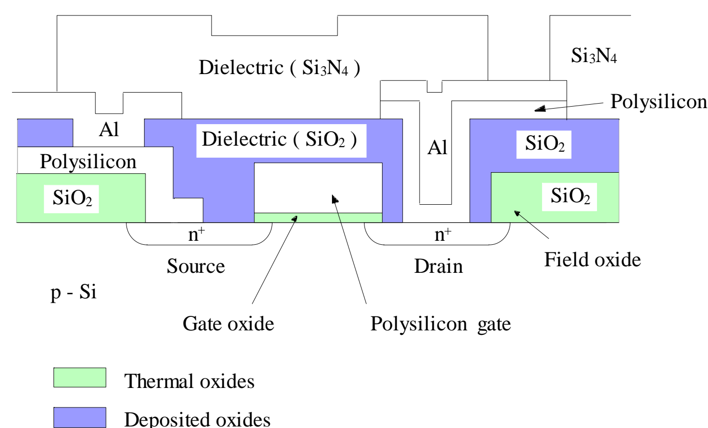


Figure 1.1: Schematic view of a n - channel MOSFET cross section. The gate oxide layer is a layer under which a conducting channel may be formed between the source and the drain. The field oxide layer provides isolation from other device structures.

1.2 - Thermal oxides

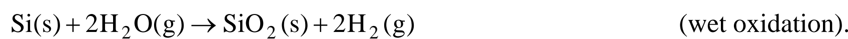
The films with best electrical properties, that is, those with higher dielectric strength[1], which is the applied electric field that will cause large amounts of current to flow in the oxide film, are formed by thermal oxidations in dry oxygen and in water-vapor-containing medium.

They are very reproducible and very stable chemically, being used as a standard for comparison with other types of oxides.

The basic structural unit of thermally grown silicon dioxide is a silicon ion (Si^{4+}) surrounded tetrahedrally by four oxygen ions (O^{2-}), as illustrated in figure 1.2a. These tetrahedra are joined together at their corners by oxygen bridges in a variety of ways to form the various phases or structures of silicon dioxide (also called silica). Silica has several crystalline structures (e.g., quartz - illustrated in figure 1.2c) and an amorphous structure. When silicon is thermally oxidized, the silicon dioxide structure is amorphous, which is illustrated in figure 1.2d. There is still considerable disagreement about the "degree of amorphicity" of this structure. However, we know that it has open spaces, reflected in the low density of amorphous silicon dioxide (a-SiO_2) compared to that of crystalline silicon dioxide (c-SiO_2) [2.27 g.cm^{-3} versus 2.65 g.cm^{-3}]. In spite of this open structure, silicon dioxide films are excellent barriers to the diffusion of several dopants (As, B, P, Sb, Ga, etc.) and metals (except Na). O_2 and H_2O have high diffusivities, an essential feature in thermal oxidation. During thermal oxidation, these oxidising species diffuse rapidly through the oxide layer growing on the silicon and continue to react with silicon at the silicon-oxide interface, according to the reactions[2]:

$$Si(s) + O_2(g) \rightarrow SiO_2(s) \qquad \text{(dry oxidation)}$$

or



$$Si(s) + 2H_2O(g) \rightarrow SiO_2(s) + 2H_2(g) \qquad \text{(wet oxidation).}$$

If the diffusivities of O_2 and H_2O in SiO_2 were low, the oxide formed at the initial stages of oxidation would have acted as a diffusion barrier and oxidation would have been self limited.

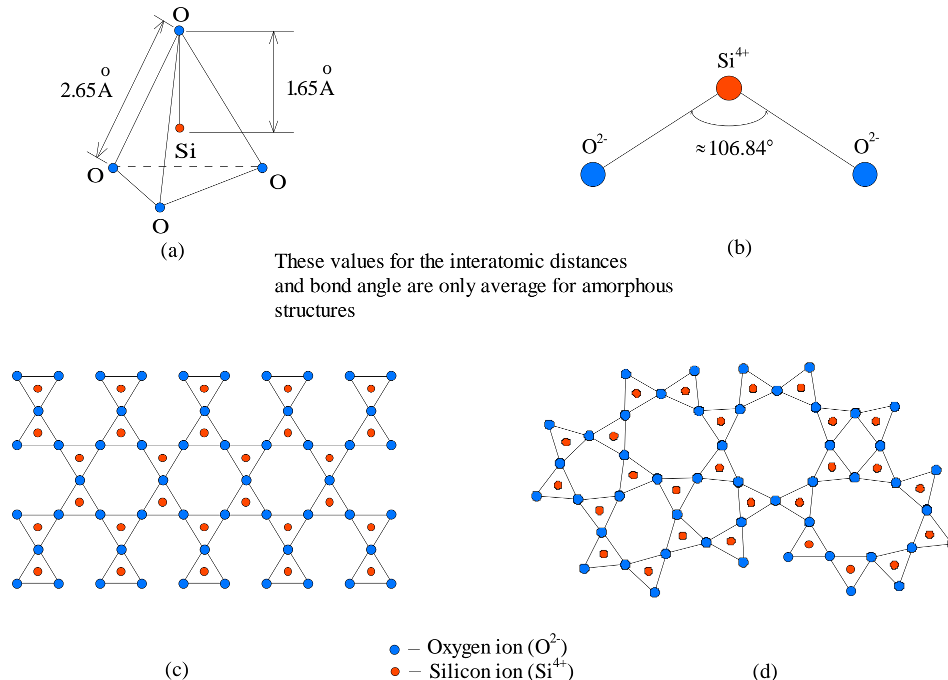


Figure 1.2: (a) Basic structural unit of silicon dioxide. (b) Bond angle. (c) Two dimensional representation of a quartz crystal lattice. (d) Two dimensional representation of amorphous structure of silicon dioxide.

[1] The dielectric or breakdown strength of the oxide film may be measured by fabricating a large number of capacitors, directly on the oxide. They are subjected to an increasing DC bias across the insulator and current is measured. Capacitors that pass more current than a preset value, usually 1-2 μA,are considered failed. For a good oxide, practically all capacitors will fail when subjected to a field of about 10^7 V/cm or more across the oxide.

[2] The silicon surface has always a native oxide with a thickness in the range 0.5 - 2.0 nm depending upon the time of exposure to atmosphere after a chemical clean.

Thermal oxidations are carried out at temperatures in the range 600-1250°C, with the oxidizing species partial pressures ranging approximately between 0.1 - 30 atm.

When the oxidation process starts, the film thickness increases linearly with time. As the oxide thickens, the diffusion of the oxidizing species through the recently formed film starts to slow down the oxidation process and the rate of film growth is no longer linear but parabolic.

Oxides grown in dry oxygen have better electrical properties since in wet oxidation of Si wafers, the hydrogen produced reacts with oxygen ions in the oxide layer, as it diffuses out into the oxidising gas, weakening the structure [6]. However, their growing rate is considerably smaller than the growing rate achieved with wet oxidation. So, to produce thick oxides (≥ 200 nm), that are needed for field and masking applications, steam oxidations are more appropriated.

As mentioned above, SiO_2 films on Si are very successfully used as diffusion barriers (or masks) for dopants. Ion implantation is used to introduce dopants in silicon. By controlling the energy and the dose rate of the implants, we will know in advance the junction depths and the surface concentrations of the created junctions. Oxide films are used as masks to prevent ion implantation in chosen regions of the silicon. The desired thickness necessary to obtain an effective ion implantation mask will depend on the dopant species and on its energy. For example, given a boron ion with an energy of 100 keV, its range in the oxide is 310.4±71.0 nm, suggesting that at least a 400.0 nm thick oxide layer should be used to stop such boron ions in a-SiO_2. A greater thickness (≈ 550 nm) will be used to allow for the diffusion in the oxide during the diffusion anneal [1]. The diffusion depths in a-SiO_2 for the most well known dopants, such as antimony, arsenic, phosphorus and the above mentioned boron, are all given in tables.

High temperatures and high partial pressure of the oxidizing species increase the oxidation rate. This rate also depends on the substrate orientation. The comercially available crystalline silicon wafers are, normally, cut with the orientation <111> or <100> [6]. The oxidation rate of <111> silicon substrates is higher than that of <100> substrates oxidized under identical conditions of substrate doping and oxidation. <111> surfaces have a higher density of silicon atoms for reaction with oxidant species which increases the reaction rate. This difference in the density of silicon atoms means that the oxides grown on <111> substrates are more dense than those grown on <100> substrates.

The oxide properties and its growth rate are also sensitive to impurities either in the oxidizing gas (such as water and chlorine) or in the silicon substrate (such as boron and phosphorus). Water as an impurity in the oxygen for dry-oxidation process increases oxidation rate constants. Chlorine is introduced into the oxidation ambient to improve both the oxide quality and the Si - SiO_2 interface property. This element, contained either in chlorine gas or in such compounds as anhydrous HCl and trichloroethylene, removes certain impurities at the Si - SiO_2 interface by changing them into volatile chlorides. The addition of chlorine also increases the oxidation rate and the dielectric breakdown strength, and reduces the interface trap density.

In silicon integrated-circuit fabrication, silicon is doped during crystalline growth and also during device processing. When doped silicon is oxidised, a redistribution of dopant between the silicon and the growing oxide occurs. The dopant could preferentially segregate in the silicon or in the oxide [4], and the dopant distribution will be affected by the relative dopant diffusivities in silicon and in SiO_2. For instance, boron is lost to the oxide, decreasing its concentration in silicon at the oxide-silicon interface. This segregation weakens the silicon dioxide bond structure, which enhances diffusion of the oxidizing species through the oxide and thus increases the oxidation rate. Phosphorus, on the other hand, piles up in silicon at the interface[3]. Arsenic and gallium behave like phosphorus and boron, respectively, except that they redistribute somewhat differently because of different relative diffusivities in silicon and SiO_2. Variation of dopant concentrations in silicon will affect the electrical characteristics of the devices [4].

There is no universally accepted theory that describes the interface a-SiO_2 / c-Si. As an example, we refer the model proposed by Lewis and Irene [7]. The transition from crystalline Si to an amorphous oxide is rather abrupt (about 3 Å), but there is a transition region of Si-rich SiO_2 which extends several monolayers into the amorphous before a stoichiometric SiO_2 structure is established. Although the oxide attains a stoichiometric SiO_2 composition within 5 - 10 Å of the interface, there is an additional transition region, extending to approximately 30 Å which does not possess the properties of bulk SiO_2. This region appears to be strained, with a distribution of tetrahedral rings different from the bulk.

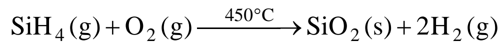
[3] At high phosphorus concentrations the oxidation rate increases.

1.3 - Deposited oxides

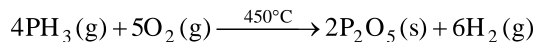
Deposited dielectric films are used for insulation between conducting layers, for diffusion and ion implantation masks, for capping doped films to prevent the loss of dopants, and for passivation to protect devices from impurities, moisture, and scratches. Their basic structural unit is equal to that of the thermally grown films. There are three commonly used deposition methods: atmospheric-pressure chemical vapor deposition (CVD), low-pressure chemical vapor deposition (LPCVD), and plasma-enhanced chemical vapor deposition (PECVD).

The deposition process is selected according to the required substrate temperature, deposition rate and film uniformity, morphology, electrical and mechanical properties, and chemical composition of the dielectric films.

Silicon dioxide films may be deposited by several methods. For low-temperature deposition (300 to 500°C), the films are formed by reacting silane, dopant, and oxygen. For example, the chemical reactions for phosphorus-doped oxides are



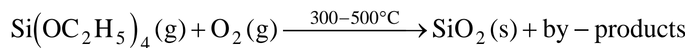
$$SiH_4(g) + O_2(g) \xrightarrow{450°C} SiO_2(s) + 2H_2(g)$$



$$4PH_3(g) + 5O_2(g) \xrightarrow{450°C} 2P_2O_5(s) + 6H_2(g)$$

The deposition process may be performed either at atmosphere pressure in a CVD reactor or at reduced pressure in a LPCVD reactor. The low deposition temperature of the silane-oxygen reaction makes it a suitable process when films must be deposited over a layer of aluminium.

Another low-temperature deposition is accomplished by using tetraethoxysilane $\left[TEOS, Si(OC_2H_5)_4\right]$ and oxygen as gas reactants:

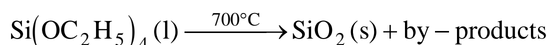


$$Si(OC_2H_5)_4(g) + O_2(g) \xrightarrow{300-500°C} SiO_2(s) + by-products$$

forming both SiO_2 and a mixture of organic and organosilicon by-products.

Low-temperature depositions will become more important as device dimensions are reduced to submicron regions because there must be a minimization of the thermal broadening of dopant distribution due to diffusion. However, their capabilities are limited, especially for large-diameter wafers.

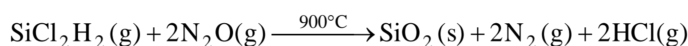
For intermediate-temperature deposition (500 to 800°C), silicon dioxide may be formed by the thermal decomposition (pyrolysis) of TEOS in an LPCVD reactor. This compound is vaporized from a liquid source. It decomposes as follows:



$$Si(OC_2H_5)_4(l) \xrightarrow{700°C} SiO_2(s) + by-products$$

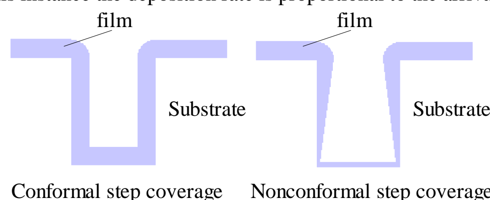
While the high temperature required for the reaction prevents its use over aluminium, it is suitable for polysilicon gates requiring a uniform insulating layer with good step coverage[4]. The oxides may be doped by adding small amounts of the dopant hydrides (phosphines, arsine, or diborane).

For high-temperature deposition (900°C),silicon dioxide is formed by reacting dichlorosilane, with nitrous oxide at reduced pressure:



$$SiCl_2H_2(g) + 2N_2O(g) \xrightarrow{900°C} SiO_2(s) + 2N_2(g) + 2HCl(g)$$

[4] Step coverage relates the surface topography of a deposited film to the various steps on the semiconductor substrate. In an ideal, or conformal, step coverage, film thicknesses are uniform along all surfaces of the step. The uniformity of film thickness, regardless of topography, is due to to the rapid migration of reactants after adsorption on the step surfaces. Nonconformal step coverage occurs when the reactants adsorb and react without significant surface migration. In this instance the deposition rate is proportional to the arrival angle of the gas molecules.



This deposition method gives excellent film uniformity and is sometimes used to deposit insulating layers over polysilicon.

Deposition methods and properties of silicon dioxide films are listed in table 1.1. In general, there is a direct correlation between deposition temperature and film quality. At higher temperatures, deposited oxide films are structurally similar to silicon dioxide that has been thermally grown.

The lower densities occur in films deposited below 500°C. Heating deposited silicon dioxide at temperatures between 600 and 1000°C causes densification, during which the oxide thickness decreases while the density increases to 2.2 g.cm^{-3}. The refractive index of silicon dioxide is 1.46, at a wavelength of 632.8 nm. Oxides with lower indices are porous, such as the oxide from the silane-oxygen deposition, which has a refractive index of 1.44. The porous nature of the oxide is also responsible for the lower dielectric strength. The etch rates of oxides in a hydrofluoric acid solution depend on deposition temperature, annealing history, and dopant concentration. Usually higher quality oxides are etched at lower rates.

Table 1-1: Properties of a-SiO_2 formed by different methods

	Method			
Property	**Thermally grown at 1000°C**	**$SiH_4 + O_2$ at 450°C**	**TEOS at 700°C**	**$SiCl_2H_2 + N_2O$ at 900°C**
Composition	SiO_2	SiO_2 (H)	SiO_2	SiO_2 (Cl)
DC resistivity (Ω -cm at 25°C)	10^{14} - 10^{16}(a)	n.a.	n.a.	n.a.
Density (g.cm^{-3})	2.2	2.1	2.2	2.2
Energy gap (eV)	~ 8 - 9	n.a.	n.a.	n.a.
Linear thermal expansion coefficient (α)	0.5×10^{-6} ppm per °C	n.a.	n.a.	n.a.
Melting point	~ 1700°C	n.a.	n.a.	n.a.
Refractive index (at optical frequencies)	1.46	1.44	n.a.	n.a.
Thermal conductivity	0.014 W/(cm. °C)	n.a.	n.a.	n.a.
Dielectric strength (10^6 V.cm^{-1})	> 10	8	10	10
Etch rate (Å /min) (100:1 H_2O:HF)	30	60	30	30
Etch rate (Å /min) (buffered HF)	440	1200	450	450
Step coverage	–	Nonconformal	Conformal	Conformal

Note. Abbreviation used: (a) Contamination may lower this value; n.a. – not available.

1.4- Infrared 'signature'

From sections 1.2 and 1.3 we saw that silicon dioxide film properties may be made to change greatly, depending on the growing process. Infrared (IR) spectroscopy is an excellent tool for the study of the oxide film characteristics. The shapes of vibrational spectra observed in IR spectra are very sensitive to the chemical bonding and nature of the atomic species interacting with electromagnetic radiation. IR spectroscopy is also very appealing because the studied sample is not destroyed.

Therefore, the derivation of IR optical equations for insulating films (on semiconductors) is of major importance. Using Fresnel formulas for wave propagation we may calculate film transmittance and reflectance as a function of its dielectric function (Chapter 2), which, in turn, may be modelled by a set of lorentzian oscillators (Chapter 3). These oscillators represent the chemical bondings vibrations.

In this section it will be given a simple overview of the vibrational energies of the SiO_2 molecule as well as its normal vibrational modes whose interpretation is still very controversial.

Even though SiO_2 is a polyatomic molecule we will start with the study of a diatomic molecule because, as we shall see, polyatomic molecules may be considered as a sum of simple oscillators if the correct axis system is chosen.

1.4.1- Vibrational energies of a diatomic molecule

1.4.1.1- Harmonic oscillator

Even though no exact mathematical expression for the potential energy curves of all molecules is known, there is a simple relation that yields curves similar to the real ones, called Morse function:

$$U(q) = D_e\left(1 - e^{-\beta q}\right)^2 \quad (1\text{-}1)$$

where q measures the distortion of the bond from its equilibrium length; D_e is the dissociation energy measured from the equilibrium position, i.e., from the minimum of the curve (figure 1.3); and β is a constant for any given molecule and may be said to determine the narrowness, or curvature, of the potential well.

A straightforward procedure for deducing the allowed vibrational energies would be to substitute the Morse function into the Schrödinger equation and look for functions that solve the resulting differential equation. Since this procedure has some mathematical difficulties, it's better to consider the following strategy: we know that the vibrations of a molecule result in only small distortions of the bond from its equilibrium position. So we are, fundamentally, only interested in an expression for the potential energy near the minimum in the potential-energy curve. This suggests that a suitable expression for the potential energy may be obtained by a series expansion about the minimum. A Maclaurin series expansion about q=0, i.e., about the equilibrium length [8], gives

$$U(q) = U_{q=0} + \left(\frac{dU}{dq}\right)_{q=0} q + \frac{1}{2!}\left(\frac{d^2U}{dq^2}\right)_{q=0} q^2 + \ldots \quad (1\text{-}2)$$

The usual choice of U = 0 at q = 0, sets $U_{q=0}$ equal to zero. Furthermore, at q = 0 the potential energy has a minimum, and therefore $(dU/dq)_{q=0}$ must be zero. If only the next higher term in the expansion is retained, we have the approximation, valid near the equilibrium position,

$$U(q) = \frac{1}{2}\left(\frac{d^2U}{dq^2}\right)_{q=0} q^2 \quad (1\text{-}3)$$

This result should be recognised as identical to that which would have been obtained if it had been assumed that a chemical bond behaves as a spring following Hooke's law, with a potential energy given by:

$$U(q) = \frac{1}{2} kq^2 \quad (1\text{-}4)$$

where k is the force constant. For a chemical bond the curve may be said to approximate a parabola only near the minimum. The force constant is then the value of d^2U/dq^2 near the minimum (figure 1.3).

When the parabolic function is used to approximate the potential energy of a chemical bond, the system is said to be treated as a harmonic oscillator.

The time-independent Schrödinger equation for a diatomic molecule with reduced mass $\mu = m_1 m_2 / (m_1 + m_2)$ that vibrates harmonically is given by the following equation:

$$-\frac{h^2}{8\pi^2\mu}\frac{d^2\psi}{dq^2} + \left(\frac{1}{2}kq^2\right)\psi = E\psi \quad (1\text{-}5)$$

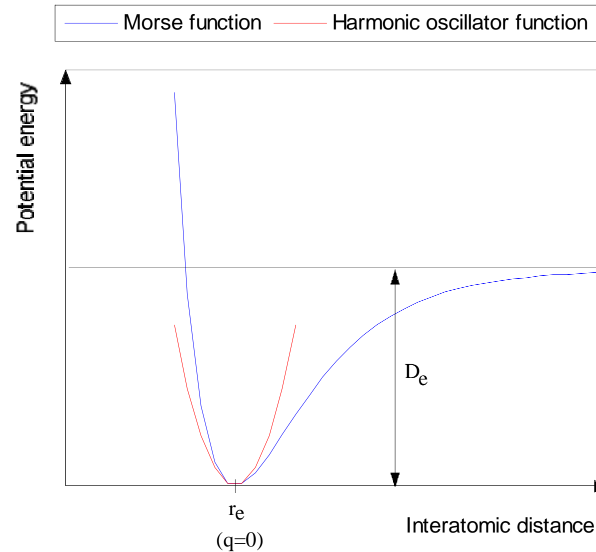


Figure 1.3: The harmonic oscillator (red line) and Morse (blue line) potential energy functions.

The wave function, ψ, only exists for certain energy values characterized by a quantum number υ which may take on the values 0, 1, 2, ...

The energies of the allowed states are given by the expression:

$$E_{\upsilon} = \left(\upsilon + \frac{1}{2}\right)\frac{h}{2\pi}\sqrt{\frac{k}{\mu}} \quad \text{(ergs)} \tag{1-6a}$$

or

$$\overline{E}_{\upsilon} = \left(\upsilon + \frac{1}{2}\right)\frac{1}{2\pi c}\sqrt{\frac{k}{\mu}} \quad (\text{cm}^{-1}) \tag{1-6b}$$

A vibrating molecule may not interact with electromagnetic radiation unless a vibrating, or oscillating, dipole moment accompanies the molecular vibration. Qualitatively, one may picture an oscillating dipole as coupling with the electric field of the radiation so that energy may be exchanged between the molecule and the radiation. This stipulation that the vibration be accompanied by a dipole-moment change implies that all homonuclear diatomic molecules, which necessarily have zero dipole moment for all bond lengths, will fail to interact with radiation and will exhibit no vibrational spectral transitions. On the other hand, a heteronuclear diatomic molecule will generally have a dipole moment and, generally, the dipole moment will depend on the internuclear distance. The vibration of a heteronuclear diatomic molecule will, therefore, be generally accompanied by an oscillating dipole moment. Such molecules may interact with radiation and may absorb energy of the radiation changing their vibrational state to one of higher energy. They may, as well, emit radiation and thereby change their vibrational state to one of lower energy.

For harmonic oscillator type systems, the vibrational quantum number may change only by one unit, i.e., $\Delta\upsilon = \pm 1$.

1.4.1.2- The anharmonicity of molecular vibrations

That the harmonic oscillator expression for potential energy is only an approximate representation of the correct potential energy function shows up in the presence of weak spectral transitions corresponding to $\Delta\upsilon = \pm 2, \pm 3, \ldots$, in violation of the $\Delta\upsilon = \pm 1$ selection rule, and in the fact that the frequencies of these overtone transitions are not exactly ± 2, ± 3,... times that of the fundamental, $\Delta\upsilon = \pm 1$, transition.

The retention of one additional term in the Maclaurin series expansion of equation (1-2) provides a better approximation to the potential energy [8]. One has, then,

$$U(q)=\frac{1}{2!}\left(\frac{d^2U}{dq^2}\right)_{q=0}q^2+\frac{1}{3!}\left(\frac{d^3U}{dq^3}\right)_{q=0}q^3 \qquad (1.7)$$

This expression for the potential energy may be used in the Schrödinger equation to deduce the energy levels of the allowed states of an anharmonic oscillator. The solution is obtained by an approximate, or perturbation, method and leads to an energy level expression that may be written as

$$\overline{E}_v=\overline{w}_e\left(\upsilon+\frac{1}{2}\right)-\overline{w}_e\overline{x}_e\left(\upsilon+\frac{1}{2}\right)^2 \quad (cm^{-1}) \qquad (1\text{-}8)$$

where $\overline{w}_e=\frac{1}{2\pi c}\sqrt{\frac{k_e}{\mu}}$ $\left[k_e=\left(\frac{d^2U}{dq^2}\right)_{q=0}\right]$ is the spacing of the energy levels, expressed in cm^{-1}, that would occur if the potential curve were a parabola with the curvature that the actual curve has at the minimum, or equilibrum, position. The coefficient $\overline{w}_e\overline{x}_e$ is known as the anharmonicity constant; it is always positive and much less than $\overline{w}_e$.

With this new approximation the selection rule $\Delta\upsilon=\pm 1$ is removed and we may have, for example, the transitions $\upsilon=0\rightarrow\upsilon=2$ or $\upsilon=0\rightarrow\upsilon=3$. The probability of these transitions is very weak and their intensity is many times negligible. So, we may consider the harmonic oscillator aproximation as a good one, despite its simplicity.

1.4.2- Vibrational energies of a polyatomic molecule

A molecule containing n atoms, with 3 cartesians coordinates used to describe each one of the n atoms, will have 3n degrees of freedom, i.e., there will be a total of 3n coordinates required for a description of the n atoms movements.

Molecular vibrations are described by 3n-6 (or 3n-5 for linear molecules) degrees of freedom because 3 degrees describe the translations and 3 (or 2) degrees describe the rotations of the molecule. We say that the molecule has 3n-6 (or 3n-5) internal degrees of freedom and 6 (or 5) external degrees of freedom [9].

The Schrödinger equation for treatment of internal motion of a molecule in terms of 3n-6 (or 3n-5) normal coordinates[5] may be deduced from the fact that the kinetic energy operator for a normal coordinate Q_i is given by an expression of the type $-h^2/8\pi^2\times\partial^2/\partial Q_i^2$.
The mass term that normally appears in the denominator of this operator is generally absorbed in the normal coordinate Q_i. The operator for the potential energy is the potential energy function itself. Since, in terms of the normal coordinates, the kinetic energy is a function of squared terms, we may write

$$-\frac{h^2}{8\pi^2}\sum_{i=1}^{3n-6}\frac{\partial^2\psi}{\partial Q_i^2}+U(Q_1,\ldots,Q_{3n-6})\psi=E\psi \qquad (1\text{-}9)$$

where the solution function ψ will, in general, be a function of the 3n-6 normal coordinates.

The potential energy is also a function of squared terms,

$$U(Q_1,Q_2,\ldots,Q_{3n-6})\psi=\frac{1}{2}\lambda_1Q_1^2+\frac{1}{2}\lambda_2Q_2^2+\ldots+\frac{1}{2}\lambda_{3n-6}Q_{3n-6}^2=\frac{1}{2}\sum_{i=1}^{3n-6}\lambda_iQ_i^2 \qquad (1\text{-}10)$$

where λ 's are coefficients that depend on the force constants of the molecule. Equation (1-9) may now be written as

[5] Normal coordinates are such that both the kinetic and potential energies of a system are expressed by the sum of terms each involving the square of a normal coordinate.

$$-\frac{h^2}{8\pi^2}\sum_{i=1}^{3n-6}\frac{\partial^2\psi}{\partial Q_i^2}+\left(\frac{1}{2}\sum_{i=1}^{3n-6}\lambda_i Q_i^2\right)\psi = E\psi \quad (1\text{-}11)$$

If we separate this equation into 3n-6 equations, each involving only one of the normal coordinates we get:

$$\psi = \psi_1\psi_2\cdots\psi_{3n-6} \quad (1\text{-}12)$$

which replaced in equation (1-11) gives 3n-6 equations of the type:

$$-\frac{h^2}{8\pi^2}\frac{\partial^2\psi_i}{\partial Q_i^2}+\left(\frac{1}{2}\lambda_i Q_i^2\right)\psi_i = E_i\psi_i \quad (1\text{-}13)$$

This equation is identical to that already encountered in the diatomic molecule oscillating harmonically. Each normal coordinate Q_i will lead to a set of energy levels with a constant spacing:

$$E_i = \left(\upsilon+\frac{1}{2}\right)\frac{h}{2\pi}\sqrt{\lambda_i} = \left(\upsilon+\frac{1}{2}\right)h\nu_i \qquad \text{(ergs)} \quad (1\text{-}14a)$$

where ν_i is the vibrational frequency of the ith oscillator, or

$$\overline{E}_i = \left(\upsilon+\frac{1}{2}\right)\frac{1}{2\pi c}\sqrt{\lambda_i} \qquad (\text{cm}^{-1}) \quad (1\text{-}14b)$$

So, a molecule containing n atoms will have 3n-6 (or 3n-5) vibrational energy-level patterns.

Since the wave functions written in terms of normal coordinates are the same as discussed in Section 1.4.1.1, the selection rules for transitions between the vibrational states for each normal coordinate are: $\Delta\upsilon = \pm 1$. We may, therefore, expect an absorption of radiation corresponding to the transition from the $\upsilon = 0$ to the $\upsilon = 1$ level of each energy-level pattern − other transitions with $\Delta\upsilon = 1$ are much more scarce because at room temperature, and according to Maxwell-Boltzmann law, the higher energy levels are much less populated.

It is worth mentioning that in general we do not know the normal coordinates of a molecule. It is, however, sufficient to know that there are coordinates in terms of which T and U are both sums of squared terms, implying that the vibrational energy of a molecule is to be treated in terms of 3n-6 (or 3n-5) energy-level patterns.

1.4.3 - Vibrational modes in a-SiO_2 films

The vibrational modes of amorphous silicon dioxide films may be characterized by oscillations in the Si-O-Si bonds. So, we will have 3 normal coordinates associated to 3 independent oscillators with frequencies ν_1, ν_2, ν_3. As the Si-O-Si bonds have no vibrational degenerescence (i.e., the number of normal vibrations is equal to the number of internal degrees of freedom), we would expect to see in a typical a-SiO_2 infrared spectrum three peaks, one for each normal frequency. Many spectra show more than three peaks, though. These additional peaks are due to the longitudinal-optic − transverse-optic (LO − TO) frequency splitting of the vibrational modes in amorphous silicon dioxide, first proposed by Galeener and Lucovsky [10].

Although the physical meaning of this frequency splitting is well understood for crystals, in the case of amorphous materials, the significance of such splitting at the atomic level is still unclear.

In crystals this splitting arise as a long-range Coulomb forces which are a consequence of the internal electric field created by the motions of the ions during the vibrations. In glasses there is also some theoretical support for this. Payne and Inkson [11] have shown that the long-range Coulomb interactions may generate a uniform effective electric field which acts as an ordering force and overcomes the effects of the disorder in the amorphous material to produce well defined longitudinal and transverse vibrational modes.

In order to explain this frequency splitting in a simple manner, we are going to consider that the macroscopic electric field inside the oxide film (or effective electric field), $\vec{E}_1$, has a plane wave solution,

$\vec{E}_1 = \vec{E}_1^0 \exp(-iwt)\exp(i\vec{k}\cdot\vec{r})$, where $|\vec{k}| = 2\pi/\lambda$ is the wave vector in the material, specifying the direction of wave propagation; and $\vec{E}_1^0$ specifies the amplitude as well as the direction of particle motion. We are also going to assume that the wavelengths of interest are long in comparison with the interatomic distance, so we may consider our material as continuos.

If we consider that the polarization is uniform within the oxide, we have that $\vec{E}_1$ is related to the external field, $\vec{E}_0$, by the relation

$$\vec{E}_1 = \vec{E}_0 + \vec{E}_d \quad (1\text{-}15)$$

where $\vec{E}_d$ is the field due to the uniform polarization. This field is called the depolarization field (or induced field) because it tends to oppose the applied field $\vec{E}_0$ (figure 1.4).

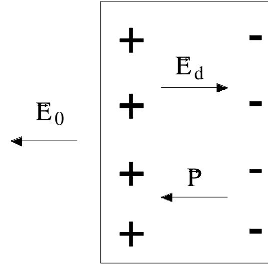


Figure 1.4: The depolarization field $\vec{E}_d$ inside the oxide film is opposite to $\vec{P}$. The fictitious surface charges (oriented dipoles) are indicated: the field of these charges is $\vec{E}_d$ within the film.

Some shapes of dielectric bodies, where the film shape is included (slab), are characterized by the following property: a uniform polarization produces a uniform depolarization field [12], that is,

$$\vec{E}_d = -\frac{L\vec{P}}{\varepsilon_0} \quad (1\text{-}16)$$

where L is the depolarization factor. Applying this equation to equation (1-15) we get

$$\vec{E}_1 = \vec{E}_0 - \frac{L\vec{P}}{\varepsilon_0} \quad (1\text{-}17)$$

For an oxide film of sufficient thickness d, such that $|\vec{k}|d >> 1$, L is given by

$$L = \begin{cases} 0, & \text{for } \vec{E}_1 \perp \vec{k} \\ 1, & \text{for } \vec{E}_1 // \vec{k} \end{cases} \quad (1\text{-}18)$$

If, on the other hand, $|\vec{k}|d << 1$, L is determined by the geometry of the material since $\vec{k}$ is perpendicular to the film surface [13]:

$$L = \begin{cases} 0, & \text{for } \vec{E}_1 \text{ parallel to the film plane} \\ 1, & \text{for } \vec{E}_1 \text{ perpendicular to the film plane} \end{cases} \quad (1\text{-}19)$$

As a result of the fact that L may take on two different values in a dielectric material, the system of electric dipole oscillators representing the film oxide may exhibit two different vibrational modes: a transverse mode ($\vec{E}_1 \perp \vec{k}$) corresponding to L=0, and a longitudinal mode ($\vec{E}_1 // \vec{k}$) corresponding to L=1.

Usually, the LO mode frequency is higher than the associated TO mode frequency because the restoring forces associated with longitudinal oscillations are greater.

Actually, equations (1-18) and (1-19) have only be proven for crystals. The “extension” of these equations to amorphous materials is not very pacific, as we shall see further below. However, the somewhat “crystal-like” behaviour of the vibrational modes of a-SiO_2 thin films suggests that this “extension” is reasonable, at least to start with.

I think it is worth mentioning that not all the amorphous materials show this frequency splitting of a vibrational mode. Only those where the disorder is not strong enough to overcome the ordering vibrational effect of the uniform field generated by the long-range forces show vibrational modes with LO-TO splitting.

From equation (1-19), LO modes are not observed in normal-incidence infrared spectra the electric field vector is always parallel to the film surface (therefore, perpendicular to vector $\vec{k}$). Figure 1.5 shows a typical normal incidence transmission spectrum for a-SiO_2 as obtained from a 342-nm-thick oxide film grown thermally at ≈ 1000°C on c-Si. Each of the three major transmission peaks may be characterized in terms of a particular vibrational mode of the oxygen atoms with respect to the silicon atom pairs which they bridge. Rocking (R) of the O atom about an axis through the two Si atoms (the oxygen atom motion is out of the plane of the Si−O−Si bond) characterizes the vibrational behaviour of the lowest-frequency peak centered at ≈ 460 cm^{-1}. Symmetrical stretching (SS) (or bending) of the O atom along a line bisecting the axis formed by the two Si atoms characterizes the vibrational mode of the middle peak centered at ≈ 810 cm^{-1}. The remaining peak and its high-frequency shoulder are due to an asymmetrical stretch (AS) motion in which the O atom moves back and forth along a line parallel to the axis through the two Si atoms. According to Kirk [14], the AS motion actually gives rise to two vibrational modes: (1) an AS_1 mode in which adjacent O atoms execute the AS motion in phase with each other, and (2) an AS_2 mode in which adjacent O atoms execute the AS motion 180° out of phase with each other. The AS_1 mode is characteristic of the vibrational behaviour of the peak centered at ≈ 1076 cm^{-1} and the AS_2 mode typifies the behaviour of the shoulder centered at ≈ 1200 cm^{-1}. However, Almeida [15], showed no evidence of this AS_2 mode in reflectance spectra. This high-frequency shoulder is not present in crystalline silicon dioxide.

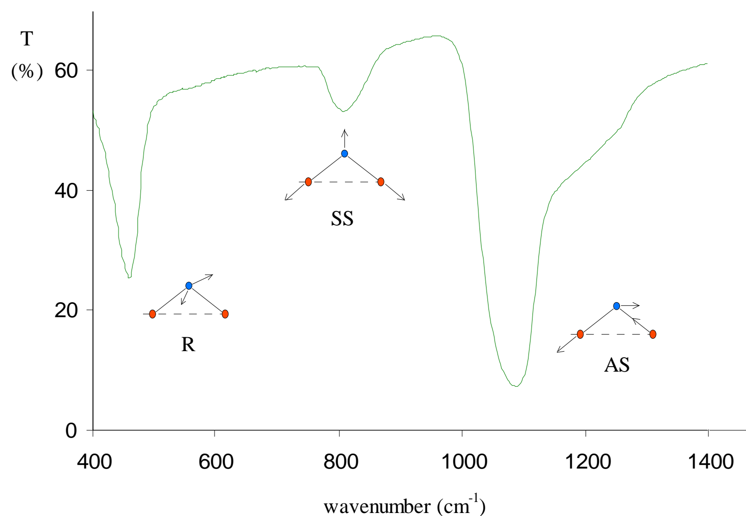


Figure 1.5: Normal incidence infrared transmission spectrum for a-SiO2 obtained by us from a 342-nm-thick oxide film grown thermally at 1000°C on a c-Si substrate showing the local vibrational motions of rocking (R), symmetrical stretching (SS), and asymmetrical stretching (AS) of the oxygen atoms (blue) with respect to the silicon atom (orange) pairs which they bridge that are associated with three major features of the transmission spectrum. The arrows represent the instantaneous distortions of atomic bonds.

We may try to shed some light into the region associated to the asymmetric stretch mode (approximately between 950 and 1300 cm^{-1}) which is so controversial, by trying to identify the number of vibrational modes present in this region, and what they represent.

The number of vibrational modes may be identified by decomposing with gaussians that spectral region. Each gaussian corresponds to a vibrational mode. A gaussian profile is a good approach because the vibrational density of states (VDOS) of a-SiO_2 have a gaussian-like continuos frequency distribution around each TO frequency (and LO frequency, as well ??).

This may be observed in the graphic of figure 1.6, taken from Ref. 14. In this graphic, Kirk [14], makes a comparison between the experimental values for the positions of the four pairs of LO-TO-mode peaks (predicted

by him) with the VDOS for a-SiO_2 as obtained by Carpenter and Price [16], from vitreous SiO_2 using coherent inelastic neutron scattering. It may be seen from the graphic that peaks in the VDOS generally tend to coincide with the peaks of the TO modes (solid vertical lines) rather than the LO-mode peaks (dashed vertical lines). This is more evident in the higher-frequency region, where there tends to be less overlapping of the VDOS bands. In fact, judging from figure 1.6, there appear to be no prominent features in the VDOS spectrum which may be associated with the LO-mode peaks.

This observation of the correspondence between the TO modes and peaks in the VDOS of a-SiO_2 agrees with theoretical results obtained by de Leeuw and Thorpe [17]. They have shown, by including long- (infinite-) range Coulomb effects in the study of the vibrational properties of a random network with AX_2 molecules, that only a single VDOS peak is associated with the splitting of a vibrational mode into a LO-TO pair, and that in the long-wavelength limit this peak corresponds closely in frequency to their calculated TO peak. Furthermore, they have also shown that in the long-wavelength limit their calculated LO peak occurs at a band edge in the VDOS. These results corroborates Kirk's theory.

Even though Kirk's theory is supported by some theorists, I think the graphic of figure 1.6 may also be reasonably interpreted by considering that the AS_2 vibrational mode does not exist, that is, by considering that the AS_1 and AS_2 modes reduce to a single AS mode. In this case, the VDOS peak previously assigned to the TO component of the AS_2 mode (TO2 in figure 1.6) should, in fact, correspond to the LO component of the AS mode, which has been predicted to lie in that wavenumber region ($\approx$1200 cm^{-1}). This different reading of the VDOS peaks may be acceptable in view of other authors' theories, as we shall see below.

The rocking and symmetric stretch (or bending) modes should also exhibit this double peak structure. However, they are not observed in the graphic of figure 1.6 because their LO-TO components are very close to each other, and their superposition makes the two peaks look like only one peak.

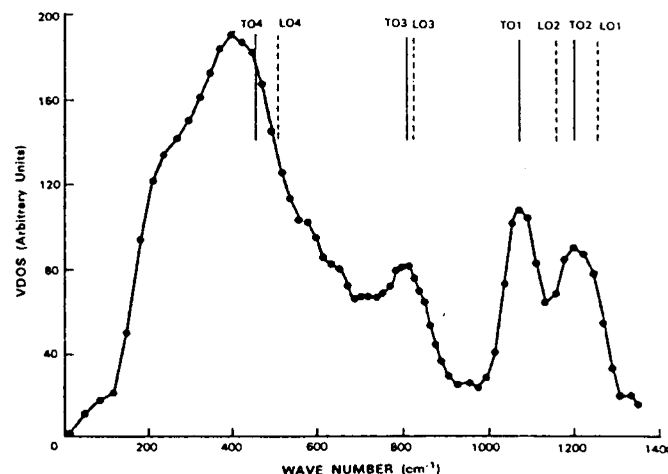


Figure 1.6: Comparison of experimental values for the four pairs of LO-TO-mode peaks of a-SiO_2 (predicted by Kirk [14]) with the VDOS spectrum for a-SiO_2 as obtained by Carpenter and Price [16]. The positions of the LO peaks are indicated by vertical dashed lines, the TO peaks by vertical solid lines, and the experimental values of the VDOS spectral plot by solid circles (•) joined by straight lines to guide the eye.

For normal incidence transmission spectra of very thin films (below $\approx$100 nm) where the high-wavenumber shoulder has not been observed, the region of interest has been fitted with two gaussians by Boyd [18]. According to this author, this double feature may be the evidence for some kind of localized order within the amorphous structure of silicon dioxide films, disclaiming the generally accepted continuos random network (CRN) arrangement, since it only predicts one peak in this wavenumber region, while some models based on microcrystalline or quasiperiodic structures do predict the two peaks.

As the thickness increases the shoulder appears and grows in intensity, so a third gaussian must be added [5].

This growth has several explanations. For example, according to Kirk [14], this results from a growth in the optical strength of the AS_2 mode due to an increasing disorder-induced mode coupling between the AS_1 and AS_2 modes, as the film thickens. For very thin films the AS_2 mode does not appear because the mode coupling is very reduced. This theory is confirmed by Lange [19], concluding that this mode coupling act as a "disorder" or "amorphicity" indicator of the a-SiO_2 film.

Devine [20], refutes this theory, saying that this effect results from the inclusion of wavenumber dependent terms. As a consequence, he also explains, the intensity of this shoulder may not be used neither as a measure of the "amorphicity" of the film (otherwise the transition region would not be ≈3 nm thick, as proposed by Lewis and Irene [7], but, at least, ≈600 nm thick) nor as a comparison between amorphous films.

In turn, Almeida and Pantano [21], suggested that this shoulder is a broadened signature of the LO mode (associated to the asymmetric stretch TO mode), that is, a mixed LO/TO mode of dominant LO character. The shoulder wavenumber falls in the LO wavenumber range and it was observed that as the incidence angle for transmission measurements changes from 0° (normal incidence) to 80° off normal, the high-wavenumber shoulder sharpens its intensity and frequency increase.

This theory also explains the evolution of the shoulder intensity with film thickness. For very thin films, the long-range Coulomb interactions are unlikely to be effective along a direction perpendicular to such thin films, thus preventing the observation of the LO mode.

So, to explain the presence of a LO mode, normally forbidden in normal incidence, Almeida and Pantano [21], as well as Goullet *et al.* [5], proposed that in amorphous materials there are no pure TO and LO modes, due to the long-range structural disorder effects on the effective electric field produced by long-range Coulomb interactions responsible for the TO-LO splittings. This fact may explain why the Raman peak positions of amorphous silicon dioxide, within the experimental error, do not appear to coincide with either the TO or LO components, except in the case of the bending mode where the coincidence might be fortuitous [15]. (The agreement is especially poor in the LO case.) In this case, equation (1-19) no longer holds.

The fact of having assigned a LO mode to the high-wavenumber shoulder agrees with the above theory that LO modes should also exhibit peaks in the VDOS spectrum of figure 1.6.

For a thermally grown oxide, the transmission peak minimum associated with the asymmetrical stretching mode has a position (in wavenumbers) which is film thickness dependent – the other peaks, that is, the transmission peaks associated to the rocking mode and to the bending mode are essentially thickness independent. Devine [20] observed this shift for absorption spectra. His results are reproduced in figure 1.7. From the graphic we see that as the thermal oxide thickens, the position of the peak maximum shifts from around 1075 cm^{-1} for very thin films (≈ 20 nm) to around 1090 cm^{-1} in thicker films (≈ 1000 nm). The latter value is comparable with that measured in bulk, amorphous silica. The reason for this shift is still open to discussion. However, I think it can not be due to any kind of "odd" strain that doesn't exist in the bulk material, because, as mentioned in section 1.2, only a transition region with approximately 3 nm into the film has a distribution of tetrahedral rings different from the bulk (strained region).

Even though the graphic of figure 1.7 suggests that below ≈ 100 nm there is no significant shift, Boyd [18] has verified that this shift also occurs for films with thicknesses in the range 2.8 - 45.0 nm.

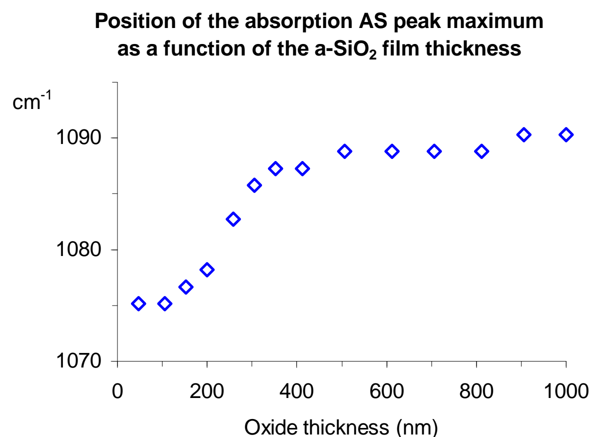


Figure 1.7: Shift of the absorption AS peak maximum with film thickness [20].

Whereas for normal incidence spectra the presence of LO modes is controversial, for obliquely incident spectra their presence is quite expected. So, through oblique incidence reflectance spectra, we may verify if the high-frequency shoulder observed in the transmittance spectra corresponds to a LO mode or, on the contrary, corresponds to the AS_2 mode. The existence of the AS_2 mode implies the presence of two LO peaks in the interval [950, 1300] cm^{-1}. To verify this presence we need a spectrum with essentially no TO peaks, which may be accomplished using the Berreman effect. This effect is based on equation (1-19). For very thin films, the

wavevector, $\vec{k}$, of the effective electric field inside the film is perpendicular to the film surface. By using the p-polarized component (parallel to the plane of incidence) of oblique incident radiation which has subcomponents parallel and perpendicular to the film surface, both TO and LO vibrational modes may be excited, respectively, as shown schematically in figure 1.8 (a).

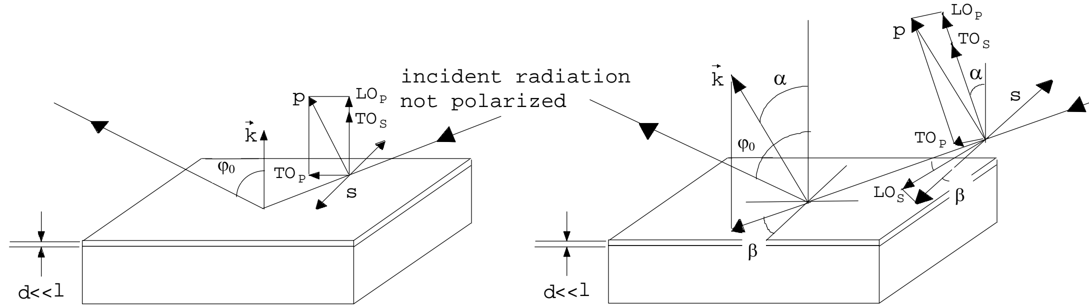


Figure 1.8: Excitation of TO and LO modes by the s (perpendicular) and p (parallel) components of radiation striking the surface of a thin film with an angle φ_0 (a) considering that LO modes are not observed in normal incidence spectra - Kirk's model and; (b) considering that LO modes are observed in normal incidence spectra, since the s component also excites these modes - Almeida's and Goullet *et al.*'s model. The subscripts in TO and LO give the radiation component which excites these modes. In both cases, film thickness, d, is much smaller than the wavelength, λ, of the incident radiation.

This effect may be enhanced by:

(i) removing with a linear polarizer the s-polarized component (perpendicular to the plane of incidence), whose spectra show the same peaks as the transmittance spectra obtained with normally incident unpolarized light. These peaks mask the LO components obtained with p-polarized light;

(ii) covering the side of the wafer opposite to the oxide with a metal. This may be arranged by the deposition of a metallic film or by mounting the oxidized wafer on a metal mirror, providing that the air gap is much smaller than the wavelength of the radiation used. Therefore, very plane mirrors and samples are necessary as well as dustfree handling [3]. The metal practically eliminates the absorption by transverse modes without reducing the absorption band associated with the longitudinal modes. The reason for this is that almost no electric field may exist adjacent and parallel to a metalic surface;

(iii) making the infrared light strike the film with a high angle of incidence, φ_0. From figure 1.8(a) we easily see that the higher this angle of incidence, the bigger the subcomponent parallel to the film surface, and smaller the subcomponent perpendicular to that surface, that is, the bigger the LO subcomponent, and smaller the TO component.

It was mentioned that if LO modes are observed for normal incidence spectra, equation (1-19) no longer holds, which means that the wavevector, $\vec{k}$, of the effective electric field inside the film is no longer perpendicular to the film surface but makes an angle α with the normal α [figure 1.8(b)]. In this case, both s and p components are capable of exciting LO modes.

Nevertheless, the Berreman effect still permits to obtain LO spectra since if we assume that the angle α is small, practically all the TO_p component disappears due to the metal back.

Table 1.2 gives the approximate values of the TO and LO vibrational modes wavenumbers for an a-SiO_2 film thermally grown.

Even though the existence of the AS_2 mode is still open to discussion, it was included in this table so that we may eventually compare its values with our own experimental data.

The AS_2 -mode LO-TO frequency splitting is inverted as a result of lying between the AS_1 -mode LO-TO pair frequencies.

Table 1-2: TO and LO wavenumbers

Oscillator mode j	$\bar{\nu}_{Tj}$ (cm^{-1})	$\bar{\nu}_{Lj}$ (cm^{-1})
1 (R)	457	507
2 (SS)	810	820
3 (AS_1)/(AS)	≈ 1076	≈ 1256/≈ 1200

4 (AS_2)	≈ 1200	≈ 1160

With gaussian decomposition of a spectral region we may identify the number of vibrational modes, their central frequencies and obtain a measure of their strength and dispersion. However, this it is not sufficient to fully characterize a vibrational mode since it “says” nothing about, for example, the reduced mass or the effective charge of the dipole responsible for that vibrational mode.

An analysis tool that permits to find these and other relevant parameters is called dispersion analysis, which is based on the dielectric dispersion theory (Chapter 3) applied to the Fresnel equations of wave propagation (Chapter 2). With dispersion analysis, the transmittance and reflectance spectra may be related directly to the physical properties of the chemical bonds.

Implicit in dispersion analysis is the determination of the a-SiO_2 frequency dependent complex dielectric function and complex refractive index defined, respectively, by $\tilde{\varepsilon}_r = \varepsilon_1 + i\varepsilon_2$, and $\tilde{n}$ =n+ik. From the obtained values for ε_1 and ε_2, the TO and LO modes frequencies may be determined since, peaks in ε_2 reveal transverse modes, while peaks in $\mathrm{Im}(-1/\tilde{\varepsilon}_r) = \varepsilon_2 / (\varepsilon_1^2 + \varepsilon_2^2)$ identify longitudinal modes. From the obtained values for n and k we may determine, respectively, the wave velocity (v = c/n) and the absorption in the oxide film.

Both Gaussian decomposition and dispersion analysis are accomplished by a nonlinear least-squares curve fitting of infrared spectra. We will return to this subject in Chapter 4.

References

[1] MURARK, Shyman P. and PECKERA, Martin C., Electronic Materials, Science and Technology

[2] MILLMAN, J., GRABEL, A., Microelectronics, 2nd Edn, Mc Graw-Hill (1987)

[3] GROSSE, P., HARBECKE, B., HEINZ, and MEYER R.; OFFENBERG, M., Appl. Phys. A 39, 257-268 (1986)

[4] SZE, S. M., Semiconductor Devices, Physics and Technology, John Wiley & Sons (1985)

[5] GOULLET, A. , CHARLES, C. , GARCIA, P. , and TURBAN, G. , J. Appl. Phys. 74, 6876 (1993)

[6] MARTINS, R., Folhas de apoio à cadeira de Microelectrónica

[7] LEWIS, E. A. and IRENE, E. A., J. Vac. Sci. Technol. A4, 916 (1986)

[8] BARROW, G. M., Introduction to Molecular Spectroscopy, Mc Graw-Hill (1964)

[9] Spectrometries Infrarouge a Transformee de Fourier et Raman, Université de Bordeaux I, Laboratoire de Spectroscopie Moléculaire et Cristalline, C.N.R.S. URA 124 (1991)

[10] GALEENER, F. L., LUCOVSKY, G., Phys. Rev. Lett. 37, 1474 (1976)

[11] PAYNE, M. C. and INKSON, J. C., J. Non-Cryst. Solids 68, 351 (1984)

[12] KITTEL, C., Introduction to Solid State Physics, 5th Edn, John Wiley & Sons (1976)

[13] BERREMAN, D. W., Phys. Rev. 130, 2193 (1963)

[14] KIRK, C.T., Phys. Rev. B 38, 1255 (1988)

[15] ALMEIDA, R. M., Phys. Rev. B 45, 161 (1992)

[16] CARPENTER, J. M. and PRICE, D. L., Phys. Rev. Lett. 54, 441 (1985)

[17] de LEEUW, S. W. and THORPE, M. F., Phys. Rev. Lett. 55, 2879 (1985)

[18] BOYD, I. W., Appl. Phys. Lett. 51, 418 (1987)

[19] LANGE, P., J. Appl. Phys. 66, 201 (1989)

[20] DEVINE, R. A., Internal report, C.N.R.S.

[21] ALMEIDA, R. M. and PANTANO, C. G., J. Appl. Phys. 68, 4225 (1990)

2 - Thin film optics

2.1- Introduction

Using the Fresnel theory of wave propagation, we may calculate the expressions that relate the transmitted and reflected intensities by a single film, with its dielectric function. Even though the silicon substrate has a finite thickness, it will be not treated as a layer, because, as we shall see in Chapter 6, its spectrum is deleted from the film-substrate spectrum. So, the optical properties may be described by the model of figure 2.1, where the silicon substrate is considered as semi-infinite.

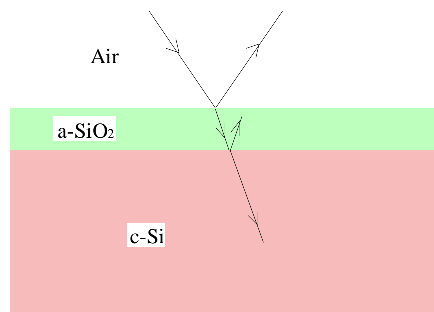


Figure 2.1- Model used to simulate the oxidized wafer.

2.2- Wave propagation in a lossy dielectric

In order to study the wave propagation in a dielectric we are going to consider a sinusoidal travelling plane wave. This simple representation is valid since:

(i) an arbitrary wave may always be represented as a spectrum of plane waves;

(ii) any periodic wave may be represented as a Fourier series of sinusoids and any nonperiodic wave may be represented as a continuos spectrum of harmonics by the Fourier integral.

There are two types of loss mechanisms which attenuate a wave in a dielectric. The first arises when the dielectric is slightly conducting. In the case of a-SiO_2, this mechanism may be neglected because its conductivity is very low. The second arises when energy is dissipated in the course of the polarization process even though the conductivity of the dielectric is zero. The dipoles experience friction as they flip back and forth in a sinusoidal field thus extracting energy from it. Because of such friction (polarization damping forces), the polarization vector $\vec{P}$ will lag behind the applied field $\vec{E}$. The phase difference between $\vec{P}$ and $\vec{E}$ is accounted for by a complex permittivity $\tilde{\varepsilon} = \varepsilon_0 \tilde{\varepsilon}_r$, where ε_0 is the permitivitty in vacuum and $\tilde{\varepsilon}_r$ is the dielectric (a-SiO_2) relative permittivity: $\tilde{\varepsilon}_r = \varepsilon_1 + i\varepsilon_2$ [6].

For a plane wave entering the plane boundary of an isotropic, absorbing medium, the planes of equal phase of the wave in the medium are perpendicular to the direction of propagation. Since the reduction in amplitude of the wave in the medium depends directly on the distance travelled in that medium, the points of equal amplitude will define planes parallel to the surface of separation (figure 2.2) [1].

[6] If we had not neglected the dielectric conductivity, we would have got a permittivity defined as follows: $\tilde{\varepsilon}_r = \varepsilon_1 + i\varepsilon_2 + i\dfrac{\sigma}{\varepsilon_0 w}$, where σ is the conductivity of the dielectric and w is the frequency of oscillation of the wave.

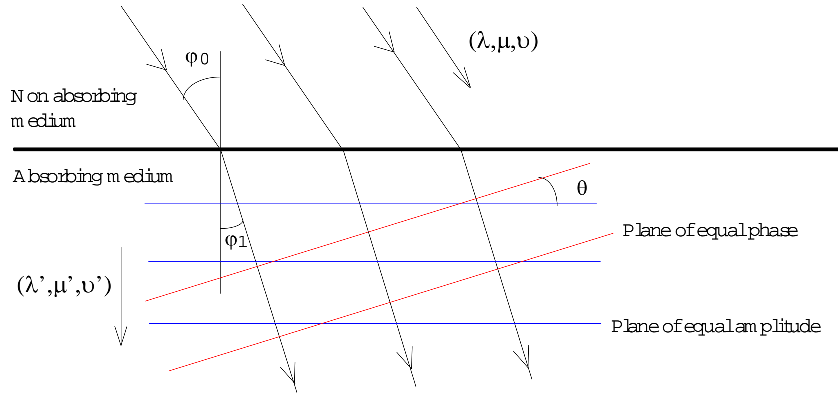


Figure 2.2- Propagation of a plane wave after being refracted into an isotropic, absorbing medium.

Thus if we consider a wave of frequency w travelling in the direction (λ, μ, υ) (see figure 2.2) in a transparent medium of refractive index n, the electric vector may be written as,

$$E = E_0 \exp\left\{-iw\left[t - \frac{n(\lambda x + \mu y + \upsilon z)}{c}\right]\right\} \tag{2-1}$$

where c is the velocity of light in vacuum.

The corresponding expression in an absorbing medium becomes [1],

$$E = E_0 \exp\left\{-iw\left[t - \frac{\alpha(\lambda x + \mu y + \upsilon z)}{c} - i\frac{\beta(\lambda' x + \mu' y + \upsilon' z)}{c}\right]\right\} \tag{2-2}$$

where $(\lambda', \mu', \upsilon')$ is the direction of maximum damping – i.e. the direction of the normal to the planes of equal amplitude.

For normal incidence, the expression for the wave may be written as

$$E = E_0 \exp\left\{-iw\left[t - \frac{(n + ik)(\lambda x + \mu y + \upsilon z)}{c}\right]\right\} \tag{2-3}$$

since, for this case, the direction of maximum damping coincides with the direction of propagation.

In equation (2-3), n is the ratio of the velocity of the wave in vacuum to that of the wave in the medium. k is known as the extinction coefficient and represents the energy absorption: the attenuation of the amplitude of the wave for a path of one vacuum wavelength in the medium is $\exp(-2\pi k)$.

If we now go to the general case of oblique incidence [equation (2-2)], we get that the values of α and β depend on the direction of propagation in the medium and hence on the angle of incidence. If the angle of incidence is φ_0 and the angle between the planes of constant phase and those of constant amplitude is θ, then it follows directly from the wave equation that [1],

$$\alpha^2 - \beta^2 = n^2 - k^2 \tag{2-4}$$

$$\alpha\beta\cos(\theta) = nk \tag{2-5}$$

$$\sin(\varphi_0) = \alpha\sin(\theta) \tag{2-6}$$

The similarity between equations (2-1) and (2-3), for waves travelling in transparent and absorbing media, respectively, has resulted in (n+ik) being termed the 'complex refractive index':

$$\tilde{n} = n + ik \tag{2-7}$$

This, in fact, suggests that in order to calculate the Fresnel equations for the transmitted and reflected intensities by an absorbing layer, we may start by assuming the layer is transparent with a real refractive index, n, and only after the equations are fully developed make the substitution of n by $\tilde{n}$. The final equations thus obtained are correct, and the mathematical difficulties implicit in equations (2-2) and (2-3) are avoided. This procedure will be adopted throughout this Chapter.

2.3- Reflection and Transmission of light at the surface of a transparent medium

For an isotropic medium, the laws of electromagnetism are given by the following equations:

$$\nabla\cdot\vec{D} = \varepsilon\nabla\cdot\vec{E} = \rho \tag{2-8}$$

$$\nabla\cdot\vec{B} = \mu\nabla\cdot\vec{H} = 0 \tag{2-9}$$

$$\nabla\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} = -\mu\frac{\partial\vec{H}}{\partial t} \tag{2-10}$$

$$\nabla\times\vec{H} = \vec{J} + \frac{\partial\vec{D}}{\partial t} = \sigma_T\vec{E} + \varepsilon\frac{\partial\vec{E}}{\partial t} \tag{2-11}$$

where the symbols have their usual meanings except σ_T, which takes into account the two loss mechanisms in a dielectric. For a medium in which there is no space charge, equations (2-8) to (2-10) may be rewritten as follows:

$$\nabla^2\vec{E} = \mu\varepsilon\frac{\partial^2\vec{E}}{\partial t^2} + \mu\sigma\frac{\partial\vec{E}}{\partial t} \tag{2-12}$$

$$\nabla^2\vec{H} = \mu\varepsilon\frac{\partial^2\vec{H}}{\partial t^2} + \mu\sigma\frac{\partial\vec{H}}{\partial t} \tag{2-13}$$

which represent the propagation of electromagnetic waves in an absorbing medium. For propagation in a transparent medium $(\sigma = 0)$, the above equations reduce to

$$\nabla^2\vec{E} = \mu\varepsilon\frac{\partial^2\vec{E}}{\partial t^2} \tag{2-14}$$

$$\nabla^2\vec{H} = \mu\varepsilon\frac{\partial^2\vec{H}}{\partial t^2}, \tag{2-15}$$

the well known simple form of wave equation, which shows that electromagnetic waves travel with velocity: $v = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{1}{\sqrt{\mu_r\varepsilon_r}}\frac{1}{\sqrt{\mu_0\varepsilon_0}} = \frac{c}{\sqrt{\mu_r\varepsilon_r}} = \frac{c}{n}$. If we consider a non-magnetic material then its relative permeability equals one $(\mu_r = 1)$, consequently,

$$n = \sqrt{\varepsilon_r} \tag{2-16}$$

The problem of determining the light reflected and transmitted at a boundary separating two media with different optical properties is dealt with by applying boundary conditions to the solutions of Maxwell's equations. These require that the tangential components of both electric and magnetic vectors be continuos at that boundary.

We will consider a plane wave incident on the surface z=0, the plane of incidence being the plane x0z, the angle of incidence φ_0 and the angle of refraction φ_1 [figure 2.3 (a)].

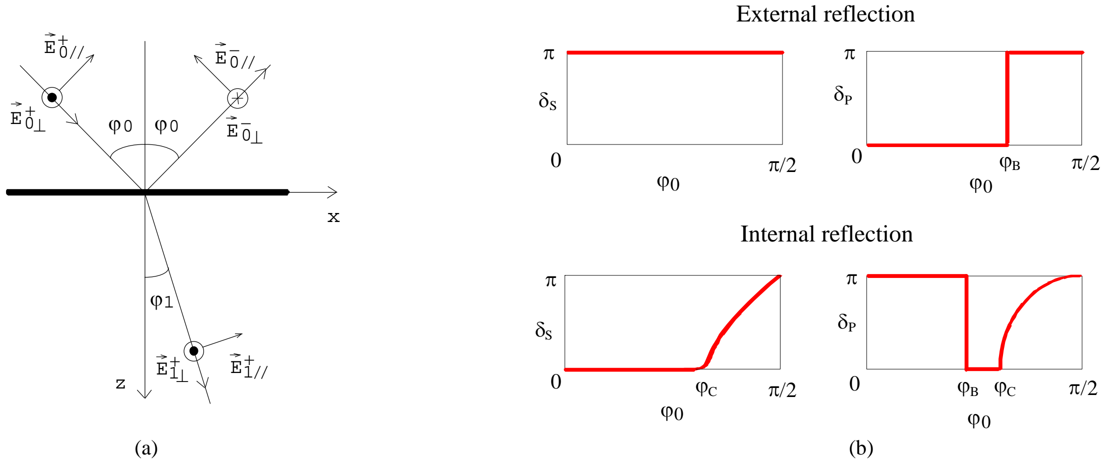


Figure 2.3: (a) An example of light reflected and transmitted at a boundary separating two media. (b) Phase changes for external , or rare to dense, reflection and internal, or dense-to-rare, reflection. δ_S stands for phase change of the s-polarized light, while δ_P stands for phase change of the p-polarized light. φ_B stands for Brewster's, or polarizing, angle and φ_C stands for critical angle.

We denote the amplitudes of the electric field vector of the wave aproaching the surface by $E^+_{0//}$ and $E^+_{0\perp}$ for the two components (the subscript // refers to the component parallel to the plane of incidence while the subscript ⊥ refers to the component normal to the plane of incidence). The reflected wave components are denoted by $E^-_{0//}$ and $E^-_{0\perp}$ and the transmitted wave components by $E^+_{1//}$, $E^+_{1\perp}$. The phase factors associated with the incident and the reflected waves are, respectively, of the form

$$\exp\left\{-i\left[wt-\frac{2\pi n_0 x\sin(\varphi_0)}{\lambda}-\frac{2\pi n_0 z\cos(\varphi_0)}{\lambda}\right]\right\}$$

$$\exp\left\{-i\left[wt-\frac{2\pi n_0 x\sin(\varphi_0)}{\lambda}+\frac{2\pi n_0 z\cos(\varphi_0)}{\lambda}\right]\right\}$$

whilst that for the transmitted wave is

$$\exp\left\{-i\left[wt-\frac{2\pi n_1 x\sin(\varphi_1)}{\lambda}-\frac{2\pi n_1 z\cos(\varphi_1)}{\lambda}\right]\right\}$$

where λ is the wavelength in vacuum.

At the boundary (z=0), and considering that the point of incidence is the origin of co-ordinates, we have that the components of the electric and magnetic field vectors in the x- and y- directions are given by equations (2-17) and (2-18), respectively, for mediounm 0 and medium 1,

$$\left.\begin{aligned} E_{0x} &= \left(E^+_{0//}+E^-_{0//}\right)\cos(\varphi_0)\\ E_{0y} &= E^+_{0\perp}+E^-_{0\perp}\\ H_{0x} &= n_0\left(-E^+_{0\perp}+E^-_{0\perp}\right)\cos(\varphi_0)\\ H_{0y} &= n_0\left(E^+_{0//}-E^-_{0//}\right)\end{aligned}\right\} \qquad (2\text{-}17)$$

$$\left.\begin{aligned} E_{1x} &= E_{1//}^{+}\cos(\varphi_1) \\ E_{1y} &= E_{1\perp}^{+} \\ H_{1x} &= -n_1 E_{1\perp}^{+}\cos(\varphi_1) \\ H_{1y} &= n_1 E_{1//}^{+} \end{aligned}\right\} \tag{2-18}$$

Applying the boundary conditions, we obtain equantions that may be solved to give the amplitudes of the transmitted and reflected vectors in terms of those of the incident vectors. We obtain,

$$\frac{E_{0//}^{-}}{E_{0//}^{+}} = \frac{n_0\cos(\varphi_1) - n_1\cos(\varphi_0)}{n_0\cos(\varphi_1) + n_1\cos(\varphi_0)} = r_{1//} \tag{2-19}$$

$$\frac{E_{1//}^{+}}{E_{0//}^{+}} = \frac{2n_0\cos(\varphi_0)}{n_0\cos(\varphi_1) + n_1\cos(\varphi_0)} = t_{1//} \tag{2-20}$$

$$\frac{E_{0\perp}^{-}}{E_{0\perp}^{+}} = \frac{n_0\cos(\varphi_0) - n_1\cos(\varphi_1)}{n_0\cos(\varphi_0) + n_1\cos(\varphi_1)} = r_{1\perp} \tag{2-21}$$

$$\frac{E_{1\perp}^{+}}{E_{0\perp}^{+}} = \frac{2n_0\cos(\varphi_0)}{n_0\cos(\varphi_0) + n_1\cos(\varphi_1)} = t_{1\perp} \tag{2-22}$$

$r_{1//}$, $r_{1\perp}$ are known as the Fresnel reflection coefficients and $t_{1//}$, $t_{1\perp}$ as the Fresnel transmission coefficients.

2.4- Reflection and transmission of light at the surface of an absorbing medium

The equations of propagation of light in a transparent medium may, as already mentioned, be applied to an absorbing medium by replacing the real refractive index by a complex term. The expressions for the Fresnel coefficients [equations (2-19) to (2-22)] also become complex. They may be written in the following form,

$$t_{1//} = \tau_{1//} e^{i\kappa_{//1}} \qquad t_{1\perp} = \tau_{1\perp} e^{i\kappa_{1\perp}} \tag{2-23}$$

$$r_{1//} = \rho_{1//} e^{i\phi_{1//}} \qquad r_{1\perp} = \rho_{1\perp} e^{i\phi_{1\perp}} \tag{2-24}$$

where $\tau_{1//}, \tau_{1\perp}$, $\rho_{1//}$ and $\rho_{1\perp}$ represent the (real) amplitudes of the transmitted and reflected waves for unit incident amplitude and $\kappa_{1//}$, $\kappa_{1\perp}$, $\phi_{1//}$ and $\phi_{1\perp}$ are the phase changes at the surface.

2.5- Reflection and transmission of light by a single film

We may apply the results of section 2.3 to the determination of the reflection and transmission coefficients for a single non-absorbing layer, bounded on either side by semi-infinite non-absorbing layers. We do this by considering a beam incident on the film which is divided into reflected and transmitted parts. Such division occurs each time the beam strikes an interface so that the transmitted and reflected beams are obtained by summing all the transmitted and reflected beams, respectivelly. When we are only considering one layer, the summation may be easily made. The results are conveniently expressed in terms of the Fresnel coefficients.

Suppose that a parallel beam of light of unit amplitude and of wavelength λ falls on a plane, parallel-sided, homogeneous, isotropic film of thickness d and refractive index n_1 supported on a substrate of index n_2 (figure 2.4) [1]. The refractive index of the first medium is n_0 and the angle of incidence is φ_0.

We may write down the amplitudes of the successively reflected and transmitted beams in terms of the Fresnel coefficients, given by equations (2-19) to (2-22). From the definitions of these coefficients, we see that they depend, for a given boundary, on the direction of propagation of light across the boundary.

In treating the problem of the single layer, we shall denote the Fresnel coefficients for propagation from medium 0 to medium 1 by r_1 and t_1 as given by (2-19) - (2-22). The corresponding coefficients for propagation from medium 1 to medium 0 will be written r_1' and t_1'. At last, the Fresnel coefficients for propagation from medium 1 to medium 2 will be denoted by r_2 and t_2. The expressions given below will be valid for either direction of polarization provided that r and t are given the appropriate values from equations (2-19) to (2-22). So the suffix // or $\perp$ will be omitted.

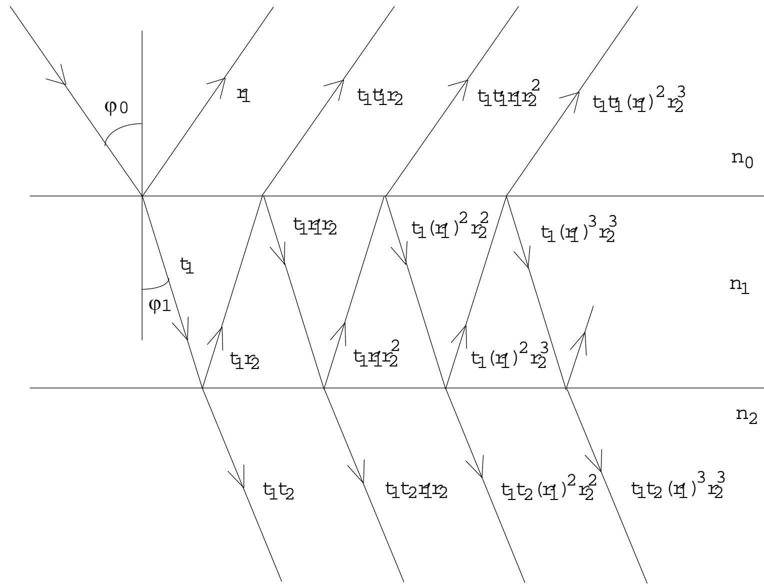


Figure 2.4: Amplitudes of the successively reflected and transmitted beams in terms of the Fresnel coefficients.

From figure 2.4 we get that the amplitudes of the successive beams reflected into medium 0 are thus given by r_1, $t_1t_1'r_2$, $t_1t_1'r_1'r_2^2$, $t_1t_1'(r_1')^2r_2^3$, ... and the transmitted amplitudes into medium 2 by t_1t_2, $t_1t_2r_1'r_2$, $t_1t_2(r_1')^2r_2^2$, $t_1t_2(r_1')^3r_2^3$, ... From equations (2-19) and (2-21) we get that r_1' is equal to $-r_1$. So the reflected and transmitted amplitudes may be written, respectively, as r_1, $t_1t_1'r_2$, $-t_1t_1'r_1r_2^2$, $t_1t_1'r_1^2r_2^3$, ..., and t_1t_2, $-t_1t_2r_1r_2$, $t_1t_2r_1^2r_2^2$, ... Writing $\delta = \frac{2\pi d}{\lambda}$ as the film thickness normalised to the wavelength, we have that $\delta n_1 \cos(\varphi_1)$ gives the beam phase change on traversing the film. The total reflected amplitude, r, will, therefore, be given by,

$$r = r_1 + t_1t_1'r_2 \exp(2i\delta n_1 \cos\varphi_1) - t_1t_1'r_1r_2^2 \exp(4i\delta n_1 \cos\varphi_1) + \ldots$$

This sum represents a converging geometrical series[7] (except for the first term). So, r may be written in a more compact form,

$$r = r_1 + \frac{t_1t_1'r_2 \exp(2i\delta n_1 \cos\varphi_1)}{1 + r_1r_2 \exp(2i\delta n_1 \cos\varphi_1)} \tag{2-25}$$

[7] Geometrical series: $\sum_{n=1}^{\infty} a_1 r^{n-1}$. $|r| < 1 \Rightarrow S = \frac{a_1}{1-r}$ (the series converges).

where the time-dependent factor is omitted. From equations (2-19) to (2-22) we have that: $t_1 t_1' = 1 - r_1^2$, so we may simplify equation (2-25):

$$r = \frac{r_1 + r_2 \exp(2i\delta n_1 \cos\varphi_1)}{1 + r_1 r_2 \exp(2i\delta n_1 \cos\varphi_1)} = \frac{r_1 + r_2 \exp(2i\delta\Gamma)}{1 + r_1 r_2 \exp(2i\delta\Gamma)} \qquad (2\text{-}26)$$

where $\Gamma = n_1 \cos\varphi_1$.

For the total transmitted amplitude, t, we have

$$t = t_1 t_2 \exp(i\delta n_1 \cos\varphi_1) - t_1 t_2 r_1 r_2 \exp(3i\delta n_1 \cos\varphi_1) + t_1 t_2 r_1^2 r_2^2 \exp(5i\delta n_1 \cos\varphi_1) - \ldots$$

This sum is also a converging geometrical series, therefore,

$$t = \frac{t_1 t_2 \exp(i\delta n_1 \cos\varphi_1)}{1 + r_1 r_2 \exp(2i\delta n_1 \cos\varphi_1)} = \frac{t_1 t_2 \exp(i\delta\Gamma)}{1 + r_1 r_2 \exp(2i\delta\Gamma)} \qquad (2\text{-}27)$$

Equations (2-26) and (2-27) are generally valid. For non-normal incidence, each takes two possible forms, depending on the state of polarization of the incident light. For light polarized with its electric field vector parallel to the plane of incidence, the reflected and transmitted amplitudes are obtained replacing r_1, r_2, t_1 and t_2 from expressions corresponding to equations (2-19) and (2-20). For light polarized with the electric field vector perpendicular to the plane of incidence, the Fresnel coefficients as given by equations (2-21) and (2-22) are used.

In order to compute the reflectance , R, and the transmittance, T, we are going to calculate the energy in each medium, which is represented by the Poynting vector $\vec{S}$ [2]:

$$\vec{S} = \vec{E} \times \vec{H} = \sqrt{\frac{\varepsilon}{\mu}} \left|\vec{E}\right|^2 \hat{u}_S, \quad (\mathrm{W/m^2}) \qquad (2\text{-}28)$$

The irradiance I is defined as the average power that crosses an unit area normal to $\vec{S}$:

$$I = \left\langle \vec{S} \right\rangle = \frac{1}{2} \sqrt{\frac{\varepsilon}{\mu}} \left|\vec{E}\right|^2. \qquad (2\text{-}29)$$

For oblique incidence of a beam on the film surface we consider that I_i, I_r, and $I_{t\text{-film}}$ give the incident, reflected and transmitted irradiances, respectively. The corresponding beams cross sections (figure 2.5a) are $A\cos(\varphi_0)$, $A\cos(\varphi_0)$, and $A\cos(\varphi_1)$, respectively, where A is the illuminated surface area. The incident power is, therefore, $I_i A\cos(\varphi_0)$. Analogously, $I_r A\cos(\varphi_0)$ gives the power of the reflected beam, and $I_{t-\mathrm{film}} A\cos(\varphi_1)$ gives the power of the transmitted beam into the film. The power transmitted to the substrate (figure 2.5b) is $I_t A\cos(\varphi_2)$, where I_t gives the transmitted irradiance to medium 2.

Film reflectance, R, is defined as the ratio of reflected to incident power:

$$R \equiv \frac{I_r A\cos(\varphi_0)}{I_i A\cos(\varphi_0)} = \frac{I_r}{I_i} = \frac{\frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left|E_0^-\right|^2}{\frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left|E_0^+\right|^2} = \left|\frac{E_0^-}{E_0^+}\right|^2 = |r|^2 \qquad (2\text{-}30)$$

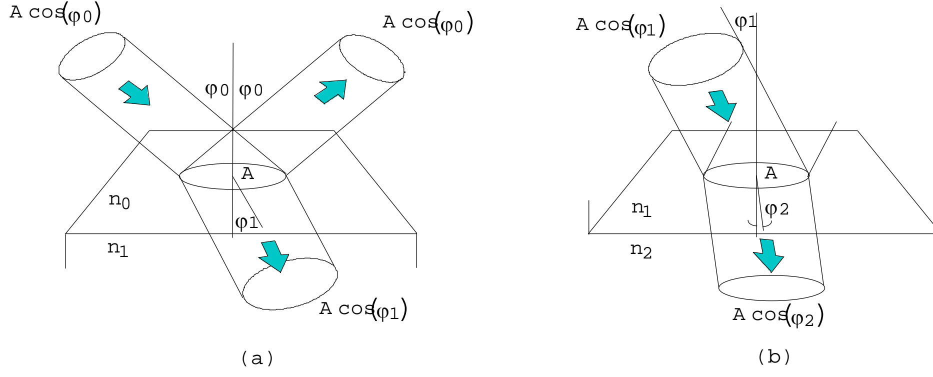


Figure 2.5: Cross sections of the incident, reflected and transmitted beams at the (a) interface medium 0 - medium 1 (air - film) and; (b) interface medium 1 - medium 2 (film - substrate).

Film transmittance, T, is defined as the ratio of transmitted to incident power:

$$T \equiv \frac{I_t A\cos(\varphi_2)}{I_i A\cos(\varphi_0)} = \frac{\frac{1}{2}\sqrt{\frac{\varepsilon_2}{\mu_2}}\left|E_2^+\right|^2 \cos(\varphi_2)}{\frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left|E_0^+\right|^2 \cos(\varphi_0)} = \frac{\sqrt{\varepsilon_2}\cos(\varphi_2)}{\sqrt{\varepsilon_0}\cos(\varphi_0)}\left|\frac{E_2^+}{E_0^+}\right|^2 \Leftrightarrow$$

$$\Leftrightarrow T = \frac{n_2\cos(\varphi_2)}{n_0\cos(\varphi_0)}|t|^2 \Leftrightarrow T = \frac{\Gamma_s}{\Gamma_0}|t|^2 \tag{2-31}$$

where $\Gamma_s = n_2\cos\varphi_2$, and $\Gamma_0 = n_0\cos\varphi_0$.

We will now make the following transformation to equations (2-26), (2-27) and (2-31):

$$n_i\cos(\varphi_i) = \sqrt{\varepsilon_i - \varepsilon_0\sin^2(\varphi_0)}\ , \qquad i = 1, 2 \tag{2-32}$$

This relation may be obtained by considering Snell´s law. From figure 2.3, we have that

$$n_i\sin(\varphi_i) = n_0\sin(\varphi_0) \Rightarrow \cos(\varphi_i) = \sqrt{1 - \frac{n_0^2\sin^2(\varphi_0)}{n_i^2}} \Leftrightarrow$$

$$\Leftrightarrow n_i\cos(\varphi_i) = \sqrt{n_i^2 - n_0^2\sin^2(\varphi_0)} \Leftrightarrow n_i\cos(\varphi_i) = \sqrt{\varepsilon_i - \varepsilon_0\sin^2(\varphi_0)}\ , \qquad i = 1, 2$$

So, we will get for Γ, in equations (2-26) and (2-27): $\Gamma = n_1\cos(\varphi_1) = \sqrt{\varepsilon_1 - \varepsilon_0\sin^2(\varphi_0)}$,and for Γ_s, in equation (2-31): $\Gamma_s = n_2\cos(\varphi_2) = \sqrt{\varepsilon_2 - \varepsilon_0\sin^2(\varphi_0)}$.

The film reflectance and transmittance may, therefore, be written, respectively, as:

$$R = \left|\frac{r_1 + r_2\exp\left(2i\delta\sqrt{\varepsilon_1 - \varepsilon_0\sin^2(\varphi_0)}\right)}{1 + r_1 r_2\exp\left(2i\delta\sqrt{\varepsilon_1 - \varepsilon_0\sin^2(\varphi_0)}\right)}\right|^2 \tag{2-33}$$

$$T = \frac{\sqrt{\varepsilon_1 - \varepsilon_0\sin^2(\varphi_0)}}{n_0\cos(\varphi_0)}\left|\frac{t_1 t_2\exp\left(i\delta\sqrt{\varepsilon_1 - \varepsilon_0\sin^2(\varphi_0)}\right)}{1 + r_1 r_2\exp\left(2i\delta\sqrt{\varepsilon_1 - \varepsilon_0\sin^2(\varphi_0)}\right)}\right|^2 \tag{2-34}$$

If we are dealing with an absorbing film, equations (2-33) and (2-34) are still valid; we only have to replace the value of n_1 by its corresponding complex quantity, $\tilde{n}_1$.

2.6- n and k as functions of ε_1 and ε_2

From equation (2-16) we may deduce the relations of n and k with the real and imaginary components of the dielectric function $\tilde{\varepsilon}_r = \varepsilon_1 + i\varepsilon_2$:

$$\tilde{n}_1 = \sqrt{\tilde{\varepsilon}_r} \Leftrightarrow n + ik = \sqrt{\varepsilon_1 + i\varepsilon_2} \tag{2-35}$$

Squaring both terms of this equation we get:

$$(n+ik)^2 = \varepsilon_1 + i\varepsilon_2 \Leftrightarrow n^2 - k^2 + i2nk = \varepsilon_1 + i\varepsilon_2$$

So,

$$\begin{cases} \varepsilon_1 = n^2 - k^2 \\ \varepsilon_2 = 2nk \end{cases} \Leftrightarrow \begin{cases} \varepsilon_1 = n^2 - \left(\dfrac{\varepsilon_2}{2n}\right)^2 \\ k = \dfrac{\varepsilon_2}{2n} \end{cases} \Leftrightarrow \begin{cases} 4\left(n^2\right)^2 - 4\varepsilon_1 n^2 - \varepsilon_2^2 = 0 \\ - \end{cases}$$

We have obtained a second degree equation, which may be easily solved by the usual formula:

$$n^2 = \frac{2\varepsilon_1 \pm \sqrt{4\varepsilon_1^2 + 4\varepsilon_2^2}}{4} = \frac{\varepsilon_1 \pm \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2}$$

Since $\sqrt{\varepsilon_1^2 + \varepsilon_2^2}$ is always greater or equal than ε_1, the minus sign has to be removed, otherwise n would be complex, which is, by definition, not true. So,

$$n^2 = \frac{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2} \Rightarrow n = \pm\sqrt{\frac{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2}} \Rightarrow n = \sqrt{\frac{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2}} \tag{2-36}$$

The minus sign was removed because n is always positive.

Now, for the extinction coefficient, k, we have that:

$$k = \frac{\varepsilon_2}{2n} = \frac{\varepsilon_2}{2\sqrt{\dfrac{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2}}} = \frac{\varepsilon_2}{\sqrt{2}\sqrt{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}} = \frac{1}{\sqrt{2}}\sqrt{\frac{\varepsilon_2^2}{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}}$$

If we multiply the numerator and the denominator of the square root by $\varepsilon_1 - \sqrt{\varepsilon_1^2 + \varepsilon_2^2}$ we get:

$$k = \frac{1}{\sqrt{2}}\sqrt{\frac{\varepsilon_2^2\left(\varepsilon_1 - \sqrt{\varepsilon_1^2 + \varepsilon_2^2}\right)}{\varepsilon_1^2 - \left(\varepsilon_1^2 + \varepsilon_2^2\right)}} = \sqrt{\frac{-\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2}} \tag{2-37}$$

From the value of n we may determine the wave velocity (v = c/n) in the oxide film. This value is in certain frequency ranges smaller than one, meaning that the wave velocity is greater than the velocity c of light in vacuum. This is, at first sight, a contradiction to one of the fundamental results of the theory of relativity, according to which c is the highest attainable velocity. There is actually no contradiction here, however, since relativity applies to the velocity with which energy (a light signal) is transmitted, and this is always less than c. We know that energy travels with the group velocity u. So, we require that it is c/u that shall be greater than unity, rather than c/v. u and v are related by the following equation

$$\frac{c}{u} = n - \lambda \frac{dn}{d\lambda} \tag{2-38}$$

where λ is the wavelength in vacuum. It was demonstrated [4] that this ratio is always greater than one. Therefore the group velocity is less than c, and there is no violation of the principle of relativity.

References

[1] Heavens, O. S., Optical Properties of Thin Solid Films, Dover Publications, Inc. (1991 - reissue of the 1955 edition)

[2] Plonus, M. A., Aplied Electromagnetics, Mc Graw-Hill (1978)

[3] Hecht, E., Óptica, Fundação Calouste Gulbenkian (1991)

[4] Jenkins, F. A. and White, H. E., Fundamentals of Optics, 4th Edn, Mc Graw-Hill (1981)

3 - Dielectric dispersion theory

3.1- Introduction

The intrinsic and extrinsic properties of a material influence its dielectric function. As these properties are related to the kind of interatomic bonds present in that material and since these bonds may be treated as oscillators, it seems reasonable to model the dielectric function with a set of independent lorentzian oscillators. It will be shown that, in fact, this is so.

This theory of modelling the dielectric function is known as Dielectric dispersion theory. It enables the direct study of the molecular structure of silicon dioxide films by specifying the oscillators parameters that allow the synthesis of curves having reasonable agreement with the optical data, which, in this work, will be transmission and reflection spectra.

The infrared spectra of amorphous silicon dioxide is very similar to that of crystalline, with the main difference being a general increase in the widths of the absorption bands and the appearance of a high-frequency "shoulder" in the band associated to the asymmetric stretch motion (see Chapter 1).

Given that the medium and long range structure of a-SiO_2 is considerably different from the ordered crystalline arrangement in c-SiO_2 , this suggests that the IR spectra are primarily dominated by the Si-O chemical bond which must be similar in both materials. We may, therefore, try to determine the a-SiO_2 dielectric function by making some simple transformations on the crystalline dielectric function, which is easier to calculate.

We will consider that a-SiO_2 is perfectly isotropic. Otherwise, the dielectric function, ε_r , would depend on the direction of the electric field, that is, it would not be a scalar but a tensor of the second rank.

3.2- Dielectric function for crystalline silicon dioxide

Consider the permanent electric dipole created by two atoms (or ions) with different electronegativities (e.g., the Si-O chemical bond). Under the influence of an external electric field, this dipole suffers a distortion x, i.e., the interatomic distance departs from its equilibrium position. Considering the mechanical equivalent of two masses connected by a spring, we have that this distortion originates a restoring force that counteracts the distortion and is directly proporcional to the last. There is also a damping force directly proporcional to the velocity dx/dt, originated by the friction that the dipole experiences when it flips back and forth. If $\mu w_0^2 x$ gives the restoring force and $\mu\gamma \frac{dx}{dt}$ the damping force, we have for the dipole the following equation of motion:

$$\mu \frac{d^2x}{dt^2} + \mu\gamma \frac{dx}{dt} + \mu w_0^2 x = e^* E_x \exp(-iwt) \quad (3\text{-}1)$$

where μ is the dipole reduced mass; γ is the damping coefficient; w_0 is the dipole natural angular TO frequency of vibration; e^* is the effective dipolar charge (we have to consider the electrons that are also orbiting around other ions which don't make part of the dipole); and $E_x \exp(-iwt)$ is the external applied electric field. The solution of equation (3-1) shows that x varies sinusoidally with an complex amplitude (due to the damping force) given by [1]:

$$x = \frac{e^* E_x / \mu}{w_0^2 - w^2 - iw\gamma} \quad (3\text{-}2)$$

The dipolar moment is then:

$$p_x = e^* x = \frac{\left(e^*\right)^2 E_x / \mu}{w_0^2 - w^2 - iw\gamma} \tag{3-3}$$

If we have N such dipoles per unit volume the polarization will be:

$$P_x = Ne^* x = \frac{N\left(e^*\right)^2 E_x / \mu}{w_0^2 - w^2 - iw\gamma} \tag{3-4}$$

We know that, $\vec{D} = \varepsilon_r \varepsilon_0 \vec{E} = \varepsilon_0 \vec{E} + \vec{P}$. From this relation we have that $\varepsilon_r = 1 + \frac{\vec{P}}{\varepsilon_0 \vec{E}}$. As we are considering an isotropic medium, we may work only in one dimension, so

$$\varepsilon_r = 1 + \frac{P_x}{\varepsilon_0 E_x}. \tag{3-5}$$

From this equation we get that the dielectric function, ε_r, is frequency dependent, and is complex, as expected:

$$\tilde{\varepsilon}_r(w) = 1 + \frac{N\left(e^*\right)^2 / \left(\mu\varepsilon_0\right)}{w_0^2 - w^2 - iw\gamma} \tag{3-6}$$

In Chapter 1 it was shown that a polyatomic molecule may be treated as a set of 3n-6 (or 3n-5 for linear molecules) oscillators, where n is the number of atoms in the molecule. We may, by analogy, make the same reasoning for the dielectric function. So, we will have the dielectric function of a polyatomic molecule with n atoms given by a sum of independent lorentzian oscillators:

$$\tilde{\varepsilon}_r(w) = 1 + \sum_{j=1}^{3n-6} \frac{N_j\left(e_j^*\right)^2 / \left(\mu_j\varepsilon_0\right)}{w_{0j}^2 - w^2 - iw\gamma_j} \tag{3-7}$$

If there is an optical electronic contribution to the polarization from the ion cores, this must be taken into account as well. This is done by replacing the constant term, 1, of $\tilde{\varepsilon}_r$ by the optical or high frequency dielectric constant, ε_∞ [2]. So, we get,

$$\tilde{\varepsilon}_r(w) = \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{N_j\left(e_j^*\right)^2 / \left(\mu_j\varepsilon_0\right)}{w_{0j}^2 - w^2 - iw\gamma_j} \tag{3-8}$$

The dielectric function may be written as a function of the wavenumber, $\bar{\nu}$, since $w = 2\pi c\bar{\nu}$, where the angular frequency is given in s^{-1}, the speed of light in vacuum is 3×10^{10} $cm.s^{-1}$, and the wavenumber is given in cm^{-1}:

$$\tilde{\varepsilon}_r(\bar{\nu}) = \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{N_j\left(e_j^*\right)^2 / \left[(2\pi c)^2 \mu_j\varepsilon_0\right]}{\bar{\nu}_{0j}^2 - \bar{\nu}^2 - i\bar{\nu}\gamma_j / 2\pi c} \tag{3-9}$$

This equation may be written in a simpler form:

$$\tilde{\varepsilon}_r(\bar{\nu}) = \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2}{\Omega_{0j}^2 - \bar{\nu}^2 - i\bar{\nu}\Omega_{lj}} \tag{3-10}$$

where $\Omega_{pj}^2 = N_j \left(e_j^*\right)^2 \Big/ \left[(2\pi c)^2 \mu_j \varepsilon_0\right]$, $\Omega_{0j} = \bar{\nu}_{0j}$, and $\Omega_{lj} = \gamma_j / 2\pi c$ are the strength, vibrational frequency, and Lorentzian width, respectively, of the materials jth lorentzian oscillator. These parameters are known as dispersion parameters.

The real and imaginary parts of the dielectric function, ε_1 and ε_2, respectively, are calculated as follows:

$$\tilde{\varepsilon}_r(\bar{\nu}) = \varepsilon_1(\bar{\nu}) + i\varepsilon_2(\bar{\nu}) = \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2}{\Omega_{0j}^2 - \bar{\nu}^2 - i\bar{\nu}\Omega_{lj}} = \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2\left[\left(\Omega_{0j}^2 - \bar{\nu}^2\right) + i\bar{\nu}\Omega_{lj}\right]}{\left(\Omega_{0j}^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} =$$

$$= \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2\left(\Omega_{0j}^2 - \bar{\nu}^2\right)}{\left(\Omega_{0j}^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} + i\sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2\Omega_{lj}\bar{\nu}}{\left(\Omega_{0j}^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2}$$

Therefore,

$$\varepsilon_1(\bar{\nu}) = \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2\left(\Omega_{0j}^2 - \bar{\nu}^2\right)}{\left(\Omega_{0j}^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} \tag{3-11}$$

and,

$$\varepsilon_2(\bar{\nu}) = \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2\Omega_{lj}\bar{\nu}}{\left(\Omega_{0j}^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} \tag{3-12}$$

3.3- Dielectric function for amorphous silicon dioxide

The accomodation required to change from a crystalline to an amorphous structure, proposed by Naiman *et al*. [3], introduces a distribution in bond strengths, and thereby a distribution in oscillator frequencies. We take this distribution to be gaussian and hence replace Ω_{pj}^2 in equations (3-11) and (3-12) by:

$$\Omega_{pj}^2 \longrightarrow \Omega_{pj}^2 \frac{1}{1.201\sqrt{\pi}\Omega_{gj}} \int_0^{+\infty} \exp\left[-\left(\frac{\bar{\nu}_i - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right] d\bar{\nu}_i \tag{3-13}$$

where Ω_{gj}, also a dispersion parameter, is the gaussian half width at half maximum for the jth oscillator distribution ($1.201\Omega_{gj} = \sqrt{2}\sigma_j$, where σ_j is the standard deviation). The ratio $1\Big/\left(1.201\sqrt{\pi}\Omega_{gj}\right)$ is the normalization factor.

A gaussian distribution is a good approach because the vibrational density of states (VDOS) spectrum of a-SiO_2 also shows a gaussian-like continuos frequency distribution around each peak (see Chapter 1).

Equations (3-11) and (3-12) now become

$$\varepsilon_1(\bar{\nu}) = \varepsilon_\infty + \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2}{1.201\sqrt{\pi}\Omega_{gj}} \int_0^{+\infty} \frac{\exp\left[-\left(\frac{\bar{\nu}_i - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)}{\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} d\bar{\nu}_i \tag{3-14}$$

and,

$$\varepsilon_2(\bar{\nu}) = \sum_{j=1}^{3n-6} \frac{\Omega_{pj}^2 \Omega_{lj} \bar{\nu}}{1.201\sqrt{\pi}\Omega_{gj}} \int_0^{+\infty} \frac{\exp\left[-\left(\frac{\bar{\nu}_i - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]}{\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} d\bar{\nu}_i \tag{3-15}$$

All the calculations necessary to compute these integrals are given in Chapter 5.

This model for amorphous materials is correct if we consider that only TO modes are observed for normal incidence spectra, which means that we are accepting Kirk's theory [4] about the existence of two asymmetric stretching modes (AS_1 and AS_2). On the other hand, if we consider that LO modes may also be observed in these spectra, then our model is incorrect, which is expected since we have considered that the only difference when going from a crystalline structure to an amorphous structure is a gaussian broadening of our peaks. If we were to consider that this model could also be used for LO modes, then these would correspond to peaks in ε_2 and not to peaks in $\mathrm{Im}(-1/\tilde{\varepsilon}_r)$, as it ought to be.

Equations (2-33), (2-34), (2-36) and (2-37) from Chapter 2, along with equations (3-11), (3-12), (3-14), and (3-15) from this Chapter, make possible the determination of some features of the a-SiO_2 molecular structure through the parameters $\varepsilon_\infty, \Omega_{pj}, \Omega_{0j}, \Omega_{lj}$ (and Ω_{gj} if we are considering the dielectric equation for amorphous silicon dioxide) that give agreement with reasonable accurate transmittance and reflectance spectra of that film.

This method of analysing the molecular structure of a material through its optical data is known as Dispersion Analysis. This method is very powerful since we may compute in a relatively easy manner some of the microscopic properties of a material, such as the effective dipolar charge, e*, the dipole reduced mass, μ, and the density of dipoles, N, with a determined moment.

A dispersion analysis is successful if it fulfills the following requirements:

(i) within the wavenumber region of interest, the theory should fit the data within reasonable estimates of the experimental error;

(ii) the number of oscillators used is the minimum number that the data requires;

(iii) the parameters $\varepsilon_\infty, \Omega_{pj}, \Omega_{0j}, \Omega_{lj}$ (and Ω_{gj} if we are considering the dielectric equation for amorphous SiO_2) are uniquely determined within reasonable uncertainties by the requirement of fitting the data.

In this method reasonable values are chosen for the dispersion parameters, the reflectance (or transmittance) is then calculated by computer, and the parameters are adjusted until agreement is obtained with experiment. This take us to Chapter 4.

References

[1] MATHIEU, M. H., Proprietes Optiques des Semiconducteurs, Universite des Sciences et Techniques du Languedoc

[2] OMAR, M. A., Elementary Solid State Physics: Principles and Applications, Addison-Wesley Publishing Company (1975)

[3] NAIMAN, M. L., KIRK, C.T., AUCOIN, R.J., TERRY, F.L., and WYATT, P.W.; SENTURIA, S.D., J. Electrochem. Soc. 131, 637 (1984)

[4] KIRK, C.T., Phys. Rev. B 38, 1255 (1988)

4 - FITIR , a nonlinear least-squares curve fitting program

4.1- Introduction

In order to find the dispersion parameters or the gaussian profiles parameters which predict correctly the measured infrared spectra, we must use a curve fitting procedure. For that we must choose a figure-of-merit function ("merit function", for short) that measures the agreement between the data (infrared spectrum) and the model (Fresnel equations or gaussian profiles) with a particular choice of parameters. The merit function is arranged so that small values represent close agreement. The parameters of the model are then adjusted to achieve a minimum in the merit function, yielding best-fit parameters. The adjustment, or curve fitting, process is thus a problem in minimization in many dimensions (as many as the number of parameters). The merit function is the chi-square function. The reason for having selected this particular function will be explained in section 4.2.

However, fitting of parameters does not ends the parameter estimation. For instance, data points are generally not exact. They are subject to measurement errors. Thus, typical data never exactly fit the model that is being used, even when that model is correct. So, we need to assess whether or not the model is appropriate, that is, we need to test the goodness-of-fit against some useful statistical standard.

We also need to know the accuracy with which parameters are determined by the data set, i.e., the likely errors of the best-fit parameters.

Finally, it is not uncommon in fitting data to discover that the merit function is not unimodal, with a single minimum. This means that when a minimum is found, it is often not obvious whether it is a local or a global minimum. This problem is generally very difficult to solve.

So, to be useful, a fitting procedure should provide [1]:

(i) parameters;

(ii) error estimates on the parameters;

(iii) a statistical measure of goodness-of-fit.

These are the main goals of program FITIR. It uses sophisticated algorithms (briefly described in section 4.3) that have been developed to fit data to models which depend nonlinearly on the set of parameters, as is our case: dispersion analysis and gaussian decomposition.

The computer language used was FORTRAN 77 with some FORTRAN 90 features. This language was chosen by two reasons:

(i) it is specially designed for mathematical calculations;

(ii) almost all the available mathematical software libraries are written in this language.

The FORTRAN compiler used was the F77L–EM/32 from Lahey Computer Systems, Inc.. This compiler had available some basic graphic routines. From these routines and from some MS-DOS interrupt functions it was possible to build a graphics library that has as main features the following points:

(i) visualize the evolution of the model curve as the parameters are adjusted, in contrast with the curve we wish to fit, as well as the difference between the two curves (residuals);

(ii) visualize the parameters values, the merit function value, the number of iterations;

(iii) possibility to change the original data we want to fit, with the objective of making it more "easy" to fit. As examples, we may refer the capacity to select the spectral region we wish to fit, the capacity to interpolate or give weights to selected regions of the curve, and the capacity of smoothing the curve (noise filtering);

(iv) possibility to change the parameters values whenever we wish to do so (before and during the fit). This is very important since it often happens that during the minimization process one or more parameters evolve to values without any physical meaning. This correction possibility will, therefore, permit to put the "stray" parameters "back into the right track";

(v) besides these possibilities to change the data or the parameters values, it is also possible to stop a fitting process in order to change a few other inputs, such as the minimization code;

(vi) easy-to-use program due to its structure of windows and pressing buttons.

Table 4-1 shows the number of lines, subroutines and functions of FITIR. “Homemade” stands for those lines, subroutines and functions specially developed for this program, while “Imported” stands for those that are of public domain, namely the minimization codes.

Table 4-1: Number of lines, subroutines and functions of FITIR

Source	Lines	Subroutines	Functions
“Homemade”	≈12140	157	4
“Imported”	≈13125	74	18
Total	≈25265	231	22

4.2- Chi-square fitting

If we know both the form of the model to be fitted and the form of the distribution of random error in the observations, each independent piece of data has a “probability” that varies with each choice of a set of values for the parameters. The product of the probabilities of all the data points, for a given set of values for all the parameters, is called the likelihood of that set. There is usually a particular set that maximizes this product; it is termed the maximum likelihood set.

When the distribution of random disturbances is normal (or gaussian), the maximum likelihood estimates may be found by a simpler procedure called the method of least-squares.

The least-squares (LS) method says: ”Find the values for the parameters in the chosen model equation that minimize the sum of the squared deviations of the observed values from those predicted by the model.”. That is,

$$\text{minimize over } p_1, \ldots, p_M : \quad \sum_{i=1}^{N} \left[y_i - y(x_i ; p_1 \ldots p_M) \right]^2 \tag{4-1}$$

where we are fitting N data points (x_i , y_i) i = 1,...,N, to a model, $y(x_i ; p_1 \ldots p_M)$, that has M adjustable parameters p_j j = 1,...,M.

The justification for this statement is given by the Gauss - Markov theorem [2], which states that the estimates obtained are the most precise unbiased estimates that are linear functions of the observations. It is assumed, of course, that the correct model has been chosen. By “most precise” we mean the estimates in (imagined) successive sets of data taken under identical circumstances scatter around the true values with minimum variance. By “unbiased” we mean that the LS estimates averages out to the true values for a sufficient big number of data sets. By “linear” minimum variance estimates we imply that, although there may be other estimates that have even smaller variances, they will not be linear functions of the observations. This property of linearity is advantageous in its simplicity, but there may be nonlinear estimates with much smaller variance. However, when the random errors are normally distributed, the LS estimates are maximum likelihood estimates and are of possible minimum variance.

The first two of the following three key assumptions of the LS method are also crucial assumptions of maximum likelihood.

The first assumption is that we know, or have chosen, the correct model. If the model is not correct the parameters values obtained are meaningless.

The second assumption is that the data are typical, that is, that they are a representative sample from the whole range of situations about which we want to generalize. It is obvious that a model which represents the typical behaviour of a system may not be derived from nontypical data.

The third key assumption of the LS method is that the y_i -observations are statistically independent. Each y_i is taken to be made up of a “true” value, η (eta), plus a random component, e. If we denote any two of these random components by e_j and e'_j, then this assumption may be stated as: if the probability of e'_j appearing in some range of values is the same, regardless of the magnitude of e_j appearing in some range, then e'_j is statistically independent of e_j.

There are also two less important assumptions for the LS method. The first assumption is that all observations on y_i have the same (though unknown) variance. (When the variance of the dependent variable changes in a known way, then the more general weighted least squares (or chi-square) method − weighting each point inversely by its variance − may be used.) The second is that all the values of the independent variables (x_1 , x_2 , ... , x_N), are known without error.

In addition to the requirements already stated for LS estimation, nonlinear estimation requires that:

(i) the model by definition be nonlinear in some of its coefficients;

(ii) preliminary estimates of all the parameters be available. If these estimates are not close "enough", the program will not converge to the best values.

All that has been said until now may be more or less summarized as follows: we are going to suppose that each data point y_i has a measurement error that is independently random and with a Gaussian distribution around the "true" model $y(x_i ; p_1 ... p_M)$. And we are also going to suppose that the standard deviations σ of these normal distributions are the same for all points. Then the probability of the data set is the product of the probabilities of each point,

$$P \propto \prod_{i=1}^{N} \left\{ \exp\left[-\frac{1}{2} \left(\frac{y_i - y(x_i ; p_1 \ldots p_M)}{\sigma} \right)^2 \right] \Delta y \right\} \tag{4-2}$$

Maximizing equation (4-2) is equivalent to maximizing its logarithm, or minimizing the negative of its logarithm, namely,

$$\left[\sum_{i=1}^{N} \frac{\left[y_i - y(x_i ; p_1 \ldots p_M) \right]^2}{2\sigma^2} \right] - N \log \Delta y \tag{4-3}$$

Since N, σ, and Δy are all constants, minimizing this equation is equivalent to minimize equation (4-1) which is the LS method.

For real data, the normal distribution is often rather poorly realized. However, if data are taken in a sufficient big number and carefully enough, the LS method gives a good approximation to the maximum likelihood set.

As mentioned above, if we know the standard deviation of each point y_i then equation (4-2) is modified only by putting a subscript i on the symbol σ. That subscript also propagates into equation (4-3), so that the maximum likelihood estimate of the model parameters is obtained by minimizing the quantity

$$\chi^2 \equiv \sum_{i=1}^{N} \left[\frac{y_i - y(x_i ; p_1 \ldots p_M)}{\sigma_i} \right]^2 \tag{4-4}$$

called the "chi-square".

Equation (4-4) may be written in a different form by defining the weight of y_i as $w_i = 1/\sigma_i^2$, and by putting the parameters in the vectorial form, $\vec{p} = (p_1, \ldots, p_M)$:

$$\chi^2 \equiv \sum_{i=1}^{N} w_i \left[y_i - y(x_i ; \vec{p}) \right]^2 \tag{4-5}$$

The difference $y_i - y(x_i ; \vec{p})$ is known as the ith point residual, and χ^2 the sum of squares of the residuals (SSR).

4.3- Mathematical software

FITIR program uses eight FORTRAN codes which solve nonlinear least squares problems: DN2G, DN2F, DN2GB, DN2FB, LMDER, LMDIF, MRQMIN, and FRPRMN.

The algorithms these codes implement are based on three methods, Levenberg-Marquardt, augmented Gauss-Newton and Conjugate Gradient.

The reason for using all these codes is due to the fact that there is no perfect minimization algorithm. They all have strong points and weak points. If, for a given set of initial parameters, the selected code fails to converge, maybe one of the other codes will work well.

All the three methods solve the unconstrained nonlinear least squares problem [except codes DN2GB and DN2FB, where there are a priori limitations on the allowed values of the parameters, that is, the parameters have upper and lower bounds (defined by the user)], which is [3,4]:

$$\text{minimize } \chi^2 = F(\vec{p})^T F(\vec{p}) = \sum_{i=1}^{N}\left[\frac{y_i - y(x_i;\vec{p})}{\sigma_i}\right]^2 = \sum_{i=1}^{N} f_i(\vec{p})^2 \tag{4-6}$$

$\vec{p} \in \Re^M, F: \Re^M \to \Re^N$, for unconstrained optimization;

$\vec{p} \in D^M \subset \Re^M, F: D^M \to \Re^N$, for constrained optimization.

by finding a zero of the gradient of χ^2, that is, an $\vec{p}$ such that

$$\nabla\chi^2(\vec{p}) = J(\vec{p})^T F(\vec{p}) = 0, \tag{4-7}$$

where J is the Jacobian of F (this is a necessary condition for a solution). The Levenberg-Marquardt and the augmented Gauss-Newton methods use variations of Newton's method to find a zero of the gradient of χ^2. The Newton step, $\Delta\vec{p}$, for improving an approximate solution $\vec{p}$ of equation (4-7) is the solution of the following system of equations

$$\nabla^2\chi^2 \cdot \Delta\vec{p} = -\nabla\chi^2 \tag{4-8}$$

or, equivalently,

$$\left(J^T J + \sum_{i=1}^{N} f_i \nabla^2 f_i\right) \cdot \Delta\vec{p} = -J^T F. \tag{4-9}$$

Here $\nabla^2\chi^2$ is the Hessian matrix of χ^2, $\left(\nabla^2\chi^2\right)_{ij} = \frac{\partial^2\chi^2}{\partial p_i p_j}$, i,j = 1,...,M, and $\nabla^2 f_i$ is the Hessian matrix of the ith residual. We are going to refer to the term $\sum_{i=1}^{N} f_i \nabla^2 f_i$, in equation (4-9) as S. The variations of Newton's method involve the approximation of this S term.

The Levenberg-Marquardt method [1] approximates S with λD, where D is a diagonal matrix and λ, a scalar, is the Levenberg-Marquardt parameter. The following codes implement the Levenberg-Marquardt method: LMDER, LMDIF, and MRQMIN.

In the augmented Gauss-Newton method [5], part of S is computed exactly and part is approximated by a quasi-Newton updating method. This approximation guarantees that $\nabla^2 f_i$ is always positive definite. The algorithm sometimes reduces to a Gauss-Newton or Levenberg-Marquardt method. Close to the minimum, the approximation approaches the true Hessian and we obtain the Gauss-Newton method. The following codes implement the augmented Gauss-Newton method: DN2G, DN2F, DN2GB, and DN2FB.

The conjugate gradient method [1], here typified by the Polak-Ribiere algorithm, says that $\Delta\vec{p}$ has the direction that is somehow constructed to be conjugate to the old gradient, and, insofar as possible, to all previous directions traversed (two vectors, $\vec{u}$ and $\vec{v}$, are said to be conjugate when $\vec{u} \cdot \nabla^2 f_i \cdot \vec{v} = 0$). The code that implements the conjugate gradient method is FRPRMN.

Obviously, all the codes need a user-supplied routine that evaluates $y(x_i;\vec{p})$ [our model]. The calculations necessary to compute the models equations are given in the next Chapter.

Inherent in the above three methods and therefore in all the codes used in FITIR is the Jacobian matrix, J, of F. The codes may be divided into two groups: those that use analytical Jacobian (Group 1: DN2G, DN2GB, LMDER, MRQMIN, FRPRMN) and those that do not (Group 2: DN2F, DN2FB, LMDIF).

The codes that use the analytical Jacobian have more accurate information about a problem than the codes that do not use it.

The main reason for dividing the codes into two groups is that Group 1 codes require the user to supply a subroutine that evaluates the Jacobian. If we had supplied to Group 1 codes, subroutines that approximate the Jacobian instead of evaluating the analytical Jacobian, then the grouping of the codes would have been arbitrary. In fact, the only difference between LMDER and LMDIF, DN2G and DN2F, DN2GB and DN2FB is that LMDER, DN2G and DN2GB calls the user-supplied routine, which evaluates the Jacobian, while LMDIF, DN2F and DN2FB calls a library routine, which calculates the finite difference approximation to the Jacobian.

Some properties of these codes are summarized in Tables 4-2 and 4-3.

The calculations needed to make the routine that computes the Jacobian matrix analitically are given in Appendix E.

In order to guarantee small rounding off errors, all the calculations performed by the codes are made in double precision (8 bytes reals).

Table 4-2: Group 1 - Codes that require the Jacobian to be supplied

Name	**Source**	**Method**	**Length of call list**	**Return criteria**	**Special remarks**
DN2G	NBER	GN+	12	8	–
DN2GB	NBER	GN+	13	8	constrained minimization
LMDER	MINPACK	LM	26	9	–
MRQMIN	NRF	LM	13[a]	2	select which parameters to fit
FRPRMN	NRF	CG	5	2	–

Note. Abbreviations used: GN+ – augmented Gauss-Newton; LM – Levenberg-Marquardt; CG – Conjugate Gradient; NRF – Numerical Recipes in FORTRAN (see References).
[a] Two of the arguments in the calling list are vectors involving many parameters.

Table 4-3: Group 2 - Codes that do not require the Jacobian to be supplied

Name	**Source**	**Method**	**Length of call list**	**Return criteria**	**Special remarks**
DN2F	NBER	GN+	12	8	–
DN2FB	NBER	GN+	13	8	constrained minimization
LMDIF	MINPACK	LM	26	9	–

Note. Abbreviations used: GN+ – augmented Gauss-Newton; LM – Levenberg-Marquardt.
[a] Two of the arguments in the calling list are vectors involving many parameters.

The implementation of the codes, that is, their call list length, varies a great deal with respect to what the code requires of the user, as may be seen in Tables 4-2 and 4-3. In FITIR, however, the inputs the user must enter are more or less the same for all codes, which are mainly the infrared spectrum data and the initial set of parameters. Convergence tolerances, maximum number of iterations, and multiplicative scale factors for the parameters are, amongst others, defined internally by FITIR. This was done in order to simplify its operation.

The codes return to the user for several reasons, that is, when they have found a solution, found an error, or performed too much work. The return criteria vary substantially among the codes used. The codes check on either the relative or absolute change in the values of the parameters from iterate to iterate and/or the size of the residuals (f_i) at the new $\vec{p}$ value. Some codes check the size of the gradient, and so on. On the basis of these checks, the codes return, claiming a solution. The other most common reason for returning is too much work: either to many function evaluations or too many iterations have been performed.

All the codes have minor changes which were necessary in order to make them “compatible” with the characteristics of FITIR. Their performance has, by no means, changed, though.

Nevertheless, it was found that the main routines of codes DN2G and DN2GB had an error which did not permit the correct determination of the Jacobian. This error, which was in fact a subroutine missing, has been corrected as follows: the original version presented the following code line:

```
50      CALL CALCJ(N, P, X, IV(NFGCAL), V(DR1), UI, UR, UF)
```

The corrected version replaced the above code line by:

```
50      MODE=35
        DINIT=38
        IF (IV(MODE) .LT. 0 .AND. V(DINIT) .EQ. 0.d0)
       &CALL DV7SCP(P, V(D1), 1.d0)
        CALL CALCJ(N, P, X, IV(NFGCAL), V(DR1), UI, UR, UF)
```

4.4- Statistical treatment

Each time the code returns or the user stops the fitting procedure, FITIR always asks if the fit is to be accepted. If the answer is negative, we may retry the fit or start a fresh new fit by changing some or all of the input variables. Conversely, if the answer is positive, FITIR enters the statistical treatment.

With the parameters found, we have to determine whether or not they are reliable because, even when the curve obtained with the model is very similar to that we wish to fit, that does not mean that the parameters are the true ones (assuming the model is correct, of course). From what was told in section 4.1, this determination may be done by the statistical measurement of goodness-of-fit and by the error estimates on the parameters. As we shall see in Chapter 7, their values play a crucial role in the analysis our IR spectra.

4.4.1- Goodness-of-fit

A quantitative measure for the goodness-of-fit of the model is given by the chi-square distribution for ν=N-M degrees of freedom. This probability function may be computed using the complement of the incomplete gamma function [1]:

$$Q\left(\chi^2|\nu\right)=Q\left(\frac{\nu}{2},\frac{\chi^2}{2}\right)=\frac{\Gamma\left(\frac{\nu}{2},\frac{\chi^2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \tag{4-10}$$

Equation (4-10) gives the probability Q that the chi-square should exceed a particular value χ^2 by chance even for a correct model.

If Q is a very small value probability for some particular data set, then the apparent discrepancies are unlikely to be chance fluctuations. Much more probably either (i) the model is wrong − may be statistically rejected, (ii) the measurement errors are larger than supposed, or (iii) the measurement errors may not be normally distributed. Possibility (iii) is fairly common. for this reason one often accepts models with Q more or less greater than 0.001.

On the other hand, if the probability Q is too large, too near to 1, it generally means that the measurement errors have been overestimated.

Care must be taken when applying the Q value to measure the model goodness-of-fit, though. The more the model is nonlinear in the parameters, the less reliable is the computed Q value. Since our model is pretty nonlinear, we must define an alternate quantity that also measures the goodness-of-fit of the model. This quantity is the standard error of the curve-fit (SCF) defined as the square root of the ratio between the chi-square function and the number of degrees of freedom:

$$SCF=\sqrt{\frac{\chi^2}{N-M}} \tag{4-11}$$

FITIR computes both quantities. It is left to the user to decide which one of them is more reliable.

4.4.2 - Error estimates on the parameters

We may estimate the errors on the parameters by assuming that near the minimum, the χ^2 function is approximately quadratic. In this case, the diagonal elements of the covariance matrix − [C] − give the variances of these parameters. This matrix is calculated by the following formula

$$C = \frac{\chi^2}{N-M}\left(\nabla^2\chi^2\right)^{-1} \tag{4-12}$$

So, the estimated standard deviation for the jth parameter (j = 1, ..., M) is given by

$$s\left(p_j\right) = \sqrt{\frac{\chi^2}{N-M}\left[\left(\nabla^2\chi^2\right)_{jj}\right]^{-1}} = \sqrt{\frac{\chi^2}{N-M}\left(\frac{\partial^2\chi^2}{\partial p_j^2}\right)^{-1}} \tag{4-13}$$

4.5 - Data saving

If the user wishes to do so, FITIR creates files where it saves the most important data obtained in the curve fitting process. Table 4-4 summarizes the name of the files created, as well as the data they store.

Table 4-4: Files created by FITIR

File name	Data stored
OUTPUT.F*xx*	Program name; spectrum analysed; type of fitting; name of the file with the experimental data; name of the file with the initial set of parameters; film thickness; angle of incidence; code used; number of points for Gauss-Legendre quadrature; code return criterium; time spent fitting; number of iterations; initial and final adjusted values for the parameters; lower and upper parameter bounds if the code used was either DN2GB or DN2FB.
PARAM.F*xx*	Final adjusted parameters (best-fit parameters).
nfedwe.F*xx*	Region of the experimental curve fitted with the transformations, if any, performed by the user.
CURVE.F*xx*	Curve obtained with the best-fit parameters.
COVAR.F*xx*	Covariance matrix (only created if this matrix was computed successfully).
STAT.F*xx*	Number of data points; number of parameters; number of degrees of freedom; χ^2 (SSR) value; Q value; SCF value; estimated standard deviation for the best-fit parameters.
RESID.F*xx*	Residuals as a function of the wavelength, that is, difference between the experimental curve and the curve obtained with the best-fit parameters.
EPS1.F*xx*[(a)]	Real component of the silicon dioxide film dielectric function as a function of the wavenumber calculated with the best-fit parameters.
EPS2.F*xx*[(a)]	Imaginary component of the silicon dioxide film dielectric function as a function of the wavenumber calculated with the best-fit parameters.
NN.F*xx*[(a)]	Real component of the silicon dioxide film complex index of refraction as a function of the wavenumber calculated with the best-fit parameters.
KK.F*xx*[(a)]	Imaginary component of the silicon dioxide film complex index of refraction as a function of the wavenumber calculated with the best-fit parameters.
GAUSS.F*xx*[(b)]	Individualized gaussian profiles as a function of the wavelength.

Note. Notation used: *xx* is a number between 0 and 99 defined by the user; *nfedwe* means name of the file with the experimental data without extension (for example if the file name is TRANS5.TRA then *nfedwe*.F*xx* will be TRANS5.F*xx*.
(a) created only in Dispersion analysis; (b) created only in Gaussian decomposition.

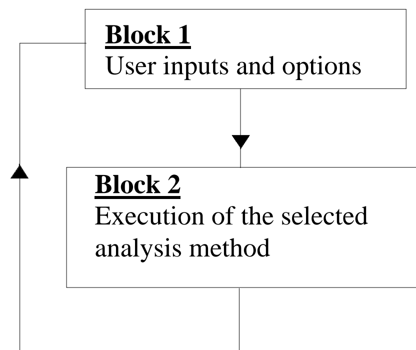
4.6 - FITIR flow chart

Basically, we may devide the software in two blocks. The first deals with the user inputs and options:

- definition of the spectrum to be studied (either transmittance or reflectance);
- selection of the analysis method (either gaussian decomposition or dispersion analysis);
- definition of the files names with the experimental data and the initial set of parameters;
- modifications on the experimental data (optional);
- modifications on the initial set of parameters (optional);
- minimization code selection.

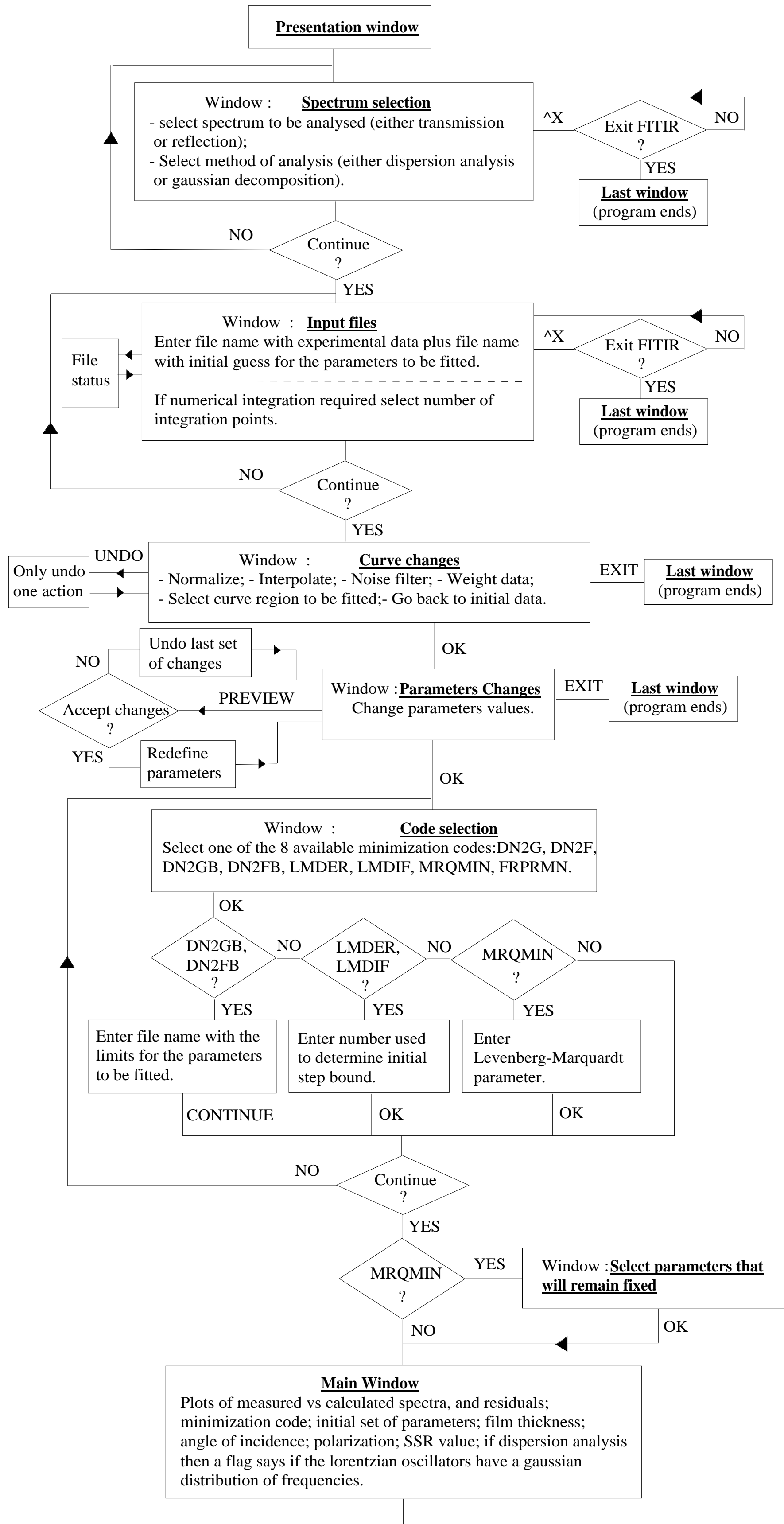
The second block deals with the execution of selected analysis method:

- curve fit (stop/pause available);
- store data (optional);
- retry fit/make new fit;
- change user inputs and options.

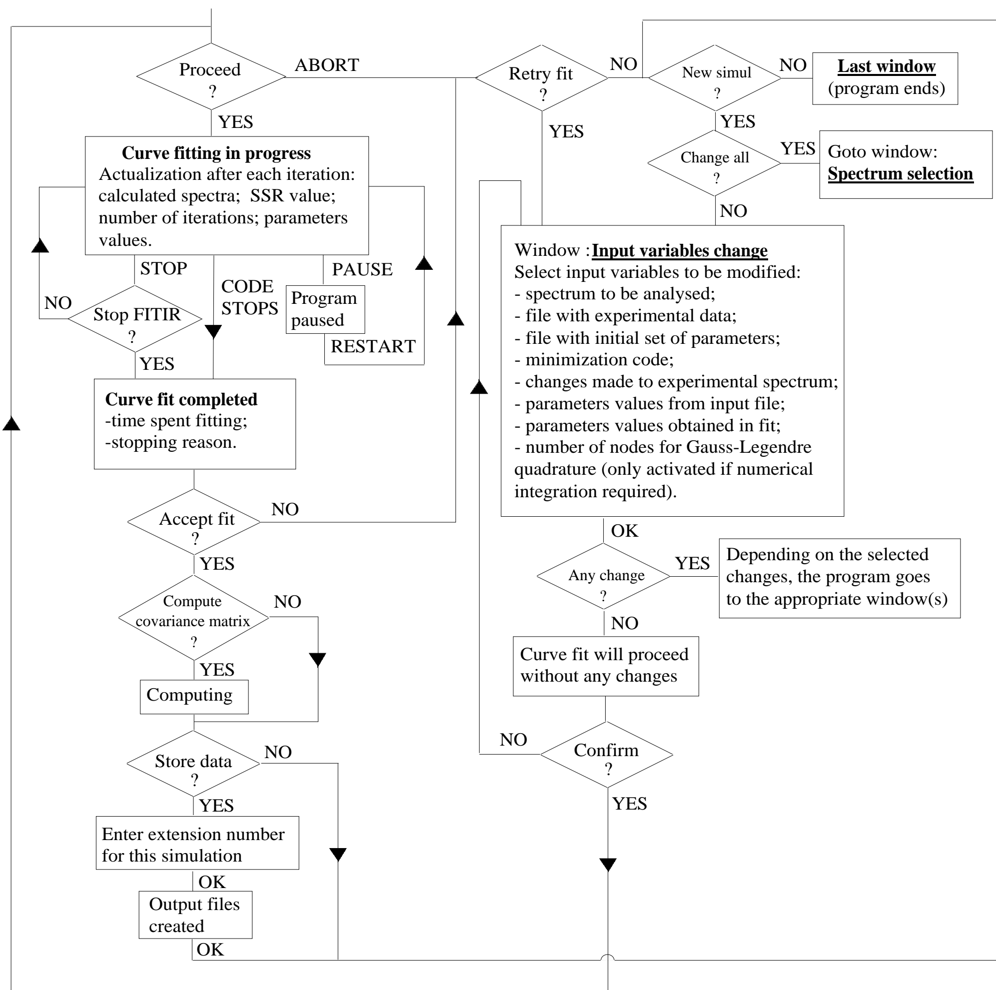


The two following flow charts describe in more detail the structure of these two blocks.

- Block 1:



- Block 2:



References

[1] PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., FLANNERY, B. P., Numerical Recipes in FORTRAN, 2nd Edn., Cambridge University Press (1992)

[2] DANIEL, C., WOOD, F. S., Fitting equations to data , 2nd Edn., John Wiley & Sons (1980)

[3] HIEBERT, K.L., ACM Trans. Math. Softw. 7, 1 (1981), pp. 1-16

[4] MORÉ, J. J., GARBOW, B. S., and HILLSTROM, K. E., ACM Trans. Math. Softw. 7, 1 (1981), pp.17-41

[5] DENNIS, J. E., GAY, D. M., and WELSCH, R. E., ACM Trans. Math. Softw. 7, 3 (1981), pp. 348-368; op. cit., pp. 369-383

5 - Models equations

5.1- Introduction

As mentioned before, program FITIR, implements two spectra analysis methods: dispersion analysis and gaussian decomposition. This Chapter presents the more important steps taken to calculate the models equations, $y(x_i;\vec{p})$, for these two methods.

Section 5.2 treats the model for dispersion analysis, which is either the film transmittance (T) or the film reflectance (R), as a function of the dispersion parameters:

$$T = T\left[\sum_i p(i)\right]; \qquad R = R\left[\sum_i p(i)\right] \tag{5-1}$$

Section 5.3 treats the model for gaussian decomposition, which are gaussian profiles.

5.2- Model for dispersion analysis

It was shown in Chapter 2, that for a film-substrate system, the complex Fresnel formulas for the transmitted and reflected intensities may be written in the form

$$T_j = \frac{\Gamma_s}{\Gamma_0}\left|t_j\right|^2, \tag{5-2}$$

$$R_j = \left|r_j\right|^2 \tag{5-3}$$

where, $t_j = \dfrac{t_{1j}t_{2j}\exp(i\delta\Gamma)}{1+r_{1j}r_{2j}\exp(2i\delta\Gamma)}$ is the transmitted amplitude; $r_j = \dfrac{r_{1j}+r_{2j}\exp(2i\delta\Gamma)}{1+r_{1j}r_{2j}\exp(2i\delta\Gamma)}$ is the reflected amplitude ($\delta = 2\pi d/\lambda$ is the film thickness d normalised to the wavelength λ, and $\Gamma = \sqrt{\tilde{\varepsilon}_r - \varepsilon_0 \sin^2\varphi_0}$ involves the complex dielectric function $\tilde{\varepsilon}_r = \varepsilon_1 + i\varepsilon_2$ of the film material and the angle of incidence φ_0); $\Gamma_s = \sqrt{\varepsilon_s - \varepsilon_0\sin^2\varphi_0}$ and $\Gamma_0 = n_0\cos\varphi_0$ are functions of the refractive indices of the two media adjacent to the film (index 0 — medium of incidence; index s — substrate) and of the angle of incidence, once again.

Owing to the oblique incidence of light, it is distinguished between perpendicular ($j = \perp$) and parallel (j = //) polarisation with respect to the plane of incidence. Here, r_{ij} and t_{ij} are the corresponding Fresnel coefficients of the two interfaces (i = 1 — medium of incidence/film, i = 2 — film/substrate), given by the equations:

$$t_{1j} = \tau_1 e^{i\kappa_1} \qquad t_{2j} = \tau_2 e^{i\kappa_2}$$

$$r_{1j} = \rho_1 e^{i\phi_1} \qquad r_{2j} = \rho_2 e^{i\phi_2};$$

We consider that both the medium of incidence (air) and the substrate (silicon[8]) have no losses, so their refractive indexes (and dielectric functions) are real:

[8] The dispersion in silicon is, for the frequency range we are interested in, negligible. The small absorption actually present in the silicon is partially compensated by the reference wafer (Chapter 6).

$$n_0 = \sqrt{\varepsilon_0} = 1 + i0$$
$$n_s = \sqrt{\varepsilon_s} = 3.42 + i0$$

Since the required calculations are very long, they will be presented by steps. First of all, we calculate the real and imaginary parts of the a-SiO_2 dielectric function (section 5.2.1), which will be needed for the transmittance and reflectance equations (calculated in sections 5.2.2 and 5.2.3, respectively).

5.2.1- Dielectric function

In program FITIR we may fit the transmittance and reflectance curves either with the dielectric dispersion theory for crystalline materials or with the dielectric dispersion theory for amorphous materials. The reason for considering the model for crystals in the study of amorphous silicon dioxide lies in the fact that there are authors that claim to have acomplished good fits using such model [1].

The relations that describe the dielectric function for crystals are given by equations (3-11) and (3-12) of Chapter 3. The only difference between these two equations and the ones we will use in FITIR is that the sum is not done over all the 3n-6 oscillators, but only over the oscillators present in the frequency range we wish to study. So, if this desired number of oscillators is nosc, we will have the following equations for the real and imaginary parts of the dielectric function of a crystalline material:

$$\varepsilon_1(\bar{\nu}) = \varepsilon_\infty + \sum_{j=1}^{nosc} \frac{\Omega_{pj}^2\left(\Omega_{0j}^2 - \bar{\nu}^2\right)}{\left(\Omega_{0j}^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} \tag{5-4}$$

and,

$$\varepsilon_2(\bar{\nu}) = \sum_{j=1}^{nosc} \frac{\Omega_{pj}^2 \Omega_{lj}\bar{\nu}}{\left(\Omega_{0j}^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} \tag{5-5}$$

For amorphous materials the same relations are given by equations (3-14) and (3-15) , also from Chapter 3. The reason presented above for replacing 3n-6 by nosc maintains, obviously, its validity. So we will have the following equations for the real and imaginary parts of the dielectric function of an amorphous material:

$$\varepsilon_1(\bar{\nu}) = \varepsilon_\infty + \sum_{j=1}^{nosc} \frac{\Omega_{pj}^2}{1.201\sqrt{\pi}\Omega_{gj}} \int_0^{+\infty} \frac{\exp\left[-\left(\frac{\bar{\nu}_i - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)}{\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} d\bar{\nu}_i \tag{5-6}$$

and,

$$\varepsilon_2(\bar{\nu}) = \sum_{j=1}^{nosc} \frac{\Omega_{pj}^2 \Omega_{lj}\bar{\nu}}{1.201\sqrt{\pi}\Omega_{gj}} \int_0^{+\infty} \frac{\exp\left[-\left(\frac{\bar{\nu}_i - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]}{\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} d\bar{\nu}_i \tag{5-7}$$

These integrals have no analytical solution. However, they may be calculated quite efficiently by the Gauss-Legendre quadrature (see Appendix A):

$$\int_a^b f(t)dt \approx \frac{b-a}{2}\sum_{j=1}^{n} w_j f\left(\frac{a+b+x_j(b-a)}{2}\right) \tag{5-8}$$

where w_k and x_k are the Gauss-Legendre weights and nodes, respectively; n gives the number of nodes (and weights) used to compute the integral.

At first sight it seems wrong to use the Gauss-Legendre quadrature to calculate an integral with the upper limit equal to + ∞, since the upper limit b must be finite. This inconsistency is removed when we consider the nature of the integrand function: it has in the numerator a gaussian, meaning that not very far away from its mean value, the integrand function is very close to zero and may, therefore, be neglected. So, equations (5-6) and (5-7) may be written as follows:

$$\varepsilon_1(\bar{\nu}) \approx \varepsilon_\infty + \sum_{j=1}^{nosc} \frac{\Omega_{pj}^2}{1.201\sqrt{\pi}\Omega_{gj}} \int_{ll1}^{ul1} \frac{\exp\left[-\left(\frac{\bar{\nu}_i - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)}{\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} d\bar{\nu}_i \qquad (5\text{-}9)$$

and,

$$\varepsilon_2(\bar{\nu}) \approx \sum_{j=1}^{nosc} \frac{\Omega_{pj}^2 \Omega_{lj} \bar{\nu}}{1.201\sqrt{\pi}\Omega_{gj}} \int_{ll2}^{ul2} \frac{\exp\left[-\left(\frac{\bar{\nu}_i - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]}{\left(\bar{\nu}_i^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2} d\bar{\nu}_i \qquad (5\text{-}10)$$

where ll1, ul1, ll2 and ul2 are such that the integrands outside the ranges defined by these limits have a very small value and may, therefore, be neglected. The determination of these limits will be given in section 5.2.1. Applying equation (5-8) to equations (5-9) and (5-10) we will have

$$\begin{aligned} \varepsilon_1(\bar{\nu}) &\approx \varepsilon_\infty + \sum_{j=1}^{nosc}\left[\frac{ul1 - ll1}{2}\sum_{k=1}^{n} w_k f_1\left(\frac{ll1 + ul1 + x_k(ul1 - ll1)}{2}\right)\right] = \\ &= \varepsilon_\infty + \sum_{j=1}^{nosc}\left[\frac{ul1 - ll1}{2}\sum_{k=1}^{n} w_k f_1(\mathrm{var}\,1)\right] \end{aligned} \qquad (5\text{-}11)$$

and,

$$\begin{aligned} \varepsilon_2(\bar{\nu}) &\approx \sum_{j=1}^{nosc}\left[\frac{ul2 - ll2}{2}\sum_{k=1}^{n} w_k f_2\left(\frac{ll2 + ul2 + x_k(ul2 - ll2)}{2}\right)\right] = \\ &= \sum_{j=1}^{nosc}\left[\frac{ul2 - ll2}{2}\sum_{k=1}^{n} w_k f_2(\mathrm{var}\,2)\right] \end{aligned} \qquad (5\text{-}12)$$

where,

$$\mathrm{var1} = \frac{ll1 + ul1 + x_k(ul1 - ll1)}{2}\ ;\quad f_1(\mathrm{var}\,1) = \frac{\Omega_{pj}^2}{1.201\sqrt{\pi}\Omega_{gj}} \frac{\exp\left[-\left(\frac{\mathrm{var}\,1 - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]\left(\mathrm{var}\,1^2 - \bar{\nu}^2\right)}{\left(\mathrm{var}\,1^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2}$$

$$\mathrm{var2} = \frac{ll2 + ul2 + x_k(ul2 - ll2)}{2}\ ;\quad f_2(\mathrm{var}\,2) = \frac{\Omega_{pj}^2\Omega_{lj}\bar{\nu}}{1.201\sqrt{\pi}\Omega_{gj}} \frac{\exp\left[-\left(\frac{\mathrm{var}\,2 - \Omega_{0j}}{1.201\Omega_{gj}}\right)^2\right]}{\left(\mathrm{var}\,2^2 - \bar{\nu}^2\right)^2 + \left(\bar{\nu}\Omega_{lj}\right)^2}$$

5.2.1.1- Finding the integration limits

The integration limits ll1, ul1, ll2, and ul2 may be found with a good approximation by the two following procedures:

(i) visualisation of the graphics that show the evolution of the ε_1 and ε_2 integrands as functions of the integration variable $\bar{\nu}_i$ and of the ratio Ω_g/Ω_1 ; the wavenumber $\bar{\nu}$, and the parameters Ω_p, and Ω_0 are constants. Some of the studied graphics are drawn in Appendix B;

(ii) comparison of the values obtained in (i) with those obtained with a FORTRAN 77 program that finds the integration limits by calculating, for each set of parameters, the values where the integrands are greater than a certain value, for example 10^{-3}. This program is printed in Appendix I.

The values thus obtained are given by the following equations:

$$\begin{aligned} ll1 &= \bar{\nu}_{mid} - \Delta\bar{\nu}_{mid} \\ ul1 &= \bar{\nu}_{mid} + \Delta\bar{\nu}_{mid} \\ ll2 &= ll1 \\ ul2 &= ul1 \end{aligned} \tag{5-13}$$

where,

$$\bar{\nu}_{mid} = \Omega_0 + \frac{\bar{\nu} - \Omega_0}{\text{factor}} \frac{\Omega_g}{\Omega_1} \tag{5-14a}$$

$$\Delta\bar{\nu}_{mid} = \text{slope} \times \Omega_g \tag{5-14b}$$

with factor = 20, and slope = 2.75.

The value of $\bar{\nu}_{mid}$ has the following constraint:

$$\left[\left(\bar{\nu} < \Omega_{0j} \wedge \bar{\nu} \geq \bar{\nu}_{mid}\right) \vee \left(\bar{\nu} > \Omega_{0j} \wedge \bar{\nu} \leq \bar{\nu}_{mid}\right)\right] \Rightarrow \bar{\nu}_{mid} = \bar{\nu}$$

The value of $\Delta\bar{\nu}_{mid}$ also has constraints, given by:

$$\begin{aligned} \left(\Delta\bar{\nu}_{mid}\right)_{minimum} &= \text{v min} = 2\text{cm}^{-1} \\ \left(\Delta\bar{\nu}_{mid}\right)_{maximum} &= \text{v max} = 100\text{cm}^{-1} \end{aligned}$$

5.2.2- Transmittance equation

In this case we are only interested in normal incidence, so $\varphi_0 = 0°$, which simplifies the calculations:

$$\begin{aligned} \Gamma &= \sqrt{\tilde{\varepsilon}_r} = n + ik \\ \Gamma_s &= \sqrt{\varepsilon_s} = n_s = 3.42 \\ \Gamma_0 &= n_0 = 1 \end{aligned}$$

For normal incidence there is no difference between perpendicular and parallel polarisation, so we may remove the index j from the Fresnel coefficients. Knowing that (n+ik) gives the film refractive index, we get for the Fresnel coefficients – equations (2-19) to (2-22) of Chapter 2 – the following results

$$t_1 = \tau_1 e^{i\kappa_1} = \frac{2}{1 + (n + ik)} \tag{5-15a}$$

$$t_2 = \tau_2 e^{i\kappa_2} = \frac{2(n+ik)}{(n+ik)+n_s} \tag{5-15b}$$

$$r_1 = \rho_1 e^{i\phi_1} = \frac{1-(n+ik)}{1+(n+ik)} \tag{5-15c}$$

$$r_2 = \rho_2 e^{i\phi_2} = \frac{(n+ik)-n_s}{(n+ik)+n_s} \tag{5-15d}$$

From the transmitted amplitude we get that,

$$t = \frac{t_1 t_2 \exp(i\delta\Gamma)}{1+r_1 r_2 \exp(2i\delta\Gamma)} = \frac{\tau_1\tau_2 \exp\{i[\kappa_1+\kappa_2+\delta(n+ik)]\}}{1+\rho_1\rho_2 \exp\{i[\phi_1+\phi_2+2\delta(n+ik)]\}} \Rightarrow$$

$$\Rightarrow |t|^2 = \left|\frac{\tau_1\tau_2 \exp\{i[\kappa_1+\kappa_2+\delta(n+ik)]\}}{1+\rho_1\rho_2 \exp\{i[\phi_1+\phi_2+2\delta(n+ik)]\}}\right|^2 =$$

$$= \left|\frac{\tau_1\tau_2 \exp(-\delta k)\exp[i(\kappa_1+\kappa_2+\delta n)]}{1+\rho_1\rho_2 \exp(-2\delta k)\exp[i(\phi_1+\phi_2+2\delta n)]}\right|^2$$

Since $|a+ib|^2 = a^2+b^2$, we have that $\left|e^{i\theta}\right|^2 = |\cos\theta + i\mathrm{sen}\theta|^2 = 1$. Applying this result to the equation for $|t|^2$ we get

$$|t|^2 = \frac{\tau_1^2\tau_2^2 \exp(-2\delta k)}{\left|1+\rho_1\rho_2 \exp(-2\delta k)\cos(\phi_1+\phi_2+2\delta n) + i\rho_1\rho_2 \exp(-2\delta k)\sin(\phi_1+\phi_2+2\delta n)\right|^2} =$$

$$= \frac{\tau_1^2\tau_2^2 \exp(-2\delta k)}{\left[1+\rho_1\rho_2 \exp(-2\delta k)\cos(\phi_1+\phi_2+2\delta n)\right]^2 + \left[\rho_1\rho_2 \exp(-2\delta k)\sin(\phi_1+\phi_2+2\delta n)\right]^2} =$$

$$= \frac{\tau_1^2\tau_2^2 \exp(-2\delta k)}{1+\rho_1^2\rho_2^2 \exp(-4\delta k) + 2\rho_1\rho_2 \exp(-2\delta k)\cos(\phi_1+\phi_2+2\delta n)} =$$

$$= \frac{\tau_1^2\tau_2^2}{\exp(2\delta k) + \rho_1^2\rho_2^2 \exp(-2\delta k) + 2\rho_1\rho_2 \cos(\phi_1+\phi_2+2\delta n)}$$

So,

$$T = \frac{\Gamma_s}{\Gamma_0}|t|^2 = \frac{n_s\tau_1^2\tau_2^2}{\exp(2\delta k) + \rho_1^2\rho_2^2 \exp(-2\delta k) + 2\rho_1\rho_2 \cos(\phi_1+\phi_2+2\delta n)} \tag{5-16}$$

Now we have to write $\tau_1, \tau_2, \rho_1, \rho_2, \phi_1$ and ϕ_2 in terms of n and k. They are as follows — derived in Appendix C:

$$\tau_1^2 = \frac{4}{(1+n)^2 + k^2} \tag{5-17a}$$

$$\tau_2^2 = \frac{4(n^2+k^2)}{\left(n+n_s\right)^2+k^2} \qquad \text{(5-17b)}$$

$$\rho_1^2 = \frac{(1-n)^2+k^2}{(1+n)^2+k^2} \Rightarrow \rho_1 = \sqrt{\frac{(1-n)^2+k^2}{(1+n)^2+k^2}} \qquad \text{(5-17c)}$$

$$\phi_1 = a\cos\frac{r_1}{\sqrt{r_1^2+i_1^2}} \quad , \; i_1 < 0 \Rightarrow \phi_1 = 2\pi - \phi_1 \qquad \text{(5-17d)}$$

$$\rho_2^2 = \frac{(n-n_s)^2+k^2}{(n+n_s)^2+k^2} \Rightarrow \rho_2 = \sqrt{\frac{(n-n_s)^2+k^2}{(n+n_s)^2+k^2}} \qquad \text{(5-17e)}$$

$$\phi_2 = a\sin\frac{i_2}{\sqrt{r_2^2+i_2^2}}, \quad r_2 < 0 \Rightarrow \phi_2 = \pi - \phi_2 \qquad \text{(5-17f)}$$

$$r_2 \geq 0 \wedge i_2 < 0 \Rightarrow \phi_2 = 2\pi + \phi_2$$

where, $r_1 = 1\text{-}n^2\text{-}k^2$; $i_1 = -2k$; $r_2 = n^2 - n_s^2 + k^2$; and $i_2 = 2n_sk$.

So, if we replace equations (5-17) into equation (5-16) we will get a transmission intensity that depends only on n and k. But our intention is to write T as a function of the parameters to fit. For that we use equations (2-33) and (2-34) from Chapter 2 (repeated below) that give n and k as functions of ε_1 and ε_2, and the equations of section 5.2.1.

$$n = \sqrt{\frac{\varepsilon_1 + \sqrt{\varepsilon_1^2+\varepsilon_2^2}}{2}} \qquad \text{(5-18)}$$

$$k = \sqrt{\frac{-\varepsilon_1 + \sqrt{\varepsilon_1^2+\varepsilon_2^2}}{2}} \qquad \text{(5-19)}$$

5.2.3- Reflectance equation

In this case we will consider any angle of incidence (between 0° and 90° , obviously) and both types of polarisation.

First of all we are going to separate the real and imaginary parts of $\Gamma = \sqrt{\tilde{\varepsilon}_r - \varepsilon_0 \sin^2 \varphi_0}$.

$$\Gamma = \sqrt{\tilde{\varepsilon}_r - \varepsilon_0 \sin^2 \varphi_0} = \sqrt{\left(n+ik\right)^2 - \sin^2 \varphi_0} = \sqrt{\left(n^2-k^2-\sin^2\varphi_0\right)+i2nk} = \beta e^{i\alpha} \Rightarrow$$

$$\Rightarrow \beta^2 e^{i2\alpha} = \left(n^2-k^2-\sin^2\varphi_0\right)+i2nk \Leftrightarrow$$

$$\Leftrightarrow \beta^2\cos(2\alpha) = n^2-k^2-\sin^2\varphi_0 \wedge \beta^2\sin(2\alpha) = 2nk$$

If we make: $sqjrl = n^2 - k^2 - \sin^2\varphi_0 \wedge sqjim = 2nk$, we will get that:

$$\beta^2\cos(2\alpha) = sqjrl \wedge \beta^2\sin(2\alpha) = sqjim \Rightarrow$$

$$\Rightarrow\begin{cases}\beta^4 = \text{sqjrl}^2 + \text{sqjim}^2\beta = \left(\text{sqjrl}^2 + \text{sqjim}^2\right)^{1/4} \\ \cos(2\alpha) = \dfrac{\text{sqjrl}}{\beta^2}\end{cases} \Leftrightarrow \begin{cases}\beta = \left(\text{sqjrl}^2 + \text{sqjim}^2\right)^{1/4} \\ \cos(2\alpha) = \dfrac{\text{sqjrl}}{\left(\text{sqjrl}^2 + \text{sqjim}^2\right)^{1/2}}\end{cases} \Leftrightarrow$$

$$\Leftrightarrow\begin{cases}\beta = \left(\text{sqjrl}^2 + \text{sqjim}^2\right)^{1/4} \\ \alpha = \dfrac{1}{2}\,\text{a}\cos\left[\dfrac{\text{sqjrl}}{\left(\text{sqjrl}^2 + \text{sqjim}^2\right)^{1/2}}\right]\end{cases}$$

Since 2nk is always greater or equal than zero, angle 2α is comprised between 0° and 180°. This implies that $\alpha \in \left[0^o, 90^o\right]$.

The equation for reflectance is given by,

$$r_j = \frac{r_{1j} + r_{2j}\exp(2i\delta\Gamma)}{1 + r_{1j}r_{2j}\exp(2i\delta\Gamma)} \Rightarrow R_j = \left|\frac{r_{1j} + r_{2j}\exp(2i\delta\Gamma)}{1 + r_{1j}r_{2j}\exp(2i\delta\Gamma)}\right|^2 =$$

$$= \left|\frac{\rho_{1j}\exp\left(i\phi_{1j}\right) + \rho_{2j}\exp\left[i\left(\phi_{2j} + 2\delta\Gamma\right)\right]}{1 + \rho_{1j}\rho_{2j}\exp\left[i\left(\phi_{1j} + \phi_{2j} + 2\delta\Gamma\right)\right]}\right|^2 = \left|\frac{\rho_{1j}\exp\left(i\phi_{1j}\right) + \rho_{2j}\exp\left[i\left(\phi_{2j} + 2\delta\beta e^{i\alpha}\right)\right]}{1 + \rho_{1j}\rho_{2j}\exp\left[i\left(\phi_{1j} + \phi_{2j} + 2\delta\beta e^{i\alpha}\right)\right]}\right|^2 =$$

$$= \left|\frac{\rho_{1j}\cos\left(\phi_{1j}\right) + i\rho_{1j}\sin\left(\phi_{1j}\right) + \rho_{2j}\exp\left\{i\left[\phi_{2j} + 2\delta\beta(\cos(\alpha) + i\text{sen}(\alpha))\right]\right\}}{1 + \rho_{1j}\rho_{2j}\exp\left\{i\left[\phi_{1j} + \phi_{2j} + 2\delta\beta(\cos(\alpha) + i\text{sen}(\alpha))\right]\right\}}\right|^2 =$$

$$= \left|\frac{\rho_{1j}\cos\left(\phi_{1j}\right) + i\rho_{1j}\sin\left(\phi_{1j}\right) + \rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\exp\left\{i\left[\phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\}}{1 + \rho_{1j}\rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\exp\left\{i\left[\phi_{1j} + \phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\}}\right|^2 =$$

$$= \frac{\left|\rho_{1j}\cos\left(\phi_{1j}\right) + i\rho_{1j}\sin\left(\phi_{1j}\right) + \rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\exp\left\{i\left[\phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\}\right|^2}{\left|1 + \rho_{1j}\rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\exp\left\{i\left[\phi_{1j} + \phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\}\right|^2}$$

We will develop separately the numerator and the denominator of the above equation. Starting with the numerator (nume) we get:

$$\rho_{1j}\cos\left(\phi_{1j}\right) + i\rho_{1j}\sin\left(\phi_{1j}\right) + \rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\exp\left\{i\left[\phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\} =$$
$$= \rho_{1j}\cos\left(\phi_{1j}\right) + i\rho_{1j}\sin\left(\phi_{1j}\right) + \rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\left\{\cos\left[\phi_{2j} + 2\delta\beta\cos(\alpha)\right] + i\sin\left[\phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\} =$$
$$= \left\{\rho_{1j}\cos\left(\phi_{1j}\right) + \rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\cos\left[\phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\} +$$
$$+ i\left\{\rho_{1j}\sin\left(\phi_{1j}\right) + \rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\sin\left[\phi_{2j} + 2\delta\beta\cos(\alpha)\right]\right\}$$

The square of the modulus of this expression is:

$$\left\{\rho_{1j}\cos\left(\phi_{1j}\right)+\rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\cos\left[\phi_{2j}+2\delta\beta\cos(\alpha)\right]\right\}^2+$$
$$+\left\{\rho_{1j}\sin\left(\phi_{1j}\right)+\rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\sin\left[\phi_{2j}+2\delta\beta\cos(\alpha)\right]\right\}^2=$$
$$=\left[\rho_{1j}\cos\left(\phi_{1j}\right)+\rho_{2j}\exp(\mathrm{var}\,1)\cos\left(\mathrm{var}\,2_j\right)\right]^2+\left[\rho_{1j}\sin\left(\phi_{1j}\right)+\rho_{2j}\exp(\mathrm{var}\,1)\sin\left(\mathrm{var}\,2_j\right)\right]^2=$$
$$=\rho_{1j}^2+\rho_{2j}^2\exp(2\,\mathrm{var}\,1)+2\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\left[\sin\left(\phi_{1j}\right)\sin(\mathrm{var}\,2)+\cos\left(\phi_{1j}\right)\cos\left(\mathrm{var}\,2_j\right)\right]\Leftrightarrow$$
$$\Leftrightarrow \mathrm{nume}=\rho_{1j}^2+\rho_{2j}^2\exp(2\,\mathrm{var}\,1)+2\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\cos\left(\phi_{1j}-\mathrm{var}\,2_j\right)$$

where $\mathrm{var}\,1=-2\delta\beta\sin(\alpha)$ and $\mathrm{var}\,2_j=\phi_{2j}+2\delta\beta\cos(\alpha)$.

For the denominator (dene) we get:

$$1+\rho_{1j}\rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\exp\left\{i\left[\phi_{1j}+\phi_{2j}+2\delta\beta\cos(\alpha)\right]\right\}=$$
$$=1+\rho_{1j}\rho_{2j}\exp\left[-2\delta\beta\sin(\alpha)\right]\left\{\cos\left[\phi_{1j}+\phi_{2j}+2\delta\beta\cos(\alpha)\right]+i\sin\left[\phi_{1j}+\phi_{2j}+2\delta\beta\cos(\alpha)\right]\right\}=$$
$$=1+\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\cos\left(\phi_{1j}+\mathrm{var}\,2_j\right)+i\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\sin\left(\phi_{1j}+\mathrm{var}\,2_j\right)$$

The square of the modulus of this expression is:

$$=\left\{1+\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\cos\left(\phi_{1j}+\mathrm{var}\,2_j\right)\right\}^2+\left\{\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\sin\left(\phi_{1j}+\mathrm{var}\,2_j\right)\right\}^2\Leftrightarrow$$
$$\Leftrightarrow \mathrm{dene}=1+r_{1j}^2r_{2j}^2\exp(2\mathrm{var}1)+2r_{1j}r_{2j}\exp(\mathrm{var}1)\cos\left(\phi_{1j}+\mathrm{var}2_j\right)$$

So, reflectance is given by the equation

$$R_j=\frac{\mathrm{nume}}{\mathrm{dene}}\Leftrightarrow R_j=\frac{\rho_{1j}^2+\rho_{2j}^2\exp(2\,\mathrm{var}\,1)+2\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\cos\left(\phi_{1j}-\mathrm{var}\,2_j\right)}{1+\rho_{1j}^2\rho_{2j}^2\exp(2\,\mathrm{var}\,1)+2\rho_{1j}\rho_{2j}\exp(\mathrm{var}\,1)\cos\left(\phi_{1j}+\mathrm{var}\,2_j\right)} \tag{5-20}$$

Now we have to write $\rho_{1j},\rho_{2j},\phi_{1j}$ and ϕ_{2j} in terms of n and k. They are as follows — derived in Appendix D.

- If the electric field vector is parallel to the plane of incidence (j = //) then:

$$\begin{cases}\rho_{1//}=\sqrt{\dfrac{v_{1//}-v_{2//}}{v_{1//}+v_{2//}}}\\ \phi_{1//}=a\cos\left[\dfrac{\left(n^2+k^2\right)\cos^2\left(\varphi_0\right)-\zeta^2}{\sqrt{v_{1//}^2-v_{2//}^2}}\right]\end{cases} \tag{5-21a}$$

$$\left[k\cos(\gamma)-n\sin(\gamma)\right]<0\Rightarrow\phi_{1//}=2\pi-\phi_{1//} \tag{5-21b}$$

where,

$$\zeta=\left[\left(r^2+s^2\right)^2+2\left(s^2-r^2\right)+1\right]^{1/4};\ \gamma=\frac{1}{2}a\cos\left[\frac{\left(1-r^2+s^2\right)}{\zeta^2}\right];\ r=\frac{n\sin\left(\varphi_0\right)}{n^2+k^2};\ s=\frac{k\sin\left(\varphi_0\right)}{n^2+k^2};$$

$$v_{1//}=\left(n^2+k^2\right)\cos^2\left(\varphi_0\right)+\zeta^2;\qquad v_{2//}=2\zeta\cos\left(\varphi_0\right)\left[n\cos(\gamma)+k\sin(\gamma)\right]$$

$$\begin{cases} \rho_{2//} = \sqrt{\dfrac{v_{3//} - v_{4//}}{v_{3//} + v_{4//}}} \\ \phi_{2//} = a\cos\left[\dfrac{n_s^2\zeta^2 - \left(n^2 + k^2\right)\cos^2(\varphi_2)}{\sqrt{v_{3//}^2 - v_{4//}^2}}\right] \end{cases} \tag{5-22a}$$

$$\left[n\sin(\gamma) - k\cos(\gamma)\right] < 0 \Rightarrow \phi_{2//} = 2\pi - \phi_{2//} \tag{5-22b}$$

where,

$$\cos(\varphi_2) = \sqrt{\frac{n_s^2 - \sin^2(\varphi_0)}{n_s^2}} ;$$

$$v_{3//} = n_s^2\zeta^2 + \left(n^2 + k^2\right)\cos^2(\varphi_2) ; \qquad v_{4//} = 2n_s\zeta\cos(\varphi_2)\left[n\cos(\gamma) + k\sin(\gamma)\right]$$

- If the electric field vector is perpendicular to the plane of incidence ($j = \perp$) then:

$$\begin{cases} \rho_{1\perp} = \sqrt{\dfrac{v_{1\perp} - v_{2\perp}}{v_{1\perp} + v_{2\perp}}} \\ \phi_{1\perp} = a\cos\left[\dfrac{\cos^2(\varphi_0) - \zeta^2\left(n^2 + k^2\right)}{\sqrt{v_{1\perp}^2 - v_{2\perp}^2}}\right] \end{cases} \tag{5-23a}$$

$$\left[k\cos(\gamma) + n\sin(\gamma)\right] > 0 \Rightarrow \phi_{1\perp} = 2\pi - \phi_{1\perp} \tag{5-23b}$$

where,

$$v_{1\perp} = \cos^2(\varphi_0) + \zeta^2\left(n^2 + k^2\right) ; \qquad v_{2\perp} = 2\zeta\cos(\varphi_0)\left[n\cos(\gamma) - k\sin(\gamma)\right]$$

$$\begin{cases} \rho_{2\perp} = \sqrt{\dfrac{v_{3\perp} - v_{4\perp}}{v_{3\perp} + v_{4\perp}}} \\ \phi_{2\perp} = a\cos\left[\dfrac{\zeta^2\left(n^2 + k^2\right) - n_s^2\cos^2(\varphi_2)}{\sqrt{v_{3\perp}^2 - v_{4\perp}^2}}\right] \end{cases} \tag{5-24a}$$

$$\left[n\sin(\gamma) + k\cos(\gamma)\right] < 0 \Rightarrow \phi_{2\perp} = 2\pi - \phi_{2\perp} \tag{5-24b}$$

where,

$$v_{3\perp} = n_s^2\cos^2(\varphi_2) + \zeta^2\left(n^2 + k^2\right) ; \qquad v_{4\perp} = 2n_s\zeta\cos(\varphi_2)\left[n\cos(\gamma) - k\sin(\gamma)\right]$$

So, if we replace equations (5-21) to (5-24) into equation (5-20) we will get a reflectance equation that depends only on n and k. But our intention is, once again, to write R as a function of the parameters to fit. For that we do the same thing as was done for transmitance, i.e., use equations (5-18) and (5-19) that give n and k as functions of ε_1 and ε_2, and the equations of section 5.2.1.

5.3- Model for gaussian decomposition

For gaussian decomposition, the calculations are much simpler. For a transmittance spectrum, decomposition with L gaussians will be given by the following equation:

$$y_T(\bar{\nu}) = T_{max} - \sum_{j=1}^{L} A_j \exp\left[-\frac{1}{2}\left(\frac{\bar{\nu} - B_j}{C_j}\right)^2\right] \tag{5-25}$$

where T_{max} gives the maximum value of T in the spectral region to be analysed, and is the gaussians baseline; A_j , B_j, and C_j are the parameters to adjust, and they give the amplitude, the central frequency, and the width (standard deviation), respectively, of the jth gaussian.

The minus sign after T_{max} means that the gaussians are inverted, that is, they point downwards, like the transmittance peaks.

For a reflectance spectra, the model is given by:

$$y_R(\bar{\nu}) = R_{min} + \sum_{j=1}^{L} A_j \exp\left[-\frac{1}{2}\left(\frac{\bar{\nu} - B_j}{C_j}\right)^2\right] \tag{5-26}$$

where Rmin gives the miniimum value of R in the spectral region to be analysed, and is the gaussians baseline. All the other values keep their previous meanings.

The plus sign after R_{min} means that the gaussians point upwards, like the reflectance peaks.

References

[1] GROSSE, P., HARBECKE, B., HEINZ, and MEYER R.; OFFENBERG, M., Appl. Phys. A 39, 257-268 (1986)

6 - Experimental procedure

To test our software, and the theories discussed in Chapter 1 related to the asymmetric stretch peak spectral region in the a-SiO_2, oxide films were prepared, characterized by ellipsometry, and their ir-transmittance and ir-reflectance spectra obtained for normally incident and obliquely incident light, respectively.

My participation in the above mentioned experimental procedure was limited to the measurement of the infrared spectra, which took place at the "Laboratoire des Plasmas et Couches Minces" in the "Institut des Matériaux de Nantes", France.

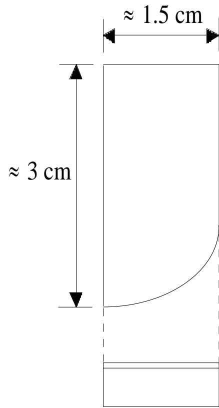


Figure 6.1- Cut portion of the oxidized wafer used to obtain the ir spectra.

The oxide films, ranging in thickness from 16 to 527.5 nm were thermally grown in water vapor (wet oxidation) at ≈ 1000°C on crystalline one-side polished silicon wafers. The diameter, thickness, orientation, and doping of these wafers in not known. The wafer diameter and thickness are not important to our work. It is sufficient to know that the wafer thickness (in the order of the millimetres) is far more greater than the oxide film thickness. The diameter is not important since the infrared spectra were obtained from a portion of the oxidized wafer cut to fit inside the spectrometer chamber where the sample support was located (figure 6.1). However, remembering Chapter 1, we realize that orientation and doping are two parameters important to our work, since they both affect the oxide film properties. Orientation modify the film density, and the dopants, if any, may modify the film structure weakening it.

The precise oxidation procedures, such as the wafers treatment prior to oxidation, the characteristics of the oxidation furnace, the temperature steps taken during oxidation and the time taken for each step, the H_2O (oxidizing species) partial pressure, the annealing done after oxidation (if any), and so on, are not known.

Elipsometry was used to characterize the oxide films in terms of their thickness. The obtained thickness values had an uncertainty of about 10 nm which is, for the thinner films, a very high error.

All ir-spectra where measured with a Nicolet 20 SXC FTIR (Fourier-transform infrared) single-beam spectrometer (see Appendix F) equipped with a DTGS detector (pyroelectric detector with a triglycine sulfate acid surface). In this detector, the variation of temperature caused by the IR radiation that strikes its surface induces a variation of the charge and gives rise to an electrical signal. A linear polarizer − a KRS-5 crystal with an aluminium grid − was added for the oblique incidence reflectance measurements. The step of the grid is calculated for 400–4000 cm^{-1}.

The spectra were taken at room temperature and at a pressure a little below 1 atm in order to decrease the peaks due to the molecules present in the atmosphere, such as H_2O − the chamber where the samples were placed was, in fact, a little vacuum chamber. With the sample placed in its position and the chamber hatch closed, we had to wait a few minutes until the pressure stabilized, before starting our measurements.

The nominal resolution was 4 cm^{-1} and 150 scans, over the spectral range 400 ... 4000 cm^{-1} , were done to average the data. In order to subtract the contribution of the substrate spectral features, we must have a reference spectrum which is the spectrum of the wafer with no oxide layer. The spectrum of the oxide film is then calculated as the ratio between the oxidized wafer spectrum and the reference spectrum (measured separately). So, the transmitted or reflected intensity by the oxide film is given by:

$$I_{SiO_2} = \frac{I_{SiO_2+Si}}{I_{Si}} \qquad (6\text{-}1)$$

Obviously, this ratio is always smaller or equal to 1, since the oxide film is an energy absorber, not an energy producer.

The reference spectrum was obtained by etching off the oxide with a HF solution. This method has the disadvantage of roughening the wafer surface. If the size of these roughenesses is of the same order of magnitude of the radiation wavelength that strikes the surface, then reflection by this surface is no longer specular but diffuse. If we remember that the oxide was grown on a polished wafer surface, which implies an interface a-SiO_2/Si with no roughnesses, we conclude that light in this interface is reflected specularly (mirror like). This difference of behaviour of light in the rough surface in contrast to the behaviour at the interface may have the following

consequence: when light is reflected diffusely by the rough surface, less radiation than predicted (by considering a polished surface) reaches the detector, which was positioned to receive the light reflected specularly. So, our reference spectrum is not as intense as it should be. From equation (6-1) we conclude that the value of I_{SiO_2} will be larger than expected, and may even reach values greater than 1 (or 100, in percentage).

In fact, this was what happened to most of our spectra. In order to try to reduce to a minimum this difference, program FITIR has an option which divides in a spectrum all the values by its maximum. We will see in Chapter 7 that this procedure is, in some cases, not enough, though.

So, the main goal of our work, the quantitative analysis of the asymmetric stretching peak of SiO_2 films, loses partially its meaning. This is specially true for dispersion analysis. However, we will assume that the difference between the real and the measured I_{SiO_2} after being corrected will only "bias" a little our results, and that on the whole they are still acceptable.

Gaussian decomposition should not be affected by this problem since it is made with respect to a line which keeps always its relative position in relation to the peaks we wish to fit.

Normal incidence beam transmission spectra of the oxidized wafers were obtained using unpolarized light. Oblique incidence reflection spectra were obtained using polarized light. In this case the angles of incidence ranged from 15° to 75°.

At oblique angles of incidence, two unique polarizations of the incident beam may be distinguished, depending on how the plane of the oxidized wafer is tilted with respect to the polarization direction of the incident beam: one is the case in which the electric field vector of the incident light lies in the plane of the oxidized wafer (s polarization). The other is the case in which the magnetic field vector of the incident beam lies in the plane of the oxidized wafer (p polarization). s-polarized oblique incidence oxidized wafer reflectance spectra show the same peaks as the oxidized wafer transmission spectra obtained using normally incident unpolarized light. p-polarized oblique incidence oxidized wafer reflectance spectra exhibit new peaks in addition to the peaks found in the s-polarized reflection spectra, due to the Berreman effect (see section 1.4 of Chapter 1).

The transmittance and reflectance data were stored in files. Tables 6.1 and 6.2 summarize the main features of these spectra. They are plotted in figures 6-2 to 6-26. From these figures we see that almost all the spectra have regions with intensities greater than 100%, which is expected by the diffuse reflectance of the etched surface. However, this "theory" does not predict the observed increase of the transmittance values with film thickness. This increase is explained if we consider the film as a Fabry-Perot interferometer. To derive the transmitted interference pattern we will start by considering the oxide film as non-absorbing (which will enable a simple mathematical treatment). In this case, the transmitted intensity is obtained replacing equations (5-17) into equation (5-16), of Chapter 5, and by making k, the extinction coefficient, equal to zero:

$$T = \frac{n_s \dfrac{4}{(1+n)^2} \dfrac{4n^2}{(n+n_s)^2}}{1+\left(\dfrac{1-n}{1+n}\right)^2 \left(\dfrac{n-n_s}{n+n_s}\right)^2 + 2\dfrac{1-n}{1+n}\dfrac{n-n_s}{n+n_s}\cos(2\delta n)} \qquad (6\text{-}2)$$

where, n_s and n are the substrate and film refractive indexes, respectively, and $\delta = 2\pi d/\lambda$ (d = film thickness; λ=incident radiation wavelength in vacuum).

From equation (6-2) we see that transmittance, T, is maximum when cos(2δn) equals zero, that is,

$$2\delta n = 2m\pi \Leftrightarrow \frac{4\pi d}{\lambda} n = 2m\pi \Leftrightarrow d = \frac{m\lambda}{2n}, \quad m = 0,1,2,3,... \qquad (6\text{-}3)$$

For an absorbing film, equation (6-3) is still valid if we consider d not as the true thickness, but as an effective thickness slightly different from the true one. From this equation we conclude that the greater the film thickness the lower will be the first frequency maximum. So, as expected, we only observe the first maximum in the thickest films (484 and 527.5 nm). The other films have their maximum at frequencies out of the range of our spectra, that is, above 4000 cm^{-1}. The thinner the more distant will be this maximum, giving the idea that the intensities decrease when the film thickness also decreases.

Obviously, this Fabry-Perot interference pattern is also observed in the reflectance spectra.

We also observe in several transmittance spectra a sharp peak at ≈ 670 cm^{-1}. This peak corresponds to an absorption band of the silicon substrate, which confirms the deficiencies associated to our reference spectra. However, since it is out of the frequency range we wish to analyse, ≈[970,1300] cm^{-1}, it will be neglected.

Some of the spectra show considerable noise. With program FITIR, it is possible to smooth the spectrum with a Savitzky-Golay filter. However, some care must be taken when applying this technique, since a bias is introduced when the underlying function has a nonzero second derivative. In the spectrometric application, a narrow spectral line has its height reduced and its width increased. Since these parameters are of interest to us, the bias introduced is distinctly undesirable.

Smoothing was only applied to the reflectance spectrum s15.ref since it was the only spectrum with a large amount of noise in the wavenumber interval ≈[970,1300] cm^{-1}.

Table 6-1: Transmittance spectra files

File name	film thickness (nm)	Polarization	Angle of incidence (°)	Spectral range (cm^{-1})	Number of data points
trans1.tra	527.5	unpolarized	0	400–4000	1867
trans2.tra	422.0	unpolarized	0	400–4000	1867
trans3.tra	484.0	unpolarized	0	400–4000	1867
trans4.tra	376.5	unpolarized	0	400–4000	1867
trans5.tra	342.0	unpolarized	0	400–4000	1867
trans6.tra	254.0	unpolarized	0	400–4000	1867
trans7.tra	311.0	unpolarized	0	400–4000	1867
trans8.tra	238.0	unpolarized	0	400–4000	1867
trans9.tra	293.0	unpolarized	0	400–4000	1867
transa.tra	197.0	unpolarized	0	400–4000	1867
transb.tra	325.0	unpolarized	0	400–4000	1867
transc.tra	218.5	unpolarized	0	400–4000	1867
transd.tra	71.0	unpolarized	0	400–4000	1867
transe.tra	276.5	unpolarized	0	400–4000	1867
transf.tra	107.5	unpolarized	0	400–4000	1867
transg.tra	165.0	unpolarized	0	400–4000	1867
transh.tra	150.5	unpolarized	0	400–4000	1867
transi.tra	16.0	unpolarized	0	400–4000	1867
transj.tra	136.0	unpolarized	0	400–4000	1867
transk.tra	53.0	unpolarized	0	400–4000	1867
transl.tra	43.0	unpolarized	0	400–4000	1867

Table 6-2: Reflectance spectra files

File name	film thickness (nm)	Polarization	Angle of incidence (°)	Spectral range (cm^{-1})	Number of data points
p15.ref	520.0	p	15	400–4000	1867
s15.ref	520.0	s	15	400–4000	1867
p50.ref	520.0	p	50	400–4000	1867
s50.ref	520.0	s	50	400–4000	1867
p65.ref	520.0	p	65	400–4000	1867
s65.ref	520.0	s	65	400–4000	1867
p75.ref	520.0	p	75	400–4000	1867
s75.ref	520.0	s	75	400–4000	1867

Transmittance spectra

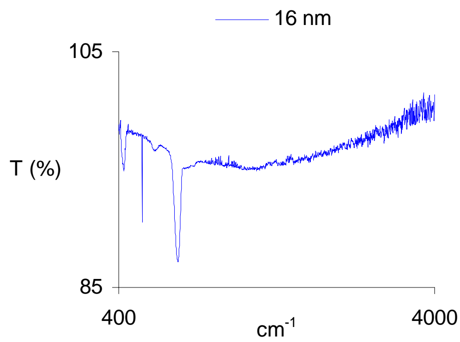


Figure 6.2 - Transmittance spectrum: d_{ox} = 16 nm

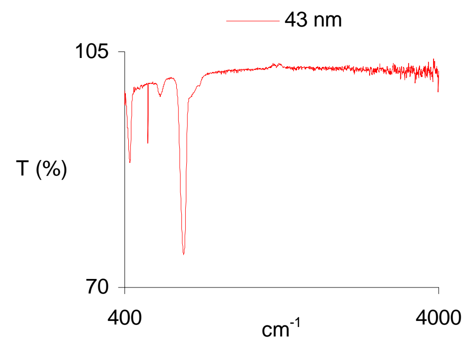


Figure 6.3 - Transmittance spectrum: d_{ox} = 43 nm

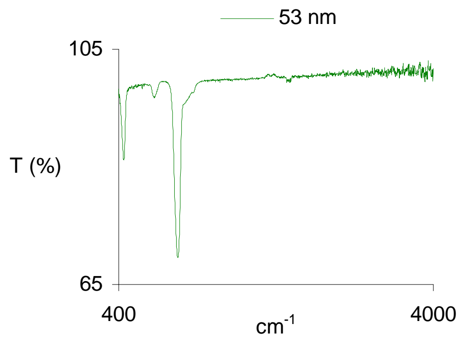


Figure 6.4 - Transmittance spectrum: d_{ox} = 53 nm

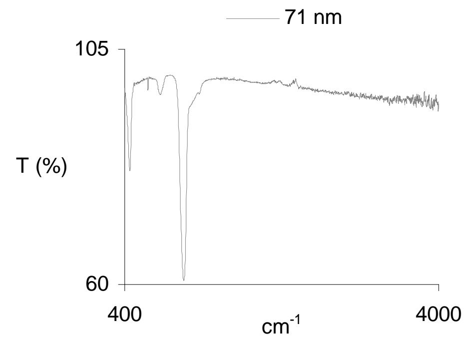


Figure 6.5 - Transmittance spectrum: d_{ox} = 71 nm

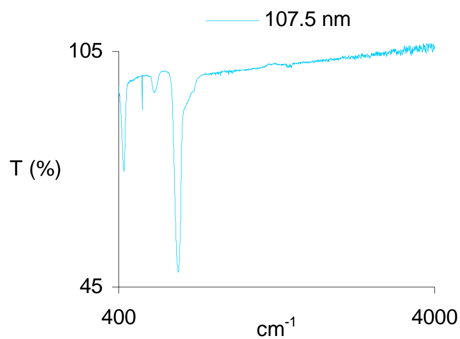


Figure 6.6 - Transmittance spectrum: d_{ox} = 107.5 nm

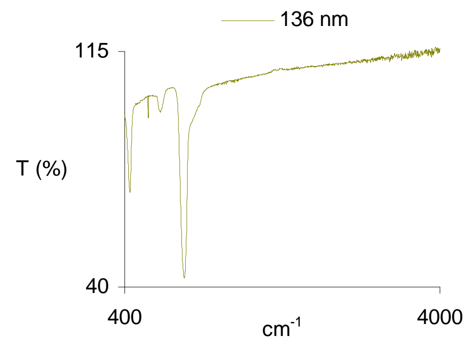


Figure 6.7 - Transmittance spectrum: d_{ox} = 136 nm

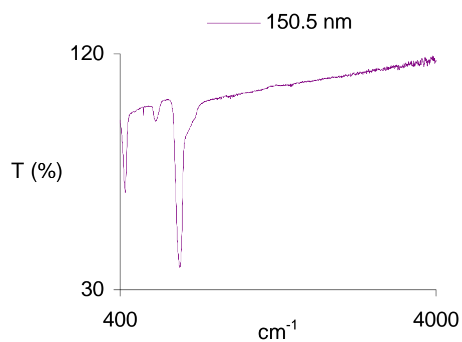


Figure 6.8 - Transmittance spectrum: d_{ox} = 150.5 nm

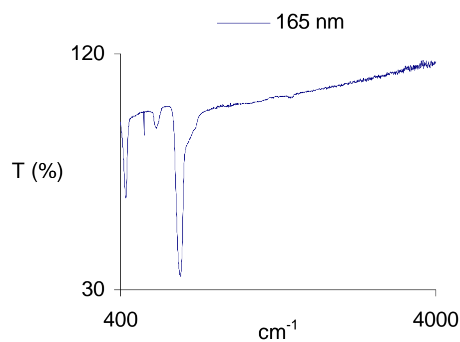


Figure 6.9 - Transmittance spectrum: d_{ox} = 165 nm

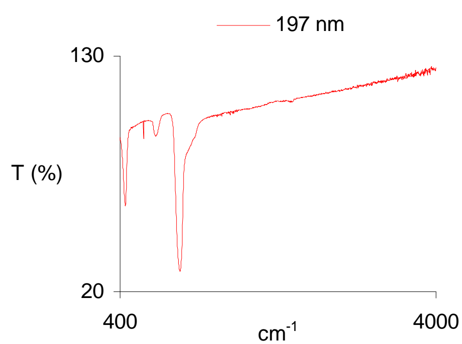


Figure 6.10 - Transmittance spectrum: d_{ox} = 197 nm

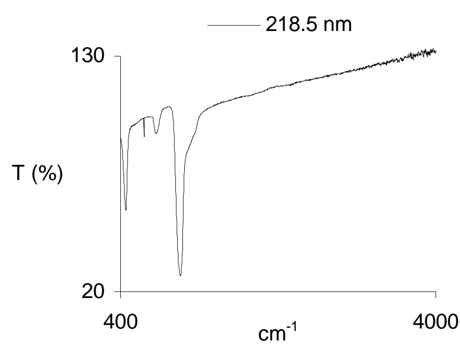


Figure 6.11 - Transmittance spectrum: d_{ox} = 218.5 nm

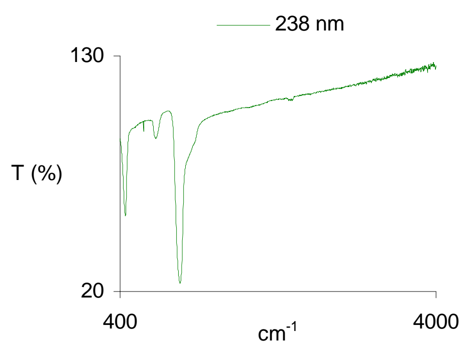


Figure 6.12 - Transmittance spectrum: d_{ox} = 238 nm

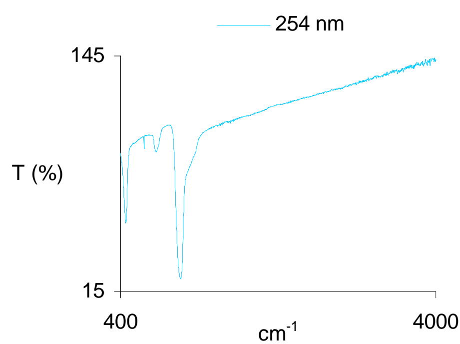


Figure 6.13 - Transmittance spectrum: d_{ox} = 254 nm

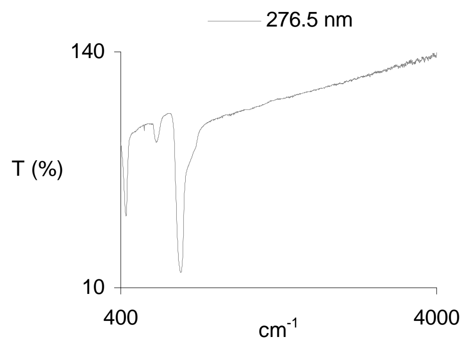


Figure 6.14 - Transmittance spectrum: d_{ox} = 276.5 nm

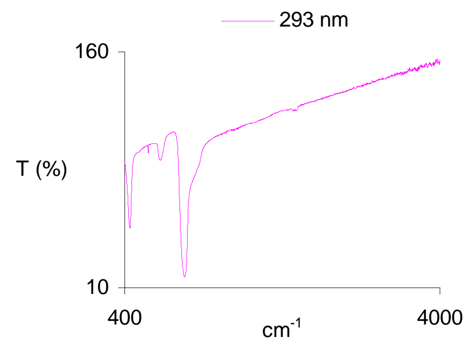


Figure 6.15 - Transmittance spectrum: d_{ox} = 293 nm

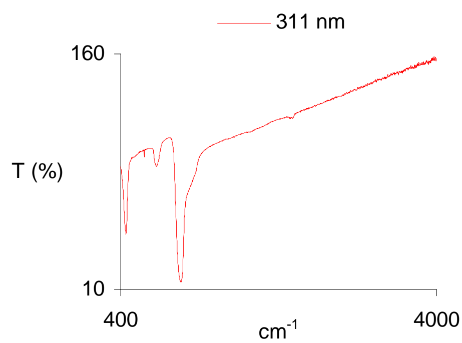


Figure 6.16 - Transmittance spectrum: d_{ox} = 311 nm

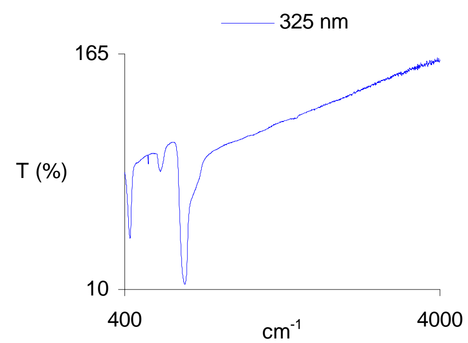


Figure 6.17 - Transmittance spectrum: d_{ox} = 325 nm

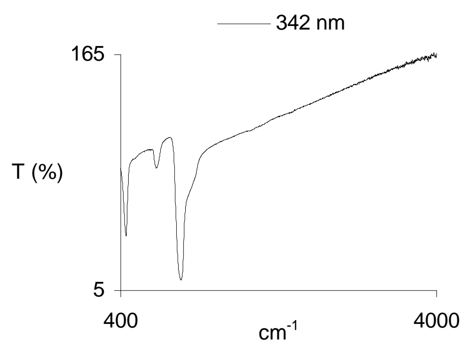


Figure 6.18 - Transmittance spectrum: d_{ox} = 342 nm

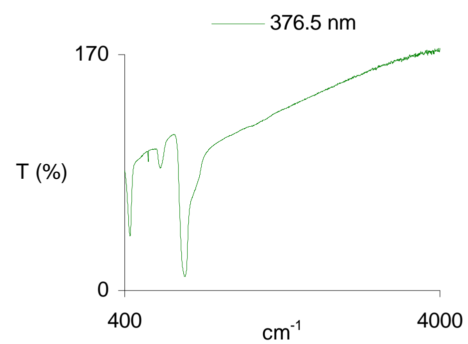


Figure 6.19 - Transmittance spectrum: d_{ox} = 376.5 nm

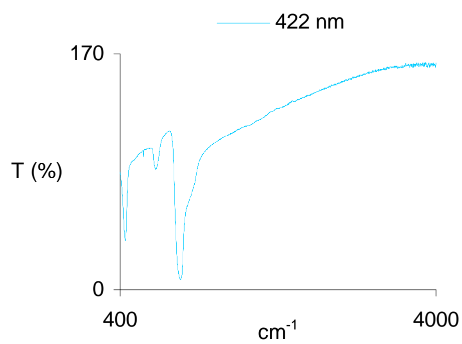


Figure 6.20 - Transmittance spectrum: d_{ox} = 422 nm

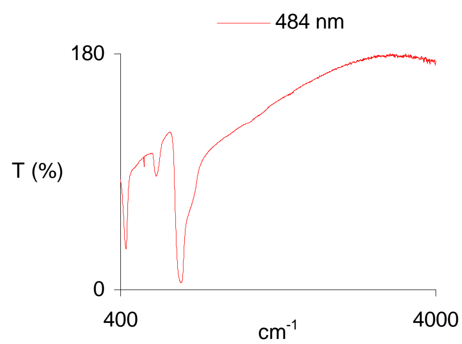


Figure 6.21 - Transmittance spectrum: d_{ox} = 484 nm

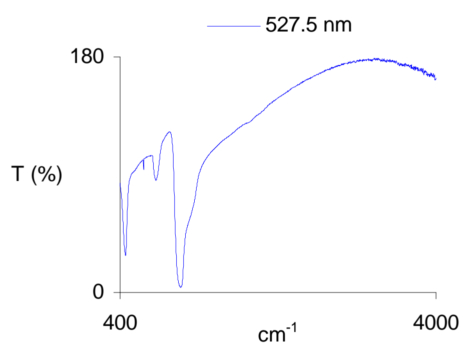


Figure 6.22 - Transmittance spectrum: d_{ox} = 527.5 nm

Reflectance spectra

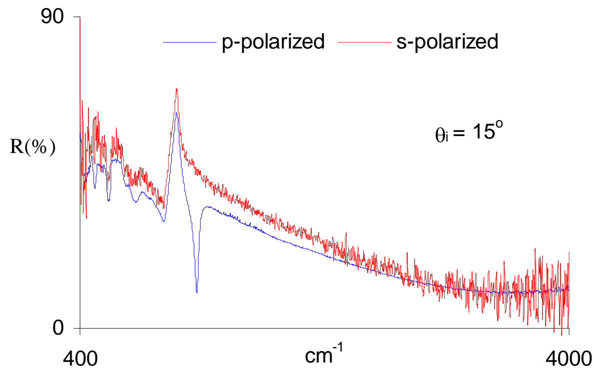


Figure 6.23 - Reflectance spectrum: $\theta_i = 15°$

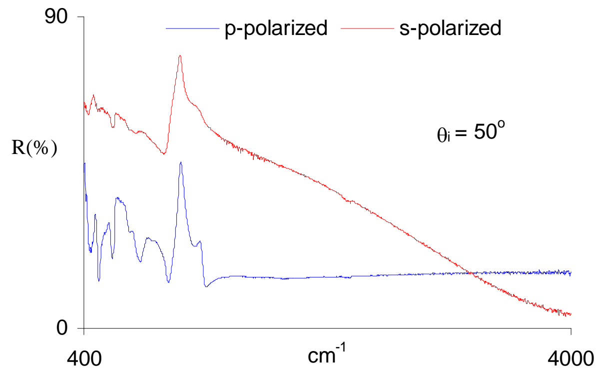


Figure 6.24 - Reflectance spectrum: $\theta_i = 50°$

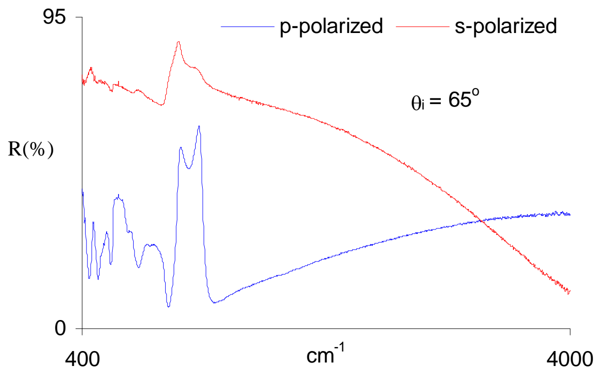


Figure 6.25 - Reflectance spectrum: $\theta_i = 65°$

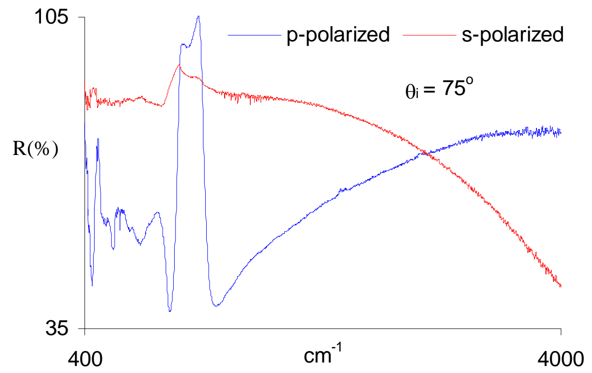


Figure 6.26 - Reflectance spectrum: $\theta_i = 75°$

7 - Spectra analysis

7.1- Introduction

All transmittance and reflectance spectra were studied by gaussian decomposition and dispersion analysis in the asymmetric stretch (AS) vibrational region.

The procedure followed was, first, try to identify, by gaussian decomposition, the number of oscillators that vibrate in the above mentioned wavenumber region. Second, see whether or not, dispersion analysis confirmed the results obtained in the first step, as well as, determine the physical properties of the vibrating dipoles.

In both steps all eight codes implemented by the program FITIR were tested. Their performance will be commented in section 7.4.

All spectra which had an intensity maximum greater than 100% were divided by this maximum. The other spectra were divided by 100.

Also, several initial sets of parameters were used in order to:

(i) make a reliable code testing;

(ii) assure that for each fit the codes are, in fact, converging to the global minimum of the χ^2 function and not to a dubious local minimum. This means that there is only one set of best-fit parameters, for each spectrum analysed.

The error bars plotted in the graphics which show the best-fit parameters correspond to their estimated standard deviations, whose values were calculated using equation (4-13) − Chapter 4.

Some of the fits performed by program FITIR and analised in this chapter are plotted in Appendix G.

7.2- Gaussian decomposition

7.2.1- Transmittance spectra

In the wavenumber interval [970,1300] cm^{-1}, all transmittance spectra were successfully fitted with three gaussians, as may be seen in the two examples of figure G.2 (Appendix G).

Fitting with two and four gaussians was also attempted. Two gaussians are insufficient to achieve a good fit, clearly evident in figure G.1 (Appendix G). Fitting with four gaussians (figure G.3 in Appendix G) has slightly reduced the SCF (standard error of the curve fit) obtained with three gaussians. The best-fit parameters and their estimated standard deviations obtained with three and four gaussians are also plotted in the graphics of figures 7.1 to 7.6, and in tables in Appendix H.

At first sight, this SCF decrease suggests that the asymmetric stretching peak decomposes not in three but four vibrational modes. However, a more careful study reveals that this is not so. Adding one gaussian to the set of three increases the estimated standard deviations of the best-fit parameters which define the two gaussians with higher mean value − referenced in the graphics as “third gaussian” and “fourth gaussian”. (The other two gaussians remain essentially the same with this addition.) This increase may be visualised by comparing the graphics of figures 7.1 to 7.3 (corresponding to the curve fits with three gaussians) with the graphics of figures 7.4 to 7.6 (corresponding to the curve fits with four gaussians). Even more, when we add the fourth gaussian, we have that the SCF reduction becomes higher as the films gets thicker. However, the estimated standard deviations of the best-fit parameters do not decrease. In fact, they increase for the “third” and “fourth” gaussian.

These two arguments are not, by themselves, quite sufficient to refuse the presence of the fourth gaussian, since even though the estimated standard deviations increase, their values are still acceptable.

The decisive argument in favour to refuse the fourth gaussian comes from the gaussian decomposition of oblique incidence s-polarised spectra (see section 7.2.2). The number of predicted peaks is the same for these reflectance spectra and for the normal incidence transmittance spectra, so the number of gaussians ought to be the

same. As we shall see in the next section, the curve fit with four gaussians fails clearly for the reflectance spectra obtained with s-polarised light. So, we must also refuse the fourth gaussian in our transmittance spectra.

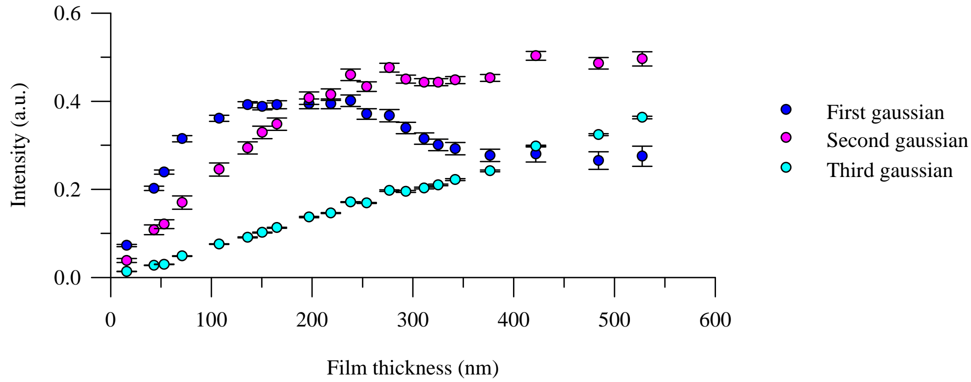


Figure 7.1 - Evolution of the gaussian profiles intensities with film thickness. Number of gaussians = 3.

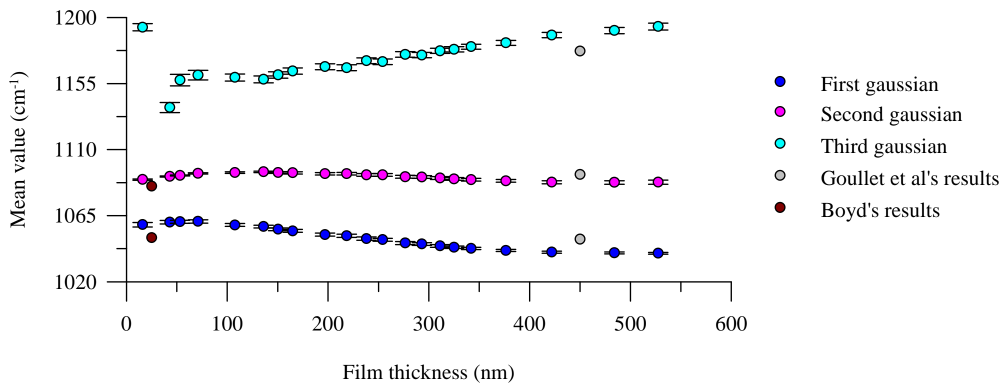


Figure 7.2 - Evolution of the gaussian profiles mean values with film thickness. Number of gaussians = 3.

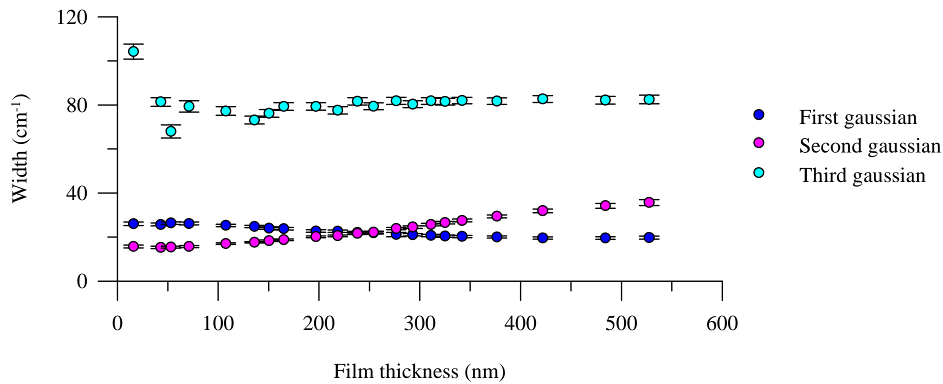


Figure 7.3 - Evolution of the gaussian profiles widths with film thickness. Number of gaussians = 3.

Evolution of the gaussian profiles intensities with film thickness

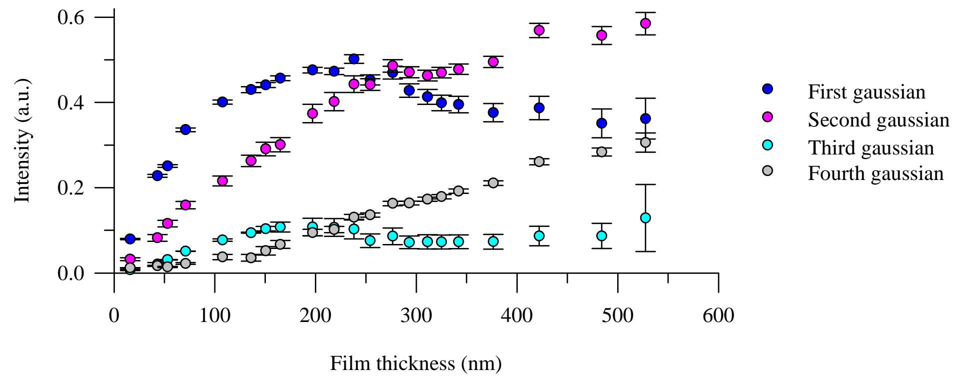


Figure 7.4 - Evolution of the gaussian profiles intensities with film thickness. Number of gaussians = 4.

Evolution of the gaussian profiles mean values with film thickness

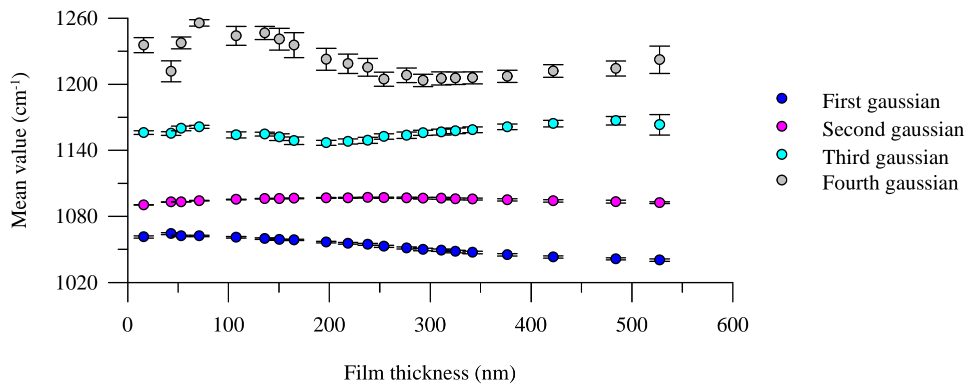


Figure 7.5 - Evolution of the gaussian profiles mean values with film thickness. Number of gaussians = 4.

Evolution of the gaussian profiles widths with film thickness

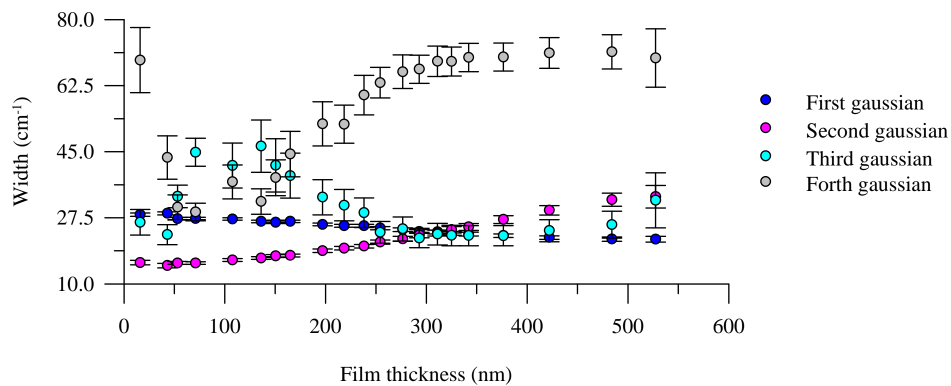


Figure 7.6 - Evolution of the gaussian profiles widths with film thickness. Number of gaussians = 4.

Now with the number of gaussian profiles confirmed to three, we will analyse in more detail the graphics of figures 7.1 to 7.3.

The gaussian profiles intensities increase with film thickness, except the lowest wavenumber gaussian (referenced in the graphics as "first gaussian") which starts decreasing for thicknesses higher than ≈136 nm. This gaussian, which is the main profile for the thinner films, loses its "leading position" to the middle and higher wavenumber gaussians (referenced in the graphics as "second gaussian" and "third gaussian", respectively).

The increase of the intensity with film thickness is expected since a greater thickness implies a higher absorption, therefore a lower transmission. So, the behaviour of the first gaussian is somewhat "odd". However, its origin is also "odd". According to Boyd [1], who was the first to signal it when studying films with thicknesses in the range 2.8 nm to 45 nm, this lower wavenumber gaussian may be the consequence of a microcrystalline or quasiperiodic structure for the silicon dioxide films thermally grown. This author has fitted the AS vibrational mode region with only two gaussians since, and contrary to us, he did not observe the high wavenumber shoulder. This may be due, for example, to different wafer oxidation conditions because we know that they may alter the oxide characteristics. (However, may these changes be so dramatic ?)

The evolution of the intensities with film thickness may be also visualised in the graphic of figure 7.7.

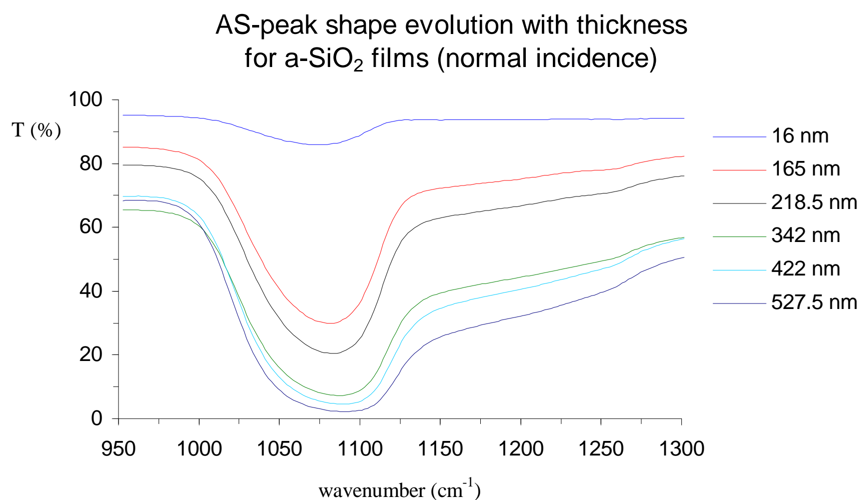


Figure 7.7: Increase of the high-wavenumber shoulder as a function of the film thickness.

For the parameter mean value we verify that for the first gaussian it remains essentially unchanged (≈ 1060 cm^{-1}) below ≈100 nm. For higher thicknesses the mean value starts decreasing reaching the final value of ≈ 1040 cm^{-1}. The second gaussian mean value only presents small variations around ≈ 1090 cm^{-1}. The third gaussian mean value shows a steady increase with film thickness [if we consider that, despite its small estimated standard deviation, the value obtained for the thinnest film studied (16 nm) is an outlier, biased by the low ratio Signal/Noise].

I do not know the explanation for this observed "movement" of the "first" and "third" gaussians. At first, one should be tempted to say that it was due to the existence of bond strains which change with film thickness. However, this may not be so, if we consider as true that silicon dioxide films have the same properties of bulk samples, except for a region approximately 3 nm thick situated at the interface wafer/film (see Chapter 1). Even more, should not changes in bond strains affect all the three peaks instead of only the lowest and highest wavenumber peaks (as well as the rocking and bending mode peaks)?

Finally, for the gaussian widths, we verify that in the case of the first gaussian, this parameter shows a steady increase with film thickness, while that for the second gaussian we observe the opposite behaviour. For the third gaussian there is almost no change. Once again, the justifications for these behaviours are not easy.

These results may be interpreted in terms of dipoles numbers, as each gaussian area is proportional to the number of dipoles that vibrate with the average mean value of the corresponding gaussian. So, we have that for very thin films the majority of the dipoles vibrate with an average wavenumber ≈ 1060 cm^{-1} (first gaussian mean value). However, as the film thickness increases, this majority decreases, "transferring" to vibrational frequencies with higher average values given by the second and third gaussians mean values. This "transfer" is so drastic that for the thickest films studied, the first gaussian represents the band with fewer dipoles, while the second and the third gaussians represent the bands with higher number of vibrating dipoles.

These changes of the gaussian parameters explain the shift of the AS peak to higher frequencies with film thickness (see Chapter 1). Figure 7.8 shows this shift in our transmittance spectra. From it, we conclude that even though the first gaussian mean value decreases with increasing film thickness, it is counteracted by the increase in the third gaussian mean value together with the relative growth in intensity of the second and third gaussian when compared with the first gaussian intensity.

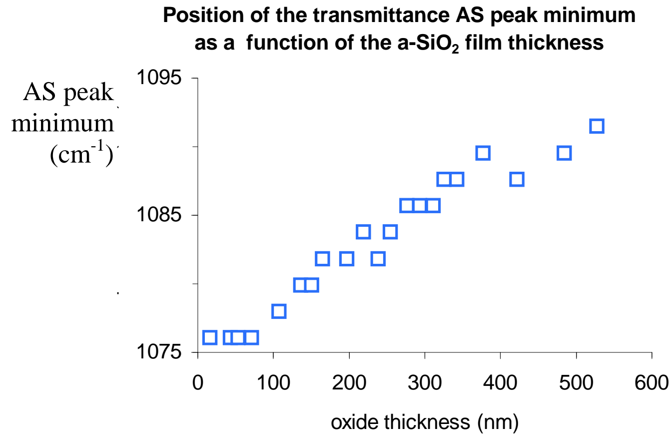


Figure 7.8: Shift of the transmittance AS peak minimum as a function of the film thickness.

The results obtained with gaussian decomposition of normal incidence transmittance spectra are in good agreement with the results obtained by Boyd [1] and by Goullet *et al*. [2].

As already mentioned, Boyd obtained two gaussians for thermally grown films with thicknesses between 2.8 nm and 45 nm. The lowest wavenumber gaussian (mean value around 1050 cm^{-1}) had a higher intensity and width than the higher wavenumber gaussian (mean value around 1085 cm^{-1}), which agrees with our results in this thickness range for the first and second gaussians. The only difference between our results and Boyd's is the absence of the third gaussian in Boyd's fits. The average mean values obtained by Boyd are plotted in the graphic of figure 7.2.

Goullet *et al*. [2] obtained three gaussian profiles for a 450 nm thick film thermally grown. Their mean values (1049 cm^{-1}, 1093 cm^{-1} , and 1177 cm^{-1})are close to ours (as may be seen in the graphic of figure 7.2). There is also agreement in the relative intensities and widths of the three gaussians.

7.2.2- Reflectance spectra

The reflectance spectra were fitted with three, four and five gaussians. Examples of these fits are plotted in figures G.4 to G.8 (Appendix G).The best-fit parameters and their estimated standard deviations are also plotted in figures 7.9 to 7.23, and in tables in Appendix H.

Some caution must be taken when analysing these fits since the adjusted parameters show some dependence on the chosen initial set of parameters. Nevertheless, we will consider them as reliable, except for the parameters obtained in the fits of the spectrum obtained with p-polarized light having an angle of incidence, φ_0, of 15°. In this particular case, the final set of parameters depended severely on their initial guess, so they had to be rejected.

We started with three gaussians (see figures G.4 and G.5 in Appendix G). Comparing the gaussian mean values obtained for s- and p-polarized light we see that they do not agree (see graphics of figures 7.10 and 7.13). We may try to explain it by considering that in the spectra obtained with p-polarized light (p-spectra) there are vibrational modes, namely the LO modes, which are not observed in the spectra obtained with s-polarized light (s-spectra). Therefore, we may try to verify the acceptability of this supposition by adding more gaussians to our fits.

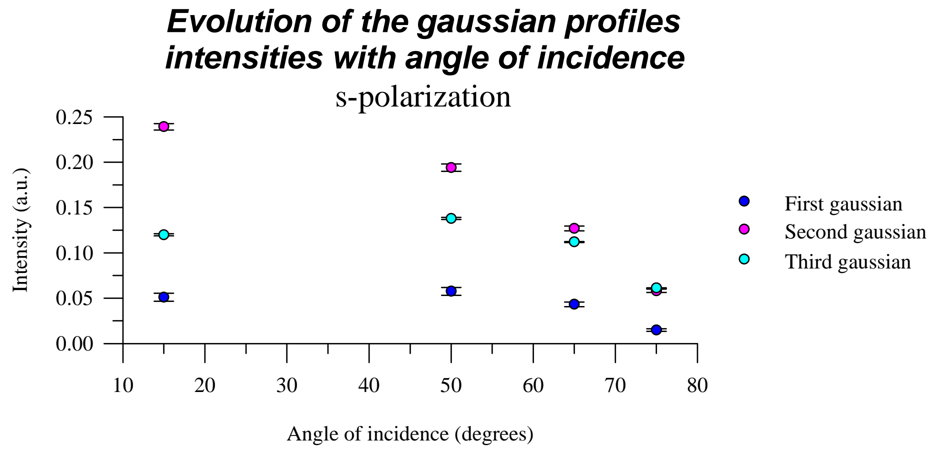


Figure 7.9 - Evolution of the gaussian profiles intensities with angle of incidence. s-polarization. Number of gaussians = 3.

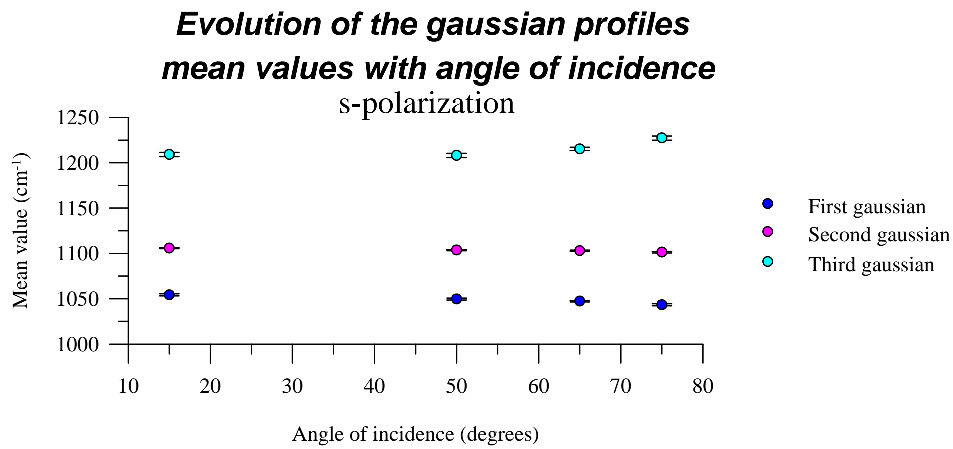


Figure 7.10 - Evolution of the gaussian profiles mean values with angle of incidence. s-polarization. Number of gaussians = 3.

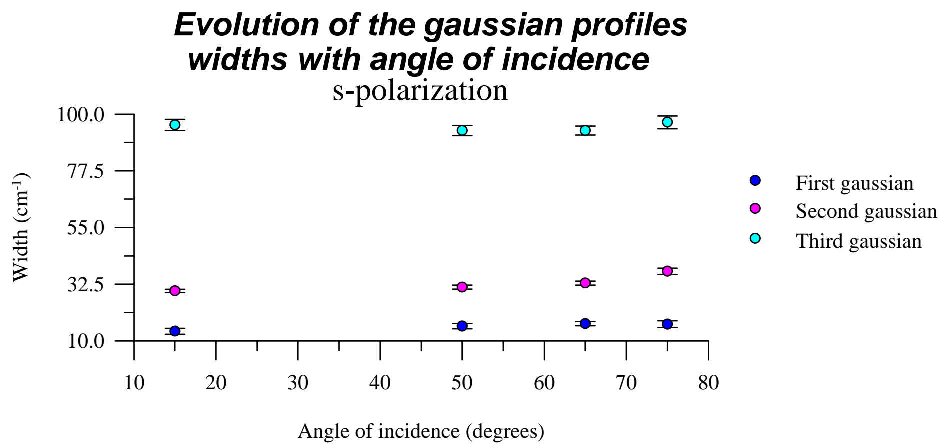


Figure 7.11 - Evolution of the gaussian profiles widths with angle of incidence. s-polarization. Number of gaussians = 3.

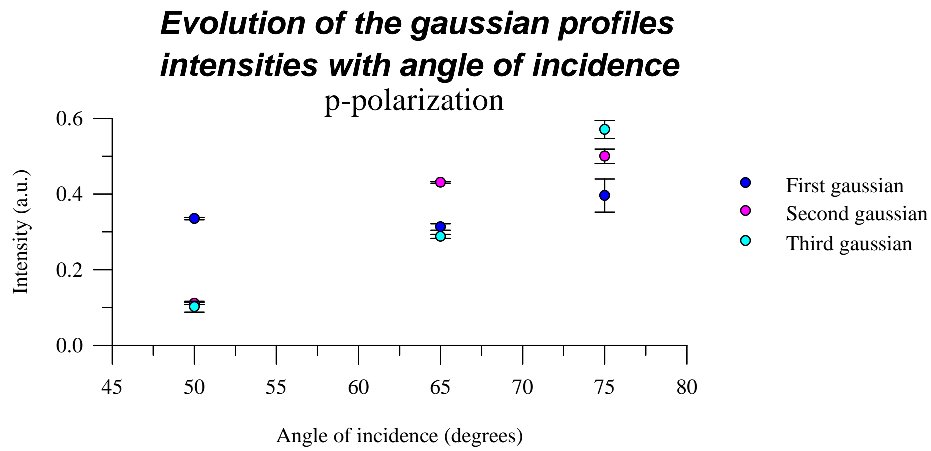


Figure 7.12 - Evolution of the gaussian profiles intensities with angle of incidence. p-polarization. Number of gaussians = 3.

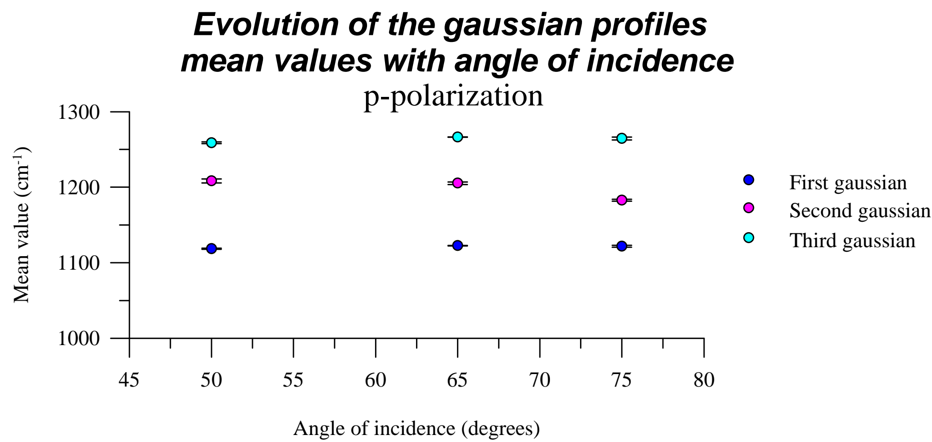


Figure 7.13 - Evolution of the gaussian profiles mean values with angle of incidence. p-polarization. Number of gaussians = 3.

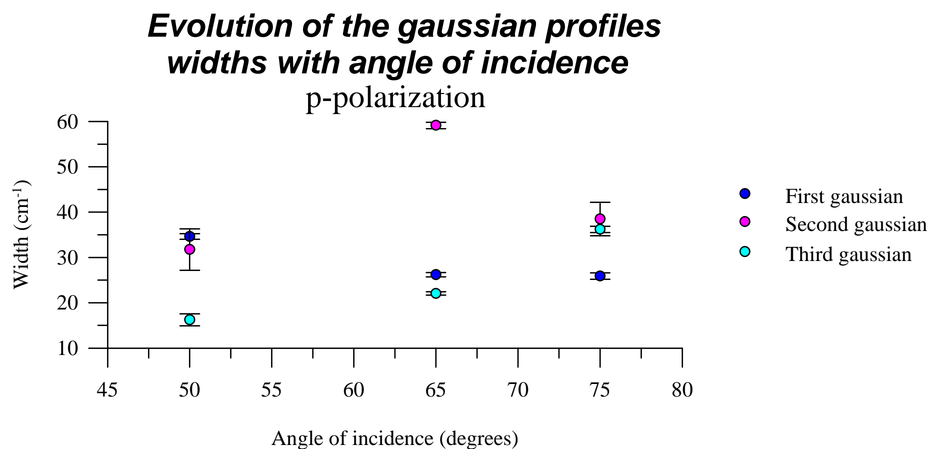


Figure 7.14 - Evolution of the gaussian profiles widths with angle of incidence. p-polarization. Number of gaussians = 3.

The fit with four gaussians of the s-spectra (figure G.6 in Appendix G) must be rejected since the estimated standard deviations of the best-fit parameters are on the whole extremely high (see graphics of figures 7.15 to 7.17). This rejection has already been mentioned in the previous section as being the strongest reason to

reject the four gaussian fit of the transmittance spectra. However, for the p-spectra (figure G.7 in Appendix G), and with the exception of the spectrum obtained with $\varphi_0 = 15°$ (rejected by the above mentioned reason), we verify that adding the fourth gaussian improves the fits obtained with three gaussians, namely in the SCF's and in the estimated standard deviations of the best-fit parameters (see graphics of figures 7.18 to 7.20).

These results seem to confirm that LO modes are only excited by oblique incidence p-polarized light.

Since, both transmitted spectra and reflectance spectra obtained with s-polarized light were adjusted with three gaussians, we conclude that there are three TO modes in the AS vibrational mode interval.

If we assume that associated to each TO mode there is a LO mode, we conclude that the p-spectra ought to be fitted by six gaussians (three gaussians for the TO modes and the other three to the LO modes). However, three problems arise. The first one is related with the fact that we do not know if all the vibrational modes show a TO-LO splitting. We only know that at least one mode shows this splitting, due to the presence of the fourth gaussian. The second problem is related with the fact that if there are really five or six vibrating modes they must be very close to each other which implies that a clear distinction of the five or six gaussians may be complicated. The third and last problem is related with the experimental procedure. The reflectance spectra were obtained with the sample positioned over a metallic mirror, which reduces the TO peaks intensities (see Chapter 1). Now, how drastic this reduction was, we do not know. We only know that it was not complete, otherwise the p-spectra would only show three modes (the LO modes) and not the four already seen.

So, due to these three problems we will try first to fit the spectra with five gaussians. Depending on the results obtained in these fits we will decide whether to proceed to the six gaussians or not.

The fits are plotted in figure G.8 (Appendix G). The fits for $\varphi_0 = 15°$ and $\varphi_0 = 75°$ were rejected. The first by the reasons already mentioned and the second because one of the gaussians had always an intensity lower than zero, which means that we may only see four modes vibrating at this angle of incidence. This is not at all unexpected if we remember, once again, that increasing the angle of incidence reduces the TO components.

From the results plotted in the graphics of figures 7.21 to 7.23 and from the obtained SCF's we accept the presence of five vibrational modes. Some of the gaussians have very small intensities, so I think that adding one more gaussian will not permits us to take any conclusion about its physical meaning. Furthermore, the best-fit parameters obtained with six gaussians are very dependent on the initial guess.

From the graphics of figures 7.10, 7.19 and 7.22, we verify that the mean values of the three gaussians obtained for the s-spectra match reasonably well some of the mean values obtained for the p-spectra when fitted by four or five gaussians. This confirms our supposition. Furthermore, based on these mean values and on their comparison we may try to distinguish the TO from the LO modes in the p-spectra, as may be seen in table 7.1. This mean values assignment is not easy, and in some cases it is very doubtfull. Nevertheless, it clearly shows that for big angles of incidence the lowest wavelength TO mode disappears which is not unexpected, since for this film thickness it is the TO mode with lowest intensity. Also plotted in this table are the gaussian mean values obtained for the normal incidence transmittance spectrum of the film 527.5 nm thick. We see that they match reasonably well the reflectance spectra values.

These results confirm Kirk's model which assigned a TO mode to the high wavenumber shoulder observed in normal incidence transmittance spectra [3] (see Chapter 1). This author predicted only two TO modes (AS_1 and AS_2) instead of the three observed by us. Some care must be taken when making this comparison since Kirk's results were obtained by dispersion analysis, and not by gaussian decomposition. However, the two methods are expected to produce similar results, at least in the number of vibrational modes.

Almeida [4] proposed that the high wavenumber shoulder was due to a mixed LO/TO mode of dominant LO character because, amongst other reasons, he did not see any evidence of a transversal mode responsible for the shoulder in reflectance spectra. However, it was shown by gaussian decomposition that the TO modes may also be present.

Table 7.1- Mean values in cm^{-1} for the TO and LO modes predicted by gaussian decomposition

φ_0	gaussians	polarization	TO_1	TO_2	TO_3	LO_1	LO_2	LO_3
0°	3	unpolarized	1039	1088	1194	–	–	–
15°	3	s	1054	1106	1209	–	–	–
50°	3	s	1050	1103	1208	–	–	–
	4	p	1062	1115	–	–	1256	1181
	5	p	1072	1115	–	1147	1263	1232
65°	3	s	1047	1104	1215	–	–	–
	4	p	–	1117	1200	1134	1261	–
	5	p	–	1117	1216	1127	1268	1232
75°	3	s	1043	1101	1227	–	–	–
	4	p	–	1117	1208	–	1262	1161
	5	p	–	–	–	–	–	–

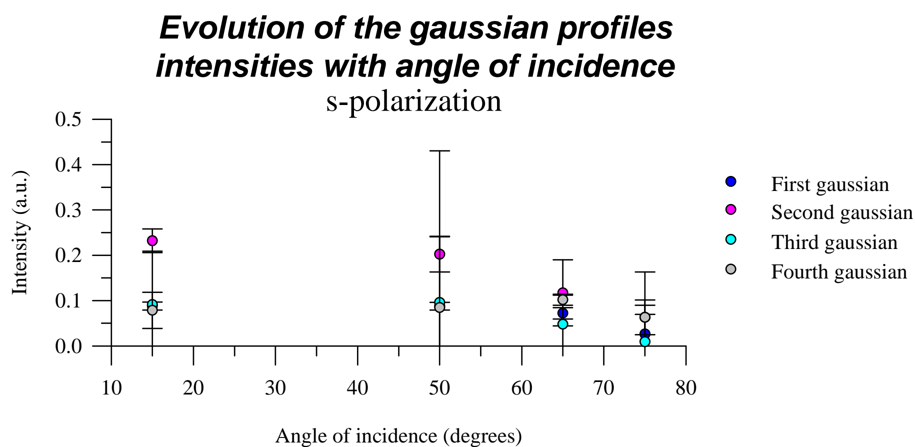


Figure 7.15 - Evolution of the gaussian profiles intensities with angle of incidence. s-polarization. Number of gaussians = 4.

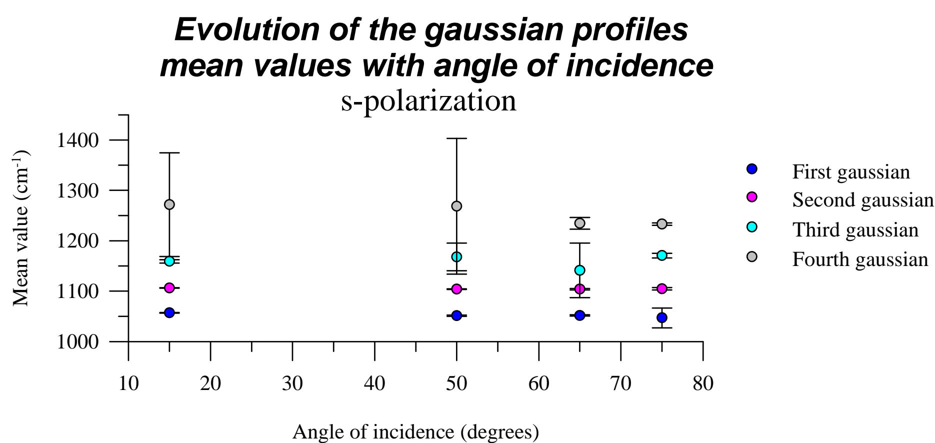


Figure 7.16 - Evolution of the gaussian profiles mean values with angle of incidence. s-polarization. Number of gaussians = 4.

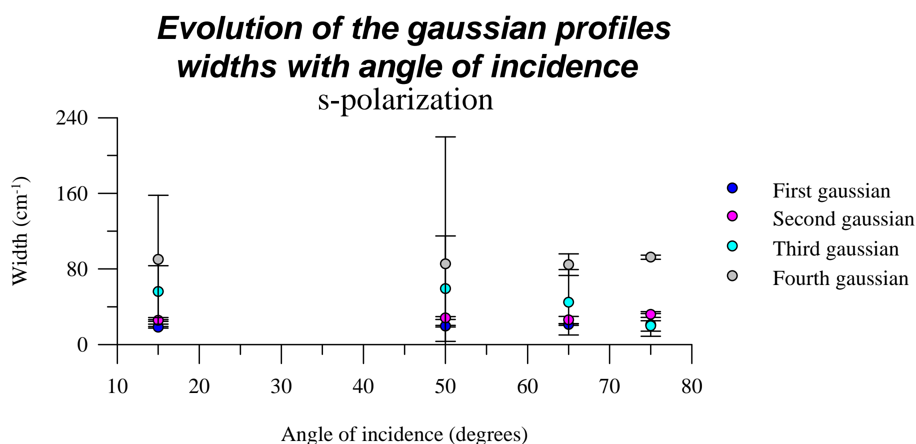


Figure 7.17 - Evolution of the gaussian profiles widths with angle of incidence. s-polarization. Number of gaussians = 4.

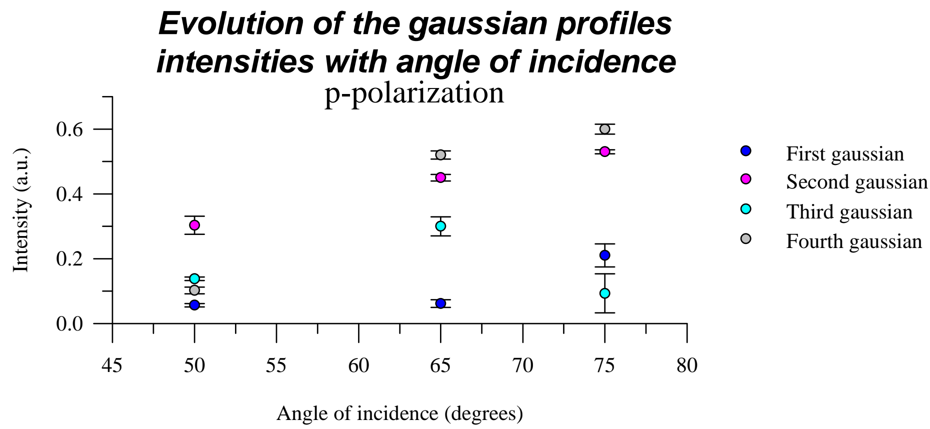


Figure 7.18 - Evolution of the gaussian profiles intensities with angle of incidence. p-polarization. Number of gaussians = 4.

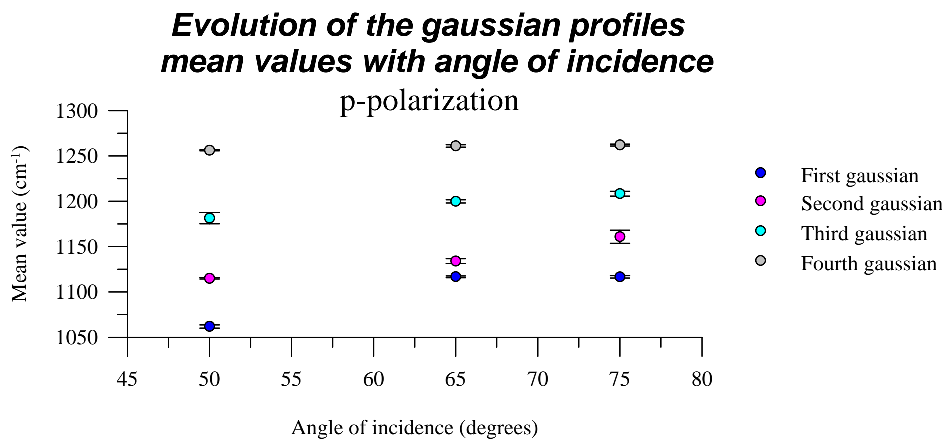


Figure 7.19 - Evolution of the gaussian profiles mean values with angle of incidence. p-polarization. Number of gaussians = 4.

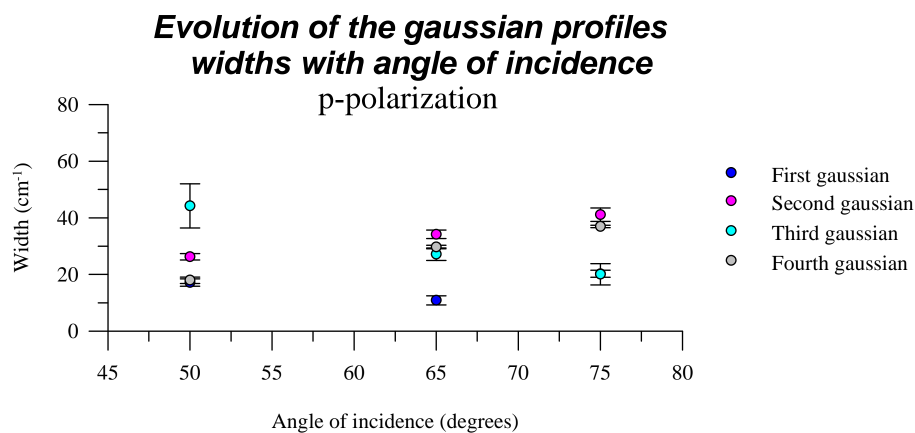


Figure 7.20 - Evolution of the gaussian profiles widths with angle of incidence. p-polarization. Number of gaussians = 4.

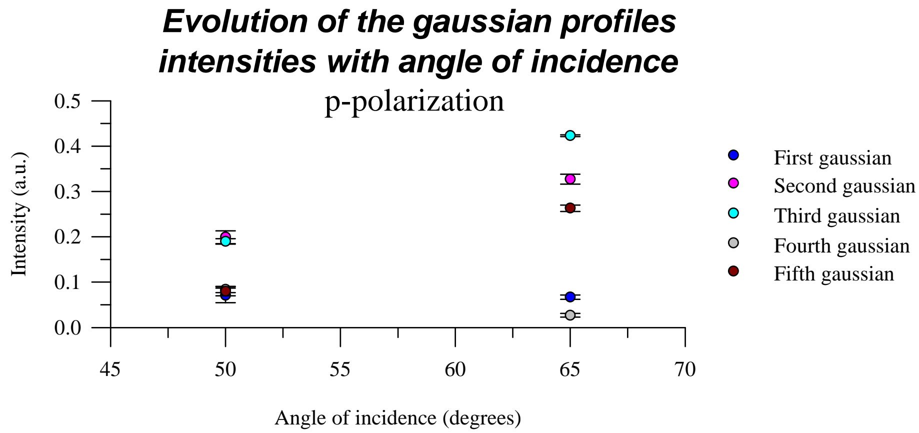


Figure 7.21 - Evolution of the gaussian profiles intensities with angle of incidence. p-polarization. Number of gaussians = 5.

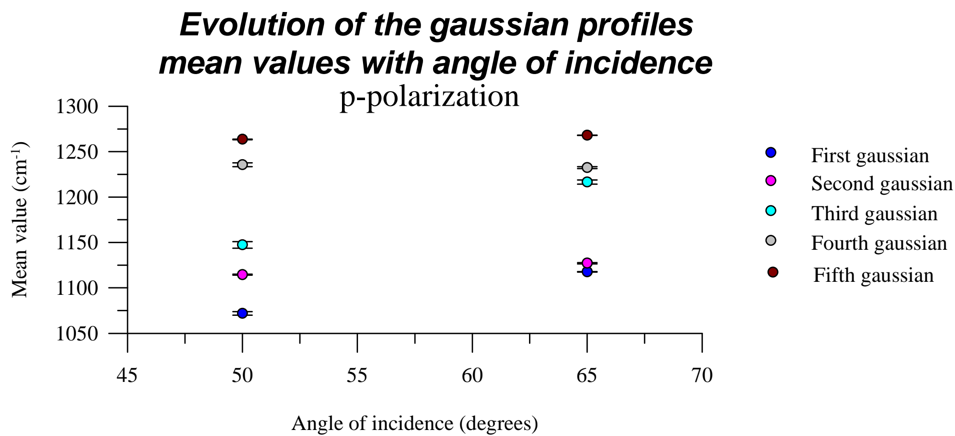


Figure 7.22 - Evolution of the gaussian profiles mean values with angle of incidence. p-polarization. Number of gaussians = 5.

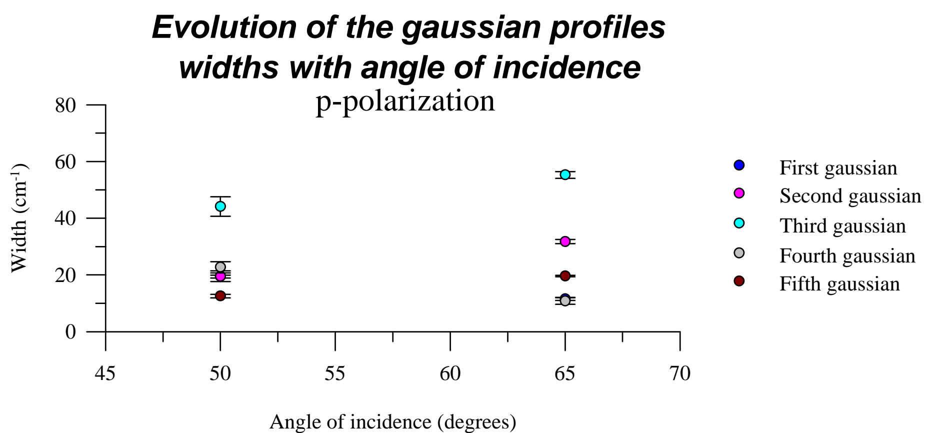


Figure 7.23 - Evolution of the gaussian profiles widths with angle of incidence. p-polarization. Number of gaussians = 5.

7.3- Dispersion analysis

The results obtained by gaussian decomposition strongly suggests that only TO modes are observed in normal incidence spectra, which means that the model for the dielectric dispersion, explained in Chapter 3, seems to be reasonably correct.

7.3.1- Transmittance spectra

Dispersion analysis was tried on these spectra using the dielectric function model for crystalline materials and for amorphous materials (see Chapter 3). As expected, the first model is not able to fit the spectral region we are interested in, so it was rejected. For the second model, good fits were accomplished for films thicker than ≈ 250 nm using three lorentzian oscillators. Some examples of these fits are plotted in figure G.9 (Appendix G). Fits with only two oscillators were also tried, but the results were not good. It confirms the results obtained by gaussian decomposition concerning the number of TO vibrational modes.

The fact that dispersion analysis does not converge for the thinner films is not at all unexpected. In Chapter 6, an explanation was given for the unusual high values of transmittance in our samples. The only procedure taken to correct this situation was to divide all the transmittance values in a spectrum by its maximum value, which was considered to be the highest value transmittance could take for that particular film. However, we know that is not true because, since the oxide film "works" like a Fabry-Perot interferometer, its transmittance maximum is, for the thinner films, far beyond the 4000 cm^{-1} limit (see Chapter 6), and this distance increases with the decrease of film thickness. Therefore, the implemented correction procedure becomes less efficient as the film thickness decreases, because the higher is the difference between the maximum value in the interval [400,4000] cm^{-1}, and the first maximum value of the interferometer. This means that the dispersion parameters are "biased", so some care must be taken when analysing them.

A transmittance value superior to the real value is equivalent to consider that the film has an effective thickness lower than its actual value (lower thickness ⇒ lower absorption ⇒ higher transmittance). Since the transmittance Fresnel equations are film thickness dependent (see Chapter 2), it is to be expected that dispersion analysis does not converge. It would have converged if we had introduced in Fresnel equations, not the actual thickness value, but a smaller one.

Analysing the graphics of figures 7.24, 7.25, 7.26, 7.30 and 7.31 (and respective tables in Appendix H) that show the evolution of the dispersion parameters with film thickness, as well as their estimated standard deviations, we verify that:

(1) the dielectric constant for optical frequencies, ε_∞ , depends on the film thickness. Its behaviour is not very steady, changing between ≈ 1.3 (d_{ox} = 276.5 nm) and ≈ 4.5 (d_{ox} = 325.0 nm). However, its value seems to stabilize at ≈ 2.4 for the thicker films. This value agrees with the value in found in literature (e.g., Kirk [3], Grosse *et al.* [5]).

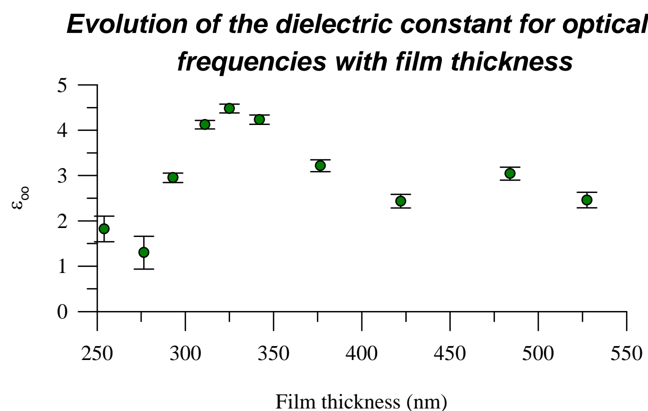


Figure 7.24 - Evolution of the dielectric constant for optical frequencies with film thickness. Number of lorentzian oscillators with gaussian distribution of frequencies = 3.

(2) The oscillators strengths, Ω_p, show a similar behaviour to the gaussian intensities (see section 7.2.1). The major difference lies in the behaviour of the high-wavenumber shoulder ("third" oscillator): while the gaussian intensity shows steady increase with film thickness, the oscillator strength increases, decreases and stabilizes for the thicker films.

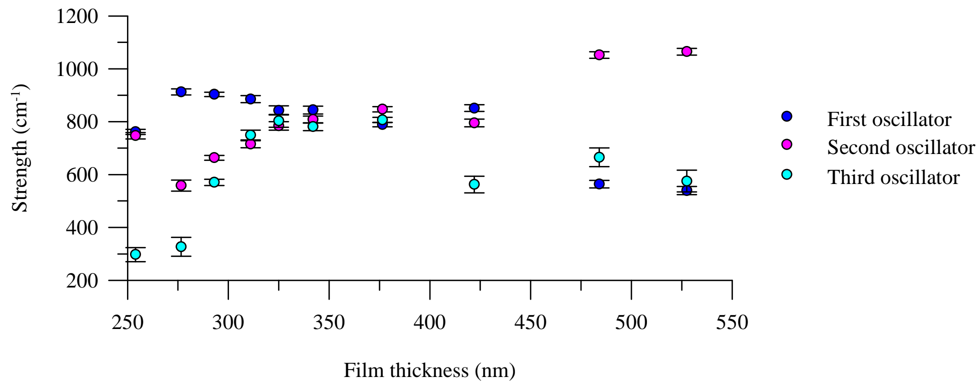


Figure 7.25 - Evolution of the strengths with film thickness, for the three lorentzian oscillators with gaussian distribution of frequencies.

(3) The oscillators resonance frequencies, Ω_o, also show a similar behaviour to the gaussias mean values, even though their values do not exactly match. The differences are plotted in the graphics of figures 7.27, 7.28, and 7.29 (values in Appendix H), for the lowest, middle, and highest wavenumber peaks, respectively. From these graphics we see that for the first peak the resonance frequency, obtained by dispersion analysis, has always a higher value than the mean value obtained by gaussian decomposition. The maximum difference is $\approx$ 17 cm^{-1} (d_{ox} = 276.5 nm) and the lowest $\approx$ 4 cm^{-1} (d_{ox} = 527.5 nm). For the second peak, and contrary to the first, the resonance frequency, obtained by dispersion analysis, has always a lower value than the mean value obtained by gaussian decomposition. The maximum difference is $\approx$ 11 cm^{-1} (d_{ox} = 527.5 nm) and the lowest $\approx$ 2 cm^{-1} (d_{ox} = 311.0 nm). The peak that shows closer agreement between the resonance frequency, obtained by dispersion analysis, and the mean value, obtained by gaussian decomposition, is the third. Except for the thinner films (d_{ox} = 254.0, 276.5 and 293.0 nm) the resonance frequency has a lower value than the gaussian mean value. The maximum difference is $\approx$ 11 cm^{-1} (d_{ox} = 527.5 nm) and the lowest $\approx$ 4 cm^{-1} (d_{ox} = 311.0 nm).

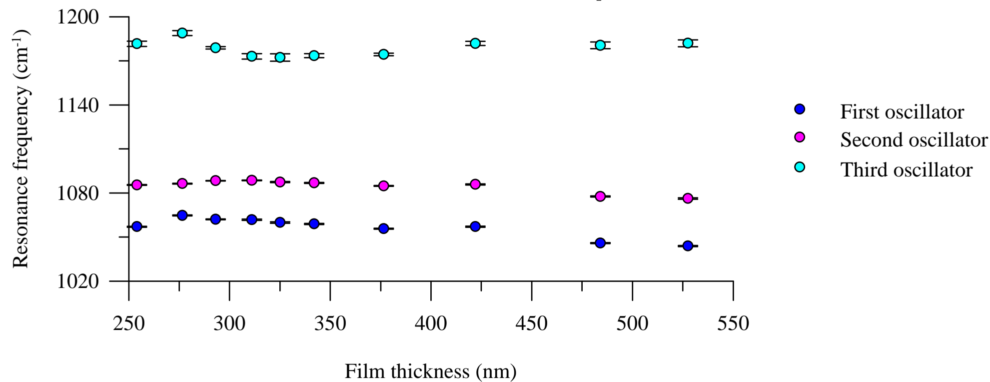


Figure 7.26 - Evolution of the resonance frequencies with film thickness, for the three lorentzian oscillators with gaussian distribution of frequencies.

So we have that the gaussians are more separated from each other than the lorentzian oscillators. We have tried to use the gaussian mean values in dispersion analysis and the other way around, that is, use resonance frequencies in gaussian decomposition (the resonance frequencies and mean values were inhibited from changing by making the fit with code MRQMIN – it permits to select the parameters that will remain unchanged during the fit). In both cases the fits did not converge.

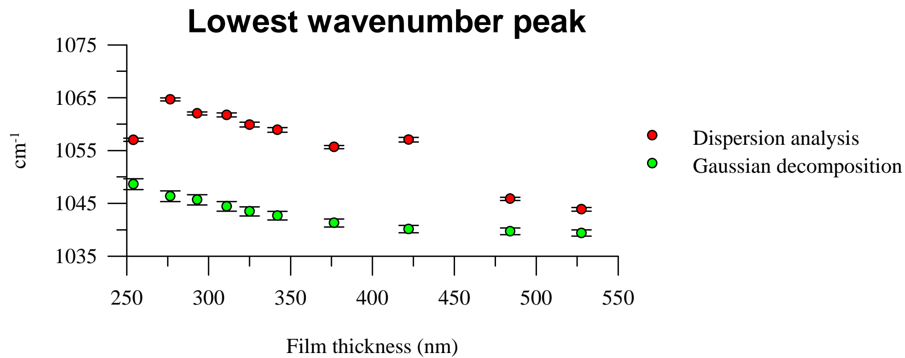


Figure 7.27 - Resonance frequencies (obtained by dispersion analysis) and mean values (obtained by gaussian decomposition) for the lowest wavenumber peak.

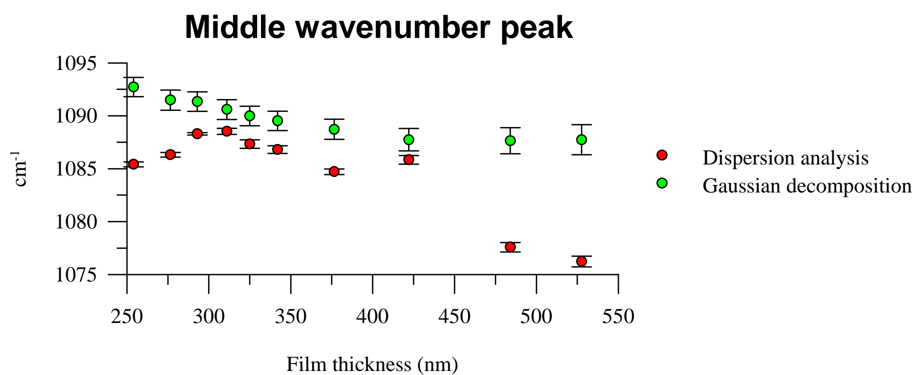


Figure 7.28 - Resonance frequencies (obtained by dispersion analysis) and mean values (obtained by gaussian decomposition) for the middle wavenumber peak.

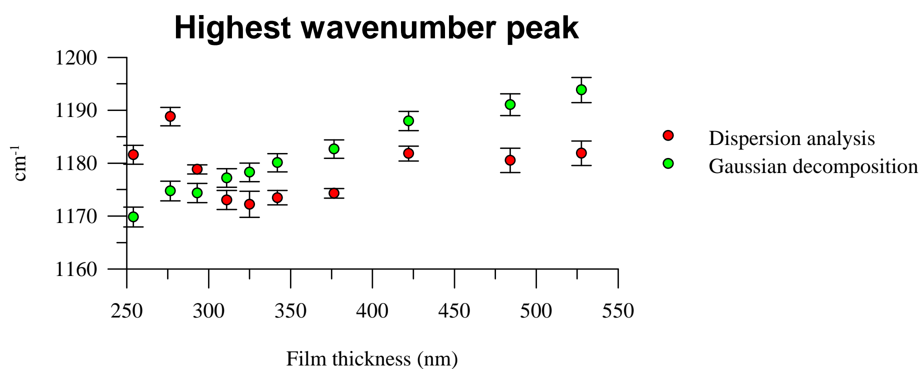


Figure 7.29 - Resonance frequencies (obtained by dispersion analysis) and mean values (obtained by gaussian decomposition) for the highest wavenumber peak.

(4) The parameter lorentzian width, Ω_l, remains essentially unchanged with film thickness. The "first" and "second" oscillators lorentzian widths are about the same, 5 cm^{-1}. The "third" oscillator shows a more erratic behaviour. Even though some values have a high estimated standard deviation, we may consider that it oscillates between ≈ 11 cm^{-1} (d_{ox} = 376.5 nm) and ≈ 17 cm^{-1} (d_{ox} = 484.0 nm).

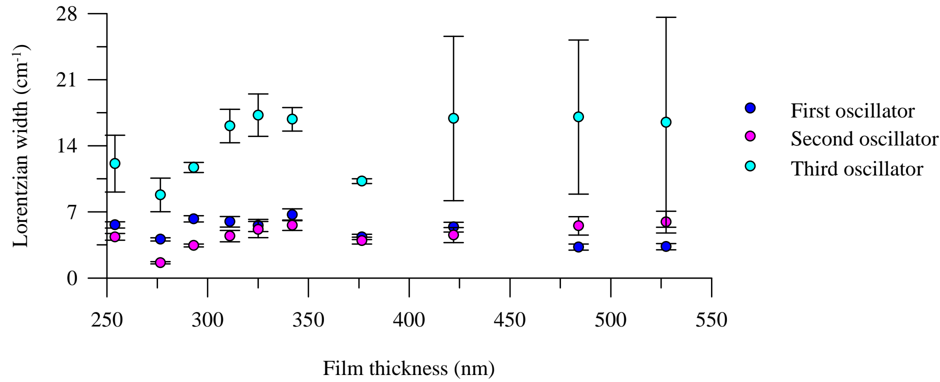


Figure 7.30 - Evolution of the lorentzian widths with film thickness, for the three lorentzian oscillators with gaussian distribution of frequencies.

(5) The "first" oscillator gaussian width, Ω_{g1}, shows a small decrease with thickness, from ≈ 23 cm^{-1} to ≈14 cm^{-1}. For the "second" oscillator this parameter is almost constant, ≈ 16 cm^{-1}. For the "third" oscillator, despite the big oscillations in this parameter (between ≈ 47 cm^{-1} and ≈ 118 cm^{-1}), it seems to stabilize at ≈ 80 cm^{-1} for the thicker films.

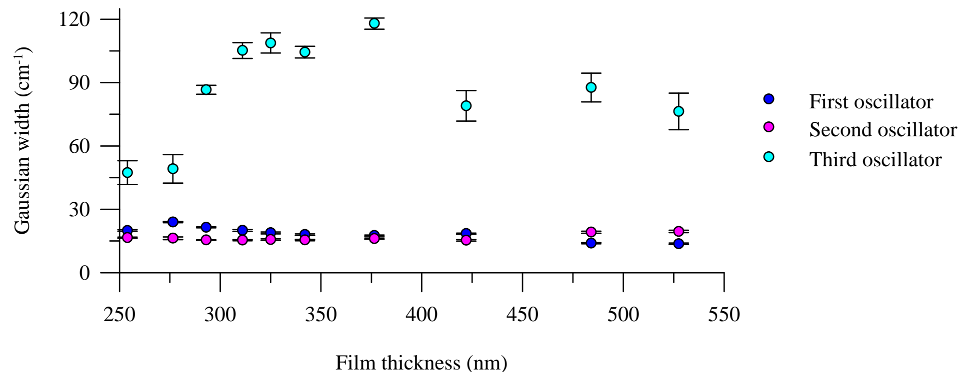


Figure 7.31 - Evolution of the gaussian widths with film thickness, for the three lorentzian oscillators with gaussian distribution of frequencies.

The results obtained by dispersion analysis are in reasonable agreement with Kirk´s results [3] except for the absence of the lower wavenumber oscillator in Kirk's fits. The TO resonance frequencies obtained by this author for films ≈ 100 nm, are 1076 cm^{-1} and 1200 cm^{-1}.

With the dispersion parameters found, it is possible to determine some intrinsic properties of the amorphous silicon dioxide vibrational modes, namely (see Chapter 3):

$$\begin{cases} \Omega_{pj}^2 = \dfrac{N_j\left(e_j^*\right)^2}{\left(2\pi c\right)^2 \mu_j \varepsilon_0} \\ \Omega_{0j} = \overline{\nu}_{0j} \\ \Omega_{lj} = \dfrac{\gamma_j}{2\pi c} \end{cases} \Leftrightarrow \begin{cases} \dfrac{N_j\left(e_j^*\right)^2}{\mu_j} = \left(2\pi c\right)^2 \varepsilon_0 \Omega_{pj}^2 \\ \overline{\nu}_{0j} = \Omega_{0j} \\ \gamma_j = 2\pi c \Omega_{lj} \end{cases} \tag{7-1}$$

where N_j is the number of dipoles per unit volume that oscillate with a resonance frequency $\overline{\nu}_{0j}$; e_j^* is the effective dipolar charge; μ_j is the reduced mass of each of the N_j dipoles; and γ_j is the N_j dipoles damping coefficient. $\varepsilon_0 = 8.854\times10^{-14}$ F/m is the permittivity of vacuum.

The values of $N_j\left(e_j^*\right)^2 / \mu_j$ and γ_j are plotted in the graphics of figures 7.32 and 7.33. The error bars were computed by the familiar error propagation formula.

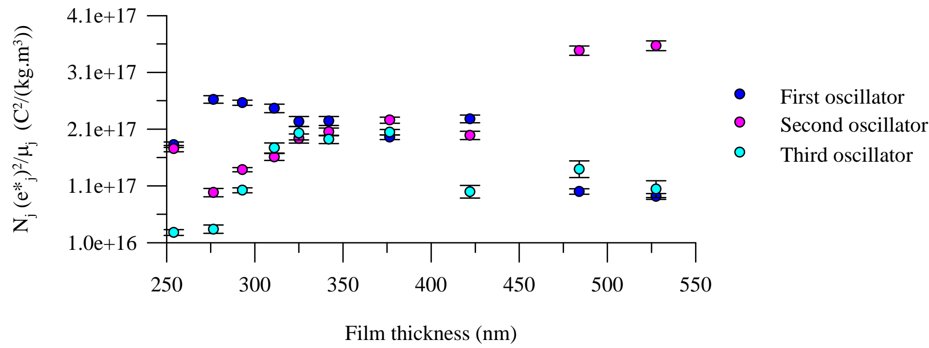


Figure 7.32 - Evolution of the ratios $N_j\left(e_j^*\right)^2 / \mu_j$, (j=1,2,3) with film thickness.

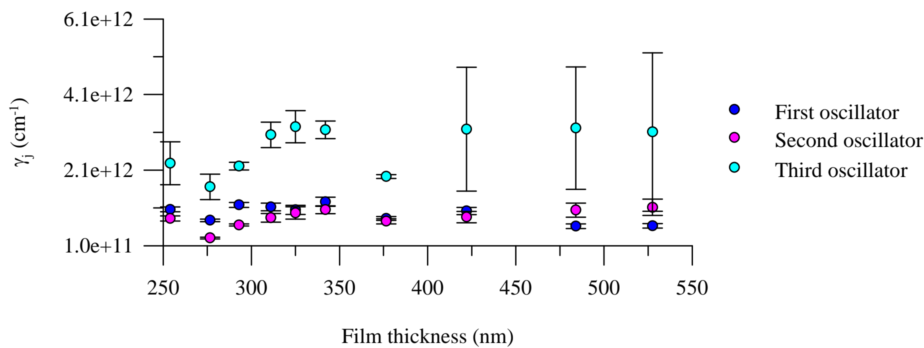


Figure 7.33 - Evolution of the damping coefficients γ_j , (j=1,2,3) with film thickness.

Implicit in dispersion analysis is the determination of the a-SiO_2 complex dielectric function and complex refractive index defined, respectively, by $\tilde{\varepsilon}_r = \varepsilon_1 + i\varepsilon_2$, and $\tilde{n}$ =n+ik (see Chapters 2 and 3). Figures 7.34 and 7.35 show plots of the real and imaginary components of $\tilde{\varepsilon}_r$ and $\tilde{n}$ for two different film thicknesses, 254 nm and 527.5 nm, respectively.

Comparing these plots we verify, as expected, that both $\tilde{\varepsilon}_r$ and $\tilde{n}$ show small changes with film thickness: the magnitudes of ε_1, ε_2 , n and k increase with film thickness. From the values of n and k we conclude, respectively, that the group velocity of light in the film changes with its thickness (see Chapter 2), and absorption increases with film thickness.

We know that peaks in ε_2 identify transverse modes, while peaks in $\mathrm{Im}\left(-1/\tilde{\varepsilon}_r\right) = \varepsilon_2 \Big/ \left(\varepsilon_1^2 + \varepsilon_2^2\right)$, which is the dielectric energy-loss function, identify longitudinal modes. Figures 7.34 and 7.35 show plots of ε_2, where the three TO modes may be identified: ≈ 1057 cm^{-1}, ≈ 1085 cm^{-1}, ≈ 1182 cm^{-1} for the film 254 nm thick; ≈1044 cm^{-1}, ≈ 1076 cm^{-1}, ≈ 1182 cm^{-1} for the film 527.5 nm thick (in both cases, the influence of the higher wavenumber mode is rather weak). Figure 7.36 shows a plot of $\varepsilon_2 / (\varepsilon_1^2 + \varepsilon_2^2)$ for a film 527.5 nm thick. The three LO modes resonance frequencies are at ≈ 1070 cm^{-1}, ≈ 1220 cm^{-1}, and ≈ 1300 cm^{-1}.

Comparing the values obtained for the film 527.5 nm thick with the values in table 7.1, we verify that there is some agreement between them, namely for the ≈ 1044 cm^{-1} TO mode, and the ≈ 1220 cm^{-1}, and ≈ 1300 cm^{-1} LO modes.

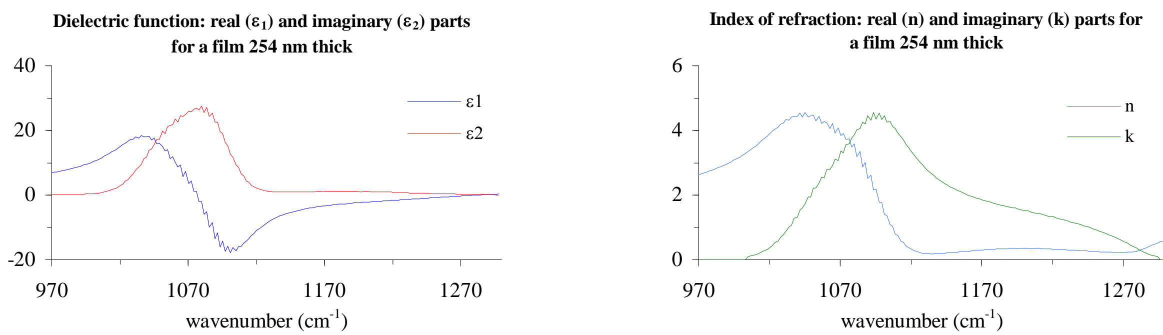


Figure 7.34 - Real and imaginary components of the dielectric function and refractive index for a film 254 nm thick.

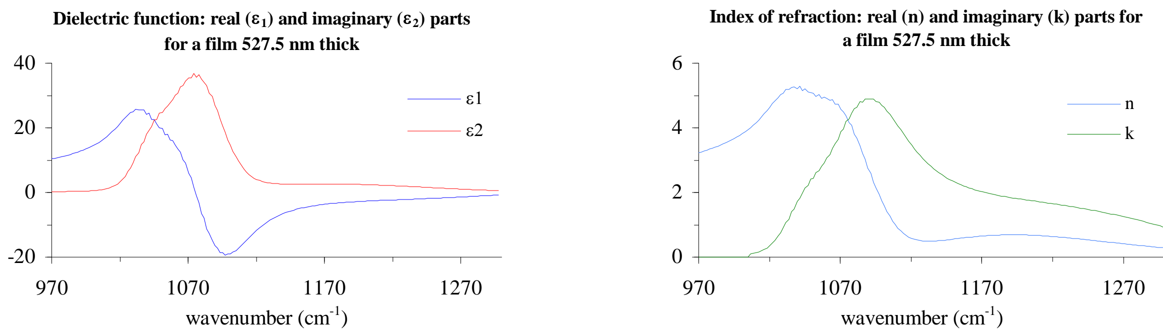


Figure 7.35 - Real and imaginary components of the dielectric function and refractive index for a film 527.5 nm thick.

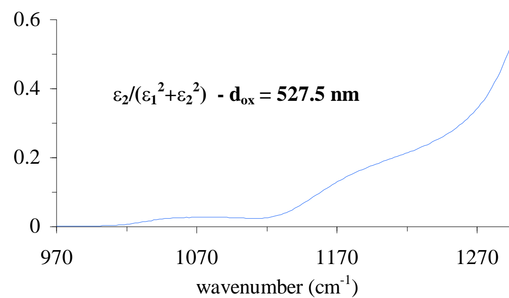


Figure 7.36 - Dielectric energy-loss function for a film 527.5 nm thick.

7.3.2- Reflectance spectra

Dispersion analysis was tried on these spectra using the dielectric function model for crystalline materials [lorentzian oscillators without Gaussian Distribution of Frequencies (GDF)] and for amorphous materials (lorentzian oscillators with GDF) − see Chapter 3. However, and opposite to what happened with transmittance spectra, some reflectance spectra were successfully fitted by the model for crystalline materials. This result, even though surprising, is not new since Grosse *et al.* [5] also claimed good fits for reflectance spectra using this model. However, they fitted the whole [400,4000] cm^{-1} spectral region, so its is hard to know if the AS-mode spectral region, [≈1000,≈1300] cm^{-1}, was really well fitted. Two modes with resonances frequencies of 1075 cm^{-1} and 1090 cm^{-1} were assigned to this region, by these authors. The thickness of the film was 840 nm.

Our results may be summarized as follows:

- polarization s; angle of incidence 15° → we are not able to distinguish, in terms of SCF's (standard error of the curve fit) and parameters estimated standard deviations, between two and three oscillators either with GDF or not. The fits obtained with two and three oscillators with GDF are plotted in figure G.10 (Appendix G).
- polarization s; angle of incidence 50° → successful fits either with two oscillators with GDF or three oscillators without GDF. They are plotted in figure G.11 (Appendix G). The fit using three oscillators with GDF was rejected since both the SCF and the parameters estimated standard deviations increased with the addition of a third oscillator.
- polarization s; angle of incidence 65° → successful fits either with two oscillators with GDF or three oscillators without GDF. They are plotted in figure G.12 (Appendix G). The fit using three oscillators with GDF was rejected since the parameters estimated standard deviations were too high.
- polarization s; angle of incidence 75° → successful fit using three oscillators without GDF (plotted in figure G.13 of Appendix G). Fits using oscillators with GDF were rejected, since one or more of the final parameters were negative, thus without any physical meaning.
- polarization p; angle of incidence 15° → no fit achieved reasonable results, so they were all rejected.
- polarization p; angle of incidence 50° → successful fit using three oscillators without GDF (plotted in figure G.14 of Appendix G). The fit using three oscillators with GDF was rejected since the estimated standard deviation for the gaussian widths, Ω_g, were overwhelming high.
- polarization p; angle of incidence 65° → successful fit using three oscillators with GDF (plotted in figure G.15 of Appendix G). Reasonable fits are not achievable using oscillators without GDF.
- polarization p; angle of incidence 75° → no fit achieved reasonable results, so they were all rejected.

The results in terms of accepted and rejected fits may be visualized in table 7.2.

Table 7.2- Accepted (🗎) and rejected (🗄) fits of reflectance spectra

		number of lorentzian oscillators			
		no GDF		GDF	
polarization	φ_0	2	3	2	3
s	15°	🗎	🗎	🗎	🗎
s	50°	🗄	🗎	🗎	🗄
s	65°	🗄	🗎	🗎	🗄
s	75°	🗄	🗎	🗄	🗄
p	15°	🗄	🗄	🗄	🗄
p	50°	🗄	🗎	🗄	🗄
p	65°	🗄	🗄	🗄	🗎
p	75°	🗄	🗄	🗄	🗄

These results are not easy to interpret because they are very inconsistent. Only the spectrum s-15° shows good agreement with the results obtained in the previous sections. But even in this case, this agreement is not guaranteed since good fits were also accomplished using three oscillators without GDF and two oscillators (with and without GDF).

In most spectra, we are only able to keep the number of TO modes to three by fitting with the model for crystalline materials. If we wish to mantain our model for amorphous materials we, in general, either get no fit at all or get only two TO modes, which is in complete disagreement with the results obtained in the previous sections.

The model for the dielectric function of amorphous materials, which reproduced quite well the normal incidence transmittance spectra, seems to fail for oblique incidence polarized reflectance spectra. However, this apparent failure may be due to reasons other than the model itself. For instance, reflectance values above the expected (see Chapter 6) and the large amount of noise especially evident below ≈ 1000 cm^{-1} may have somewhat "biased" the spectra. This "bias" would affect more deeply dispersion analysis than gaussian decomposition because the first method is quite more "sensitive" than the second.

In order to solve this problem, new reflectance spectra should be taken, and dispersion analysis repeated. Depending on the results, we would take the appropriate conclusions.

7.4- Mathematical software performance

The most robust of the eight codes implemented by the program FITIR were DN2G and DN2F. Due to the complexity of their algorithm – augmented Gauss-Newton method (see Chapter 4) – it is also one of the heaviest (that is, with more lines of code), along with codes DN2GB and DN2FB. These last two codes, despite using the same minimization method, converged some times to odd values due, undoubtedly, to the limits imposed on the parameters.

The less reliable code was FRPRMN, based on the Conjugate Gradient method, which, in order to converge, required an initial guess very close to the solution.

The codes based on the Levenberg-Marquardt method, namelly LMDER, LMDIF and MRQMIN, have also proven to be robust, even though not as robust as DN2G and DN2F.

The only difference between the codes that used an exact (analytical) Jacobian and the codes that used an approximate (finite difference approximation) Jacobian is that the first codes are quite faster (1.5 to 3 times faster) than the second codes, with the exception of FRPRMN which was the slowest of them all. However, the results obtained in the curve fit are equal, which is somewhat surprising since we are comparing codes that use approximate calculations with codes that use exact calculations.

References

[1] Boyd, I. W., Appl. Phys. Lett. 51, 418 (1987)

[2] Goullet, A. , Charles, C. , Garcia, P. , and Turban, G. , J. Appl. Phys. 74, 6876 (1993)

[3] Kirk, C.T., Phys. Rev. B 38, 1255 (1988)

[4] Almeida, R. M., Phys. Rev. B 45, 161 (1992)

[5] Grosse, P., Harbecke, B., Heinz, and Meyer R.; Offenberg, M., Appl. Phys. A 39, 257-268 (1986)

8- Conclusions

In this work we made a quantitative infrared analysis of the asymmetric stretching peak of SiO_2 films thermally grown, by two different methods: gaussian decomposition and dispersion analysis. They are both accomplished by a nonlinear least-squares curve fitting of infrared spectra. Due to the calculations complexity and extent, a specific software was developed (FITIR), which has proven to work very well, in the sense that it has permitted us to obtain some interesting results.

Gaussian decomposition strongly suggests that, for films with thicknesses ranging from ≈16 nm to ≈527.5 nm, the asymmetric stretching peak is composed by three vibrational modes, instead of the one or two modes suggested in literature, and that at least two of these modes show a longitudinal-optic − transverse-optic (LO−TO) frequency splitting. This implies that LO modes are only observed in oblique incidence spectra of radiation polarized with the electric field vector parallel to the plane of incidence, refuting the theories that assigned a LO mode to the asymmetric stretching peak high wavenumber shoulder.

The shift of the asymmetric stretching peak with film thickness, observed in normal incidence transmittance spectra, is understood in terms of the variation of the gaussian parameters with film thickness, namely their intensities and mean values.

When the dielectric function is modelled by a set of lorentzian oscillators with a gaussian distribution of frequencies (model for amorphous materials), dispersion analysis of transmittance spectra corroborates quite reasonably the results obtained in gaussian decomposition, namely in the number of vibrational modes and their respective relative intensities. However, this model for amorphous materials seems to fail for reflectance spectra. In fact, the model for crystalline materials seems to fit better these spectra than the model for amorphous materials. This failure may be due to reasons other than the model itself. So, new reflectance spectra should be taken, and dispersion analysis repeated.

The dispersion parameters, following the example of the gaussian parameters, also show a high dependence on film thickness.

There is some temptation to justify these parameters changes by assuming the existence of bond strains in the oxide, which in turn would also change with film thickness. However, if this is true, its effect is rather irregular since there are parameters that change, with film thickness, but not for all peaks.

With dispersion analysis, some properties of the dipoles responsible for the vibrational peaks were determined, as well as the SiO_2 films complex dielectric function and complex refractive index.

The attempt to assign a resonance frequency to each TO and LO mode was rather difficult and incomplete, therefore further studies should be made in order to get a more clear picture about their distribution. For instance, Raman spectroscopy is widely used to study the vibrational modes of a molecule. So, a comparison of results obtained by this method and by IR spectroscopy should be very enlightening.

Another interesting continuation of this work would be to apply the models here developed to the study of the infrared spectra of deposited silicon dioxide films, which are being increasingly used in the microelectronic industry.

Appendixes

A
Gauss-Legendre quadrature

Gaussian quadrature or Gaussian numerical integration is a very powerful tool for integrals that may not be performed, or which are very difficult to perform, analitically. It works very well when the integrand is smooth, in the sense of being "well-approximated by a polynomial" [1]. In this case, it is superior to other quadrature methods[1]. With Gaussian quadratures we have the freedom to choose not only the location of the abscissas at which the function is to be evaluated (they may not be equally spaced), but also the weighting coefficients that affect that function. Even more, we may arrange the choice of weights and abscissas to make the integral exact for a class of integrands "polynomials times some known function W(x)", where W(x) is the weighting function, which may be chosen as to remove integrable singularities from the desired integral. In other words, given W(x), and given an integer n, we may find a set of weights, w_j, and abscissas (or nodes), x_j, such that the approximation

$$\int_a^b W(x)f(x)dx \approx \sum_{j=1}^{n} w_j f(x_j) \qquad \text{(A-1)}$$

is exact if f(x) is a polynomial.

In our specific case, the integrands are not polynomials but are smooth, though.

The nodes, x_j, and the weights, w_j, may be determined using orthogonal polynomials. To introduce these orthogonal polynomials, we fix the interval of interest to be (a,b). We may define the "scalar product of two functions f and g over a weight function W" as

$$\langle f|g\rangle \equiv \int_a^b W(x)f(x)g(x)dx \qquad \text{(A-2)}$$

The scalar product is a number, not a function of x. Two functions are said to be orthogonal if their scalar product is zero. A function is said to be normalized if its scalar product with itself is unity. A set of functions that are mutually orthogonal and also all individually normalized is called an orthonormal set.

We may find a set o polynomials (i) that includes exactly one polynomial of order j, called $p_j(x)$, for each j = 0,1,2,..., and (ii) all of which are mutually orthogonal over the specified weight function W(x). This set of polynomials may be found using the following recurrence relation

$$\begin{aligned} &p_{-1}(x) = 0 \\ &p_0(x) = 1 \\ &p_{j+1}(x) = \left(x - a_j\right)p_j(x) - b_j p_{j-1}(x)\ , \qquad j = 0,1,2,\ldots \end{aligned} \qquad \text{(A-3)}$$

where,

$$\begin{aligned} a_j &= \frac{\langle xp_j|p_j\rangle}{\langle p_j|p_j\rangle}\ , \qquad j = 0,1,\ldots \\ b_j &= \frac{\langle xp_j|p_j\rangle}{\langle p_{j-1}|p_{j-1}\rangle}\ , \qquad j = 1,2,\ldots \end{aligned} \qquad \text{(A-4)}$$

The coefficient b_0 is arbitrary; we may take it to be zero.

The polynomials defined by equation (A-3) are monic, i.e., the coefficient of their leading term $\left[x^j \text{ for } p_j(x)\right]$ is unity. If we divide each $p_j(x)$ by the constant $\left[\langle p_j|p_j\rangle\right]^{1/2}$ we may render the set of polynomials orthonormal [1]. The orthogonal polynomial $p_j(x)$ may be shown to have exactly j distinct roots in the interval (a,b). These roots are very important because the abscissas of the n - point Gaussian quadrature formula (A-1) with

[1] e.g. Trapezoidal rule or Simpson's rule.

weighting function W(x) in the interval (a,b) are precisely the roots of the orthogonal polynomial $p_n(x)$ for the same interval and weighting function. This is the fundamental theorem of Gaussian quadratures.

Once the abscissas $x_1,..., x_n$ are known, the weights w_j , j =1,...,n, may be evaluated by the equation

$$w_j = \frac{\langle p_{n-1} | p_{n-1} \rangle}{p_{n-1}(x_j) p'_n(x_j)} \tag{A-5}$$

where $p'_n(x_j)$ is the derivative of the orthogonal polynomial at its zero x_j.

The computation of Gaussian quadrature rules thus involves two distinct phases: (I) the generation of the orthogonal polynomials $p_0,..., p_n$, i.e., the computation of the coefficients a_j and b_j in equation (A-3); the determination of the zeros of $p_n(x)$ and the computation of the associated weights.

We are only interested in the Gauss-Legendre quadrature, which has the following weight function, interval and recurrence relation

$$W(x) = 1 \qquad -1 < x < 1$$

$$(j+1)P_{j+1} = (2j+1)xP_j - jP_{j-1} \tag{A-6}$$

where P_j is the Legendre polynomial of degree j[2]. So the nodes are equal to the zeros of the degree n Legendre polynomial $P_n(x)$ on $[-1,+1]$, and the weights are given by

$$w_j = \frac{-2}{(n+1)P_{n+1}(x_j)P'_n(x_j)} \quad , \qquad j = 1,2,\ldots,n \tag{A-7}$$

Equation (A-1) may now be rewritten as

$$\int_{-1}^{+1} f(x)dx \approx \sum_{j=1}^{n} w_j f(x_j) \tag{A-8}$$

For integrals on other finite intervals with weight function W(x) = 1, we use the following linear change of variables,

$$\int_a^b f(t)dt = \frac{b-a}{2} \int_{-1}^{+1} f\left(\frac{a+b+x(b-a)}{2}\right)dx \tag{A-9}$$

Using this equation on equation (A-8) we get

$$\int_a^b f(t)dt \approx \frac{b-a}{2} \sum_{j=1}^{n} w_j f\left(\frac{a+b+x_j(b-a)}{2}\right) \tag{A-10}$$

References

[1] ATKINSON, K. E., An Introduction to Numerical Analysis, 2nd Edn., John Wiley & Sons (1989)

[2] Legendre polynomial: $P_n(x) = \sum_{k=0}^{n/2} (-1)^k \frac{(2n-2k)!}{2^n k!(n-k)!(n-2k)!} x^{n-2k}$