Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Chimiche

Scuola di Dottorato di Ricerca in Scienza e Ingegneria dei Materiali

XXIV ciclo

Design of nanodevices for sensing applications

Direttore della Scuola: Ch.mo Prof. Gaetano Granozzi

Supervisore: Ch.mo Prof. Filippo Romanato

Dottorando: Michele Massari

to Riffola, Tato Patato and Patatino col Pollo

Contents

Abstract **5**

Riassunto **7**

Introduction **9**

Chapter 1 PLASMONICS **13**

Introduction 13

1.1 Surface Plasmon Polaritons at metal/insulator interfaces 13

1.2 Excitation of Surface Plasmon Polaritons at planar interfaces 23

1.2.1 Prism coupling 23

1.2.2 Grating coupling 25

Chapter 2 EXTRAORDINARY OPTICAL TRANSMISSION **29**

Introduction 29

2.1 Transmission through single apertures 29

2.2 Transmission through arrays of apertures 31

Chapter 3 SERS **39**

Introduction 39

3.1 Raman Spectroscopy 40

3.2 The SERS Enhancement Factor 41

3.3 Definition of the SERS Enhancement Factors 42

3.3.1 The SERS Substrate Enhancement Factor - experimental approach 43

3.3.2 The SERS cross-section and single-molecule EF 44

3.3.3 The SERS Substrate Enhancement Factor - formal definition 45

3.4 Experimental measurement of SERS Enhancement Factors 46

3.4.1 Link between SSEF definition and experiments 46

3.5. Overview of the main EM effects in SERS 48

3.5.1 Analysis of the EM problem of SERS 48

3.5.2 Local field enhancement 49

3.5.3 Radiation enhancement 51

3.5.4 The common $|\boldsymbol{E}|^4$-approximation to SERS enhancements 52

3.6. Modified Spontaneous Emission 54

3.6.1. The link between spontaneous emission and dipolar emission 54

3.6.2. Modification of dipole emission: definitions of enhancement factors 56

3.6.3. Spontaneous emission and the optical reciprocity theorem 57

3.7. Formal derivation of SERS EM enhancements 60

3.7.1. SERS EM enhancements in the backscattering configuration 64

Chapter 4 ELECTRON BEAM LITHOGRAPHY 67

Introduction 67

4.1 Electron-solid interactions 67

4.1.1 Electron scattering 68

4.1.2 The Proximity Effect 73

4.1.3 Resolution and Process Latitude 79

Chapter 5 RESIST MATERIALS AND PROCESSES 83

Introduction 83

5.1 Resist materials 83

5.1.1 Positive resists 84

5.1.2 Negative resists 87

5.2 Resist processes 88

5.2.1. Resist adhesion and substrate priming 88

5.2.2 Resist coating 89

5.2.3 Soft baking 91
5.2.4 Exposure 91
5.2.5 Sensitivity of resist materials 92
5.2.6 Contrast of resist materials 93
5.2.7 Resolution enhancement processes 94
5.2.8 Multiple-layer resist strategies 97
5.2.9 Pattern transfer 97
Chapter 6 ELECTROPLATING 99
Introduction 99
6.1 Electrolytic metal deposition 99
6.1.1 Direct Current electrodeposition 100
6.2 Electrodeposition electrolytes 102
6.2.1 Electroplating electrolytes 102
Chapter 7 ELLIPSOMETRY AND SPM MICROSCOPY 105
7.1 Ellipsometry 105
7.2 Scanning Probe Microscopy 108
7.2.1 Atomic Force Microscopy 109
7.2.2 Scanning Near-field Optical Microscopy 110
Chapter 8 DESIGN 113
8.1 Design of the nanodevice 113
8.1.1 The Finite Element Method 113
8.1.2 FEM simulations 113
Chapter 9 NANOFABRICATION 119
Introduction 119
9.1 The EBL system 119
9.2 The nanofabrication process 121

9.2.1 Spin coating and soft baking 122

9.2.2 EBL patterning 125

9.2.3 Resist development 131

9.2.4 Electrolytic growth 132

9.2.5 Stripping 133

9.2.6 Characterization of the growth 134

Chapter 10 SERS MEASUREMENTS 137

Introduction 137

10.1 Functionalization 137

10.2 SERS characterization 138

10.2.1 Experimental set-up 138

10.2.2 SERS measurements varying incident light polarization 139

10.3 Experimental SERS Enhancement Factor 141

10.4 Device homogeneity quality and accuracy 142

Chapter 11 EXPERIMENTAL RESULTS - EOT 145

Chapter 12 EXPERIMENTAL RESULTS - SERS 153

Chapter 13 METAMATERIALS 161

Introduction 161

13.1 Metamaterials as sensors 162

13.1.1 Design, fabrication and characterization of SRRs using X-Ray Lithography 162

13.1.2 Nanogap Ring Antennae as Plasmonically coupled SERRS substrates 167

Conclusions 173

Projects and Publications 175

Acknowledgments 177

Bibliography 179

Abstract

Plasmonically active SERS substrates, such as metallic gratings or MetaMaterials like Split Ring Resonators, have sub-wavelength features that enable direct coupling of normally incident electromagnetic radiation to surface plasmons. The high detection sensitivity of these methods exploits the EM field enhancements which arise in the proximity of the metallic nanostructures.

The object of the PhD research is the design, fabrication and characterization of plasmonic nanostructures to be used as biosensors.

In order to fabricate SERS-active substrates suitable for sensing applications, a well-tested nanofabrication process for the realization of easily manufacturable and reproducible devices has been developed, exploring the different mechanisms of light transmission through 1D digital metallic gratings in order to optimize the optical response and efficiency of the devices.

1D digital gold grating devices, able to concentrate the electromagnetic radiation inside the slits, have been fabricated. The SERS Enhancement Factor of the realized nanostructures and its dependence on a generic hybrid polarization have been measured and compared with the theoretical estimations.

A good agreement between the experimental and numerical estimations, added up to a well-tested nanofabrication process able to guarantee a good uniformity and reproducibility of the device, has confirmed the potentiality of the nanofabrication technique in substrate engineering to provide local fields of the desired intensity and location.

Finally, novel types of sensing nanodevices, coupling biosensing to MetaMaterials properties have been realized, to explore the peculiar properties of such kind of structures in combination with plasmonic nano-optics.

These type of sensors would likelihood represent a promising step toward future reproducible single-molecule detection using engineered plasmonic substrates.

Riassunto

Substrati SERS plasmonicamente attivi, come grating metallici o MetaMateriali quali Split Ring Resonators, hanno caratteristiche sub-wavelength che consentono l'accoppiamento diretto tra radiazione elettromagnetica normalmente incidente e plasmoni di superficie. L'elevata sensibilità di rivelazione di questi metodi sfrutta l'incremento del campo EM che si origina in prossimità di nanostrutture metalliche.

L'oggetto della ricerca di dottorato è la progettazione, fabbricazione e caratterizzazione di nanostrutture plasmoniche per applicazioni biosensoristiche.

Allo scopo di fabbricare substrati SERS-attivi adatti per applicazioni di tipo sensoristico, è stato sviluppato un collaudato processo di nanofabbricazione in grado di consentire la realizzazione di dispositivi facilmente fabbricabili e riproducibili. Ciò ha richiesto lo studio e la comprensione dei diversi meccanismi di trasmissione della luce attraverso reticoli metallici digitali unidimensionali al fine di ottimizzare la risposta ottica e l'efficienza dei dispositivi.

Sono stati fabbricati dispositivi costituiti da grating d'oro unidimensionali in grado di concentrare la radiazione elettromagnetica all'interno delle fenditure ed è stato misurato e confrontato con stime teoriche il fattore di enhancement SERS delle nanostrutture realizzate e la sua dipendenza da una polarizzazione ibrida generica.

Un buon accordo tra le stime numeriche e sperimentali, assieme ad una tecnica di nanofabbricazione collaudata in grado di garantire una buona uniformità e riproducibilità del processo, conferma le potenzialità della tecnica nell'ingegnerizzazione di substrati per la generazione di campi locali dell'intensità e della distribuzione desiderata.

Sono stati infine realizzati nuovi tipi di nanosensori allo scopo di accoppiare le proprietà biosensoristiche con quelle dei MetaMateriali, in modo da studiare le proprietà peculiari di questo tipo di strutture se combinate con nano-ottiche plasmoniche.

Questo tipo di sensori rappresentano un passo promettente verso una possibile futura rivelazione di singole molecole tramite ingegnerizzazione di substrati plasmonici.

Introduction

The sensing field has emerged as a topic of great interest for the wide number of applications both in the environmental and in the agroindustrial sector for the detection of pathogens and chemicals, and in the medical field for the detection of genetics and proteomics. Moreover, the constant health danger posed by new strands of microbic organisms and spread of infectious diseases is another concern requiring sensing for their rapid detection and identification.

Plasmonic sensors provide a means of detecting chemical and biological species through the observation of spectral features by measurement techniques such as Surface Enhanced Raman Spectroscopy (SERS) or Surface Plasmon Resonance (SPR). Plasmonically active SERS substrates, such as metallic gratings or periodic metal nanowire arrays, have sub-wavelength features that enable direct coupling of normally incident electromagnetic radiation to surface plasmons. The high detection sensitivity of these methods exploits the EM field enhancements which arise in the proximity of the metallic nanostructures. In order to develop a reliable plasmonic sensor, the periodic nanostructures must be replicated with adequate uniformity and fidelity. Furthermore, the process definition must enable a fine-tuning of the nanostructures in order to optimize field enhancement and reproducibility for the final sensor device configuration.

The design and manufacture of plasmonic nano-optics exploits physical phenomena, called plasmons, which have the peculiarity to govern the coupling of light with the electrons on the surface of a nanostructured material. The plasmon control at the nanometer level allows to develop highly sensitive biological sensors, able in principle to detect even few molecules of DNA.

The thesis project was mainly devoted to the study, design and fabrication of plasmonic nano-optics to be used as biosensors. In particular, the aim was to create plasmonic crystals able to concentrate the light within so-called 'hot spots', which, due to the strong increase of the local electromagnetic field, result in an amplification of the Raman signal of several orders of magnitude.

The focus of the work has been the exploration of the different mechanisms of light transmission through 1D digital metallic gratings and the study of the SERS response of a functionalized digital gold grating to a normally impinging monochromatic wave, at the exchange of the polarization of the incident light.

A purpose of the project was also the development of a well-tested nanofabrication process for the realization of a sensing device easily manufacturable and reproducible and characterized by a good uniformity of the response.

The innovation of the project consist in the combination and use of a Finite Element Method (COMSOL Multiphysics) to study and project the devices to be used as sensors, nanofabrication techniques (Electron Beam Lithography (EBL), X-Ray Lithography (XRL)) for the development and realization of the nanostructures, and nano-optical characterization techniques such as SERS, SNOM (Scanning Near-Field Microscopy) and Ellipsometry to analyze the response and compare it with theoretical simulations.

The plasmonic nano-optics, if carried out by Metamaterials [e.g. Split Ring Resonators (SRRs)], can match the biosensing properties to the peculiar properties characteristic of these new type of materials: part of the thesis work was thus devoted to the study, design and fabrication of plasmonic nano-optics realized through SRRs, to explore the new applications offered by such kind of hybrid devices in the biosensing field.

Summing up:

- we explored, in two published papers the different mechanisms of light transmission through 1D digital metallic gratings, both numerically and experimentally;
- in a further published paper we theoretically optimized and fabricated 1D digital gold grating devices for sensing applications, measured the SERS Enhancement Factor (SSEF) of the realized nanostructures and its dependence on a generic hybrid polarization and compared it with the theoretical estimations. A good agreement between the experimental and numerical values of the $SSEF_{TM}$ to $SSEF_{TE}$ ratio confirms the validity of our approach to the problem, together with the reliability of the nanofabrication process in the realization of easily manufacturable and reproducible devices;
- a further part of the project has concerned the realization of two kind of MetaMaterials (MMs) to be used as biosensors:
  - through the development of a well-tested nanofabrication process combining EBL and XRL tecniques, we obtained arrays of Split Ring Resonators (SRRs) MMs characterized by high aspect ratio and large patterned areas;
  - by using Electron Beam Lithography, arrays of circular Multiple-Split Ring antennas (MSRRs) with very small feature sizes have been engineered into nanophotonic ring structures, thus creating concentrated areas of high localized field coupling exploitable for enhanced plasmonic applications.

The research work is organized as follows.

Part I: Introduction

Chapter 1. Plasmonics

An introduction to the theory involving SPPs is presented. Their main physical properties are described, as well as the most common ways to excite them.

The content of this chapter is a literature review on SPPs.

Chapter 2. Extraordinary Optical Transmission

The transmission of electromagnetic fields through 1D metallic gratings is addressed. Theoretical study and experimental evidence supporting the EOT phenomenon are shown; a theoretical survey exploring the different mechanisms of light transmission through digital metallic gratings is also presented.

The content of this chapter is a literature introduction to the EOT topic.

Chapter 3. SERS

In this chapter, the fundamentals of the SERS effect are outlined. After a review of some possible definition of the SERS enhancement factor and a discussion of some pratical aspects of their experimental measurements, the SERS EM problem is addressed. The rest of the chapter is dedicated to

a more rigorous justification of the concept introduced in the first part, reviewing the most important aspects of modified spontaneous emission and applying them to a formal description of the EM-SERS ehancements.
This chapter is a literature review of Surface Enhanced Raman Spectroscopy.

Chapter 4. Electron Beam Lithography

In this chapter, the fundamentals of the Electron Beam Lithography technique are presented. After an introduction about EBL technology and applications, the physics of the electron-beam exposure process is outlined through a thorough description of the electron-solid interactions and of the proximity effect.
This chapter is a review of the fundamental aspects of Electron Beam Lithography.

Chapter 5. Resist Materials and Processes

In this chapter, a description of resist materials and processes is outlined. After an introduction about e-beam resists, a special survey of PMMA resist and PMMA developers is provided; the most important resist properties and processes are then discussed.
This chapter is a review of the fundamental aspects of resist processing.

Chapter 6. Electroplating

In this chapter, a description of the electrolytic metal deposition process is presented. An overview of the process is provided, together with the main mechanisms in the deposition of metal layers and a brief description of electroplating electrolytes.
This chapter is a review of the fundamental aspects of electroplating.

Chapter 7. Ellipsometry and SPM Microscopy

In this chapter, a description of the Ellipsometry, AFM (Atomic Force Microscopy) and SNOM (Scanning Near-field Optical Microscopy) characterization techniques is provided.
This chapter is a review of some characterization techniques used.

Part II: Biosensor Design

Chapter 8. Design

In this chapter, a design of the biosensing device is outlined, together with a simulation of the EM field configuration under illumination with a normal incident monochromatic wave. After an introduction about the FEM simulation technique, the configuration providing the solution of the addressed ‘hot spot’ excitation problem is provided, together with an analysis of the transmittance properties of the simulated grating.
The optical design of the device was carried on through the use of COMSOL Finite Element Method (FEM).

Part III: Device fabrication

Chapter 9. Nanofabrication

This chapter discusses the nanofabrication process used to realize the designed plasmonic gratings. After an introduction on the EBL machine used, the whole nanofabrication process is thoroughly described, and experimental evidence of the various steps provided.
Electron Beam Lithography was performed at LaNN Laboratory. Electroplating was carried on at TASC Laboratory. SEM characterizaton was performed both at LaNN and TASC laboratories.

Part IV: Device Characterization and Experimental Results

Chapter 10. SERS Measurements

In this chapter the functionalization of the fabricated device and the determination of the SERS enhancement factors are addressed; the experimental setup is described and the collected data analysed.
The SERS measurements were performed at Dipartimento di Scienze Chimiche in Padova.

Chapter 11. Experimental Results - EOT

In this chapter the result of our research on Extraordinary Optical Transmission is outlined through the presentation of two published papers on the argument. A thorough description of the mechanisms of light propagation through 1D metallic gratings is provided, and numerical simulations and experimental results compared; the link between far- and near-field regimes of EM wave propagation is also investigated.
The ellipsometric characterization was performed at LaNN Laboratory. The SNOM measurements were performed at NanoFAB Laboratory.

Chapter 12. Experimental Results - SERS

In this chapter the theoretical optimization and nanofabrication of 1D digital gold gratings to be used as biosensors is provided. The SERS Enhancement Factor (SSEF) of the realized nanostructures and its dependence on a generic hybrid polarization are calculated and compared with theoretical estimations. The reliability of the developed nanofabrication process is also investigated through the characterization of a suitable number of samples.

Chapter 13. Metamaterials

In this chapter the nanofabrication of two different MetaMaterials to be used as biosensors is presented. First, the realization of Split Ring Resonators (SRRs) with high aspect ratio, through the use of combined EBL and XRL techniques, is outlined; then the engineering of Multiple-Split Ring antennas (MSRRs) with very small feature sizes is discussed.
Electron Beam Lithography and SEM characterization were performed at LaNN Laboratory. X-ray lithography, electroplating and evaporation were carried on at TASC Laboratory. SEM characterizaton was performed both at LaNN and TASC laboratories.

PLASMONICS

Introduction

The first documented observation of surface plasmons dates back to 1902, when Wood illuminated a metallic diffraction grating with polychromatic light and noticed narrow dark bands in the spectrum of the diffracted light, which he referred as to anomalies [1]. Theoretical work by Fano [2] concluded that these anomalies were associated with the excitation of electromagnetic surface waves on the surface of the diffraction grating. In 1958 Thurbadar observed a large drop in reflectivity when illuminating thin metal films on a substrate [3], but did not link this effect to surface plasmons. In 1968 Otto explained Turbadar's results and demonstrated that the drop in the reflectivity in the attenuated total reflection method is due to the excitation of surface plasmons [4]. In the same year, Kretschmann and Raether reported excitation of surface plasmons in another configuration of the attenuated total reflection method [5]. The pioneering work of Otto, Kretschmann, and Raether established a convenient method for the excitation of surface plasmons and their investigation, and introduced surface plasmons into modern optics.

1.1 Surface Plasmon Polaritons at metal/insulator interfaces

Surface plasmon polaritons are electromagnetic excitations propagating at the interface between a dielectric and a conductor, evanescently confined in the perpendicular direction. These electromagnetic surface waves arise via the coupling of the electromagnetic fields to oscillations of the conductor's electron plasma.

The electrons at the surface of the metal, which are usually subject to random fluctuations, in presence of a suitable electromagnetic field give rise to coherent oscillations parallel to the surface, forming a wave attenuated in the perpendicular direction and with the maximum intensity in proximity of the surface (see Fig. 1.1).

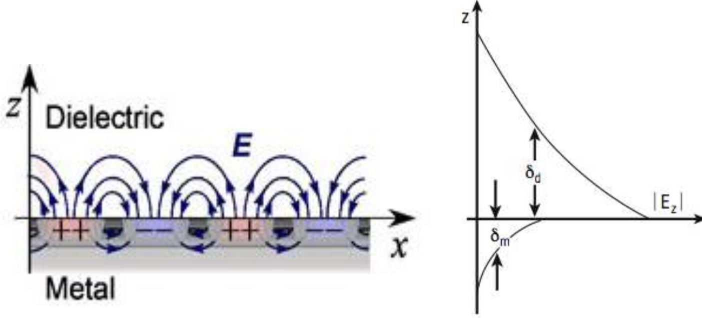


Fig. 1.1. On the left: charge and electromagnetic field of a surface plasmon propagating along a metal-dielectric interface in the x direction; on the right: exponential decay of the electric field in the z direction.

The starting point to investigate the physical properties of surface plasmons polaritons is looking for a wavelike solution of Maxwell's equations at the flat interface between a conductor and a dielectric.

In the absence of external charge and current densities the equations to solve are:

$$\nabla \cdot \mathbf{D} = 0 \tag{1.1a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.1b}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.1c}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \tag{1.1d}$$

which can be combined and rewritten as:

$$\nabla\left(-\frac{1}{\varepsilon}E \cdot \nabla\varepsilon\right) - \nabla^2 E = -\mu_0\varepsilon_0\varepsilon\frac{\partial^2 E}{\partial t^2} \tag{1.2}$$

using the identities $\nabla \times \nabla \times E \equiv \nabla(\nabla \cdot E) - \nabla^2 E$ and $\nabla \cdot (\varepsilon E) \equiv E \cdot \nabla\varepsilon + \varepsilon(\nabla \cdot E)$ and the constitutive relations for linear, isotropic and nonmagnetic media:

$$D = \varepsilon_0 \varepsilon E \tag{1.3a}$$

$$B = \mu_0 H \tag{1.3b}$$

where μ_0 is the free-space permeability, ε the relative permittivity (dielectric constant) of the medium and ε_0 the free-space permittivity.

For negligible variation of the dielectric profile $\varepsilon = \varepsilon(\mathbf{r})$ over distances on the order of an optical wavelength, (1.2) simplifies to:

$$\nabla^2 E - \frac{\varepsilon}{c^2}\frac{\partial^2 E}{\partial t^2} = 0 \tag{1.4}$$

This equation has to be solved separately in regions of constant ε, and the obtained solutions have to been matched using appropriate boundary conditions.

If we assume a harmonic time dependence $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega t}$ of the electric field, we obtain:

$$\nabla^2 E + {k_0}^2 \varepsilon E = 0 \tag{1.5}$$

where $k_0 = \frac{\omega}{c}$ is the wave vector of a propagating wave in vacuum.

To define the propagation geometry, we assume for simplicity a one-dimensional problem, in which the waves propagate along the x-direction of a cartesian coordinate system and show no spatial variation in the in-plane y-direction, therefore $\varepsilon = \varepsilon(z)$.

By orienting cartesian axes in this way, the plane $z = 0$ coincides with the interface sustaining the propagating waves, which can now be written as:

$$E(r,t) = E(z)e^{i(\beta x - \omega t)} \tag{1.6}$$

The complex parameter $\beta = k_x$ is called the propagation constant of the traveling waves and corresponds to the component of the wave vector in the direction of propagation.

Inserting this expression into (1.5) yields:

$$\frac{\partial^2 E(z)}{\partial z^2} + \left({k_0}^2 \varepsilon - \beta^2\right)E = 0 \tag{1.7}$$

A similar equation exists for the magnetic field H:

$$\frac{\partial^2 H(z)}{\partial z^2} + \left({k_0}^2 \varepsilon - \beta^2\right)H = 0 \tag{1.8}$$

Equations (1.7) and (1.8) are the starting point for the general analysis of guided electromagnetic modes in waveguides.

Substituting now (1.6) and the corresponding equation for the field H into the curl equations (1.1c) and (1.1d), allow us to obtain explicit expressions for the different field components of E and H:

$$\frac{\partial E_y}{\partial z} = -i\omega\mu_0 H_x \tag{1.9a}$$

$$\frac{\partial E_x}{\partial z} - i\beta E_z = i\omega\mu_0 H_y \tag{1.9b}$$

$$i\beta E_y = i\omega\mu_0 H_z \tag{1.9c}$$

$$\frac{\partial H_y}{\partial z} = i\omega\varepsilon_0\varepsilon E_x \tag{1.9d}$$

$$\frac{\partial H_x}{\partial z} - i\beta H_z = -i\omega\varepsilon_0\varepsilon E_y \tag{1.9e}$$

$$i\beta H_y = -i\omega\varepsilon_0\varepsilon E_z \tag{1.9f}$$

It can easily be shown that this system allows two linearly independent sets of modes with different polarization properties of the propagating waves. The first set, with only the field components E_x, E_z and $H_y \neq 0$, are the transverse magnetic (**TM** or **p**) modes; the second set, with only the field components H_x, H_z and $E_y \neq 0$, are the transverse electric (**TE** or **s**) modes.

For TM modes we obtain:

$$E_x = -i\frac{1}{\omega\varepsilon_0\varepsilon}\frac{\partial H_y}{\partial z} \tag{1.10a}$$

$$E_z = -\frac{\beta}{\omega\varepsilon_0\varepsilon}H_y \tag{1.10b}$$

with the wave equation:

$$\frac{\partial^2 H_y}{\partial z^2} + \left({k_0}^2\varepsilon - \beta^2\right)H_y = 0 \tag{1.10c}$$

For TE modes the analogous set is:

$$H_x = i\frac{1}{\omega\mu_0}\frac{\partial E_y}{\partial z} \tag{1.11a}$$

$$H_z = \frac{\beta}{\omega\mu_0}E_y \tag{1.11b}$$

with the wave equation:

$$\frac{\partial^2 E_y}{\partial z^2} + \left({k_0}^2\varepsilon - \beta^2\right)E_y = 0 \tag{1.11c}$$

Let's now consider the case of a flat interface (Fig. 1.2) between a dielectric, non-absorbing semi-infinite space (z $>$ 0) with positive real dielectric constant ε_d and an adjacent conducting semi-infinite space (z $<$ 0) described via a dielectric function $\varepsilon_m(\omega)$, where we want to look for propagating wave solutions confined to the interface, i.e. with evanescent decay in the perpendicular z-direction.

Fig. 1.2. A metal–dielectric waveguide.

Let us first look at TM solutions. Using the equation set (1.10) in both semi-infinite spaces, yields:

$$H_y(z) = A_2 e^{i\beta x} e^{-k_d z} \tag{1.12a}$$

$$E_x(z) = iA_2\frac{1}{\omega\varepsilon_0\varepsilon_d}k_d e^{i\beta x} e^{-k_d z} \tag{1.12b}$$

$$E_z(z) = -A_2 \frac{\beta}{\omega \varepsilon_0 \varepsilon_d} e^{i\beta x} e^{-k_d z} \tag{1.12c}$$

for $z > 0$ and

$$H_y(z) = A_1 e^{i\beta x} e^{k_m z} \tag{1.13a}$$

$$E_x(z) = -iA_1 \frac{1}{\omega \varepsilon_0 \varepsilon_m} k_m e^{i\beta x} e^{k_m z} \tag{1.13b}$$

$$E_z(z) = -A_1 \frac{\beta}{\omega \varepsilon_0 \varepsilon_m} e^{i\beta x} e^{k_m z} \tag{1.13c}$$

for $z < 0$, where $k_m \equiv k_{z,m}$ and $k_d \equiv k_{z,d}$ are the components of the wave vector perpendicular to the interface in the two media. Their reciprocal values define the evanescent decay lengths of the fields perpendicular to the interface, which quantify the confinement of the wave.

The boundary conditions of Maxwell's equations require that the components E_x and H_y of the electric and magnetic field intensity vectors parallel to the boundaries of the interface are continuous at the boundaries. From the set of equations (1.12) and (1.13) we get $A_1 = A_2$ and

$$\frac{k_d}{k_m} = -\frac{\varepsilon_d}{\varepsilon_m} \tag{1.14}$$

Since $\varepsilon_d > 0$, confinement to the surface demands $\mathrm{Re}[\varepsilon_m] < 0$: the surface waves exist only at interfaces between materials with opposite signs of the real part of their dielectric permittivities, i.e. between a conductor and an insulator.

For metals following the free-electron model:

$$\varepsilon_m = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma\omega} \tag{1.15}$$

where γ is the collision frequency and ω_p is the plasma frequency:

$$\omega_p = \sqrt{\frac{ne^2}{\varepsilon_0 m_e}} \tag{1.16}$$

where n is the free electron density and m_e is the effective optical mass of the electrons, the requirement $\mathrm{Re}[\varepsilon_m] < 0$ is fulfilled for frequencies lower than the plasma frequency of the metal. As shown in Fig. 1.3, metals such as gold, silver and aluminum exhibit a negative real part of permittivity in the visible and near infrared region of the spectrum.

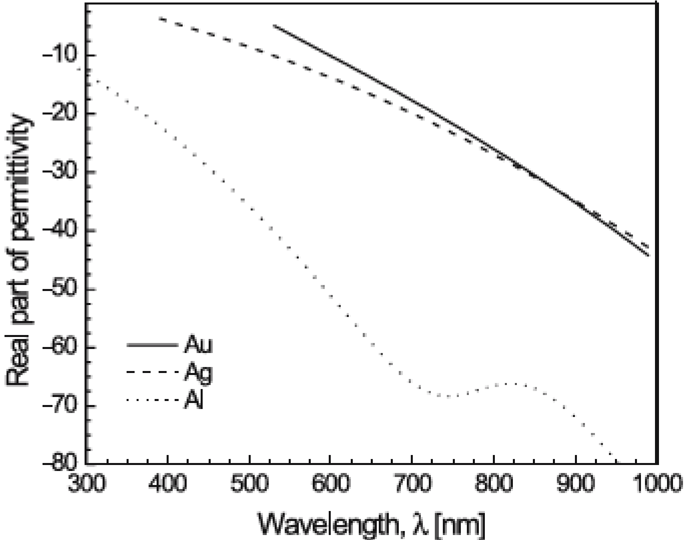


Fig. 1.3. Real part of the permittivity of gold, silver and aluminum as a function of wavelength. [6]

Now, from (1.12a), (1.13a) and (1.10c) we obtain:

$$k_m^2 = \beta^2 - k_0^2 \varepsilon_m \quad (1.17a)$$

$$k_d^2 = \beta^2 - k_0^2 \varepsilon_d \quad (1.17b)$$

Which, combined with (1.14) yields:

$$\beta = k_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}} \quad (1.18)$$

This expression, which is valid for both real and complex ε_m (i.e. for conductors without and with attenuation) represents the dispersion relation of the SPPs propagating at the interface between the two semi-infinite spaces.

Let us now analyze the possibility of having TE surface modes. Using the equation set (1.11), the respective expressions for the field components are:

$$E_y(z) = A_2 e^{i\beta x} e^{-k_d z} \quad (1.19a)$$

$$H_x(z) = -iA_2 \frac{1}{\omega \mu_0} k_d e^{i\beta x} e^{-k_d z} \quad (1.19b)$$

$$H_z(z) = -A_2 \frac{\beta}{\omega \mu_0} e^{i\beta x} e^{-k_d z} \quad (1.19c)$$

for $z > 0$ and

$$E_y(z) = A_1 e^{i\beta x} e^{k_m z} \tag{1.20a}$$

$$H_x(z) = iA_1 \frac{1}{\omega\mu_0} k_m e^{i\beta x} e^{k_m z} \tag{1.20b}$$

$$H_z(z) = -A_1 \frac{\beta}{\omega\mu_0} e^{i\beta x} e^{k_m z} \tag{1.20c}$$

for $z < 0$. Continuity of field components E_y and H_x at the interface leads to the condition:

$$A_1(k_m + k_d) = 0 \tag{1.21}$$

Since confinement to the surface requires $\mathrm{Re}[k_m] > 0$ and $\mathrm{Re}[k_d] > 0$, this condition is only fulfilled if $A_1 = 0$, so that $A_2 = A_1 = 0$. Thus, no surface modes can exist for TE polarization. Surface plasmon polaritons only exist for TM polarization.

In order to examine the properties of the SPPs, let's take a look at their dispersion relation.

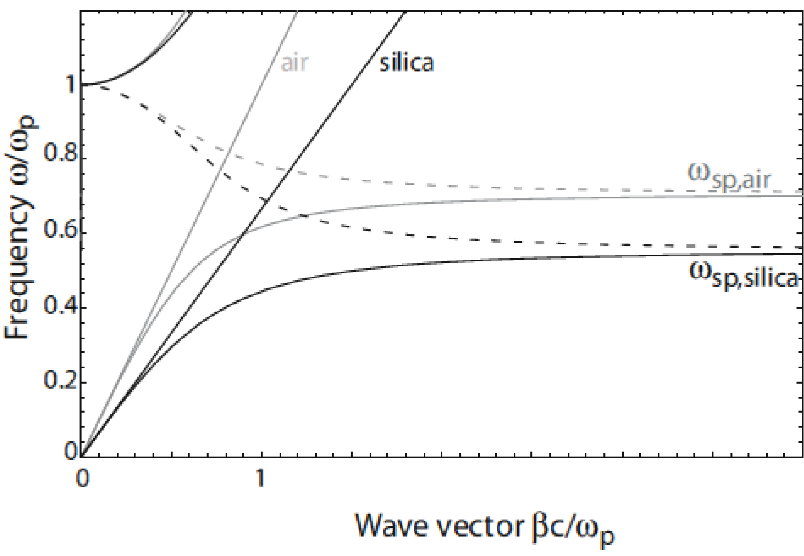


Fig. 1.4. Dispersion relation of the SPPs at the interface between a Drude metal with negligible collision frequency and air (gray curves) and silica (black curves). [7]

Fig. 1.4 shows plots of (1.18) for a metal with negligible damping of the conduction electron oscillation ($\mathrm{Im}[\varepsilon_m(\omega)] = 0$) described by the real Drude dielectric function (obtained from 1.15 in the limit of $\gamma \to 0$)

$$\varepsilon_m(\omega) = 1 - \frac{\omega_p^2}{\omega^2} \tag{1.22}$$

for an air and a fused silica interface, where both the real (continuous curves) and the imaginary part (broken curves) of the wave vector β are shown. The bound modes, corresponding to the SPP

excitations, are described by the part of the dispersion curves lying to the right of the light lines. Between the regime of bound $(\omega < \omega_p)$ and radiative $(\omega > \omega_p)$ modes, a frequency gap region with purely imaginary β prohibiting propagation exists.

For small wave vectors corresponding to low (mid-infrared or lower) frequencies, the SPP propagation constant is close to $k_0 = \frac{\omega}{c}$, and the waves extend over many wavelengths into the dielectric space. In the opposite regime of large wave vectors, the frequency of the SPPs approaches the characteristic surface plasmon frequency:

$$\omega_{sp} = \frac{\omega_p}{\sqrt{1+\varepsilon_d}} \quad (1.23)$$

In the limit of negligible damping of the conduction electron oscillation, the wave vector β goes to infinity as the frequency approaches ω_{sp}, and the group velocity $v_g \to 0$. The mode thus acquires electrostatic character, and is known as the surface plasmon.

Up to now we have assumed an ideal conductor with $(Im[\varepsilon_m(\omega)] = 0)$. Real metals however suffer both from free-electron and interband damping, therefore $\varepsilon_m(\omega)$ and the SPP propagation constant β are complex. The traveling SPPs are damped with an energy attenuation length (also called propagation length), defined as the distance in the direction of propagation at which the energy of the surface plasmon decreases by a factor of 1/e, given by:

$$L = \frac{1}{2\mathrm{Im}[\beta]} \quad (1.24)$$

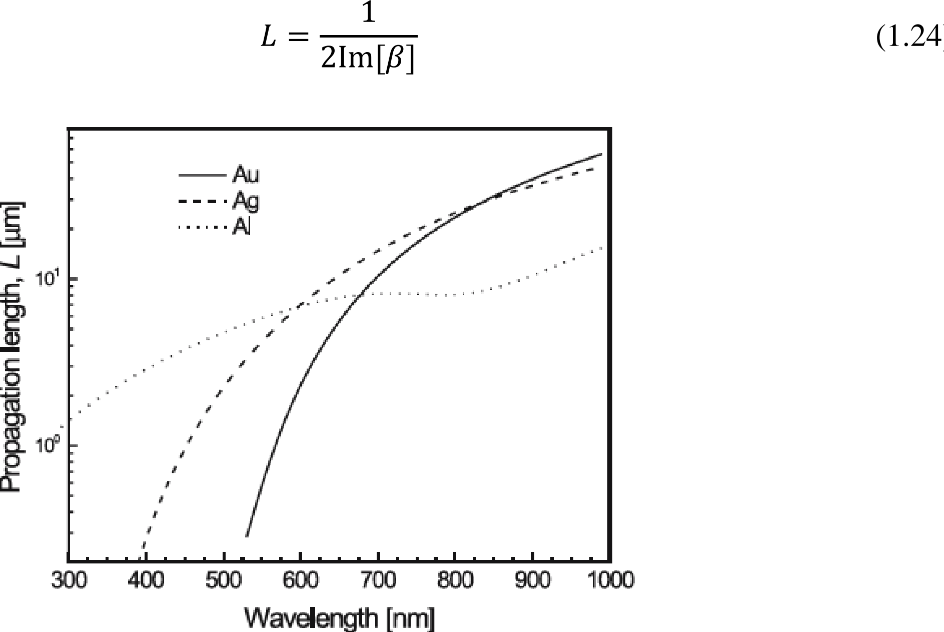


Fig. 1.5. Propagation length of a surface plasmon at the interface between a dielectric (refractive index 1.32) and a metal, calculated for gold (Au), silver (Ag), and aluminum (Al). [6]

Spectral dependencies of the propagation length of a surface plasmon supported by gold, silver and aluminum are shown in Fig. 1.5.

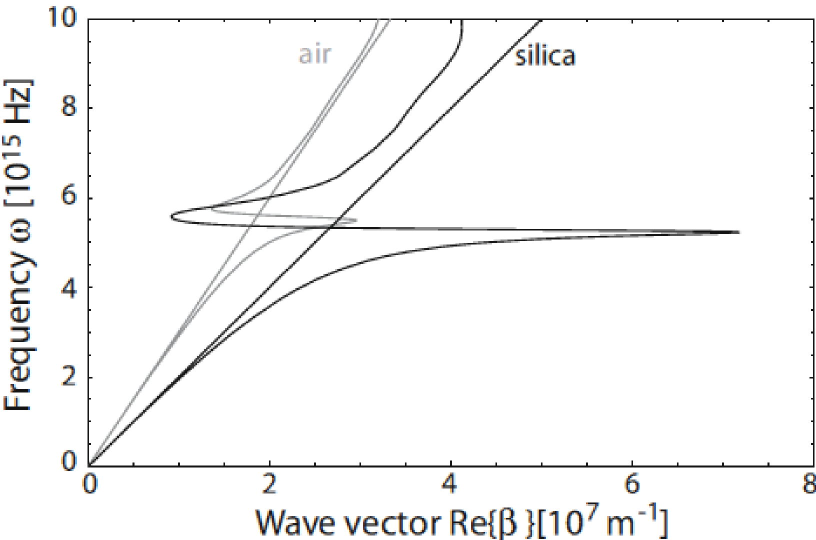


Fig. 1.6. Dispersion relation of the SPPs at a silver/air (gray curve) and silver/silica (black curve) interface. Due to the damping, the wave vector of the bound SPPs approaches a finite limit at the surface plasmon frequency. [7]

Fig. 1.6 shows plots of the dispersion relation of the SPPs propagating at a silver/air and silver/silica interface. Compared with the dispersion relation of completely undamped SPPs (Fig. 1.4.), the bound SPPs approach now a maximum, finite wave vector at the surface plasmon frequency ω_{sp} of the system.

This limitation puts a lower bound both on the wavelength $\lambda_{sp} = \frac{2\pi}{Re[\beta]}$ of the surface plasmon and on the amount of mode confinement perpendicular to the interface, since the SPP fields fall off as $e^{-|k_i||z|}$, $i = k, d$

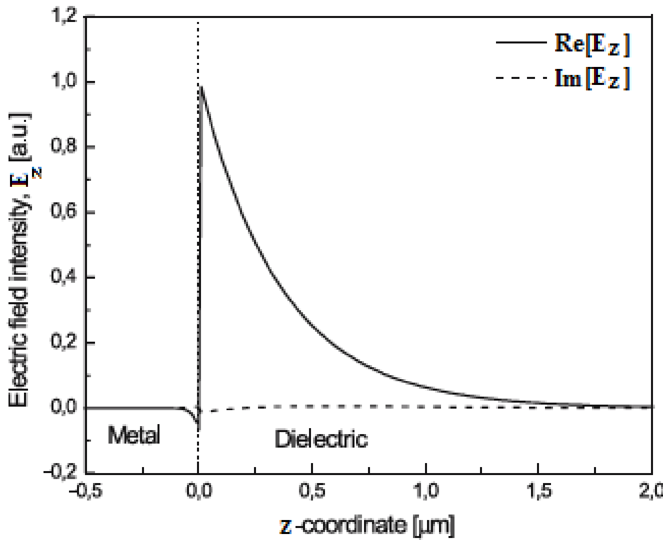


Fig. 1.7. Distribution of the z component of the electric field of a surface plasmon at the interface between gold ($\varepsilon_m = 25 - 1.44i$) and a dielectric medium with refractive index = 1.32 for the wavelength λ=800 nm. [6]

As follows from Fig. 1.7, the electromagnetic field of a surface plasmon reaches its maximum at the metal–dielectric interface and decays into both media. The field decay in the direction perpendicular to

the metal–dielectric interface is characterized by the penetration depth L_p, defined as the distance from the interface at which the amplitude of the field decreases by a factor of 1/e:

$$L_{pm} = \frac{1}{|k_m|} \tag{1.25}$$

$$L_{pd} = \frac{1}{|k_d|} \tag{1.26}$$

The spectral dependence of the penetration depth of a surface plasmon at the interface between gold and a non-dispersive medium with refractive index = 1.32 is shown in Fig. 1.8.

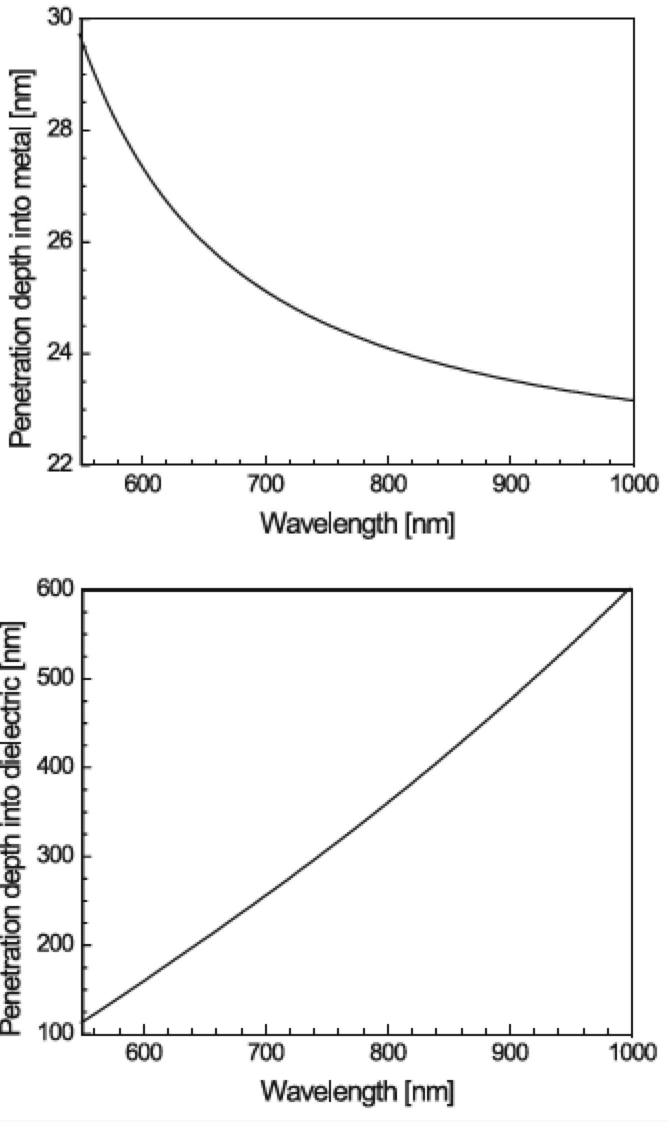


Fig. 1.8. Penetration depth of a surface plasmon into metal (upper plot) and dielectric (lower plot) as a function of the wavelength for a surface plasmon propagating along the interface between gold and a dielectric medium with refractive index = 1.32. [6]

As follows from Fig. 1.8, with an increasing wavelength the portion of the electromagnetic field carried in the dielectric increases and the field of the surface plasmon extends farther into the dielectric.

Both the propagation length (1.24) and the energy confinement (1.25), (1.26) show a strong dependence on frequency. The better the confinement, the lower the propagation length: this characteristic trade-off between localization and loss is typical for plasmonics.

1.2 Excitation of Surface Plasmon Polaritons at planar interfaces

Surface plasmon polaritons propagating at the flat interface between a conductor and a dielectric are essentially two-dimensional electromagnetic waves.

Confinement is achieved since the propagation constant β is greater than the wave vector k in the dielectric, leading to evanescent decay on both sides of the interface.

As a consequence, the projection along the interface of the momentum $k_x = k \sin\theta$ of photons impinging under an angle θ to the surface normal is always smaller than the SPP propagation constant β, even at grazing incidence, prohibiting phase-matching

Since SPP dispersion curve lies to the right of the light line of the dielectric (given by $\omega = ck$), excitation by three-dimensional light beams is not possible unless special techniques for phase-matching are employed, such as grating or prism coupling.

1.2.1 Prism coupling

The most common approach to excitation of surface plasmons is by means of a prism coupler and the attenuated total reflection method (ATR).

Phase-matching to SPPs can in fact be achieved by means of a three-layer system consisting of a thin metal film sandwitched between two insulators of different dielectric constants.

A beam reflected at the interface between the insulator of higher dielectric constant ε_P, usually in the form of a prism (see Fig. 1.9), and the metal will have an in-plane momentum $k_x = k\sqrt{\varepsilon_P}\sin\theta$, which is sufficient to excite SPPs at the interface between the metal and the lower-index dielectric, usually air.

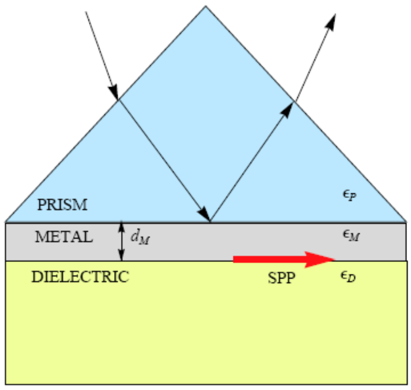


Fig. 1.9. ATR coupling scheme.

What happens in this case is that SPPs can be excited having propagation constants β between the light lines of air and the higher-index dielectric (Fig. 1.10).

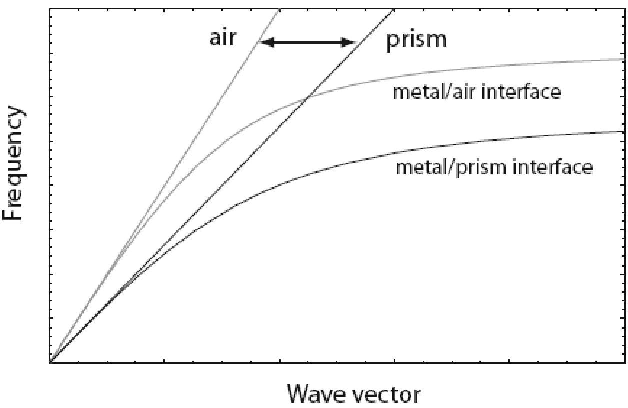


Fig. 1.10. Prism coupling and SPP dispersion. Propagation constants between the light lines of air and prism are accessible to excite SPPs at the interface between metal and air. [7]

When the light wave propagating in the prism is made incident on the metal film, part of the light is reflected back into the prism and a part propagates in the metal in the form of an evanescent electromagnetic wave which decays exponentially in the direction perpendicular to the prism–metal interface. If the metal film is sufficiently thin (less than 100 nm for light in visible and near infrared part of spectrum), the evanescent wave penetrates through the metal film and couples with a surface plasmon at the outer boundary of the metal film.

SPP excitation manifests itself as a minimum in the reflected beam intensity from the metal (Fig. 1.11). For each wavelength, the matching condition is satisfied for a single angle of incidence on the metal film, the coupling angle, which increases with decreasing wavelength.

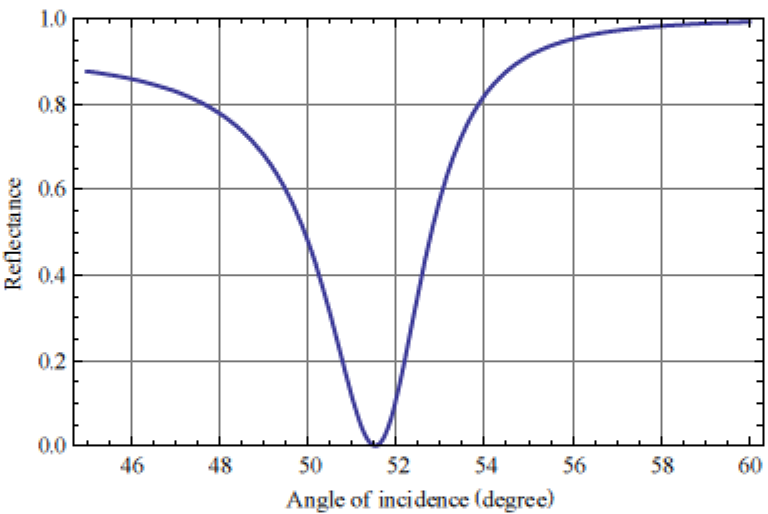


Fig. 1.11. Example of reflectivity for TM polarization as a function of the angle of incidence.

This coupling scheme - also known as attenuated total internal reflection - therefore involves tunneling of the fields of the excitation beam to the metal/air interface where SPP excitation takes place.

Two configurations of the ATR method are possible, the Kretschmann and the Otto geometry.

The most common configuration is the Kretschmann geometry, depicted in Fig. 1.12, in which a thin metal film is evaporated on top of a glass prism. Photons from a beam impinging from the glass side at

an angle greater than the critical angle of total internal reflection tunnel through the metal film and excite SPPs at the metal/air interface.

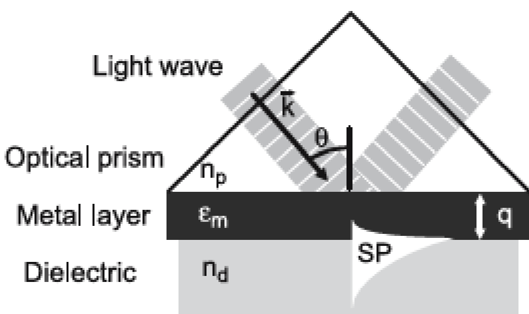


Fig. 1.12. Excitation of surface plasmons in the Kretschmann geometry of the attenuated total reflection (ATR) method.

In the Otto geometry, depicted in Fig. 1.13, the prism is separated from the metal film by a thin air gap. Total internal reflection takes place at the prism/air interface, exciting SPPs via tunneling of the evanescent wave to the air/metal interface. For the coupling to occur, the thickness of the air layer has to be chosen properly (typically few microns), in order to match the propagation constant of the evanescent wave with that of the surface plasmon.

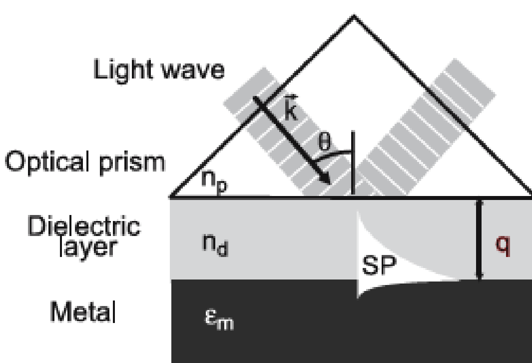


Fig. 1.13. Excitation of surface plasmons in the Otto geometry of the attenuated total reflection (ATR) method.

The dip in the angular reflectivity spectrum of Fig. 1.11 is associated with a transfer of energy from the incident light wave into the surface plasmon and its subsequent dissipation in the metal film. The interaction between the incident light wave and the surface plasmon also affects the phase of the reflected light, which exhibits an abrupt phase jump. The depth of the reflectivity dip depends on the thickness of the metal film, with a maximum for a characteristic metal film thickness. The dip can be observed not only in the angular domain, but also when the angle of incidence is kept constant and the wavelength is varied.

1.2.2 Grating coupling

Another approach to overcome the mismatch in wave vector between the in-plane momentum $k_x = k \sin\theta$ of impinging photons and the propagation constant β is based on patterning the metal surface with a diffraction grating.

In this method, a light wave incident from a dielectric medium on a metal grating with period Λ gives rise to a series of diffracted waves (Fig. 1.14 and Fig. 1.15). Phase matching takes place whenever the condition

$$\beta = k \sin\theta \pm mG \tag{1.27}$$

is fulfilled, where G is the reciprocal vector of the grating, and m = (1, 2, 3 ...).

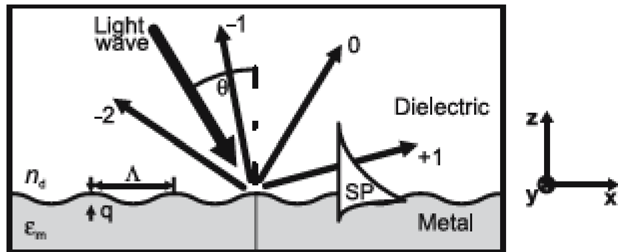


Fig. 1.14. Excitation of surface plasmons by diffraction of light on a diffraction grating.

The grating vector $G = \frac{2\pi}{\Lambda}$ lies in the plane of the grating and is perpendicular to the grooves. Therefore the component of the wavevector of the diffracted light perpendicular to the plane of the grating k_z is equal to that of the incident wave, while the component of the wavevector in the plane of the grating k_x is diffraction altered.

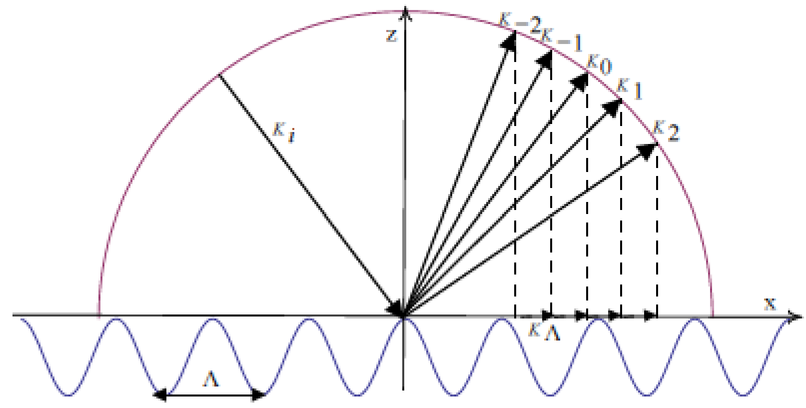


Fig. 1.15. Reflection of a planar incident wave by a sinusoidal diffraction grating.

As with prism coupling, excitation of SPPs is detected as a minimum in the reflected light.

The coupling condition (1.27) can be fulfilled for various combinations of the angle of incidence, grating pitch, and diffraction order.

The characteristic absorption dip can be observed not only in the angular domain, but also when the angle of incidence is kept constant and the wavelength is varied (Fig. 1.16). The interaction between the light wave and the surface plasmon results also in a change in the phase of the reflected light.

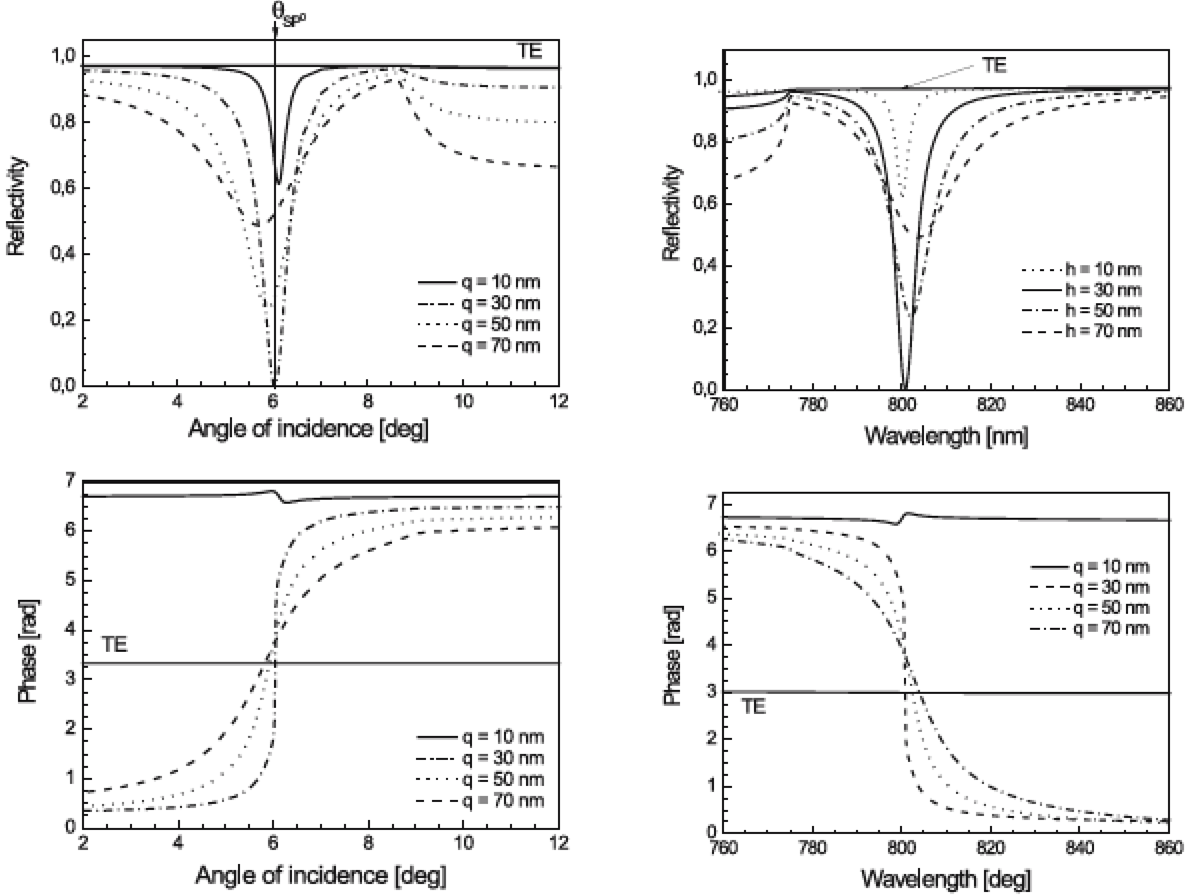


Fig. 1.16. Reflectivity (upper plots) and phase (lower plots) as a function of the angle of incidence (left) and of the wavelength (right) for four different depths of a metallic sinusoidal grating and TM polarization. Reflectivity and phase of the TE polarization are shown for comparison. [6]

The strongest absorption occurs for a single depth of the grating. The width and asymmetry of the reflectivity dip increase with an increasing depth of the grating.

For one-dimensional gratings, significant changes to the SPP dispersion relation occur if the gratings are sufficiently deep so that the modulation can no longer be treated as a small perturbation of the flat interface.

SPPs can also be excited by random surface roughness or localized scatterers. Momentum components Δk_x are in this case provided via scattering, so that the phase-matching condition:

$$\beta = k \sin\theta \pm \Delta k_x \tag{1.28}$$

can be fulfilled.

The reverse process can also take place: SPPs propagating along a surface modulated with a grating can couple to light and thus radiate. Equation (1.28) implies that random surface roughness also constitutes an additional loss channel for SPP propagation via coupling to radiation.

As it has been shown by Offerhaus et al. [8], suitably designing the shape of the grating it is possible to influence the propagation direction of the SPPs and even focusing can be achieved.

EXTRAORDINARY OPTICAL TRANSMISSION

Introduction

In Chapter 1 our discussion has focused on excitation of surface plasmon polaritons along a planar interface. We will now move to the perpendicular direction and take a look at the Extraordinary Optical Transmission (EOT) phenomenon of light propagation through subwavelength apertures.

If a metallic film is patterned with a regular array of holes or slits, enhanced transmission can occur, resulting in more light passing through the apertures than might be expected on the basis of classical diffraction theory. This unusual optical property is due to the coupling of light with plasmons - electronic excitations - on the surface of the periodically patterned metal film. This effect has triggered an enormous amount of interest ever since its first description in 1998.

Extraordinary optical transmission is an optical phenomenon in which a structure containing subwavelength apertures in an opaque screen transmits more light than might naively be expected on the basis of either ray optics or even knowledge of the transmission through individual apertures. This phenomenon was first discovered by Ebbesen et al., 1998 [9]. In the original experiment, a two-dimensional 2D periodic array of subwavelength holes was milled in optically thick metal films. What would be expected is that the electromagnetic waves could only tunnel through the apertures in the transmission process. Surprisingly, such arrays may, for certain wavelengths, exhibit transmission efficiencies, normalized to the total area of the holes, that exceed unity. In other words, for these wavelengths a periodic array of subwavelength holes transmits more light than a large macroscopic hole with the same area as the sum of all the small holes. The surprise is increased by the fact that a single subwavelength aperture generally transmits light with an efficiency that is substantially below unity, due to the evanescent decay of the EM fields inside the holes. With evidence, in the EOT phenomenon the drilled metal film must be an active participant in the transmission process. In physical terms, a collective response of the periodic array of holes must occur in order to boost the transmission. The physical origin of this collective behavior lies in the key role played by surface EM modes in the appearance of EOT.

2.1 Transmission through single apertures

One-dimensional aperture: single slit

Let's start to analyze the transmission properties of subwavelength apertures by considering the simplest case: a single 1D slit of width a perforated on a metal film of thickness h, illuminated by a plane wave with the wave vector lying in the x-z plane (see Fig. 2.1).

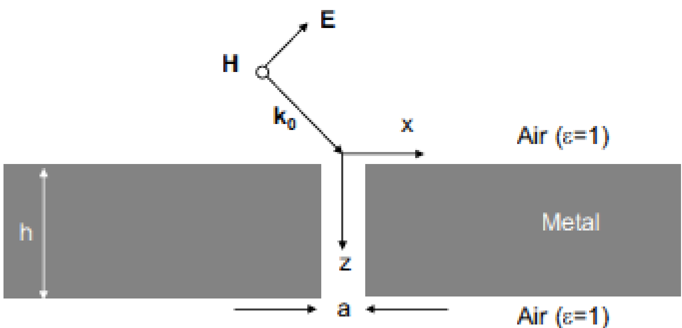


Fig. 2.1. Schematic of a single slit of width a perforated on a metal film of thickness h illuminated by p-polarized radiation. For the case of s-polarized incident radiation, vectors **E** and **H** are interchanged.

Since the system presents translational symmetry along the direction parallel to the slit, the analysis can be restricted to the perpendicular plane where the two light polarizations **s** and **p** are decoupled. For p-polarized light, the magnetic field is parallel to the slit y direction and the electric field has nonzero x and z components. For s-polarized light, the **E** field is pointing to the y direction and the **H** field lies in the x-z plane.

Analyzing the optical response of a single slit on a PEC (Perfect Electrical Conductor) film under p-polarized illumination, several studies making use of different theoretical approaches predict the appearance of slit-cavity transmission resonances. In particular, Garcia-Vidal et al. [10] show that in the extreme subwavelength regime $(a \ll \lambda)$ and in the limit of extremely narrow slits, a good approximation to the final transmittance spectrum can be obtained by taking into consideration the presence of only a single TM waveguide mode inside the slit (TE modes do not couple to the incident p-polarized radiation), which does not present cutoff (this is, technically speaking, a TEM mode, i.e. a mode where both electric and magnetic fields are tranverse). In this case the theory predicts the appearance of transmission peaks close to the Fabry-Perot resonances of the slit, showing that cavity resonances are responsible for the emergence of transmission peaks for p-polarized light.

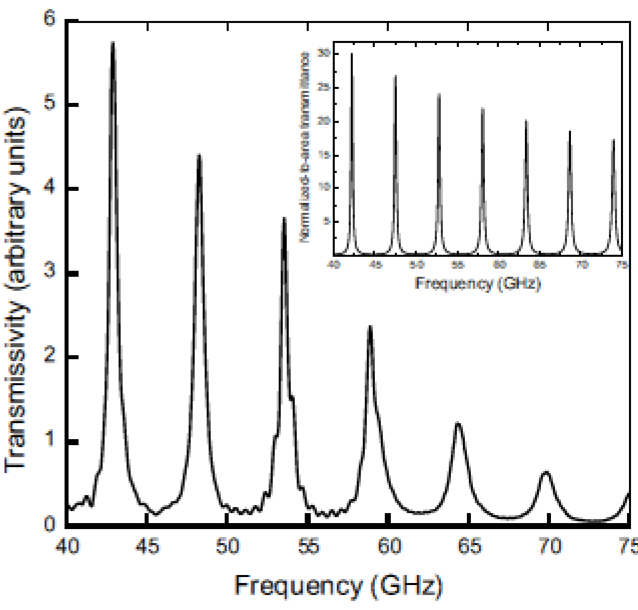


Fig. 2.2. Transmission as a function of frequency for a single slit of width 75 µm perforated on a metal film of thickness 28.2 mm. Inset: Normalized-to-area transmittance theoretical calculation. [11]

An estimation for the normalized-to-area transmittance at resonance T_{res} in the limit $a \ll \lambda$ leads to:

$$T_{res}^{slit} = \frac{\lambda}{\pi a} \tag{2.1}$$

This analytical result predicts a linear increase of T_{res} with the resonant wavelength, as experimentally observed in Fig. 2.2.

Under s-polarized illumination, only the TE modes inside the slit need to be considered. Unlike TM modes, these TE waveguide modes present cutoff. In Fig. 2.3 are shown the normalized to-area transmittance spectra for slits with the same width a, but different thicknesses h. Note that, in contrast to the p-polarized case, transmission process is dominated by the existence of a cutoff wavelength for TE modes, $\lambda_C = 2a$. For $\lambda > \lambda_C$ and thick enough films, the transmission is strongly attenuated due to the evanescent character of the EM fields inside the slit. Fabry-Perot resonances similar to those found for p-polarized light appear in the spectrum for $\lambda < \lambda_C$.

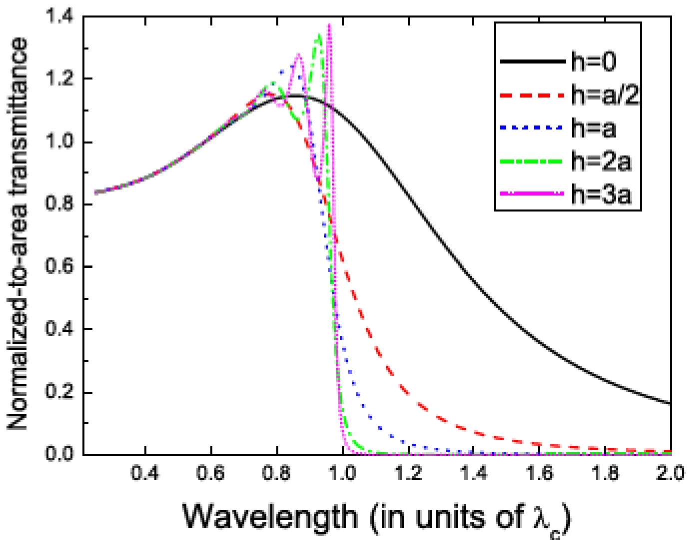


Fig. 2.3. Normalized-to-area transmittance vs wavelength in units of the cutoff wavelength of the fundamental TE mode $\lambda_C = 2a$ through a single slit of width a perforated on a PEC film of different thicknesses h. The normally incident plane wave is s-polarized. [10]

2.2 Transmission through arrays of apertures

One-dimensional periodic arrays of slits

After the discovery by Ebbesen et al. of the EOT phenomenon in periodic arrays of subwavelength holes, many theoretical studies devoted to EOT research have been made. Schroter and Heitmann [12] presented in 1998 the first numerical study in the optical regime on the resonant transmission features of the so-called transmission gratings, i.e., periodic arrays of 1D slits perforated through metallic films (see Fig. 2.4), finding that the transmission spectrum presented peaks whose spectral locations indicated that SPPs were involved in the resonant transmission process.

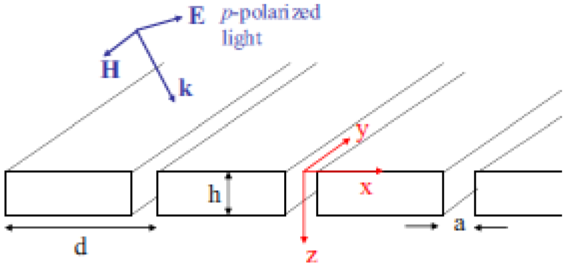


Fig. 2.4. Schematic of a 1D periodic array of slits (period d, width of the slit a) perforated on a metallic film of thickness h. The structure is illuminated by p-polarized radiation, with the **E** field lying in the x-z plane.

Porto et al. [13] later demonstrated that there are two possible ways of transferring p-polarized light efficiently from the upper surface to the lower one: by the coupling of the incident plane wave with cavity resonances located inside the slits (similar to those found in the single slit case) or by the excitation of coupled surface EM modes on both surfaces of the grating. The existence of these two channels for enhancing the transmittance has been corroborated by many subsequent theoretical studies. The two resonant mechanisms are clearly differentiated only for some ranges of the geometrical parameters. In general, due to the existence of a propagating TEM mode inside the slits which always couples both interfaces, the two mechanisms are mixed in a way that depends strongly on the geometrical parameters of the structure. Marquier et al. [14] in 2005 theoretically characterized this hybridization between cavity modes and surface EM modes in the transmission resonances (see Fig. 2.5). Figure 2.5 (a) shows transmission versus both frequency and parallel momentum, while Figure 2.5 (b) gives the dispersion relation of the mode responsible for the enhanced transmission (branch B-C). Figure 2.5 (c) shows the surface-cavity (SC) ratio as a function of parallel momentum. This SC ratio is calculated by comparing the EM energies associated with the cavity and SPP components of the guided mode. This is done quantitatively integrating the EM energy within the slit (cavity component) and on both interfaces (SPP component). As shown in Fig. 2.5 (c), for $k_x = 0$ the character of the mode is mainly cavity-like. As the parallel momentum is increased, the character is evolving to be SPP-like.

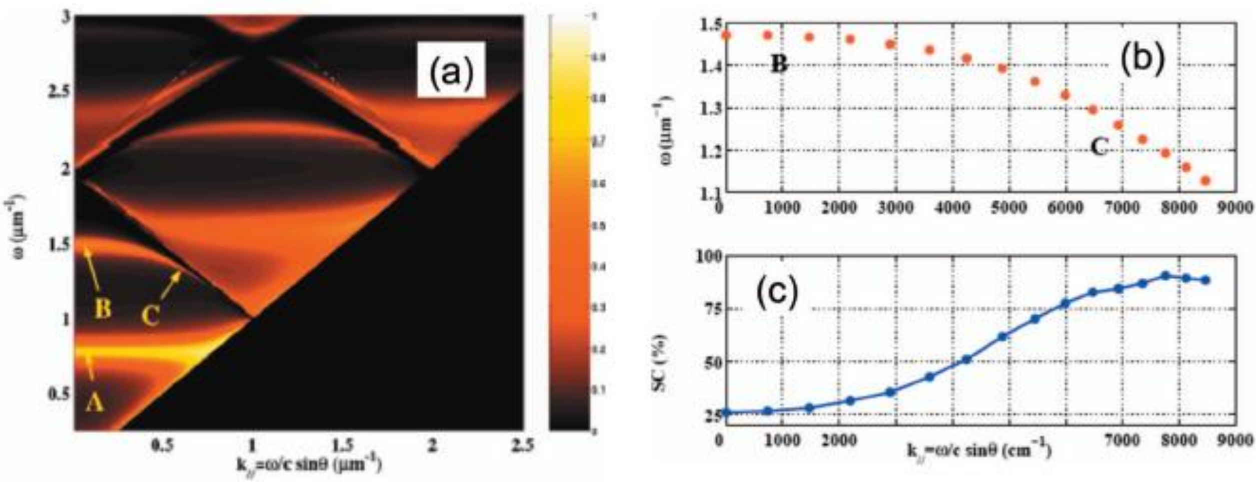


Fig. 2.5. Transmission resonances in 1D transmission gratings for a silver film. (a) Transmission efficiency as a function of both ω and k_x (period of the grating d = 500 nm, width of the slits a = 50 nm, thickness h = 400 nm); (b) branch B-C of the dispersion relation; (c) Surface-cavity (SC) ratio for the transmission branch displayed in (b). [14]

Garcia-Vidal et al. [10], analyzing theoretically the transmission resonances appearing in a periodic array of 1D slits illuminated by p-polarized radiation, show that, regardless of the precise form of the coupling, EM coupled arrays of indentations in a metal film can present bound modes. This occurs even if the metal is a PEC, where the EM coupling between indentations is purely through diffraction modes and does not contain contributions from SPPs (SPPs vanish in the limit of a perfect electrical conductor). They first consider the case of a perforated PEC film: TM waveguide modes inside the slits can be written as a discrete sum over p-polarized diffracted modes. In Fig. 2.6 are shown the transmittance spectra for a 1D array of slits of period d, width $a = 0.2d$, and different values of h.

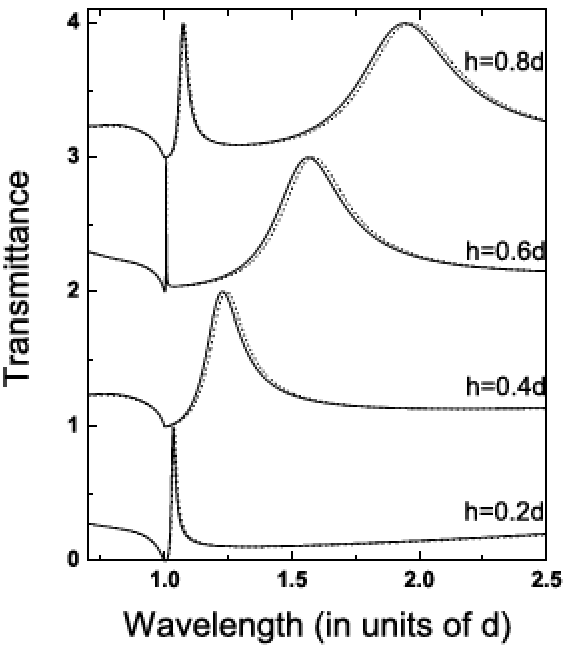


Fig. 2.6. Total transmittance (normalized to the unit cell of the array) of a normal-incidence p-polarized plane wave vs wavelength (in units of d) for periodic arrays of slits of fixed width, $a = 0.2d$, and several thicknesses $h = 0.2d, 0.4d, 0.6d$ and $0.8d$. [14]

Two types of transmission resonances can be distinguished: very narrow transmission peaks appearing at a wavelength close to d (with strong E-fields at the horizontal surfaces) and wider peaks emerging at longer wavelengths (related to Fabry-Perot resonances, with strong E fields inside the slits). In fact, as a function of the metal thickness there is a continuous shift of the resonances and, accordingly, a continuous modification of their character, as discussed before. An interesting point is that the transmission peak emerging close to d also appears in a perforated PEC film which, in principle, does not support SPPs. The physical origin of this resonant feature stems from the excitation of geometrically induced surface EM modes that mimic the behavior of SPPs in perforated PEC structures.

Let's now put ourselves in the single-mode approximation (only a single TM (TEM) waveguide mode is considered inside the slits). Characterizing the reflection coefficient ρ by a modulus R and a phase θ, we can define a total phase:

$$\phi = 2\theta + 2k_{\omega}h \tag{2.2}$$

as the sum of the scattering (associated with ρ) and geometrical (determined by the optical path) phases. As a function of this total phase, the transmission coefficient can be expressed as:

$$\tau = \frac{\tau_{12} e^{ik_\omega h} \tau_{23}}{1 - R^2 e^{i\phi}} \tag{2.3}$$

where τ_{ij} is the amplitude of the transmitted field between media i and j. Equation (2.3) resembles the formula governing the transmission properties of a Fabry-Perot interferometer. In contrast to the canonical case in which the phase of the EM fields inside the cavity is determined by the optical path $2k_\omega h$, equation (2.2) also contains the scattering phase shift, which sometimes presents a strong spectral dependence. Given that for narrow slits: $R \approx 1$, there is a transmission peak, associated with a resonant condition for the denominator in Eq. (2.3), whenever ϕ is an integer times 2π. For thin PEC films, the total phase is dominated by the scattering phase; correspondingly, the spectral location of the transmission resonance appears at a wavelength slightly longer than d (see the case $h = 0.2d$ in Fig. 2.6). When h is increased, the geometrical phase $2k_\omega h$ can also lead to the appearance of transmission peaks, emerging at $\lambda \gg d$. The physical origin of these transmission resonances stems from the excitation of cavity resonances similar to the ones emerging in single slits, whereas the physical origin of the transmission peak located at $\lambda \approx d$ is related to the excitation of an EM bound mode. As fields at the two horizontal metal surfaces of the film are always connected via a propagating wave (the TEM mode), this mode has an E field that may oscillate inside the slits. For thicker films and large enough h, Fabry-Perot resonances dominate the transmission spectrum.

For real metals, the existence of a SPP is accounted for by the presence of the deep transmission minimum, which does not emerge at $\lambda = d$, as in the case of a PEC film, but at a slightly longer wavelength. The transmission peak is associated with the excitation of a guided EM mode, whose linewidth is much smaller than the Fabry-Perot ones.

In 2010 Zilio et al. [15] of our research group published a theoretical study exploring the different mechanisms of light transmission through 1D digital metallic gratings by means of the Finite Elements Method (FEM) and a modal analysis.

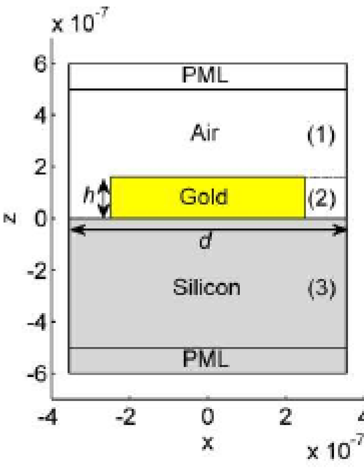


Fig. 2.7. Schematic of the structure analyzed. [15]

The focus of the work was the study of the optical response of a digital gold grating on a semi-infinite silicon substrate to a normally impinging 1000nm-monochromatic wave, varying systematically both thickness (h) and period (d) of the grating.

Only TM polarized (i.e. magnetic field parallel to the metal stripes) incident light was taken into account, being the only polarization that can give rise to SPPs in a 1D metallic grating.

If the slit width is smaller than the vacuum wavelength of the incident radiation, only the fundamental propagating eigenmode is allowed in the region within the slit (see Fig.2.7).

Imposing at the horizontal interfaces the surface impedance boundary conditions, the transmittance of the structure results to be:

$$T = \frac{1}{\sqrt{\varepsilon_3}} \frac{|\tau_{12}|^2 \sum_{i=-\infty}^{+\infty} \cos\theta_i \left|\tau_{23,i}^2\right|^2}{|1 - |\rho_{12}||\rho_{23}|e^{i\phi_{tot}}|^2} \tag{2.4}$$

where τ and ρ are the single-interface (1-2 and 2-3) transmission and reflection coefficients, ε_3 is the dielectric constant of the substrate, θ_i is the *i-th* diffraction order angle and ϕ_{tot} is the total phase accumulated by the single propagating waveguide mode travelling back and forth in the slit:

$$\phi_{tot} = arg(\rho_{12}) + arg(\rho_{23}) + 2k_0 N_{eff} h \tag{2.5}$$

being k_0 the vacuum wave vector and N_{eff} the effective refractive index of the medium inside the slits, introduced to take into account the effective wave vector of the fundamental waveguide propagating mode inside the slits, which results to be higher than the vacuum wave vector. The model is valid as long as $|\varepsilon_m|$ is not too low and/or the slit width isn't very small with respect to the wavelength of the incident light, since these situations correspond to high penetration of the fields in the metal and the plasmonic contributions to the fields are relevant.

Equation (2.4) allows to identify the main features of the optical response of the 1D plasmonic grating, that turn out to be the following:

1) ***Cavity mode (CM) resonances***. These resonances result from the multiple scattering of the single propagating mode inside the slits. The mode is partially reflected and transmitted at the slit ends. When the phase difference ϕ_{tot} between the waves transmitted in the substrate is a multiple of 2π, a peak in transmittance is observed, like in a Fabry-Perot resonator. The interesting fact is that, when this resonance takes place, a great fraction of the incident light is channeled within the slits and is transmitted in the substrate [Fig. 2.8 (a)]. This phenomenon is commonly referred to as extraordinary optical transmission.

2) ***Surface Plasmon Polariton (SPP) resonances***. SPPs are surface EM waves which propagate at a metal-dielectric interface and are evanescent in the normal direction [(Fig. 2.8 (b)]. They are excited in a 1D grating when the in-plane component of the incident p-polarized radiation and that one of the scattered waves sum up to match the SPP momentum, i.e.:

$$k_0 \sin\alpha + mG = k_0 Re\left(\sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}}\right) \tag{2.6}$$

where α is the incidence angle, $G = 2\pi/d$ (d is the period of the grating) and ε_d the dielectric constant of the facing insulator (vacuum for the 1-2 interface, silicon for the 2-3 interface) and

m an integer (1,2,...). This kind of resonance is associated to a transmittance extinction and to a high field enhancement in proximity of the grating.

3) ***Wood-Rayleigh anomalies (WR)***. Abrupt changes in transmittance as a function of period are observed in configurations for which a diffraction order lies in the plane of the grating [Fig. 2.8 (c)], i.e. at periods:

$$d_m = \frac{m\lambda}{n_d} \tag{2.7}$$

being n_d the refractive index of the dielectric medium. These configurations mark a discontinuity, since for $d > d_m$ the *m-th* diffraction order does exist, while for $d < d_m$ it does not. It is generally accepted that the peak in transmission is due to the abrupt redistribution of energy among the allowed orders passing through the $d = d_m$ configurations. WR anomalies are not resonant phenomena being due to pure geometrical reasons and are fully independent of CM and SPPs.

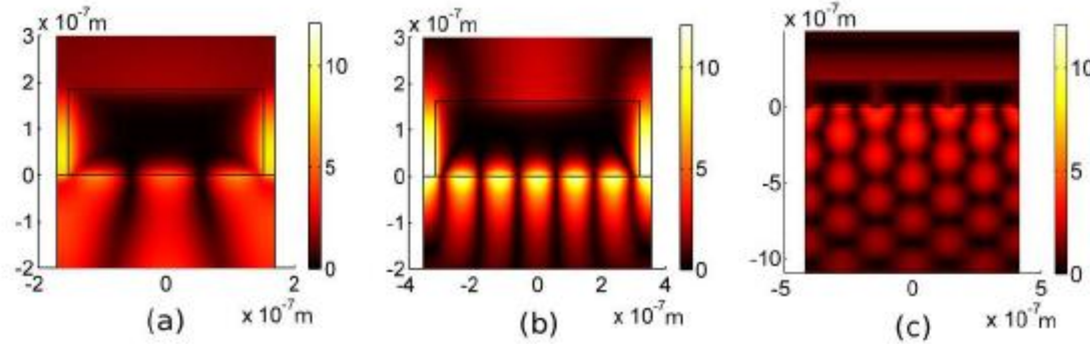


Fig. 2.8. Magnetic field enhancements in correspondence of: (a) CM resonance; (b) SPP-CM resonance; (c) WR-CM resonance. [15]

It is worth noting that CMs are local resonances, since they would appear as well in a single illuminated slit without any periodicity. On the other hand SPPs are global resonances, since they can exist thanks to the coherent periodicity of the structure. As many authors pointed out, however, CMs and SPPs are not independent to each other. Actually one should better speak of a hybrid mode which presents both CM and SPP resonant characteristics. The SPP character dominates the transmission properties of the grating as the pitch gets proximate to one of those satisfying Eq. (2.5). On the other hand the EOT efficiency is greatly enhanced in presence of a periodic structure.

In Fig. 2.9 are shown the maps of transmittance and of absorption within 300nm in silicon, together with resonant (h, d) configurations, obtained with FEM for a slit-to-period ratio of 10%. We see that cavity modes resonances (black lines) are clearly associated with an enhanced transmission in the (h, d) configurations far from the SPP resonant ones (grey lines) in which the dominant SPP character determines a transmission extinction. Considerably large transmittances can be obtained optimizing the duty cycle.

Cavity modes are also correlated to absorption enhancements. Within 300nm in Silicon [Fig. 2.9 (b)], the largest enhancements are obtained when both the conditions for SPP and cavity mode excitation are satisfied at the same time (at the crossing between gray black lines). In these configurations CM resonances canalize the whole incident power into an SPP mode rather than into a z-propagating mode, giving rise to an extraordinary SPP [Fig. 2.8 (b)]. By conversion, in Fig. 2.8 (a) it

is shown the typical much weaker field enhancement in a configuration where only CM is excited without SPP resonant coupling.

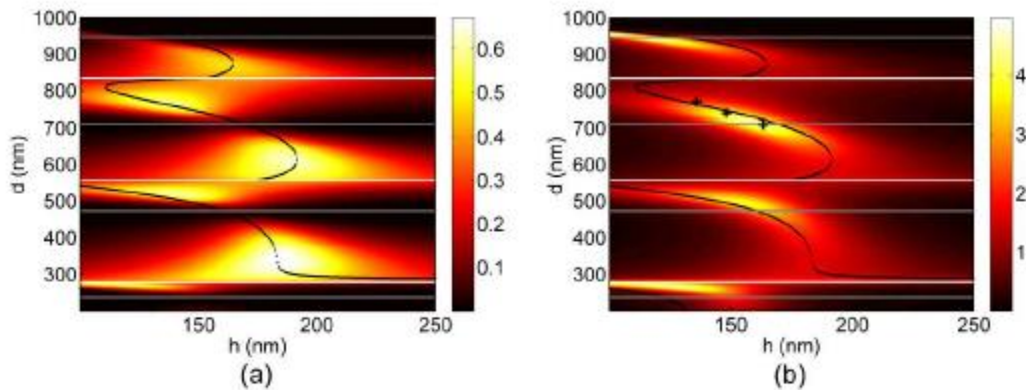


Fig. 2.9. FEM-calculated transmittance (a) and absorption enhancement (b) map within 300nm in silicon. Overplotted black and grey lines mark respectively CM and SPP resonances according to the analytical model. Crosses mark configurations whose absorption profile is reported in Fig. 2.10. [15]

Looking at the absorption profiles (Fig. 2.10) we observe that, in strongly SPP-CM-resonant configurations (black line), most of the enhancement is confined in the metal proximity and decays exponentially within the typical plasmonic decay length. Far from the grating (Fig. 2.10, inset) the absorption is almost zero in SPP resonances since SPPs typically produce transmission extinctions. On the other hand, strong CM but not SPP resonant configurations (light grey lines) show absorption profiles that are much less confined close to the surface, but are higher in depth due to EOT.

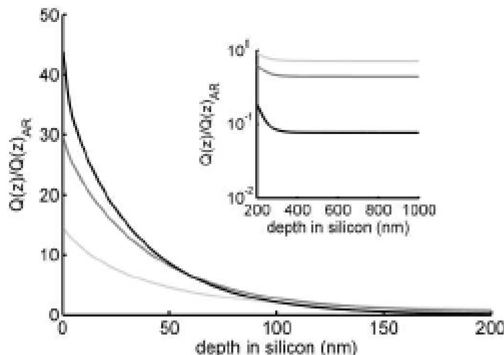


Fig. 2.10. Absorption profile enhancement in configurations marked with crosses in Fig. 2.9 (b) (coloured in light gray, medium gray and black for increasing h); inset: absorption profiles within 1 μm depth. [15]

Barbara et al. [16] presented in 2002 the first experimental study in the optical regime with gold-coated silica gratings and they were able to interpret their results by invoking the excitation of both SPPs and slit-cavity resonances. Recently, Pang et al. [17] measured the dispersion relation of the transmission resonances appearing in 1D arrays of subwavelength slits in both the visible and near infrared (NIR) ranges of the EM spectrum. Their experimental transmission spectra, shown in Fig. 2.11, confirm the theoretical predictions regarding the existence of two different channels, SPPs and slit-cavity resonances, that enhance the transmission through a 1D periodic array of slits. The two main transmission peaks emerging in the spectrum result from the strong interplay between these two types of resonances, as described previously.

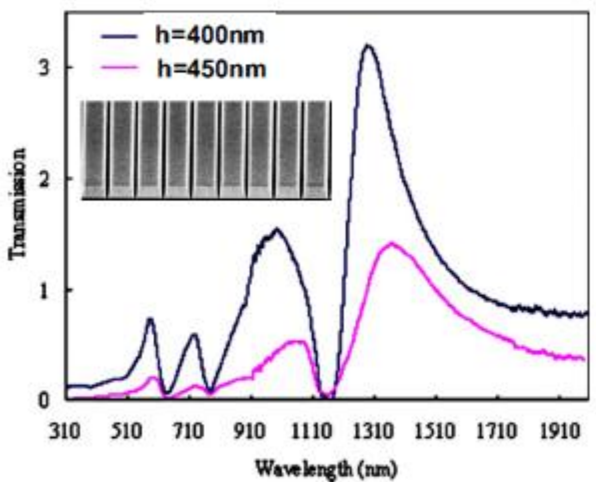


Fig. 2.11. Transmission through 1D periodic arrays of slits. Inset: SEM image of one of the 1D arrays analyzed. The period of the array is $d = 750nm$, the width $a = 1050nm$ and two different thicknesses are analyzed, $h = 400nm$ and $h = 450nm$. Main panel: Zero-order transmittance of a 1D slit array illuminated by a normal incidence p-polarized plane wave. [17]

SERS

Introduction

SERS (Surface-Enhanced Raman Scattering) effect consists in the Raman signals' amplification coming mainly through the electromagnetic interaction of light with metals and the excitation of plasmons. To exploit this effect, the emitting molecules must typically be adsorbed on the metal surface, or at least very close to it (10 nm maximum).

The denomination of SERS well summarizes the three characteristics of the effect:

- *Surface* (S): SERS is a surface spectroscopy technique; the molecules must be on (or close to) the surface. One must ensure that the molecules to be detected can attach to (or at least be in close proximity to) the surface of the metal substrate.
- *Enhanced* (E): the signal enhancement is provided by plasmon resonances in the metal substrates, which can go from metallic colloids in solution to substrates fabricated by nano-lithography or self-organization.
- *Raman* (R): the technique consists in measuring the Raman signals of molecules. Raman spectroscopy is the study of inelastic light scattering and, when applied to molecules, it provides an insight into their chemical structure (in particular their vibrational structure).
- *Scattering* or *Spectroscopy* (S), depending on whether one prefers to emphasize the optical effect (scattering) or the technique and its applications (spectroscopy).

Among the many parameters that can be varied in a SERS experiment, two stand out naturally: the molecular species to be detected (the probe), and the metallic structures onto which it adsorbs (the SERS substrate).

Good SERS substrates are those that support the 'strongest' plasmon resonances; in other words, those that provide the largest enhancement or amplification. In this respect, one should in addition distinguish between those that provide a relatively uniform enhancement on the surface and those with large variations. The latter typically exhibit some highly localized positions of very high enhancement (hot-spots), particularly suited for single-molecule detection. Nevertheless, the former should be preferred for reproducibility in applications. Moreover, because the SERS enhancements arise from a resonant response of the substrate, they are typically strongly wavelength dependent, i.e. they vary with the excitation wavelength (and to a lesser degree with the Raman shift of the modes). A given SERS substrate will, therefore, typically exhibit good enhancements in a limited excitation wavelength range. Most SERS substrates are designed to operate with visible/near-infrared excitation (~400-1000 nm), which is the typical range of interest for molecular Raman scattering experiments, and typically arise from structures made of gold or silver, the two metals most used for SERS and plasmonics in general.

Not all molecules are good SERS probes, even though the technique can be used with a remarkable variety of analytes. The two major characteristics of a SERS probe are: the intensity of Raman scattering (characterized by the Raman cross-section) and the efficiency of adsorbtion on the SERS substrate to be used, i.e. typically on gold or silver surfaces. Some molecules (e.g. molecules with thiol or triazole moieties in their structure) have a strong chemical affinity for such metal surfaces (for example forming strong covalent bonds) and are therefore easier to work with.

Before continuing, it is worth having a short overview of the source of the commonly used expression for the SERS enhancement power, together with a brief look at typical data to convey a direct visual impression of the amplification of Raman signals achieved in SERS.

3.1 Raman Spectroscopy

Raman Spectroscopy is a spectroscopic technique based on the inelastic scattering process between a photon and a molecule, mediated by a fundamental vibrational or rotational mode of the molecule. Due to energy exchange between the scattering partners, the incoming photon of energy $h\nu_L$ is shifted in energy by a characteristic energy of vibration $h\nu_V$. This shift can be in both directions, depending on whether the molecule in question is in its vibrational ground state or in an excited state. In the first case, the photon loses energy by excitation of a vibrational mode (Stokes scattering). In the second case, additional energy gain by de-excitation of such a mode (anti-Stokes scattering) is possible.

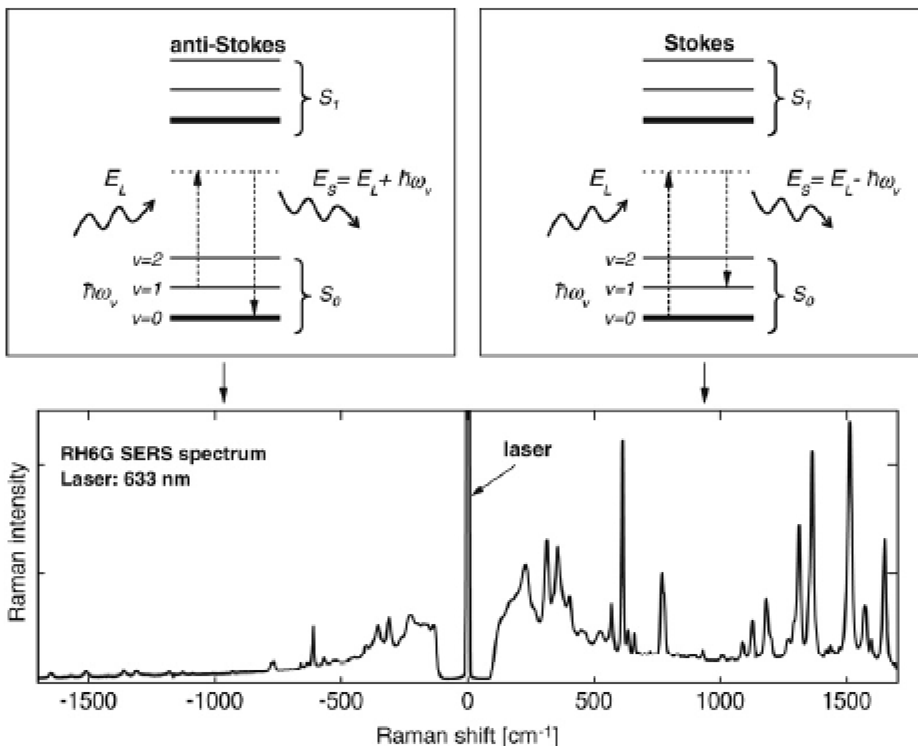


Fig. 3.1 Diagrams of the anti-Stokes (top left) and Stokes (top right) scattering. Bottom: a typical Raman spectrum (Raman intensity vs Raman shift) for rhodamine 6G, with several Raman peaks on the Stokes side and their (weaker) anti-Stokes counterparts. [18]

As a consequence, the molecule produces a Raman spectrum with a series of peaks (which are fingerprints of a specific molecule), an example of which, for the dye rhodamine 6G (RH6G), is shown in Figure 3.1.

3.2 The SERS Enhancement Factor

The Raman scattering is a linear process: the total power of the inelastically scattered beam scales linearly with the intensity of the incoming excitation beam. In the Stokes process the power of the scattered beam can be expressed as

$$P_S(\nu_S) = N\sigma_{\mathrm{RS}} I(\nu_L) \tag{3.1}$$

where N is the number of Stokes-active scatterers within the excitation spot, σ_{RS} is the scattering cross section, and $I(\nu_L)$ the intensity of the excitation beam.

SERS describes the enhancement of this process, accomplished by placing the Raman-active molecules within the near-field of a metallic nanostructure. The enhancement of P_S is mainly due to the increased electromagnetic field produced by excitations of surface plasmons at the metal interface. This leads to an enhancement of both the incoming and emitted light fields, expressed via

$$L(\nu) = \frac{|E_{\mathrm{Loc}}(\nu)|}{|E_0|} \tag{3.2}$$

where $|E_{\mathrm{Loc}}|$ is the local field amplitude at the Raman active site.

$L(\nu)$ is called the electromagnetic enhancement factor. The total power of the Stokes beam under SERS conditions is

$$P_S(\nu_S) = N\sigma_{\mathrm{SERS}} L(\nu_L)^2 L(\nu_S)^2 I(\nu_L) \tag{3.3}$$

Since $|L(\nu_L)| \approx |L(\nu_S)|$, this brings to the important result that the electromagnetic contribution to the total SERS enhancement is proportional to the fourth power of the field enhancement factor.

The commonly used expression for the enhancement of the power of the Stokes beam is therefore

$$\mathrm{EF} = \frac{|E_{\mathrm{Loc}}|^4}{|E_0|^4} \tag{3.4}$$

In Fig. 3.2 the Raman (non-SERS) and SERS spectra of a very common (and widely used) SERS dye, rhodamine 6G (RH6G), are shown for two different conditions. The spectrum at the bottom is obtained in a solution (100 μM) of RH6G in water (no SERS amplification) with an integration time of 400 s. The number of molecules contributing to the signal can be obtained from the concentration and the knowledge of the scattering volume of the system. The signal can be compared to that of a single RH6G molecule under SERS conditions, here in an aggregated Ag colloidal solution, obtained in the same system, with the same experimental conditions, but with an integration time of only 0.05 s. If we now compare the signals of both cases and normalize for the different integration times and the number of molecules, we conclude that the signal of the single molecule in SERS conditions has been amplified

by a factor $\sim 7 \times 10^9$ (the enhancement factor) compared to that of the same single molecule in normal Raman conditions. This figure should suffice to demonstrate visually the SERS capability of boosting the sensitivity of Raman experiments to the point of detecting a single molecule.

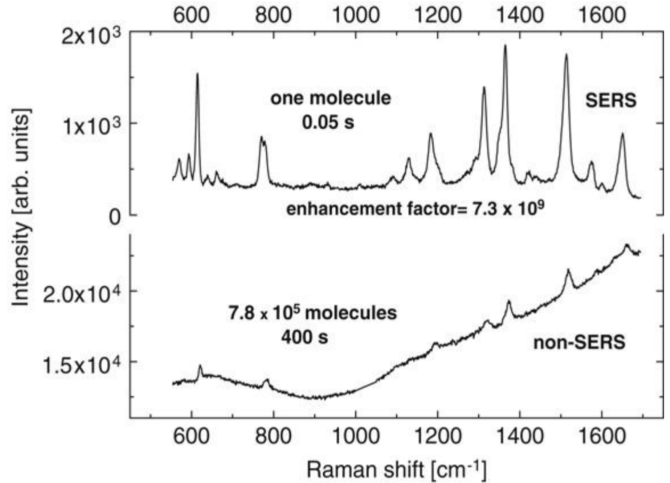


Fig. 3.2 Raman (non-SERS) and SERS spectra at 633nm laser excitation for rhodamine 6G molecules (RH6G). Bottom spectrum: signal of 7.8×10^5 RH6G molecules with 400 s integration time. Top spectrum: signal from a single RH6G molecule with 0.05 s integration time. The enhancement factor is 7.3×10^9. [18]

3.3 Definition of the SERS Enhancement Factors

Electromagnetic and chemical enhancements

The SERS EFs are traditionally separated into two main multiplicative contributions:

the electromagnetic (EM) enhancement *factor,* F_{EM}, is thought to be, by far, the main contribution. It is due to the coupling of the incident and Raman electromagnetic fields with the SERS substrate and it can usually be separated into two multiplicative EFs, one for the incident field, and one for the re-emitted (Raman) field, as in Eq. (3.3). The EM enhancement relies on the large local field enhancements that occur close to metallic surfaces when localized surface plasmon (LSP) resonances are excited. To profit from these, the molecule must then be in close proximity to the surface (typically within 10 nm from the surface, at most). In most implementations, the molecule is actually directly adsorbed on the surface, either through physisorption or chemisorption (i.e. through the formation of a chemical bond with the metal).

Another (multiplicative) contribution to the EF is the so-called *chemical enhancement (CE) factor,* F_{Chem}. Its existence are still subject to controversy [19] and its contribution is, in any case, believed to be much smaller than the EM effect. The CE factor is sometimes viewed as a modification of the electronic polarizability of the probe, which can induce resonant-Raman scattering (and therefore enhanced signals) at wavelengths where the non-adsorbed molecules would not be resonant [19]. The most widely accepted explanation for this is the so-called charge-transfer (CT) mechanism [20]. This requires the molecule to be chemically adsorbed on the surface (hence the name chemical enhancement).

Factors influencing the SERS enhancements

The SERS process (and therefore the enhancement factor) depends on a long list of parameters, including:

- *Characteristics of the laser excitation*, in particular: wavelength, polarization, angle of incidence (for a planar substrate), etc.;
- *Detection setup*, in particular: scattering configuration (e.g. backscattering geometry), solid angle for collection, polarized and/or unpolarized detection, etc.;
- *SERS substrate*, in particular: material (usually silver or gold), geometry, orientation with respect to incident beam direction and polarization, and refractive index of the environment n_M;
- *Intrinsic properties of the analyte*, in particular intrinsic Raman cross-sections;
- *Analyte adsorption properties*, in particular: adsorption efficiency and analyte concentration (surface coverage), distance from the surface, adsorption orientation (fixed or random), and the possible modification of the intrinsic Raman cross-section induced by adsorption. This latter aspect is essentially the chemical contribution to SERS.

We note here that it is difficult to account for all of these parameters, and many of them may, in fact, be unknown.

It is useful to emphasize an important characteristic of most SERS substrates. The local field enhancements depend strongly on the exact position of the molecule on the surface. As a result, the SERS EFs on the surface of a typical SERS substrate are highly non-uniform, even at a molecular scale of nanometers. Points of very large enhancements, the so-called ‘hot-spots’, are generally highly localized, and can be within tens of nanometers of points with little or negligible enhancement.

As a consequence, many applications are interested instead in the average enhancement factor, i.e. for molecules with random positions on the substrate.

3.3.1 The SERS Substrate Enhancement Factor - experimental approach

The *SERS Substrate Enhancement Factor* (SSEF) can be defined as:

$$\mathrm{SSEF} = \frac{I_{\mathrm{SERS}} / N_{\mathrm{Surf}}}{I_{\mathrm{RS}} / N_{\mathrm{Vol}}} \tag{3.5}$$

where $N_{\mathrm{Vol}} = c_{\mathrm{RS}} V$ is the average number of molecules in the scattering volume V for the Raman (non-SERS) measurement, and N_{Surf} is the average number of adsorbed molecules in the same scattering volume for the SERS experiments. This expression is usually considered as the best estimate of the average SERS EF for a given SERS substrate. It is, however, necessary to include in this definition an additional constraint to make it more accurate: the surface coverage must remain smaller than one monolayer when using Eq. (3.5). This is because the SERS effect is distance-dependent and the SERS signals from molecules on the second monolayer and beyond are therefore typically reduced. This additional constraint ensures that the average SERS intensity is proportional to the average number of adsorbed molecules, which is necessary for Eq. (3.5) to remain independent of N_{Surf}. This is not a

limitation since the interest for many practical applications is in the low (or even ultra-low) concentration limit.

From an experimental point of view, we note that $I_{\mathrm{RS}}/N_{\mathrm{Vol}}$ is simply the average (over molecular orientations) Raman intensity per free molecule, denoted $\langle I_{\mathrm{RS}}^{SM}\rangle$. It is directly related to the differential Raman cross-section $\mathrm{d}\sigma_{\mathrm{RS}}/\mathrm{d}\Omega$ by:

$$\langle I_{\mathrm{RS}}^{SM}\rangle = \frac{\mathrm{d}\sigma_{\mathrm{RS}}}{\mathrm{d}\Omega} S_{\mathrm{Inc}}\,\Delta\Omega_{\mathrm{Det}} \tag{3.6}$$

where S_{Inc} is the incident power density at the molecule position and $\Delta\Omega_{\mathrm{Det}}$ is the solid angle for detection.

The difficulty in the practical estimation of the SSEF lies in the determination of N_{Surf}, and to a lesser extent of N_{Vol} (or equivalently of the scattering volume V).

3.3.2 The SERS cross-section and single-molecule EF

As mentioned already, SERS enhancements on most SERS substrates are highly non-uniform, even at a molecular scale of nanometers. The SSEF is a spatially-averaged property that ignore this non-uniformity. To account for this, it is therefore useful to define a single-molecule enhancement factor (SMEF). This is the SERS enhancement felt by a given molecule at a specific point. It is in general dependent on the symmetry of the Raman tensor of the mode and on the probe orientation on the SERS substrate. It is also dependent on the orientation of the SERS substrate with respect to the exciting polarization. It therefore theoretically requires the exact knowledge of the SERS substrate geometry, and the exact position and orientation of the probe on it. The SMEF could be defined simply as the ratio of the differential SERS intensity of a single molecule $\mathrm{d}P_{\mathrm{SERS}}/\mathrm{d}\Omega$ to the differential Raman intensity $\mathrm{d}P_{\mathrm{RS}}/\mathrm{d}\Omega$ of the same molecule under the exact same conditions, but in the absence of the metallic SERS substrate. The definition as stated above, however, presents a major shortcoming because the SERS intensity is compared to that of the free molecule in a fixed position and orientation. Furthermore, under SERS conditions, the differential scattered SERS power $\mathrm{d}P_{\mathrm{SERS}}/\mathrm{d}\Omega$ depends in general both on the incident polarization (with respect to the SERS substrate) and on the detection direction. For these reasons, it is not possible to define a general differential SERS cross-section as is the case for normal Raman. However, for a given incident polarization (IP), scattering configuration (SC), molecule position (**r**) and molecule orientation (MO), one can still define the *absolute differential SERS cross-section* (or SERS cross-section for short) $\mathrm{d}\sigma_{\mathrm{SERS}}/\mathrm{d}\Omega$ as:

$$\frac{\mathrm{d}P_{\mathrm{SERS}}}{\mathrm{d}\Omega}(\mathrm{IP},\mathrm{SC},\mathbf{r},\mathrm{MO},\omega_L,n_M) = \frac{\mathrm{d}\sigma_{\mathrm{SERS}}}{\mathrm{d}\Omega}(\mathrm{IP},\mathrm{SC},\mathbf{r},\mathrm{MO},\omega_L,n_M)S_{\mathrm{Inc}} \tag{3.7}$$

Like the Raman cross-section, the SERS cross-section depends on excitation wavelength or frequency ω_L, the refractive index of the environment n_M, and obviously the vibrational mode under consideration. It depends in addition on a collection of additional parameters: IP, SC, **r**, and MO. Its utility is therefore limited unless these are clearly stated or further assumptions are made.
In a similar fashion, the *single-molecule enhancement factor* (SMEF) can be defined as:

$$\mathrm{SMEF} = \frac{\mathrm{d}\sigma_{\mathrm{SERS}}/\mathrm{d}\Omega}{\mathrm{d}\sigma_{\mathrm{RS}}/\mathrm{d}\Omega} = \frac{\mathrm{d}P_{\mathrm{SERS}}/\mathrm{d}\Omega\,(\mathrm{IP},\mathrm{SC},\mathbf{r},\mathrm{MO},\omega_L,n_M)}{\langle \mathrm{d}P_{\mathrm{RS}}/\mathrm{d}\Omega\,(90°,\text{or BS},\text{or FS})\rangle} \tag{3.8}$$

where

$$\langle \frac{\mathrm{d}P_{\mathrm{RS}}}{\mathrm{d}\Omega}(90°, \mathrm{or\ BS}, \mathrm{or\ FS}) \rangle = \frac{\mathrm{d}\sigma_{\mathrm{RS}}}{\mathrm{d}\Omega} S_{\mathrm{Inc}} \tag{3.9}$$

and $\langle ... \rangle$ denotes the molecular-orientation averaging (for 90° scattering, back-scattering (BS) or forward-scattering (FS) configurations).

Note that the normal Raman cross-section should be that measured in the same dielectric environment (typically air, water, or another solvent) as that used for the SERS experiment.
Although it appears here as a simple ratio between cross-sections, it is important to realize the differences in the meaning of the SERS and non-SERS cross-sections. The former depends on many external parameters, while the latter is an intrinsic property of the analyte (for a given environment and excitation wavelength).

Orientation-averaged SMEF

For a given substrate configuration (IP and SC) and a given molecule position/orientation, the definition of the SMEF reflects directly the SERS intensity. However, the molecule orientation is not necessarily known or even fixed during the time of a measurement. It can therefore also be useful to define the *orientation-averaged SMEF* (OASMEF) at a given position as:

$$\mathrm{OASMEF} = [\mathrm{SMEF}] = \frac{[\mathrm{d}P_{\mathrm{SERS}}/\mathrm{d}\Omega\,(\mathrm{IP}, \mathrm{SC}, \mathbf{r}, \mathrm{MO}, \omega_L, n_M)]}{\langle \mathrm{d}P_{\mathrm{RS}}/\mathrm{d}\Omega\,(90°, \mathrm{or\ BS}, \mathrm{or\ FS}) \rangle} \tag{3.10}$$

where the average denoted by $[...]$ is taken over all allowed orientations of the molecule under SERS conditions.

3.3.3 The SERS Substrate Enhancement Factor - formal definition

Having defined rigorously the differential SERS cross-section and the SMEF (and OASMEF), we can now provide a definition of the SERS substrate enhancement factor (SSEF) from a more formal (and rigorous) perspective.

Rigorous definition of the SSEF

Let us consider a uniform coverage of a probe molecule on the surface of a SERS substrate. The SERS EF of each molecule can be characterized by the OASMEF at its position **r**. The SSEF can therefore be rigorously defined as the spatially-averaged OASMEF on the surface, i.e.

$$\mathrm{SSEF} = \frac{1}{A_M} \int_{A_M} \mathrm{OASMEF}(\mathbf{r})\, dS \tag{3.11}$$

where A_M represents the surface area of the metallic substrate. This last expression can be rewritten as:

$$\text{SSEF} = \{\text{OASMEF}\} = \{[\text{SMEF}]\} \qquad (3.12)$$

where $\{\ldots\}$ denotes the spatial averaging (on the substrate surface) and $\{[\text{SMEF}]\}$ is therefore the spatial-and-allowed-orientation-averaged single-molecule EF.

3.4 Experimental measurement of SERS Enhancement Factors

3.4.1 Link between SSEF definition and experiments

The formal definition of the SSEF, as a spatial average of the OASMEF, was given in Eq. (3.11). Although rigorous, it is somewhat distant from the more commonly-used expression given in Eq. (3.5). We will make the connection here between these two, focusing on the most relevant case of planar SERS substrates. This is also the most difficult case, because of the problems of comparing signals originating from a 2D structure to that of a 3D volume. This connection will give a rigorous meaning to N_{Vol} and N_{Surf} in Eq. (3.5), therefore allowing for a meaningful experimental estimation of the SSEF.

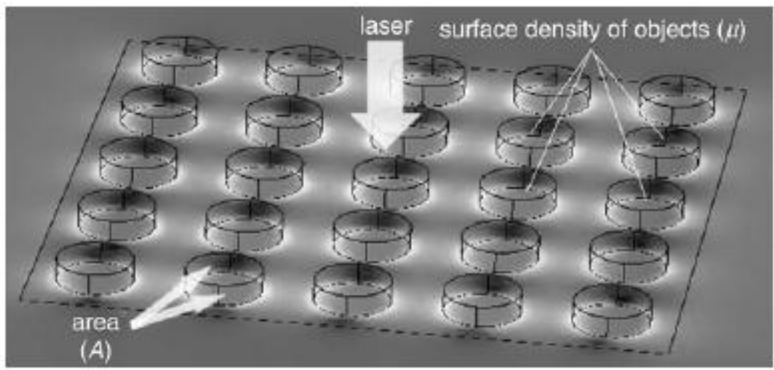


Figure 3.3 A planar SERS substrate formed by a repeating pattern of individual 3D subwavelength metallic objects on an underlying planar substrate. The surface density of individual nano-structures is μ_M and their metallic surface area is A_M (each) (which includes both the top and side areas of each object). The array is illuminated from above with a laser.

Let us consider a SERS experiment on a typical planar SERS substrate consisting of a repeating pattern of individual 3D sub-wavelength metallic structures, as illustrated in Fig. 3.3.
Let μ_M be the surface density of the individual nano-structures with respect to the main plane forming the substrate, and A_M be the metallic surface area in each structure (usually larger than the projected surface area or shadow of the structure on the main plane). The excitation is usually non-uniform and characterized in the focal plane by an intensity profile $S_{\text{Inc}}(\rho)$ and a corresponding power P_{Inc}, which can be obtained by integration of $S_{\text{Inc}}(\rho)$. We assume that the excitation area is much larger than the individual structures forming the substrate and that the exciting intensity is approximately uniform for a given single structure.

Let us now calculate the SERS and non-SERS intensities: I_{SERS} and I_{RS} used in Eq. (3.5).

SERS intensity

One molecule at a given position $\mathbf{r}'$ on the surface of an individual metallic object at a position ρ within the beam, emits a SERS signal equal to:

$$I(\mathbf{r}') = \mathrm{OASMEF}(\mathbf{r}')\frac{\mathrm{d}\sigma_{\mathrm{RS}}}{\mathrm{d}\Omega}S_{\mathrm{Inc}}(\rho)\Delta\Omega_{\mathrm{Det}} \tag{3.13}$$

Assuming a surface density of molecules μ_S on the metal, the SERS signal from this single object is then:

$$I_{\mathrm{SERS}}^{NP} = \mu_S A_M\{\mathrm{OASMEF}\}\frac{\mathrm{d}\sigma_{\mathrm{RS}}}{\mathrm{d}\Omega}S_{\mathrm{Inc}}(\rho)\Delta\Omega_{\mathrm{Det}} \tag{3.14}$$

where $\{\ldots\}$ denotes spatial averaging. Adding the contribution of all the metallic objects in the beam (by integrating over ρ), we obtain the total SERS signal as:

$$I_{\mathrm{SERS}} = \mu_M\mu_S A_M\{\mathrm{OASMEF}\}\frac{\mathrm{d}\sigma_{\mathrm{RS}}}{\mathrm{d}\Omega}P_{\mathrm{Inc}}\Delta\Omega_{\mathrm{Det}} \tag{3.15}$$

Non-SERS intensity

We now compare this SERS signal to that obtained with the same setup under non-SERS conditions I_{RS} from a solution of concentration c_{RS} [molecules m^{-3}]. We must relate I_{RS} to $\mathrm{d}\sigma_{\mathrm{RS}}/\mathrm{d}\Omega$ and P_{Inc}. For a small element of volume $\rho \mathrm{d}\rho \mathrm{d}z$, we have $\mathrm{d}I_{\mathrm{RS}} = \mathrm{c}_{\mathrm{RS}}(\mathrm{d}\sigma_{\mathrm{RS}}/\mathrm{d}\Omega)S_{\mathrm{Inc}}(\rho, z)\rho \mathrm{d}\rho \mathrm{d}z\Delta\Omega_{\mathrm{Det}}$. Integration over ρ simply gives P_{Inc}, while integration over $\mathrm{d}z$ gives the height H of the scattering volume, resulting in: $I_{\mathrm{RS}} = \mathrm{c}_{\mathrm{RS}}(\mathrm{d}\sigma_{RS}/\mathrm{d}\Omega)HP_{\mathrm{Inc}}\Delta\Omega_{\mathrm{Det}}$.

In practice, one must define (and measure) the height of the scattering volume. In fact, the collection efficiency is not uniform along the beam axis and depends on the experimental setup (confocal depth). This, together with the possible non-uniform excitation along the beam axis and the 3D nature of the scattering volume, complicates the definition of the H value.

Finally:

$$I_{\mathrm{RS}} = c_{\mathrm{RS}}H_{\mathrm{Eff}}\frac{d\sigma_{\mathrm{RS}}}{d\Omega}P_{\mathrm{Inc}}\Delta\Omega_{\mathrm{Det}} \tag{3.16}$$

where H_{Eff} is the effective height of the scattering volume. It is precisely this length scale that allows the connection between 2D and 3D measurements. H_{Eff} will be better define in chapter 10, along with its experimental evaluation.

Combining Eq. (3.15) with Eq. (3.16), and recalling Eq. (3.11), we can write:

$$\mathrm{SSEF} = \frac{I_{\mathrm{SERS}}/(\mu_M\mu_S A_M)}{I_{\mathrm{RS}}/(c_{\mathrm{RS}}H_{\mathrm{Eff}})} \tag{3.17}$$

This expression reduces to the 'non-rigorous' definition of the SSEF Eq. (3.5), provided that:

$$N_{\mathrm{Surf}} = \mu_M \mu_S A_M A_{\mathrm{Eff}} \tag{3.18}$$

and

$$N_{\mathrm{Vol}} = c_{\mathrm{RS}} V = c_{\mathrm{RS}} H_{\mathrm{Eff}} A_{\mathrm{Eff}} \tag{3.19}$$

where A_{Eff} is the effective surface area of the scattering volume (the equivalent of a scattering volume, but for 2D). The exact definition of A_{Eff} is not important since it cancels out in any case in the expression for the SSEF. It is only introduced here in order to interpret the meaning of Eq. (3.17) in terms of number of molecules (on the surface or in the volume).

3.5. Overview of the main EM effects in SERS

As mentioned already, the SERS enhancement is often separated into two contributions, the first electromagnetic, the second chemical. Within this approach, electromagnetic enhancements apply to all analytes, while chemical enhancements are probe-dependent and in general require chemical bonding of some sort to the metal substrate. If both mechanisms coexist, then their effect are in general assumed to be cumulative (i.e. the enhancement factors are multiplicative). The relative contribution, or even the existence of the chemical enhancement, has been the subject of much debate over the years, and is still arguably not fully resolved. It is however widely accepted that the main origin of the SERS enhancement is electromagnetic. Understanding the electromagnetic contributions to the SERS effect is therefore paramount. In the following, we will describe in detail the origin of the EM-enhancement contribution to the SERS effect and we will introduce the local field and radiation enhancement factors, as well as the common $|E|^4$-approximation to the single-molecule EF (SMEF).

3.5.1 Analysis of the EM problem of SERS

The study of the EM-SERS enhancement factor is equivalent to the EM problem of Raman scattering by a molecule in the vicinity of a metallic surface. The simplest approach, and arguably the one that gives most insight into the SERS effect is the phenomenological description of Raman scattering within classical EM theory. Hence, we will almost exclusively focus on this approach.

The phenomenological approach to Raman scattering

Within this approach, the Raman response of a given Raman mode of a molecule is fully characterized by its Raman polarizability tensor $\hat{\alpha}_R$. Under monochromatic light excitation at a frequency ω_L, the electric field $\mathbf{E}_{\mathrm{Inc}}$ at the molecule position induces a Raman dipole $\mathbf{p}_R = \hat{\alpha}_R \cdot \mathbf{E}_{\mathrm{Inc}}$, oscillating at the Raman frequency ω_R. Here we will ignore for simplicity the tensorial character of $\hat{\alpha}_R$ and simply write $\mathbf{p}_R = \alpha_R \mathbf{E}_{\mathrm{Inc}}$, equivalent to considering an isotropic tensor. This oscillating Raman

dipole radiates a power proportional to $|\mathbf{p}_R|^2$ at frequency ω_R, and it is this radiation that we detect in the far-field as the Raman signal.

Application to SERS

The same phenomenological description can be applied to a Raman scatterer under SERS conditions. However, the presence of the metal will affect these processes in a number of ways. Firstly, the electromagnetic field at the molecule position is modified. In cases of interest, its amplitude can be dramatically enhanced. This results in a possible local field enhancement. Secondly, the radiation properties of the Raman dipole $\mathbf{p}_R$ are modified. This results in a possible radiation enhancement. Thirdly, the Raman polarizability tensor $\hat{\alpha}_R$ may be modified. Such a modification would typically be classified as the chemical enhancement (CE), although it may be electromagnetic in origin. This aspect is secondary in most cases and we will neglect it here. The SERS EM problem therefore consists mainly in understanding the first two effects: modification of the EM field and of the radiation properties of a dipole.

The majority of SERS experiments consists in studying the Raman signal emitted by one or more molecules, in close proximity of a metallic substrate with sub-wavelength features. A number of parameters needs to be given to study this problem:

- *Metallic substrate*: exact geometry and optical properties (depending on the material used);
- *Probe molecule*: exact orientation and position (including distance to surface). Also optical properties of the probe, such as Raman polarizability tensor;
- *Optical setup*: excitation wavelength, polarization, angle of incidence, power density, and beam profile (plane-wave approximation or Gaussian beam, for example). Also important is the scattering configuration, which de nes the direction from which the emitted light is collected with respect to excitation. Standard configurations include back-scattering (when the delivering optics is also used for collection), forward-scattering, or 90°-scattering. Moreover, polarizers can also be used in the detection.

It is clear that it is difficult to actually know accurately all of these parameters. Some of them can be estimated or measured; others, like the adsorption geometry of a probe or its distance from the surface, are more difficult to measure, and in general some 'reasonable' assumptions have to be made.

3.5.2 Local field enhancement

The local field intensity enhancement factor

The electromagnetic field is strongly modified in the vicinity of metallic objects. This is particularly true when the excitation wavelength ω_L is close to the electromagnetic resonances of the system (in particular localized surface plasmon resonances for metals). This means that the electric field $\mathbf{E}_{\mathrm{Loc}}$ at the molecule position can be very different, both in magnitude and orientation, to the incident field $\mathbf{E}_{\mathrm{Inc}}$. $\mathbf{E}_{\mathrm{Loc}}$ is usually called the local field and is the macroscopic electric field felt by the molecule. It depends on excitation wavelength and polarization, and can also vary dramatically with position. In

cases of interest to SERS, the magnitude $|\mathbf{E}_{\text{Loc}}|$ on the metallic surface can be much larger than $|\mathbf{E}_{\text{Inc}}|$. This increase can sometimes be particularly spectacular at some specific (localized) positions on the surface, the so-called 'hot-spots'. The local field induces a Raman dipole $\mathbf{p}_R = \alpha_R \mathbf{E}_{\text{Loc}}(\omega_L)$, whose magnitude is therefore enhanced by a factor $|\mathbf{E}_{\text{Loc}}(\omega_L)|/|\mathbf{E}_{\text{Inc}}|$. If such a dipole radiates in free-space (no metal), then the energy radiated, which is proportional to $|\mathbf{p}_R|^2$, would then be enhanced by a factor:

$$M_{\text{Loc}}(\omega_L) = \frac{|\mathbf{E}_{\text{Loc}}(\omega_L)|^2}{|\mathbf{E}_{\text{Inc}}|^2} \tag{3.20}$$

This factor, which we will call the local field intensity enhancement factor is associated with the excitation of the Raman dipole. Note that it characterizes the enhancement of the electric field intensity, but ignores any modification of the electric field polarization.

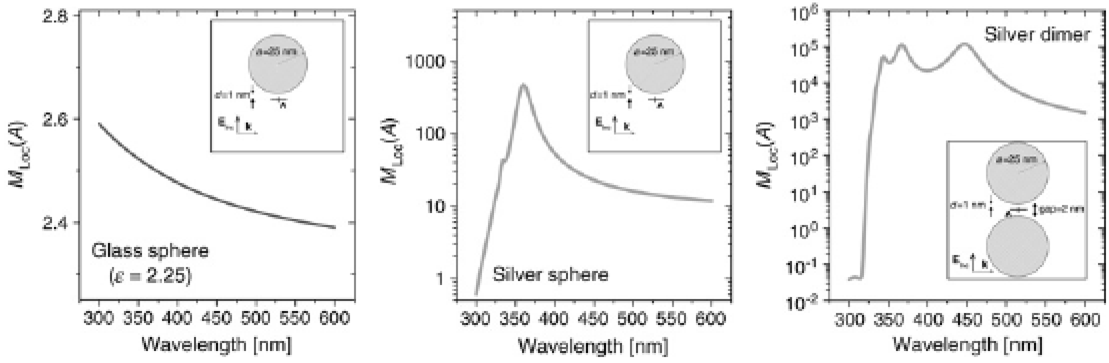


Figure 3.4. Examples of modification of the local field intensity, calculated for a molecule in air at a distance d = 1 nm from the surface at point A (shown in the insets) using Mie theory. The cases of a 50 nm-diameter dielectric glass or metallic silver spheres are shown, along with that of a molecule at the gap between two closely-spaced silver spheres (a dimer). For a glass sphere, there is a small enhancement, hardly dependent on wavelength. For a silver sphere, $M_{\text{Loc}}(A)$ varies widely with wavelength, showing resonances associated with the localized surface plasmon resonances of the silver sphere. Such effects are even more striking in the case of a molecule in between two silver spheres separated by a 2 nm gap (note the log-scales for the silver sphere and dimer). This case is a good representation of a SERS hot-spot. [18]

An example of the local field enhancement effect is illustrated in Fig. 3.4 by considering the wavelength dependence of M_{Loc} at a specific point at a distance d = 1 nm from the surface of a sphere. In the figure are compared the results for two materials: a non-absorbing dielectric such as glass, and a metal like silver. The case of a dipole in the gap of a silver dimer is also shown. These examples illustrate that large local field intensity enhancement factors are possible close to the surface of a metallic object. They are associated with a coupling to localized surface plasmon resonances, as evidenced by the resonant response as a function of wavelength. The presence of the metal in this case acts as 'funnel' to concentrate the electromagnetic fields on the surface. Under the right conditions, as in the gap between two metallic objects, huge local field enhancements of the order of $\sim 10^5$ can be obtained.

3.5.3 Radiation enhancement

The radiation enhancement factor

Under SERS conditions, the Raman dipole radiates, not in free-space, but in close proximity to the metal. Dipole radiation is strongly affected by the metal, in a similar fashion as the exciting field is modified. This phenomenon is well known in the case of atomic spontaneous emission, and will be referred to as modified spontaneous emission (MSE). The environment affects the dipole emission in two ways:

- the radiation pattern $\mathrm{d}P_{Rad}/\mathrm{d}\Omega$ (power radiated in the far-field per unit solid angle in a given direction) may be modified;
- the total power radiated by the dipole, P_{Rad}, can also be enhanced or quenched compared to that in free-space, P_0. We can therefore define a radiation enhancement factor as $M_{\mathrm{Rad}} = P_{\mathrm{Rad}}/P_0$. The emission process is hence directly affected by the environment. It means that for a given fixed oscillating dipole amplitude, more or less energy will be extracted from this dipole depending on its electromagnetic environment.

These effects will depend on several factors including the substrate geometry and optical properties, the dipole position, orientation, and its emission frequency ω_R. Overall, this results in a modification by a factor $M^d_{\mathrm{Rad}}(\omega_R)$ of the collected Raman signal compared to what would be collected if the Raman dipole radiated in free-space.

The modifications of the radiation intensity and pattern of a dipole are illustrated in Fig. 3.5 for a few simple cases of a dipole in air close to the surface of a sphere (at a distance d = 1 nm). We again compare the results for two materials: a dielectric such as glass (absorbing or not), and a metal like silver. The case of a dipole in the gap of a silver dimer, which is relevant to SERS, is also shown. These simple examples illustrate a number of important features:

- the total power radiated by the dipole can be either quenched or enhanced, depending on the relative dielectric function $\varepsilon(\omega)$ of the objects, their geometry, and the dipole position/orientation;
- when $\mathrm{Re}(\varepsilon(\omega)) < 0$, as is the case of metals in the plasmonics region, large enhancements in the radiated power can be obtained. These are associated with a coupling to radiative localized surface plasmon resonances of the metallic objects. The presence of the metal in this case 'forces' the dipole to radiate more energy;
- under the right conditions (like in the gap of a dimer), huge radiation enhancements of the order of $\sim 10^5$ can be obtained.
- finally, as seen by comparing the emission from dipoles parallel and perpendicular to the surface, these effects depend dramatically on the orientation of the dipole (and also obviously on its position).

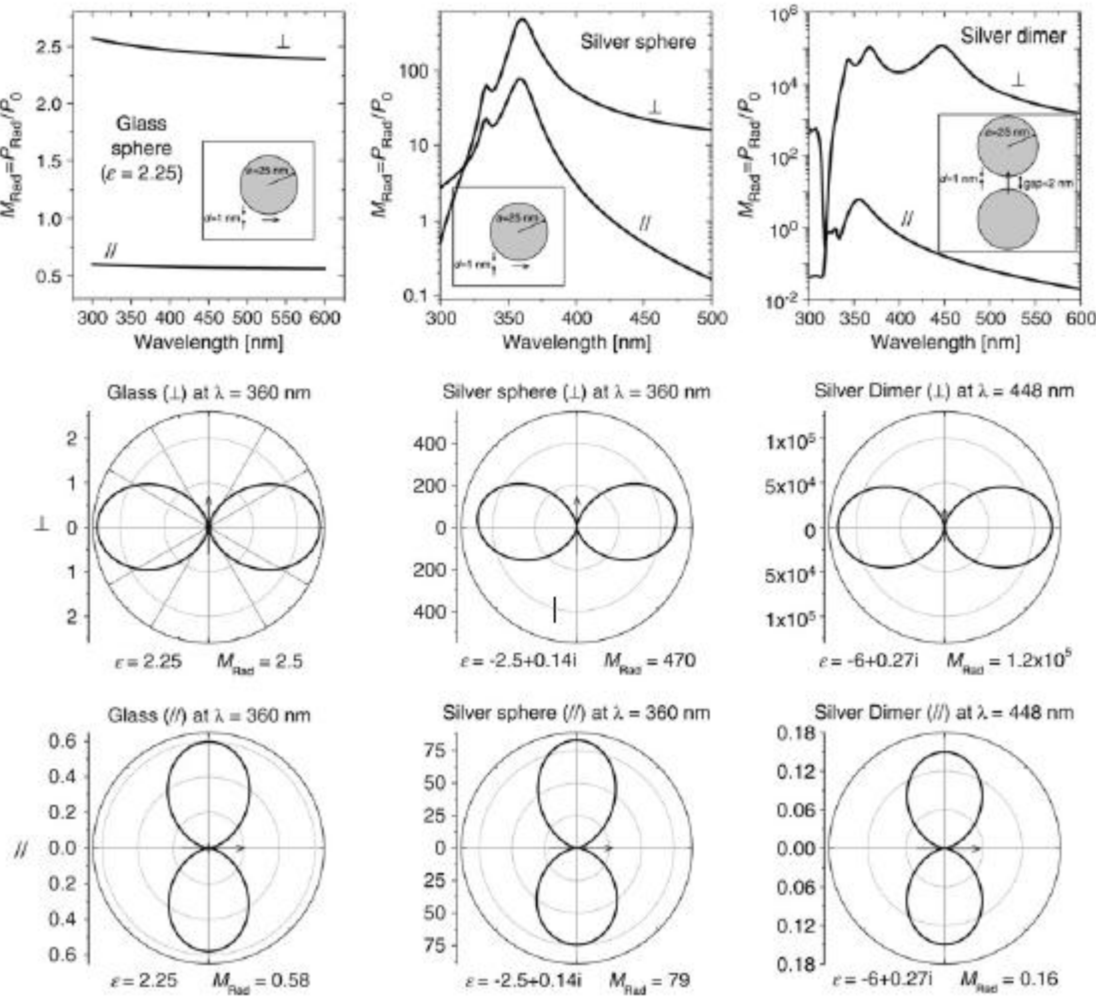


Figure 3.5. Examples of modification of dipolar emission in air at a distance d = 1 nm from a glass sphere (first column), a silver sphere (second column), or in the gap of a silver dimer (third column). The first row shows the wavelength dependence of the radiation enhancement, $M_{\text{Rad}} = P_{\text{Rad}}/P_0$ for a dipole perpendicular (⊥) or parallel (∥) to the surface. This characterizes how much more (or less) energy is radiated in the far-field compared to the same dipole in free-space. For a silver sphere or dimer, M_{Rad} varies widely with wavelength, showing resonances associated with the localized surface plasmon resonances of the system. Second and third rows show examples of radiation patterns at chosen wavelengths. They represent the angular dependence of $\mathrm{d}P_{\text{Rad}}/\mathrm{d}\Omega$. [18]

3.5.4 The common $|E|^4$-approximation to SERS enhancements

SMEF in the $|E|^4$-approximation

When considering only the two main mechanisms of enhancement, namely local field enhancement (excitation) and radiation enhancement (re-emission), the SERS EM enhancement for a single molecule can be simply expressed as:

$$\text{SMEF} \approx M_{\text{Loc}}(\omega_L) M_{\text{Rad}}^{d}(\omega_R) \tag{3.21}$$

Note that what is important here is the radiation enhancement factor in the direction of detection of the SERS signal, which we have denoted M^d_{Rad} and will call the *directional radiation enhancement factor*. This is closely related to M_{Rad}, the total radiation enhancement factor discussed earlier, but also to the radiation pattern. M_{Loc} can be found by solving the electromagnetic problem under specific external excitation conditions with an incident field $\mathbf{E}_{\text{Inc}}$, which yields the local field $\mathbf{E}_{\text{Loc}}$ everywhere. Estimating M_{Rad} is a priori a more difficult task. Because of this, it is often assumed that $M^d_{\text{Rad}}(\omega) \approx M_{\text{Loc}}(\omega)$, which means that the SERS enhancement can then be expressed simply as:

$$\text{SMEF}(\omega_L, \omega_R) \approx M_{\text{Loc}}(\omega_L) M_{\text{Loc}}(\omega_R) \approx \frac{|\mathbf{E}_{\text{Loc}}(\omega_L)|^2}{|E_{\text{Inc}}|^2} \frac{|\mathbf{E}_{\text{Loc}}(\omega_R)|^2}{|E_{\text{Inc}}|^2} \tag{3.22}$$

This expression has been used extensively in the literature, and we will refer to it as the $|E|^4$-*approximation*. It provides a fairly simple way of estimating the single-molecule enhancement factor from a calculation of the local field at the excitation and Raman frequencies. Moreover, in many cases, the Raman shift is also small and one can make the additional approximation that $\omega_R \approx \omega_L$. This leads to the even more famous expression of the SERS enhancement for zero-Stokes shift in the $|E|^4$-approximation as:

$$\text{SMEF}(\omega_L) \approx \frac{|\mathbf{E}_{\text{Loc}}(\omega_L)|^4}{|E_{\text{Inc}}|^4} \tag{3.23}$$

One could argue that this is the most important expression in SERS. In many instances, this approximation is sufficient to obtain the right order of magnitude of the SMEF. Average SERS EFs (like the SSEF) can also be derived within this approximation by surface averaging. However, this factor remains an approximation and it may be necessary for some studies to go back to the original description of local field + radiation enhancement; one example of this situation is the study of polarization effects or surface selection rules in SERS.

To justify the assumption: $M^d_{Rad}(\omega) \approx M_{\text{Loc}}(\omega)$, a few arguments can be used:

- firstly, the local field enhancement and radiation enhancement have the same physical origin: the electromagnetic field (of the incident beam in one case or emitted by the dipole in the other) couples to electromagnetic resonances of the metallic substrate, namely localized surface plasmon resonances. Although the coupling may not be identical, the resonances should at least be qualitatively similar for both enhancement mechanisms. This is already evident in the similarities of some of the plots in Figs 3.4 and 3.5;
- finally, in simple cases where analytical calculations are possible, the total EM enhancement was found to reduce to the $|E|^4$ factor [21,22].

3.6. Modified Spontaneous Emission

We will now go into details of the phenomenon of *modified spontaneous emission* (MSE), since it is directly related to the radiation enhancement contributing to SERS. We will focus in particularly on extending the optical process of spontaneous emission (SE) to the case of an emitter in the presence of (and usually close to) optically active objects (or surfaces).

The emission properties of a dipole, or any emitter, are affected by its environment. This effect can be particularly strong in close proximity of metallic structures. It is for example evident in the modification (quenching or enhancement) of spontaneous emission (SE) rates [23].

Spontaneous emission is an intrinsically quantum effect. However, because of the difficulties in the quantization of the electromagnetic field in complex structures, the modification of SE rates is most often studied within a classical model of an oscillating electromagnetic dipole. This leads (in most cases) to results in agreement with experiments and with the full quantum-mechanical treatment. We will describe an approach based on the optical reciprocity theorem, considering the emission of a dipolar point source (in short, a dipole). It is therefore necessary to first understand the link between dipolar emission and spontaneous emission.

In the following, a free-space emitter will be considered as an emitter embedded in an infinite non-absorbing dielectric of refractive index n_M, and we will study the modifications arising from the presence of nearby objects (the emitter remaining locally embedded in the same dielectric at all times). The subscript M will be used to refer to this free-space situation. The case of an emitter in vacuum can easily be recovered by taking $n_M = 1$, and the subscript 0 will be used when referring specifically to vacuum or air.

3.6.1. The link between spontaneous emission and dipolar emission

A dipolar emitter is one of the simplest (and most common) sources of radiation in classical electromagnetic theory. It is defined by its position $\mathbf{r}_0$, the angular frequency of the oscillation ω, and its dipole moment, a vector denoted $\mathbf{p}$ in complex notations.

What is is important to us is the radiation or far-field properties, in particular:

- the polarized differential power radiated in the far-field per unit solid angle, $\mathrm{d}P_{\mathrm{Rad}-P}/\mathrm{d}\Omega$, where P indicates detection with a polarizer along unit vector $\mathbf{e}_P$ (implicitly perpendicular to the detection direction);
- the differential radiated power (i.e. for unpolarized detection), $\mathrm{d}P_{\mathrm{Rad}}/\mathrm{d}\Omega$; which is simply the sum of $\mathrm{d}P_{\mathrm{Rad}-P1}/\mathrm{d}\Omega$ and $\mathrm{d}P_{\mathrm{Rad}-P2}/\mathrm{d}\Omega$ where $P1$ and $P2$ refer to two mutually orthogonal polarization detections;
- the total radiated power, P_{Rad}.

Let's consider a dipole $\mathbf{p} = p\mathbf{e}_p$ in free-space and denote by Ω the direction of observation (along unit vector $\mathbf{e}_r$). The polarized differential radiated power in this direction is then [24,25]:

$$\frac{\mathrm{d}P_{\mathrm{Rad}-P}^{M}}{\mathrm{d}\Omega}(\Omega) = \frac{n_M \omega^4 |\mathbf{p}|^2}{32\pi^2 \varepsilon_0 c^3} \left| \mathbf{e}_p \cdot \mathbf{e}_P \right|^2 \tag{3.24}$$

Summing this expression over any two orthogonal detection polarizations, $P1$ and $P2$, results in the unpolarized differential radiated power:

$$\frac{\mathrm{d}P_{\mathrm{Rad}}^{M}}{\mathrm{d}\Omega}(\Omega) = \frac{n_M \omega^4 |\mathbf{p}|^2}{32\pi^2 \varepsilon_0 c^3} \left[1 - \left| \mathbf{e}_p \cdot \mathbf{e}_r \right|^2 \right] \tag{3.25}$$

Note that the term in brackets [...] can be expressed simply as $sin^2\,\theta$, with θ the angle between the dipole $\mathbf{e}_p$ and the detection direction $\mathbf{e}_r$.

Integrating over all directions (over the full solid angle Ω), we obtain the total radiated power as:

$$P_{\mathrm{Rad}}^{M} = \frac{n_M \omega^4 |\mathbf{p}|^2}{12\pi \varepsilon_0 c^3} \tag{3.26}$$

Spontaneous emission may look at first very distant from the phenomenon of dipole emission described above. However, it can be shown formally using quantum mechanics that the two processes are equivalent [24,26].

For an atom in vacuum (the only 'easy' case for a quantum-mechanical treatment), the spontaneous emission rate can be calculated as:

$$\Gamma_{\mathrm{Rad}}^{0} = \frac{\omega^3 p_{\mathrm{qu}}^2}{3\pi \varepsilon_0 \hbar c^3} \tag{3.27}$$

where p_{qu} is the dipole moment of the transition between the upper and lower electronic states.

Moreover, the energy is emitted as quanta of energy (photons), each with an energy $\hbar\omega$, equal to the difference in energy between the two electronic states involved in the transition. A given atom in an excited state can only emit one such photon, unless there is a process which re-excites it back to the excited state. If a source of energy is constantly available to re-excite the atom, then the (average) total power radiated for an atom in vacuum is simply obtained from:

$$P_{\mathrm{Rad}}^{0\ \mathrm{qu}} = \Gamma_{\mathrm{Rad}}^{0}\,\hbar\omega = \frac{\omega^4 p_{\mathrm{qu}}^2}{3\pi \varepsilon_0 c^3} \tag{3.28}$$

which differs from the expression obtained within the classical treatment by a factor of 4. Nevertheless, this factor is independent of the environment of the emitter, and therefore disappears when studying modification to the spontaneous emission rates within a given approach. Modified spontaneous emission can therefore be studied within the framework of classical dipole emission, as long as we do not seek exact values for the absolute SE rate.

In summary, modifications to the absolute vacuum spontaneous emission rate $\Gamma_{\mathrm{Rad}}^{0}$ (and to the radiation profile) can then be estimated within the classical EM theory of dipole emission.

We therefore deduce:

$$\Gamma_{\mathrm{Rad}}^{M} = n_M \Gamma_{\mathrm{Rad}}^{0} \tag{3.29}$$

Note that non-radiative decay (without photon emission) may also occur, especially in molecules, with a decay rate Γ_{NR}^{M}. Therefore the total decay rate $\Gamma_{\mathrm{Tot}}^{M}$ for an emitter in free-space is therefore the sum of the radiative and intrinsic non-radiative decay rates: $\Gamma_{Tot}^{M} = \Gamma_{\mathrm{Rad}}^{M} + \Gamma_{NR}^{M}$.

3.6.2. Modification of dipole emission: definitions of enhancement factors

The environment can affect the electromagnetic field around the source and in particular at the source position itself. For an emitting dipole, this field, created by the dipole at its own position through interaction with the environment, is called the *self-reaction field*. In the classical oscillating dipole picture, the self-reaction field can either oppose or add to the dipole oscillations. Because the dipole amplitude is fixed, this results in more or less power needed to maintain the oscillations, and therefore, by conservation of energy, more or less power being extracted from the dipole to the electromagnetic field. This effect plays a major role in the radiation enhancements of SERS.

Absorption and non-radiative emission

In addition, many materials, metals in particular, absorb electromagnetic energy (which is then dissipated in the form of heat). For dipole emission in the presence of such media, a proportion of the total power extracted from the dipole (P_{Tot}) is absorbed and is therefore missing from the modified radiated power (P_{Rad}) in the far-field. This is usually referred to as non-radiative emission, and it cannot be detected in the far-field. Denoting P_{NR} the emitted power lost through non-radiative emission (absorption in the metal), energy conservation leads to:

$$P_{\mathrm{Tot}} = P_{\mathrm{Rad}} + P_{\mathrm{NR}} \tag{3.30}$$

The different types of emitted power, radiative, non-radiative, and total, that can be considered for modified dipolar emission, can be conveniently characterized by a number of enhancement factors, defined with respect to the free-space (in a dielectric of refractive index n_M) radiated power P_{Rad}^{M} given in Eq. (3.26) as:

$$M_{\mathrm{Rad}} = P_{\mathrm{Rad}} / P_{\mathrm{Rad}}^{M} \tag{3.31}$$

$$M_{\mathrm{Tot}} = P_{\mathrm{Tot}} / P_{\mathrm{Rad}}^{M} \tag{3.32}$$

$$M_{\mathrm{NR}} = P_{\mathrm{NR}} / P_{\mathrm{Rad}}^{M} \tag{3.33}$$

The advantage of working with enhancement factors M's instead of powers P's (or rates Γ's) is that exact results can be obtained from the classical EM theory for the M's, while the exact absolute values of the P's (or Γ's) would require the quantum treatment.

Directional radiative enhancement factor

In many cases of interest, radiation is detected in only one direction (over a small solid angle $\Delta\Omega_{\mathrm{Det}}$), and possibly analyzed with a polarizer (along $\mathbf{e}_P$, implicitly perpendicular to the detection direction). It is therefore best characterized by the polarized differential power, $\mathrm{d}P_{\mathrm{Rad}-P}/\mathrm{d}\Omega$, radiated in the detection direction defined by Ω. It is therefore convenient to define also the *polarized directional radiative enhancement factor*, $M^d_{\mathrm{Rad}-P}(\Omega)$, and the corresponding *directional radiative enhancement factor* for unpolarized detection, $M^d_{\mathrm{Rad}}(\Omega)$. Due to the fact that $\mathrm{d}P^M_{\mathrm{Rad}-P}/\mathrm{d}\Omega$ is not constant and can be zero in some directions (along the dipole axis for example), we cannot use it as a denominator. We will therefore use the following definition for polarized detection:

$$\begin{aligned} M^d_{\mathrm{Rad}-P}(\Omega) &= \left(\frac{\mathrm{d}P_{\mathrm{Rad}-P}}{\mathrm{d}\Omega}(\Omega)\right)\Big/\left(\frac{\mathrm{d}P^M_{\mathrm{Rad}}}{\mathrm{d}\Omega}\right)_{\mathrm{Max}} \\ &= \left(\frac{\mathrm{d}P_{\mathrm{Rad}-P}}{\mathrm{d}\Omega}(\Omega)\right)\Big/\left[\frac{3}{8\pi}P^M_{\mathrm{Rad}}\right] \end{aligned} \tag{3.34}$$

and the corresponding expression for unpolarized detection:

$$M^d_{\mathrm{Rad}}(\Omega) = \left(\frac{\mathrm{d}P_{\mathrm{Rad}}}{\mathrm{d}\Omega}(\Omega)\right)\Big/\left[\frac{3}{8\pi}P^M_{\mathrm{Rad}}\right] \tag{3.35}$$

This, however, implies that there is a constant appearing when relating the radiative and directional radiative enhancement factors:

$$M_{\mathrm{Rad}} = \frac{3}{8\pi}\int M^d_{\mathrm{Rad}}(\Omega)\mathrm{d}\Omega \tag{3.36}$$

Finally, as for the unpolarized dfferential radiated power, M^d_{Rad} can be derived from the sum of the polarized directional radiative enhancement factors for two perpendicular detection polarizations $(P1, P2)$:

$$M^d_{\mathrm{Rad}}(\Omega) = M^d_{\mathrm{Rad}-P1}(\Omega) + M^d_{\mathrm{Rad}-P2}(\Omega) \tag{3.37}$$

3.6.3. Spontaneous emission and the optical reciprocity theorem

The *optical reciprocity theorem* (ORT) states that the field $\mathbf{E}$ created at a given point M by a dipole $\mathbf{p}$ (at point O) is related to the field $\mathbf{E}_2$ at O created by a dipole $\mathbf{p}_2$ at M according to:

$$\mathbf{p} \cdot \mathbf{E}_2 = \mathbf{p}_2 \cdot \mathbf{E} \tag{3.38}$$

This theorem, sometimes also referred to as the *Lorentz reciprocity theorem*, allows to derive the far-field properties of an emitter in a given direction from the solution of two plane-wave-excitation PWE) problems (one for each possible detection polarization), *without any source singularities*. This will be of great importance in the formal description of EM-SERS EFs.

We consider a dipole $\mathbf{p} = p\mathbf{e}_p$ at O and focus on its far-field emission at a point M at a distance R in the direction $\mathbf{e}_r$ defined by the angles (θ, ϕ) (see Fig. 3.6).

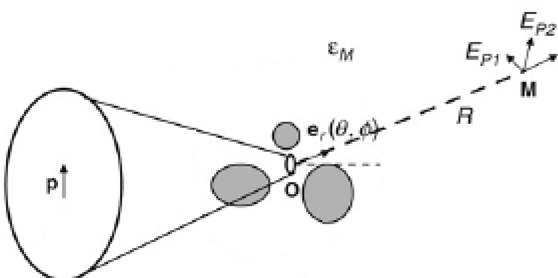


Fig. 3.6. Schematic representation of the problem of an emitter (dipole) in close proximity to a metallic environment.

The radiation field is transverse in the far-field (no radial component) and its electric field at point M, denoted $\mathbf{E}$, can be decomposed into two perpendicular polarizations along unit vectors $\mathbf{e}_{P1}$ and $\mathbf{e}_{P2}$: $\mathbf{E} = E_{P1}\mathbf{e}_{P1} + E_{P2}\mathbf{e}_{P2}$.

To apply the ORT, we then consider the (separate) problem of a dipole $\mathbf{p}_2 = p_2\mathbf{e}_2$ situated at point M and with $\mathbf{e}_2 \perp \mathbf{e}_r$. The ORT then yields:

$$p_2\mathbf{e}_2 \cdot \mathbf{E} = p\mathbf{e}_p \cdot \mathbf{E}_2(\mathrm{O}) \tag{3.39}$$

where $\mathbf{E}_2(\mathrm{O})$ is the field created by $\mathbf{p}_2$ at O. To determine $\mathbf{E}_2(\mathrm{O})$, let us first consider the field $\mathbf{E}_2(\mathbf{r})$ created by $\mathbf{p}_2$ at a point $\mathbf{r}$ in the absence of any objects. For sufficiently large R (and we can take R as large as we like), the field of this dipole in the region of interest (i.e. the region where the objects are, with $|\mathbf{r}| \ll R$) can be approximated by expanding the free-space dipole field:

$$\mathbf{E}_2(\mathbf{r}) \approx E_p e_2 e^{-ik_M \mathbf{e}_r \cdot \mathbf{r}} \tag{3.40}$$

with

$$E_p = \frac{k_M^2 p_2 e^{ik_M R}}{4\pi\varepsilon_0\varepsilon_M R} \tag{3.41}$$

This is the field of a plane wave with wave-vector k_M, propagating along $-\mathbf{e}_r$, polarized along $\mathbf{e}_2$, and with complex amplitude E_p. The problem of finding $\mathbf{E}_2(0)$, created by the dipole $\mathbf{p}_2$ at O in the presence of the objects is therefore equivalent to finding the local field at O, $\mathbf{E}_{\mathrm{Loc}}(0)$ for a plane-wave excitation with the characteristics given above. Moreover, using the ORT expression in Eq. (3.39) we get:

$$\mathbf{e}_2 \cdot \mathbf{E} = \frac{E_p}{p_2} p \frac{\mathbf{e}_p \cdot \mathbf{E}_{\mathrm{Loc}}(O)}{E_p} \tag{3.42}$$

By choosing $\mathbf{e}_2 = \mathbf{e}_{P1}$ (or $\mathbf{e}_2 = \mathbf{e}_{P2}$), we can therefore, using Eq. (3.42), determine the polarization component E_{P1} (or E_{P2}), of the radiation field of the dipole $\mathbf{p}$ in the direction defined by $\mathbf{e}_r$, by computing the local field $\mathbf{E}_{\mathrm{Loc}}^{\mathrm{PW}-P1}$ (or $\mathbf{E}_{\mathrm{Loc}}^{\mathrm{PW}-P2}$) created at O by a plane wave of amplitude E_p incident along $-\mathbf{e}_r$ and polarized along $\mathbf{e}_{P1}$ (or $\mathbf{e}_{P2}$).

The EM problem of dipolar radiation in a given direction can therefore be replaced by two EM problems of plane-wave excitation (PWE) *from this same direction.*

Note that for the PWE problems, $\tilde{\mathbf{E}}_{\mathrm{Loc}}^{\mathrm{PW}-P1} = \mathbf{E}_{\mathrm{Loc}}^{\mathrm{PW}-P1}/E_p$ is the relevant physical quantity (and is independent of E_p). It is a unit-less vector characterizing the direction of the local electric field, and its magnitude enhancement (or quenching) due to the presence of the objects. Using Eq. (3.42) and the expression for E_p in Eq. (3.41), we therefore have:

$$E_{P1} = \frac{k_M^2 e^{ik_M R}}{4\pi\varepsilon_0\varepsilon_M R} \mathbf{p} \cdot \tilde{\mathbf{E}}_{\mathrm{Loc}}^{\mathrm{PW}-\mathrm{P1}} \tag{3.43}$$

A similar expression is obtained for E_{P2}, but with the solution of a different PWE problem (with a different polarization). We can hence deduce the complex Poynting vector for the radiation field at M: $\mathbf{S} = (n_M\varepsilon_0 c/2)(|E_{P1}|^2 + |E_{P2}|^2)\mathbf{e}_r$, where the contributions for each detection polarization are clearly distinguished.

The time-averaged differential power radiated per unit solid angle in the direction $\mathbf{e}_r$ is then:

$$\frac{\mathrm{d}P_{\mathrm{Rad}}}{\mathrm{d}\Omega}(\Omega) = R^2 \mathrm{Re}(\mathbf{S}\cdot\mathbf{e}_r) = \frac{n_M\omega^4|\mathbf{p}|^2}{32\pi^2\varepsilon_0 c^3} \left|\mathbf{e}_p \cdot \tilde{\mathbf{E}}_{\mathrm{Loc}}^{\mathrm{PW}-P1}\right|^2 + \left|\mathbf{e}_p \cdot \tilde{\mathbf{E}}_{\mathrm{Loc}}^{\mathrm{PW}-P2}\right|^2 \tag{3.44}$$

Similarly, the polarized (along $\mathbf{e}_P$) differential radiated power is:

$$\frac{\mathrm{d}P_{\mathrm{Rad}-\mathrm{P}}}{\mathrm{d}\Omega}(\Omega) = \frac{n_M\omega^4|\mathbf{p}|^2}{32\pi^2\varepsilon_0 c^3} \left|\mathbf{e}_p \cdot \tilde{\mathbf{E}}_{\mathrm{Loc}}^{\mathrm{PW}-P}\right|^2 \tag{3.45}$$

It is easy to verify (by taking $\tilde{\mathbf{E}}_{\mathrm{Loc}}^{\mathrm{PW}-P} = \mathbf{e}_P$) that the above expression is fully consistent with that of an isolated dipole in free-space, yielding the same expression as the standard EM approach [Eq. (3.24)].

Radiation enhancement factors within the ORT approach

From its definition in Eq. (3.34), the *polarized directional radiative enhancement factor* is then given simply within the ORT approach by:

$$M^{d}_{\mathrm{Rad}\,-\mathrm{P}}(\Omega) = \left|\mathbf{e}_p \cdot \tilde{\mathbf{E}}^{\mathrm{PW}\,-P}_{\mathrm{Loc}}\right|^2 \tag{3.46}$$

where $\tilde{\mathbf{E}}^{\mathrm{PW}\,-P}_{\mathrm{Loc}}$ can be obtained from the solution of a PWE problem.

M^{d}_{Rad} follows naturally from Eq. (3.37), but requires the solution of two different *independent* PWE problems (for $P1$ and $P2$).

The important conclusion is that *the radiative property in a given direction of a dipole at any position in a complex environment can be obtained by modeling two (one for each polarization detection) PWE problems, without a dipolar singularity*. This is a valuable result when modeling a typical experimental situation where only radiative properties are of interest, and detection is carried out in a given fixed direction, a situation particularly common for SERS. The solution of these two PWE problems is sufficient to calculate the radiation properties in this particular direction for any dipoles at any possible positions and with any orientations.

Equation (3.46) also shows that the far-field emission of a dipole in a given direction is in some way related to the local field intensity enhancement factor: $M^{\mathrm{PW}\,-P}_{\mathrm{Loc}} = \left|\tilde{\mathbf{E}}^{\mathrm{PW}\,-P}_{\mathrm{Loc}}\right|^2$ for PWE from this direction. This will be the basis for the generalization of the $|E|^4$-approximation of SERS enhancements in the next section.

3.7. Formal derivation of SERS EM enhancements

We will now describe how the electromagnetic contribution to the SERS enhancement factors (EFs) can be calculated from standard EM theory. We will mainly focus on the SERS cross-section and single-molecule enhancement factor (SMEF) as defined in Eq. 3.7 and Eq. 3.8. Note that these definitions refer to *one* given *vibrational mode* of the molecule under study. The Raman properties of this mode are characterized by its vibrational energy $\hbar\omega_v$ and its Raman polarizability tensor $\hat{\alpha}$, which depends on the excitation frequency ω_L. The Raman photons are emitted at the Raman frequency ω_R ($\omega_R = \omega_L - \omega_v$ for Stokes processes). A major assumption we will make is that the Raman polarizability tensor is the same under normal and SERS conditions. Moreover, since the SERS EFs were defined with respect to the non-SERS properties of the same molecule embedded in the same environment, any local field correction factor cancels out.

Let us consider a fixed molecule at a given position O, embedded in a dielectric (of refractive index n_M), and in the vicinity of a number of interfaces (dielectric and/or metallic objects). The situation is schematically depicted in Fig. 3.7.

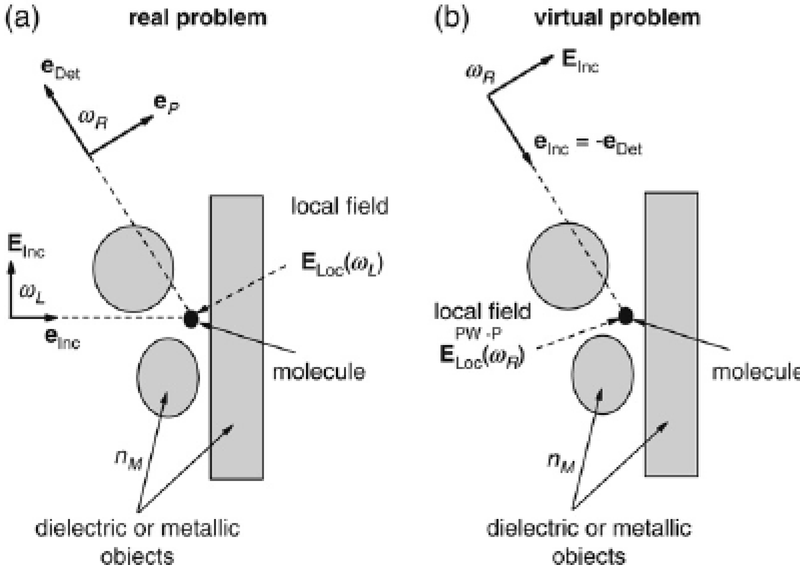


Fig. 3.7. Schematic representation of the SERS EM problem (for polarized detection), which is solved by considering both the real excitation problem (a) and the "virtual" planewave-excitation problem for re-emission (b). Using the optical reciprocity theorem, the "virtual" problem (b) allows the calculation of the SERS EF without solving the EM problem of emission from a localized dipole at the molecule position.

The system is excited by an incident laser at frequency ω_L, incoming along a direction $\mathbf{e}_{\text{Inc}}$ and polarized along $\mathbf{e}_E \perp \mathbf{e}_{\text{Inc}}$. This will be modeled as a plane wave with complex electric field $\mathbf{E}_{\text{Inc}} = E_{\text{Inc}} e^{ik_M \mathbf{e}_{\text{Inc}} \cdot \mathbf{r}} \mathbf{e}_E$. The incident power density S_{Inc} is related to the field amplitude E_{Inc} by $S_{\text{Inc}} = n_M \varepsilon_0 c |E_{\text{Inc}}|^2/2$. This excitation interacts with the objects and creates at the molecule position a local electric field $\mathbf{E}_{\text{Loc}}(\omega_L)$, which can be normalized against the incident amplitude to give: $\tilde{\mathbf{E}}_{\text{Loc}}(\omega_L)$=$\mathbf{E}_{\text{Loc}}(\omega_L)/E_{\text{Inc}}$.

We will study the SERS (Raman) signal emitted by the molecule in a given direction Ω_{Det}, along the unit vector $\mathbf{e}_{\text{Det}}$. This signal can also be analyzed along two perpendicular polarizations defined by $\mathbf{e}_{P1}$ and $\mathbf{e}_{P2}$, themselves perpendicular to $\mathbf{e}_{\text{Det}}$. The corresponding differential radiated SERS powers will be denoted by $\text{d}P^P_{\text{SERS}}/\text{d}\Omega$, with $P = P1, P2$. To apply the ORT, we will also need to consider the "virtual" problem of a plane-wave excitation with frequency ω_R incoming along $-\mathbf{e}_{\text{Det}}$ and polarized along $\mathbf{e}_P$. We will denote $\tilde{\mathbf{E}}^{\text{PW}-P}_{\text{Loc}}(\omega_R)$ the local field solution to this problem at the molecule position (normalized against the incident plane-wave amplitude as before). We remind that this "virtual" problem has no relation whatsoever with the real excitation (which is along $\mathbf{e}_{\text{Inc}}$), but is just used as a mathematical trick to study the re-emission problem.

SMEF derivation

The local field at the molecule position induces a Raman dipole:

$$\mathbf{p} = \hat{\alpha} \cdot \mathbf{E}_{\text{Loc}}(\omega_L) \tag{3.47}$$

oscillating at the Raman frequency $\boldsymbol{\omega_R}$.

The differential radiative power emitted at frequency $\boldsymbol{\omega_R}$ in direction $\boldsymbol{\Omega_{\mathrm{Det}}}$ is modified by the environment, according to the directional radiative enhancement factor $\boldsymbol{M^{d}_{\mathrm{Rad}-P}}(\boldsymbol{\Omega_{\mathrm{Det}}}, \boldsymbol{\omega_R}, \mathbf{e}_p)$. From the definition of $\boldsymbol{M^{d}_{\mathrm{Rad}-P}}$ in Eq. (3.34), we can write:

$$\frac{dP^{P}_{\mathrm{SERS}}}{d\Omega} = \frac{n_M \omega_R^4}{32\pi^2 \varepsilon_0 c^3} |\mathbf{p}|^2 M^{d}_{\mathrm{Rad}\,-P}(\Omega_{\mathrm{Det}}, \omega_R, \mathbf{e}_p) \tag{3.48}$$

From Eq. (3.7) and Eq. (3.47), the differential SERS cross-section for P-polarized detection is then obtained as:

$$\frac{d\sigma^{P}_{\mathrm{SERS}}}{d\Omega} = \frac{\omega_R^4}{16\pi^2 \varepsilon_0^2 c^4} \left|\hat{\alpha} \cdot \tilde{\mathbf{E}}_{\mathrm{Loc}}(\omega_L)\right|^2 M^{d}_{\mathrm{Rad}\,-P}(\Omega_{\mathrm{Det}}, \omega_R, \mathbf{e}_p) \tag{3.49}$$

By using Eq. (3.46), derived from the ORT, to determine $M^{d}_{\mathrm{Rad}\,-P}$, we deduce:

$$\begin{aligned} |\mathbf{p}|^2 M^{d}_{\mathrm{Rad}\,-P}(\Omega_{\mathrm{Det}}, \omega_R, \mathbf{e}_p) &= \left|\tilde{\mathbf{E}}^{\mathrm{PW}\,-P}_{\mathrm{Loc}}(\omega_{\mathrm{R}}) \cdot \mathbf{p}\right|^2 \\ &= \left|\tilde{\mathbf{E}}^{\mathrm{PW}\,-P}_{\mathrm{Loc}}(\omega_{\mathrm{R}}) \cdot \hat{\alpha} \cdot \mathbf{E}_{\mathrm{Loc}}(\omega_L)\right|^2 \end{aligned} \tag{3.50}$$

We then obtain:

$$\frac{\mathrm{d}\sigma^{P}_{\mathrm{SERS}}}{\mathrm{d}\Omega} = \frac{\omega_R^4}{16\pi^2 \varepsilon_0^2 c^4} \left|\tilde{\mathbf{E}}^{\mathrm{PW}\,-P}_{\mathrm{Loc}}(\omega_{\mathrm{R}}) \cdot \hat{\alpha} \cdot \tilde{\mathbf{E}}_{\mathrm{Loc}}(\omega_L)\right|^2 \tag{3.51}$$

The single-*molecule enhancement factor for P-polarized detection* then follows simply as the ratio of Eq. (3.51) to the non-SERS differential cross-section $\mathrm{d}\sigma_R/\mathrm{d}\Omega = \omega_R^4 \tilde{\alpha}^2/16\pi^2 \varepsilon_0^2 c^4$, giving:

$$\mathrm{SMEF}^{P} = \left|\tilde{\mathbf{E}}^{\mathrm{PW}\,-P}_{\mathrm{Loc}}(\omega_{\mathrm{R}}) \cdot \hat{\alpha}_N \cdot \tilde{\mathbf{E}}_{\mathrm{Loc}}(\omega_L)\right|^2 \tag{3.52}$$

where $\hat{\alpha}_N = \hat{\alpha}/\tilde{\alpha}$ is the *normalized Raman polarizability tensor*. Note that this SMEF only depends on the symmetry of the Raman tensor, not on its magnitude, as expected for an enhancement factor. Similar expressions can be obtained for the differential SERS cross-section and SMEF in the case of unpolarized detection, by summing the results for the two perpendicular polarization detections, $P1$ and $P2$.

From this expression of the SMEF, it is clear that in the most general case, the local field enhancement (for excitation at ω_L) and the radiative enhancement (for re-emission at ω_R) cannot be entirely decoupled. This is because the radiative enhancement depends on the induced dipole orientation, itself directly related to the local field polarization. Nevertheless, except for these orientation/polarization effects, the magnitude of the enhancement remains primarily dominated by the magnitude of both the local field and radiative enhancements. To highlight this, we can write the normalized local field as:

$$\tilde{\mathbf{E}}_{\text{Loc}}(\omega) = \tilde{\text{E}}_{\text{Loc}}(\omega)\mathbf{e}_{\text{Loc}}(\omega) \quad (3.53)$$

where $\mathbf{e}_{\text{Loc}}(\omega)$ is, by definition, a unit vector (possibly complex), characterizing the *local field polarization*. The local field intensity enhancement factor then takes the form:

$$M_{\text{Loc}}(\omega) = \left|\tilde{\mathbf{E}}_{\text{Loc}}(\omega)\right|^2 = \left|\tilde{\text{E}}_{\text{Loc}}(\omega)\right|^2 \quad (3.54)$$

A similar expression can be written for the "virtual" PWE problems:

$$\tilde{\mathbf{E}}_{\text{Loc}}^{\text{PW}-P}(\omega) = \tilde{\text{E}}_{\text{Loc}}^{\text{PW}-P}(\omega)\mathbf{e}_{\text{Loc}}^{\text{PW}-P}(\omega) \quad (3.55)$$

and for the corresponding $M_{\text{Loc}}^{\text{PW}-P}(\omega)$.

Inserting these expressions into Eq. (3.52), the SMEF for polarized detection can then be rewritten as:

$$\text{SMEF}^P = M_{\text{Loc}}(\omega_L)M_{\text{Loc}}^{\text{PW}-P}(\omega_R)T\left(\hat{\alpha}_N, \mathbf{e}_{\text{Loc}}(\omega_L), \mathbf{e}_{\text{Loc}}^{\text{PW}-P}(\omega_R)\right) \quad (3.56)$$

where

$$T(\hat{\alpha}_N, \mathbf{e}_L, \mathbf{e}_R) = |\mathbf{e}_R \cdot \hat{\alpha}_N \cdot \mathbf{e}_L|^2 \quad (3.57)$$

In Eq. (3.56):

- $M_{\text{Loc}}(\omega_L)$ represents the *local field intensity enhancement factor* (LFIEF) for excitation of the Raman dipole. It characterizes the enhancement in the intensity of the *"real" local electric field* felt by the molecule;
- $M_{\text{Loc}}^{\text{PW}-P}(\omega_R)$ characterizes the magnitude of the radiation enhancement for the Raman dipole. Thanks to the ORT, it can be written here as the LFIEF that would be obtained for a "virtual" problem with plane-wave excitation incoming from the direction of detection and polarized along the polarization of detection;
- $T\left(\hat{\alpha}_N, \mathbf{e}_{\text{Loc}}(\omega_L), \mathbf{e}_{\text{Loc}}^{\text{PW}-P}(\omega_R)\right)$ is an additional factor, which couples the excitation and re-emission problems through the normalized Raman polarizability tensor (characterizing the symmetry of the Raman tensor and the molecular orientation). In a first approximation, T has only a secondary influence on the overall magnitude of the SMEF, except for some specific combinations of Raman tensor symmetry and molecule orientation for which it would cancel out to zero. It can nevertheless account for changes in the relative intensities of Raman peaks of different symmetries under SERS conditions. It is called *surface selection rules factor (SSRF)*.

The surface selection rules factor (SSRF) is admittedly a fairly complicated factor in general, involving the polarization of the local field for both the "real" and "virtual" problems. In fact, in most SERS studies, experimental or theoretical, it is largely ignored. For the case of a fixed molecule, when considering all possible Raman tensors, one can show that: $0 \leq T(\hat{\alpha}_N, \mathbf{e}_L, \mathbf{e}_R) \leq 15/4$; in the case of full orientation averaging, we have: $0 \leq \langle T(\hat{\alpha}_N, \mathbf{e}_L, \mathbf{e}_R) \rangle \leq 1$.

SSEF derivation

The orientation-averaged SMEF (OASMEF) can be obtained from Eq. (3.56) by averaging over the allowed molecule orientation. Note that this averaging should only affect the surface selection rules factor, which must then be replaced by $[T]$. This can be case specific, since some molecules may have a fixed orientation, while others may have one (for planar adsorption for example), or even two or three (for fully random orientation) degrees of rotational freedom.

Moreover, the average SERS substrate defined as the SERS substrate EF (SSEF) can in principle be obtained by spatial averaging of the OASMEF over the entire metallic surface as in Eq. (3.11). Note that all three factors in Eq. (3.56), $M_{\text{Loc}}(\omega_L)$, $M_{\text{Loc}}^{\text{PW}-P}(\omega_R)$, and $T\left(\hat{\alpha}_N, \mathbf{e}_{\text{Loc}}(\omega_L), \mathbf{e}_{\text{Loc}}^{\text{PW}-P}(\omega_R)\right)$, depend on position r. They cannot therefore be decoupled in the spatial averaging, and the SSEF for polarized detection is therefore:

$$\text{SSEF}^P = \Big\{ M_{\text{Loc}}(\mathbf{r}, \omega_L) M_{\text{Loc}}^{\text{PW}-P}(\mathbf{r}, \omega_R) \times \left[T\left(\hat{\alpha}_N, \mathbf{e}_{\text{Loc}}(\mathbf{r}, \omega_L), \mathbf{e}_{\text{Loc}}^{\text{PW}-P}(\mathbf{r}, \omega_R)\right)\right]\Big\} \tag{3.58}$$

where $\{\ldots\}$ denotes spatial averaging, and $[\ldots]$ represents averaging over allowed molecular orientations. The SSEF for unpolarized detection can be obtained simply as before by summing the SSEF^P for two perpendicular polarization detections $P1$ and $P2$.

Notice that all the expressions so far relate to the differential SERS power, i.e. for detection in a single direction.

3.7.1. SERS EM enhancements in the backscattering configuration

In order to link the above derived expressions with the commonly-used $|E|^4$-approximation, we will now restrict our ourselves to a specific configuration, the backscattering configuration. This is done for three main reasons:

- it is one of the most widely used for SERS experiments (in particular in Raman microscopes, where the delivery and collection optics are the same);
- it makes the theoretical description much simpler since the "virtual" (re-emission) problem corresponds to plane-wave excitation along the same direction as the real excitation;
- it allows a formal justification of the $|E|^4$-approximation.

SMEF in the backscattering configuration

In the backscattering (BS) configuration, excitation and detection are along the same direction. This corresponds in the real problem of Fig. 3.7(a) to $\mathbf{e}_{\text{Inc}} = -\mathbf{e}_{\text{Det}}$. We choose them along the x-axis in the following. The incident wave is taken as a plane wave polarized along z ($\perp$ to x). To fully characterize the emitted SERS signal in the backscattering direction, we have to consider two polarizations for detection (which is also along x): we choose them either parallel to the incident polarization (along z) or perpendicular to it (along y). This re-emission problem then requires the solution of two "virtual" PWE problems, with incoming beam along x and polarization along z and y. The first one is in fact the same as the real excitation problem, while only the polarization changes for the second. Let us therefore denote $M^{P}_{\text{Loc}}(\omega)$ and $\mathbf{e}^{P}_{\text{Loc}}(\omega)$ the LFIEF and local field polarization for PWE incoming along x and polarized along $P = Z$ or $P = Y$. $P = Z$ applies to the real excitation and "virtual" problem for parallel detection, while P = Y is for the "virtual" problem for perpendicular detection. The SMEF in the BS configuration for parallel and perpendicular detections then take the form:

$$\text{SMEF}^{\parallel}_{\text{BS}} = M^{Z}_{\text{Loc}}(\omega_L) M^{Z}_{\text{Loc}}(\omega_R) T^{\parallel}_{BS} \tag{3.59}$$

with $T^{\parallel}_{BS} = |\mathbf{e}^{Z}_{\text{Loc}}(\omega_R) \cdot \hat{\alpha}_N \cdot \mathbf{e}^{Z}_{\text{Loc}}(\omega_L)|^2$, and

$$\text{SMEF}^{\perp}_{\text{BS}} = M^{Z}_{\text{Loc}}(\omega_L) M^{Y}_{\text{Loc}}(\omega_R) T^{\perp}_{BS} \tag{3.60}$$

with $T^{\perp}_{BS} = |\mathbf{e}^{Y}_{\text{Loc}}(\omega_R) \cdot \hat{\alpha}_N \cdot \mathbf{e}^{Z}_{\text{Loc}}(\omega_L)|^2$.

Note also that for unpolarized detection, the SMEF is simply the sum of the two:

$$\text{SMEF}_{\text{BS}} = \text{SMEF}^{\parallel}_{\text{BS}} + \text{SMEF}^{\perp}_{\text{BS}} \tag{3.61}$$

Comparison with the $|E|^4$-approximation

The $|E|^4$-approximation to the SERS EF appearing in Eq. (3.22) would be written here as:

$$\text{SMEF}_{\text{BS}} = M^{Z}_{\text{Loc}}(\omega_L) M^{Z}_{\text{Loc}}(\omega_R) \tag{3.62}$$

The exact expressions given above have some similarities with Eq. (3.22), but differ from it in at least two aspects:

- firstly, the factor $M^{Z}_{\text{Loc}}(\omega_R)$ appears only for parallel polarized detection. For perpendicular polarized detection, it should be replaced by $M^{Y}_{\text{Loc}}(\omega_R)$, which may be very different;
- secondly, the surface selection rules factor, T, does not appear in the $|E|^4$-approximation.

For the $|E|^4$-approximation to be exact, we therefore require that $\text{SMEF}_{\text{BS}}^{\perp} = 0$ (or $\ll \text{SMEF}_{\text{BS}}^{\parallel}$) and $T_{BS}^{\parallel} = 1$, a situation which is rarely achieved in practice. Fortunately, the discrepancies between the approximation and the exact results are most of the time secondary, at least for the backscattering configuration. This happens for the following reasons:

- In general, the SMEF for parallel detection $\text{SMEF}_{\text{BS}}^{\parallel}$ is much larger or at least of the same order as $\text{SMEF}_{\text{BS}}^{\perp}$. It is therefore a good estimate of the SMEF, even for unpolarized detection (a very common configuration). The main reason for this is that, in general, $M_{\text{Loc}}^{Z}(\omega)$ is larger (sometimes much larger) than $M_{\text{Loc}}^{Y}(\omega)$;
- in many instances, the SSRF T k BS (or its orientation-averaged equivalent) is of the order of 1.

In summary, the $|E|^4$-approximation provides in most situations a good (and "simple") estimate of the magnitude of the SMEF in the backscattering configuration.

ELECTRON BEAM LITHOGRAPHY

Introduction

Electron Beam Lithography (EBL) is a technology for the transfer of computer designed patterns from a digitally stored format (a computer data file) to an imaging layer (electron beam resist) applied to the substrate. Its main characteristics are high resolution, due to the short electron wavelength, and flexibility, due to the easily modifiable pattern data file. The patterned electron beam resist can be subsequently used as a mask for the permanent transfer of the pattern image from the resist to the substrate. For high-resolution patterning, electron beams typically have effective wavelengths significantly smaller than the pattern minimum feature size and are therefore not diffraction limited. This contrasts with optical lithography, in which the minimum resolution is comparable to the exposing light wavelength, resulting in resolution-limiting diffraction effects.

The most important limitation in electron beam lithography is writing speed (or throughput): the serial scanning manner of EBL cannot compete with the parallel projection of photons in optical lithography.

The largest application for electron beam lithography is in the fabrication of masks for projection optical lithography (or other lithographic techniques, such as X-ray lithography). For this application, the EBL process generates the master mask in a serial manner, while projection optical lithography equipment (steppers) subsequently replicate it directly onto resist-coated substrates in a parallel manner. Moreover, electron beam lithography is used in the low-volume advanced manufacture of ultra-small features for very high performance devices.

Electron beam lithography appeared in early 1960s evolving from scanning electron microscope (SEM). The SEM could evolve to e-beam lithography only because of the discovery of an electron-sensitive polymer material, called polymethylmethacrylate (PMMA). Exposure of PMMA resist by e-beam is very much like exposure of photoresists by photons. Because of the fine beam enabled by electron optics, e-beam lithography using PMMA resist immediately demonstrated much higher resolution capability than optical lithography could dream of around the time. Nowadays, the state-of the-art e-beam lithography system, combined with special e-beam resist materials and processes, can fabricate structures less than 10 nm.

4.1 Electron-solid interactions

The physics of the electron-beam exposure process is dictated by the nature of electron-solid interactions, both in the resist and in the underlying substrate materials. The main parameters influencing these interactions are those of the beam (primarily the beam energy) and those of the materials (the resist and substrate types and the resist thickness). The electrons in the beam interact with the resist-coated substrate through elastic and inelastic scattering by the resist and substrate atoms.

This results in the deposition of energy in the resist and substrate, and in the spreading of the electrons from the point of incidence. The inelastic scattering events are the means by which the electrons deposit energy in the resist and substrate and determine the absorbed energy distribution. It is this absorbed energy distribution in the resist that determines the developed resist pattern.

The electrons generated when an electron beam with an energy of several kilo-electronvolts bombards a substrate can be divided into two categories:

1. The *secondary electrons* are those ejected from the substrate material itself. They are of low energy, and their energy distribution has a peak at an energy of a few electron volts. Due to their low energy, they are readily absorbed by the resist, contributing significantly to the exposure. Fortunately, the low energy also limits their effective range in the resist to a few nanometers, thus limiting the secondary electron contribution to beam broadening to approximately 10 nm [27]. By convention, it is assumed that electrons with energies below 50 eV are secondaries.
2. The *backscattered electrons* are primaries that have been reflected from the substrate. For a substrate of silicon (atomic number Z=14), their mean energy is approximately 60% of that of the primary beam [28].

4.1.1 Electron scattering

The scattering events are classified as *forward scattering* and *backscattering*.

Forward scattering is characterized by primarily small-angle scattering, less than 90° from the primary beam direction. The main impact of forward scattering on the lithography process is broadening of the incident beam as it passes from the resist surface to the substrate. Since the scattering angles are characteristically small, the forward-scattered electrons that reach the substrate do not have further statistically significant interactions with the resist. The beam broadening effects introduced by forward scattering are reduced as resist thickness decreases or accelerating voltage increases.

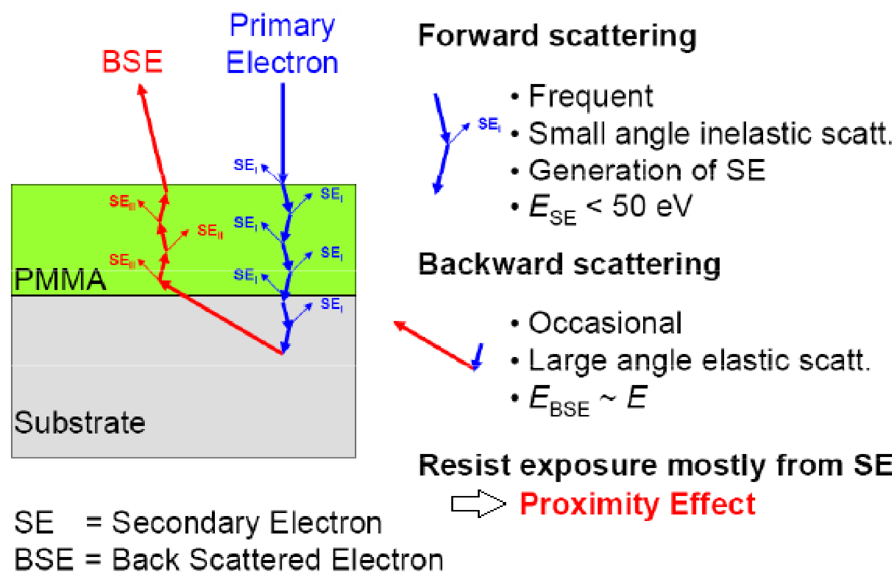


Fig. 4.1. Electron scattering events.

Backscattering is characterized by large-angle scattering, to nearly 180° from the primary beam direction. Unlike forward scattering, these large scattering angles make it possible for electrons reaching the substrate to return to the resist. Backscattered electrons may also originate from the resist layer without ever reaching the substrate. This contribution is usually much smaller than that from the substrate, since the magnitude of this scattering is strongly dependent on the atomic number of the material, and typically the ratio of resist to substrate atomic numbers is small. The practical implications are that subsequent scattering events from substrate backscattered electrons can deposit energy in the resist away from the primary beam (i.e., away from the pattern), leading to pattern distortions (the so-called proximity effect). The proximity effect is the major obstacle in the process of achieving high-resolution nanofabrication by charged beams.

The electron collisions can be modeled by Monte Carlo simulations [29,30]. Monte Carlo method makes use of mathematic descriptions of elastic and inelastic scattering cross-sections and electron energy loss in a solid material to calculate how a primary electron from an e-beam is scattered inside a solid material and how it loses its energy. The electron scattering is a random process: Fig. 4.2 shows two electron trajectories (primary electron A and B), when they travel inside a resist layer on top of substrate material. The resist is PMMA of 100 nm thickness and the substrate is silicon. Each dot along a trajectory represents a scattering event. One can see that the position, direction, and free travel distance of each scattering event are totally random.

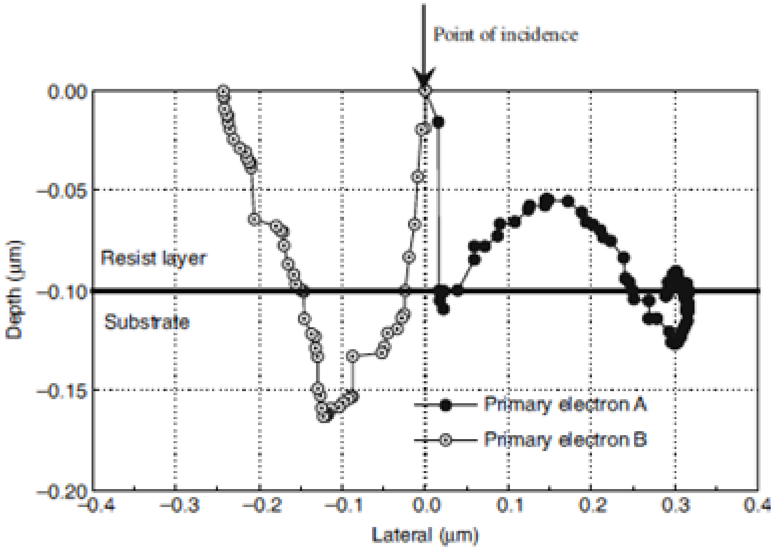


Fig. 4.2. Two electron trajectories showing random scattering events

Although each individual scattering is random, simulation of a large number of electron scatterings will generate the general picture of the extent of energy deposition due to the collision and loss of energy from the primary electrons.

Fig. 4.3 plots out 100 electron trajectories. The electron energy is 20 keV with 1-µm resist layer on top of a silicon substrate. This plot gives a rough idea how deep and how wide the electrons travel inside resist and substrate. Most of electrons scatter forward into substrate but some electrons scatter back into resist layer. These forward scatterings and backward scatterings are the cause of broadening of exposure area in resist. To generate statistically correct and reliable results, the number of electrons to be simulated has to be between 50,000 to 100,000.

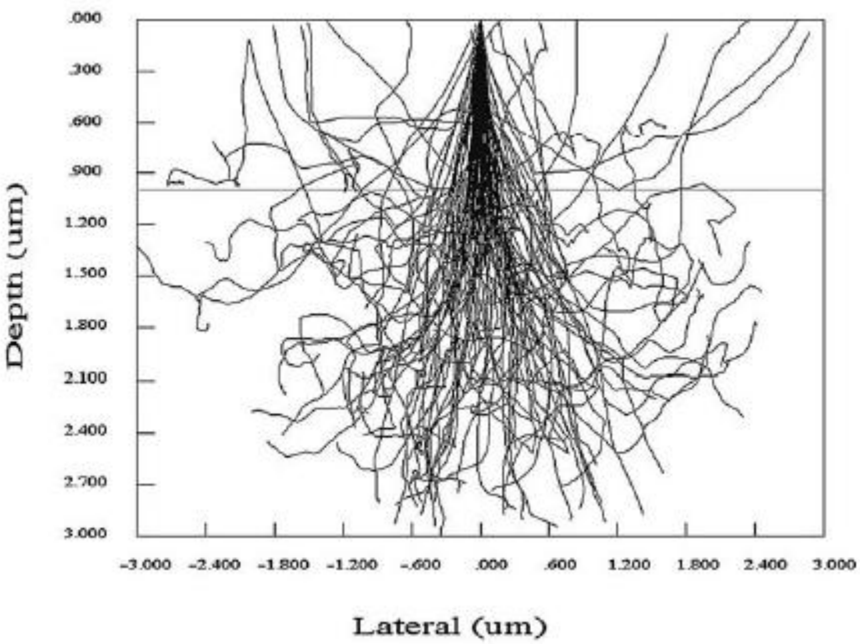


Fig. 4.3. Hundred electron trajectories showing the extent of electron scattering.

Instead of counting scattering electrons, forward and backward scattering can also be expressed by a *point spread function* (PSF) which represents deposited energy distribution in resist from single point of incidence. The typical PSF is shown in Fig. 4.4, where forward scattering part and backscattering part of energy spread are indicated. The figure also compares the PSFs at 10% and 90% depth of resist. It clearly displays a significant increase of broadening in deposited energy by forward scattering, as well as accumulated energy by backscattering, as electrons going deeper.

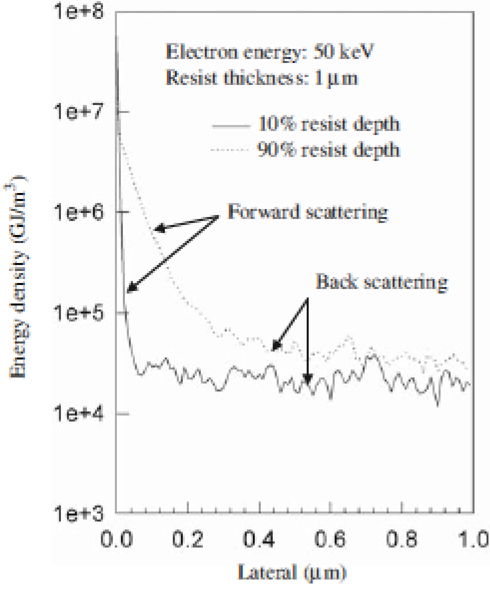


Fig. 4.4. Point spread function of electron energy distribution in resist. [31]

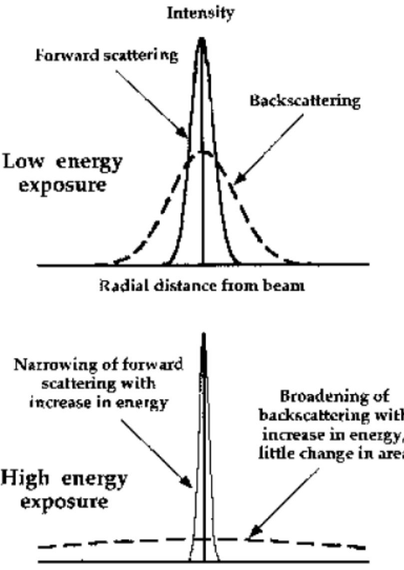


Fig. 4.5. Forward scattering and backscattering distributions.

The lateral distribution of deposited energy in the resist from the point of incidence is often approximated by Gaussian distributions, assuming initially that the primary beam is a delta function. A high-amplitude, narrow Gaussian describes the forward-scattered incident (primary) beam broadening due to elastic scattering in the resist. Because the atomic number of the resist (essentially carbon) is low, electrons are only scattered through a small angle in the resist and the exposure area is a narrow "skirt" around the incident beam. A second Gaussian, due to elastic scattering from the substrate, represents the contribution from backscattered electrons. Its amplitude is orders of magnitude less than the first Gaussian, but it is much broader, as shown in Fig. 4.5 [32]. The forward-scattered Gaussian distribution is dependent on the beam energy, resist type, and resist thickness. It increases in width with either increasing resist thickness or atomic number, and decreases in width with increasing beam energy. Broers [33] has noted that the standard deviation of the Gaussian may be expressed as σ_f (measured in μm), where

$$\sigma_f = \left(\frac{9.64z}{V}\right)^{1.75} \tag{4.1}$$

and where V (measured in keV) is the energy of the incident electron beam and z (measured in μm) is the thickness of the resist film.

The total quantity of backscattered electrons is characterized by a backscattering coefficient η. This coefficient is strongly dependent on the atomic number of the substrate and only weakly dependent on the incident beam energy. The backscattered Gaussian distribution is therefore highly dependent on the substrate atomic number, increasing in magnitude with increasing atomic number. Although not indicated by η, the width of the backscattered distribution is highly dependent on incident beam energy. The radial spread of backscattered electrons is comparable to the Bethe range, which is the characteristic path length for an electron in the solid that has given up all of its energy. The Bethe range increases with increasing incident beam energy. Thus, the characteristic width of the backscattered electron Gaussian distribution increases with increasing beam energy. Since η is weakly dependent on the beam energy, as the characteristic widths of the backscattered electron Gaussian distribution increase with beam energy, the nearly constant value of η dictates that the integrals of these

distributions remain approximately constant. This subsequently implies a decrease in the backscattered electron Gaussian peak height with increasing beam energy.

The comparatively diffuse backscattered energy distribution determines the contrast of the latent image in the resist, and the more compact forward-scattered energy distribution determines the ultimate resolution.

Summing up, there are three factors that strongly influence the electron scatterings:

1. *Electron energy*. Higher energy of primary electrons have much smaller scattering angle, therefore, less broadening of the beam due to small-angle forward scattering. Figure 4.6 compares the PSFs from 50 keV to 100 keV electron energies. The higher the electron energy the narrower the PSF. This is another proof that high-energy e-beam is preferred in e-beam lithography, in addition to benefits in lowering chromatic and space charge aberrations.

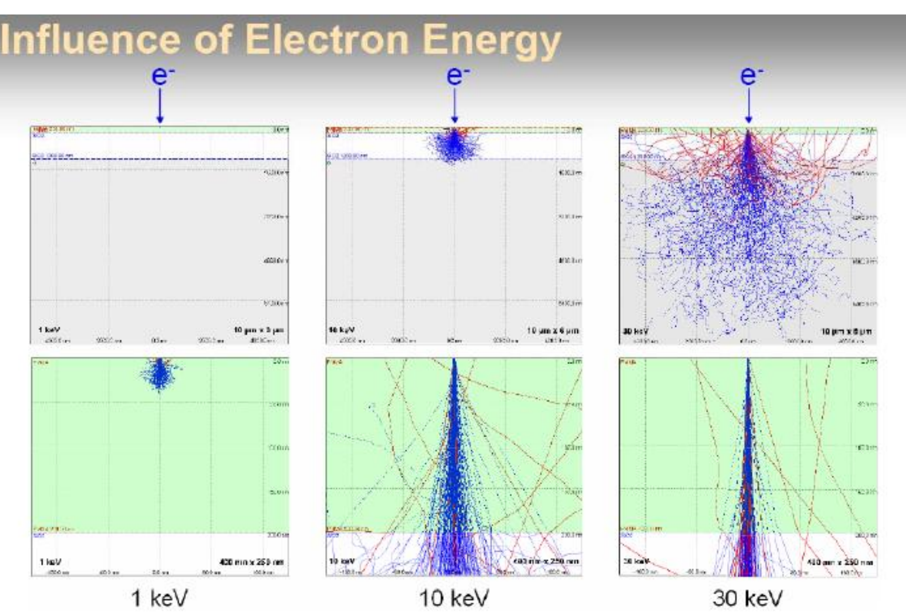


Fig. 4.6. Influence of electron energy on the electron scattering.

2. *Density of substrate material*. More electrons will scatter back from substrate if the substrate material is of higher density. Fig. 4.7 compares the PSFs for silicon and GaAs substrates. Electron energies are both at 50 keV and the resist thickness is the same as 0.5 μm. However, the deposited energy level due to backscattering electrons is significantly higher for GaAs substrate than for silicon substrate, because the density of GaAs is 2.3 times higher than that of silicon. Even though the substrate is a light material, such as silicon, the backscattering can still be significant if there is a thin film of heavy metal, such as gold or tungsten, coated at the interface between silicon and resist layer. Therefore, it is more difficult to achieve high-resolution ebeam lithography on high-density substrate.

3. *Resist thickness*. It has been apparent from Figs. 4.3 and 4.4 that electrons scatter wider forwardly as they go deeper. Therefore, for high-resolution e-beam lithography thin resist layer is preferred.

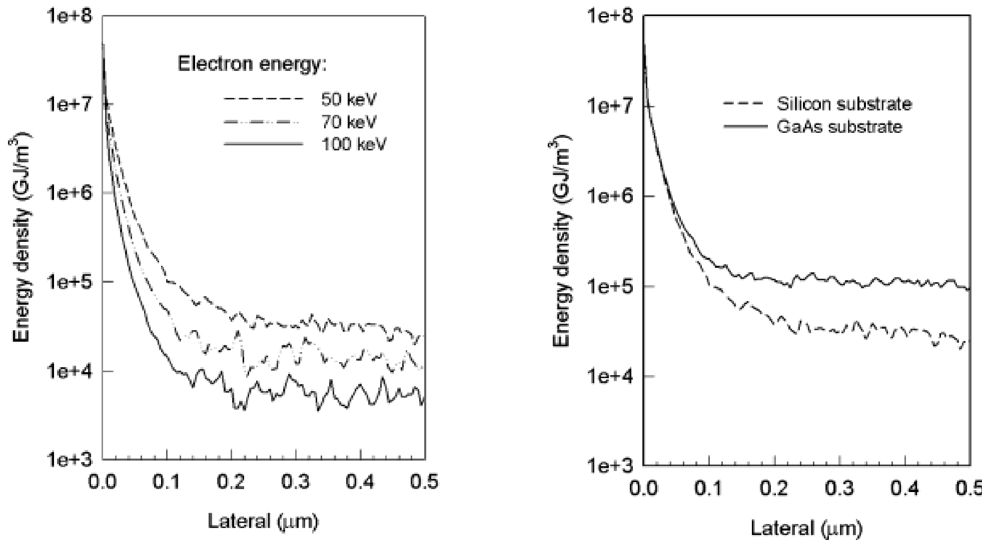


Fig. 4.7. Left: comparison of PSF at different electron energies (silicon substrate, 0.5-mm resist thickness). Right: comparison of PSF for different substrate materials (50-keV beam energy, 0.5-mm resist thickness). [31]

4.1.2 The Proximity Effect

During an e-beam exposure of materials, the region of resist interaction with electrons is larger than the size of the incident probe. This is due to both elastic and inelastic electron scattering in the resist and the substrate. The proximity effect is the exposure of resist by electrons backscattered from the substrate, constituting a background where the pattern is superimposed. Since this background is not constant, pattern distortions arise.

Fig. 4.8 illustrates the way in which the proximity effect affects pattern dimensions.

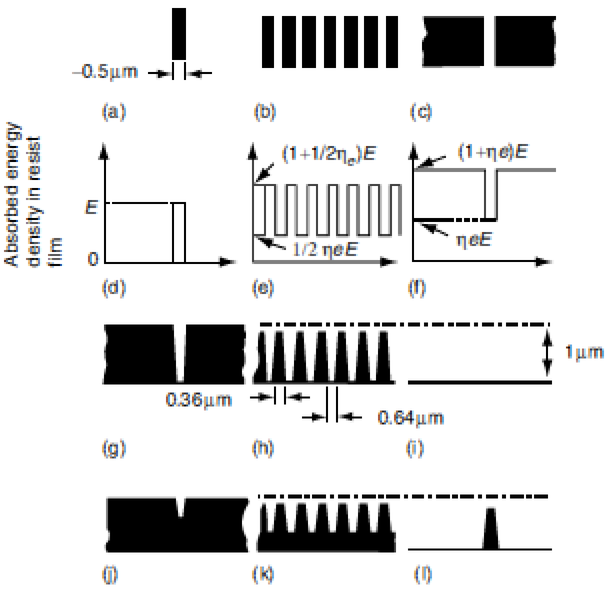


Fig. 4.8. The influence of the proximity effect on the exposure of (a) an isolated line, (b) equally spaced lines and spaces, and (c) an isolated space (exposed regions are denoted by shading). The absorbed energy densities are shown in (d), (e), and (f). The profiles of the resist patterns are shown in (g), (h), and (i) after development time appropriate for the isolated line. The profiles shown in (j), (k), and (l) occur after a development time that is appropriate for the isolated space.

In Fig. 4.8 (a), it is assumed that an isolated narrow line (for example, of width 0.5 μm) is to be exposed. The corresponding energy density distribution is shown in Fig. 4.8 (d). The energy density deposited in the resist by forward-traveling electrons is E. Because the width of the isolated line is a fraction of 1 μm, and the characteristic diameter of the backscattered electron distribution (σ_b) is several μm, the backscattered energy density deposited in the resist is negligibly small.

This, however, is not the case for the exposure of the isolated space depicted in Fig. 4.8 (c). Because the width of the space is much smaller than σ_b, it is a good approximation to assume that the pattern is superimposed on a uniform background energy density $\eta_e E$, as is shown in Fig. 4.8 (f). Figs. 4.8 (b) and (e) deal with an intermediate case: an infinite periodic array of equal lines and spaces. Here, the pattern is superimposed on a background energy density of $\eta_e E/2$. Thus, the energy densities corresponding to exposed and unexposed parts of the pattern are strongly dependent on the nature of the pattern itself. This leads to difficulties in maintaining the fidelity of the pattern during development. In the drawings in Figs. 4.8 (g) through (i), it has been assumed that a development time has been chosen that allows the isolated line (g) to develop out to the correct dimensions. However, because the energy densities associated with the isolated space are appreciably higher, this pattern may well be so overdeveloped at the end of this time that it has completely disappeared (i). The problem may not be so drastic for the equal lines and spaces, for which the energy densities have lower values. Nevertheless, the lines will be overdeveloped. The situation cannot be improved by reducing the development time. Figure 4.8 (j) through (l) illustrate what happens if the patterns are developed for a time appropriate for the isolated space. Although this is now developed out to the correct dimension, the isolated line and equal line and space patterns are inadequately developed.

In general, because of the proximity effect, anyone of these types of patterns may be made to develop out to the correct dimensions by choosing an appropriate development time. However, it is impossible to simultaneously develop them all out to the correct dimensions.

Let's see in some further details how the proximity effect lead to difficulties in maintaining the fidelity of a pattern.

In considering an arbitrary point in the resist, this point will be developed out as part of the resist image if the deposited energy at the point exceeds a certain accumulated dose [see Fig. 4.9 (and chapter 5 for details about the process)].

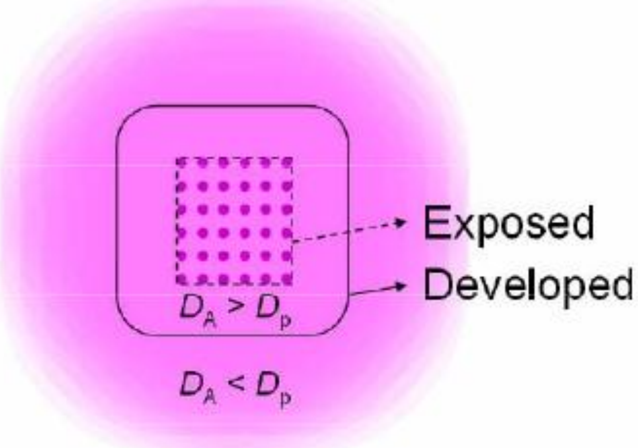


Fig. 4.9. Unwanted exposure due to the proximity effect.

We have seen that since the total contribution at a point in the resist is pattern-dependent, dense features will become overexposed and very isolated features will be underexposed, creating the so-called intershape and intrashape proximity effect.

Consider a specific example of the exposure of an isolated square pattern of dimensions comparable or larger than the Bethe range for the given exposure conditions. The exposure conditions will be fixed such that the primary energy deposited in the resist will be equal for all points in the pattern. At the center of the pattern there is a total contribution of deposited energy that is the sum of the primary energy plus backscattered energy contributions from nearby points exposed in all directions. However, in the corners of the square there are nearby backscattered contributions only over one quadrant, reducing the total deposited energy relative to the center point. The result, after development to a constant energy contour, is the rounding of the corners. This is an example of *intraproximity effect*, the variation of total deposited energy *within* the pattern (see Fig. 4.10).

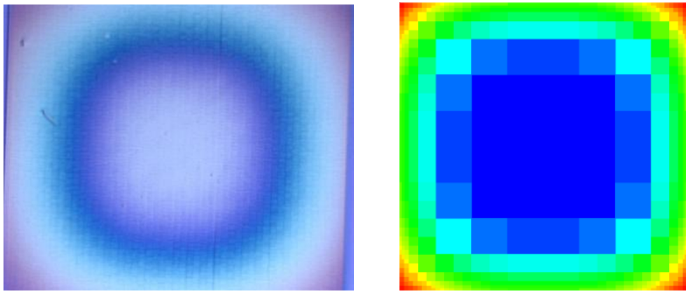


Fig. 4.10. Left: optical microscope image of a developed square pattern with an equal dose assigned to all points. Right: proximity effect can be corrected assigning different doses (shown in the image by different colors) to different points.

Next consider the exposure of an isolated cross pattern of dimensions comparable or larger than the Bethe range for the given exposure conditions, as shown in Fig. 4.11. At the tips, the outside corners will have backscattered contributions integrated over one quadrant (similar to the square example above) and will thus exhibit the characteristic intraproximity rounding. Now consider a point just outside of the pattern area, at one of the inside corners where the two cross arms intersect. At some points sufficiently close to but still outside the pattern, the integrated backscattered electron contribution will exceed the energy threshold for development. This will result in a fillet or web across the corner. This is one example of the *interproximity effect*, the variation of total deposited energy *outside* the pattern.

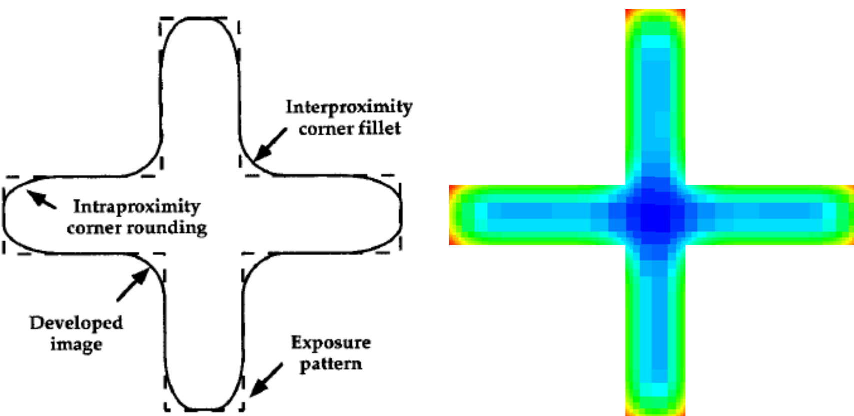


Fig. 4.11. Left: cross pattern distortion due to proximity effects. Right: proximity effect correction for a cross pattern (different colors indicate different doses).

In considering the impact of *exposure energy* on proximity effects, the peak intensity of the backscatter is reduced with increasing energy. This reduction in intensity can reduce the total integrated energy at a point outside the pattern to below the development threshold, improving pattern integrity by reducing interproximity effects. Since the backscatter contribution is made more diffuse, the variation of dose within the pattern is also reduced with increasing energy, leading to reduced intraproximity effects.

A simulation of how the proximity effect affects the resist profile at different *exposure doses* is shown in Fig. 4.12. It shows the cross-section profile of PMMA resist of 1-μm thickness, exposed by 20-keV e-beam. The pattern is 100-nm line grating with 500-nm pitch. The ideal profile is shown by the shaded area which is unexposed resist with 100-nm line openings. The actual resist profile is shown by the contours which represent the evolution of resist cross-section at different exposure doses. The higher the exposure dose, the deeper it goes until it reaches the full depth of 1 μm. It is apparent that the actual resist profile is significantly deviated from the ideal profile, because of the proximity effect. The result manifests the effect of electron forward scattering, which becomes wider as the electrons go deeper in the resist.

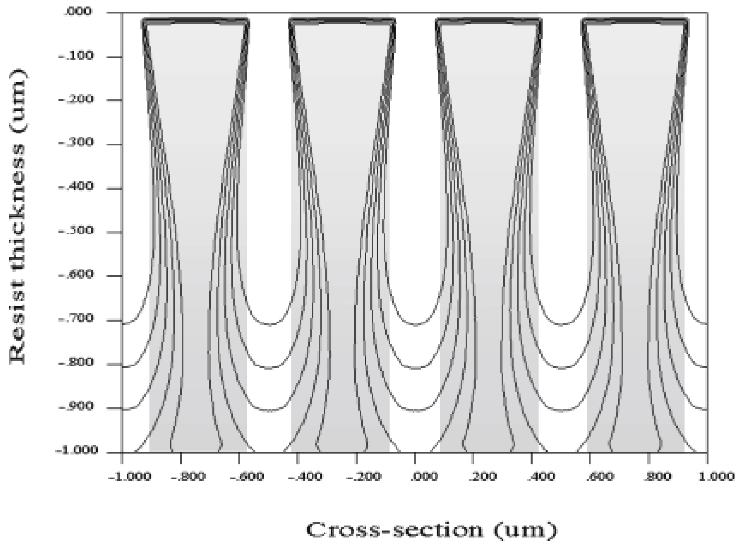


Fig. 4.12. Simulated resist cross-section profiles for 20-keV e-beam exposure of 1-μm PMMA resist layer on silicon substrate; the shaded area represents the ideal resist profile and the contours represent the actual resist profiles at exposure doses from 80 μC cm^{-2} to 120 μC cm^{-2} with 10 μC cm^{-2} increment. [31]

Although the proximity effect poses serious problems, it does not constitute a fundamental resolution limit to electron lithography. A number of methods for compensating for the effect, or at least reducing the gravity of the problems, has been devised. These are described below.

4.1.2.1 Methods of compensating for the Proximity Effect

Dose correction

The most popular form of proximity effect compensation is probably *dose correction*. To implement this method, the dose delivered by the lithography instrument is varied in such a way as to deposit the same energy density in all exposed regions of the pattern. Referring to Figure 4.8, this situation occurs

if the dose delivered to the isolated line (a) is increased by the factor $1 + \eta_e$ and that to the lines of the line and space pattern (b) by a factor of $(1 + \eta_e)/(1 + \eta_e/2)$. If this is done, all three types of patterns will develop out to approximately the correct dimensions after the same development time (that appropriate for the isolated space in Figure 4.8 (c).

Examples of dose correction applied to a square and a cross pattern with the GenISys Layout BEAMER™ software are shown in Fig. 4.10 and 4.11.

Biasing

In some cases the proximity correction can be done by resizing the pattern geometry. The resizing, also called biasing, is best suitable for simple and highly repeated patterns, such as lines and squares.

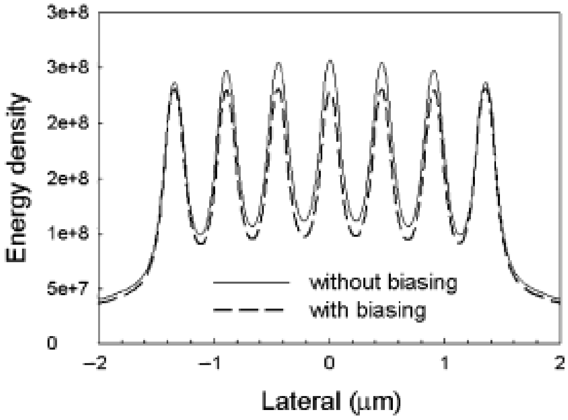


Fig. 4.14. Comparison of pattern biasing for proximity effect correction (150-nm lines and 450-nm pitch, exposed with 20-keV beam energy into 0.5-μm resist on silicon). [31]

For a group of repeated lines, the lines in the middle of pattern area can be resized narrower (negative biased) than nominal design to compensate the higher-than-average energy contributed from neighboring lines, or the lines at the outer rim of pattern area can be resized wider (positive biasing) to compensate the lower-than-average energy due to electron scattering. The overall result is a relatively uniform distribution of energy across the whole pattern area. The amount of biasing can be determined empirically or from simulation of electron energy distribution. An example is shown in Fig. 4.14.

The figure shows deposited energy for a group of seven lines with linewidth of 150 nm and pitch of 450 nm exposed by 20-keV e-beam into 0.5-μm thick resist on silicon. If all lines are exposed at equal width, the lines in the middle receive higher energy (solid line plot) because electron scatterings from both sides contribute energy to the middle lines. If the middle lines are negatively biased, all the lines receive more or less the same energy (dash line plot).

GHOST

The third method for proximity correction uses background exposure to compensate the nonuniformity of electron energy distribution. As already seen, electron scattering always causes excess of energy at dense pattern region and less energy at sparse pattern region. If a second exposure is carried out at all non-pattern regions (background), the pattern density for the second exposure is exactly reversed. The originally sparse pattern region now becomes dense pattern region and receives

more energy. Meanwhile, the original dense pattern region where there are less blank areas now receives less energy. The superimposition of the two exposures can well balance the energy distribution so that it becomes uniform everywhere.

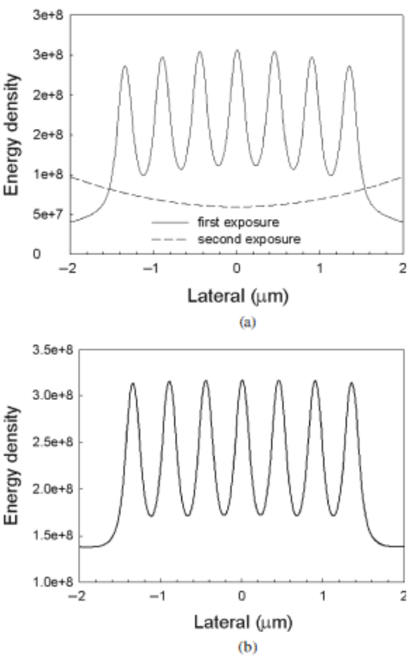


Fig. 4.15. (a) Energy profiles of first and second exposures. (b) Superimposed energy profile. [31]

In Figure 4.15 (a) the first exposure produces an unbalanced energy profile, with the lines in the middle receiving higher energy. If second background exposure, with energy profile shown by the dash line plot Fig. 4.15 (a), is added to the first exposure, the overall energy distribution is well balanced, as the energy profile plot shown in Fig. 4.15 (b). The background exposure method has a name called ‘‘GHOST’’ [34]. It is simple, does not need special proximity correction software and does not need to calculate the electron energy distribution. All it needs is to have a second exposure of reverse polarity of the original pattern design. The drawbacks of this method are increased pattern data volume, increased total exposure time, and deterioration of image contrast due to the background exposure. In addition, this technique is only practical in a raster scan e-beam system.

The role of beam energy

All the above proximity correction techniques require additional process and special software tools. As we have seen discussing interproximity and intraproximity effects, a simple and effective measure to reduce the proximity effect is to increase beam energy as well as to decrease resist thickness. High beam energy can greatly reduce the beam expansion caused by forward scattering and reduce localized backscattering. As forward scattering is wider with electrons going deeper, a thin resist layer can also minimize the effect of beam expansion. An example is shown in Fig. 4.16, which is a direct comparison with Fig. 4.12. The pattern design and resist thickness are the same in both cases. The only difference is the beam energy for the exposure. While in Fig. 4.12 the resist is exposed at 20 keV, the beam energy in Fig. 4.16 is 100 keV. No proximity effect is seen for 100-keV exposure and resist profile is almost the same as the ideal profile (shaded area).

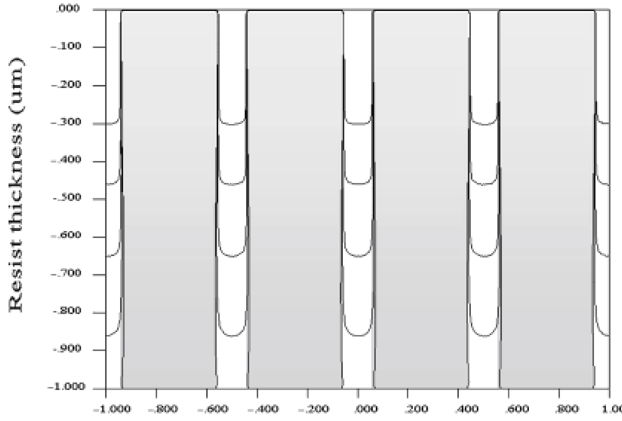


Fig. 4.16. Simulated resist cross-section profiles, to compare with those in Fig. 4.12, for 100-keV e-beam exposure of 1-μm PMMA resist layer on silicon substrate; the shaded area represents the ideal resist profile and the contours represent the actual resist profiles at exposure doses from 150 μC cm^{-2} to 350 μC cm^{-2} with 50 μC cm^{-2} increment. [31]

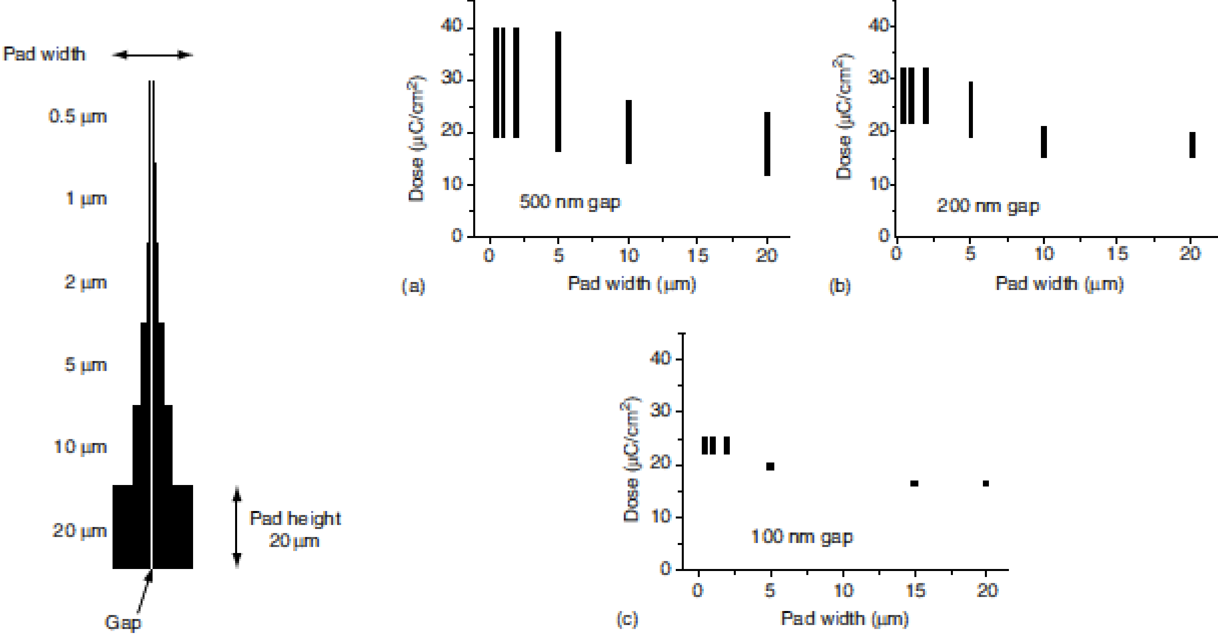
All the above discussion on electron scattering and proximity effect has favored using high-energy e-beam. High beam energy results in narrower forward scattering and lower backscattering. However, one must be aware that high beam energy requires high exposure doses. The exposure dose increases linearly with the increase of beam energy, because majority of primary electrons have gone deep into substrate and very few of them contribute to the exposure of resist.

4.1.3 Resolution and Process Latitude

In nanolithography, the critical dimension (CD) of a pattern (defined as the size of the smallest feature), is comparable to the broadened forward Gaussian. The broader the primary electron spot, the poorer the image edge contrast. In addition, the existence of a backscattered electron "fog" tends to reduce the exposure contrast at critical edges. These effects reduce the *process latitude* [35], which is already very narrow for nanolithography, and contribute to line edge roughness [36].

For an experimental demonstration of the reduction in dose latitude, as nanometer-scale dimensions are approached, consider the pattern in Figure 4.17. The size of the rectangles (pads) in each pattern varies in width from 0.5 to 20 µm. The size of the CD unexposed stripe (gap) varies in different patterns from 100 to 500 nm. Shown in Figure 4.18 is the dose latitude data to form the specified stripe dimension, ± 10% CD, for the patterns shown in Figure 4.17. The pattern was defined in the 150 nm thick SAL-601 resist on Si, with a 50 keV e-beam and a 10-nm-σ Gaussian probe. The gap width in the center of the length was measured as a function of dose and pad width. Figure 4.18 shows data in separate plots for three different gap widths: 500, 200, and 100 nm. The data shows effects of backscattered electron proximity effects in the differences in dose latitude between the different pad sizes for a given gap width. The measured primary Gaussian standard deviations for the exposure and resist process conditions was 27 nm [36], approximately one-fourth the size of the minimum CD. For each stripe width, there is a shift in centroid of the dose window with increasing pad width. There is a transition in the width of the dose window at 5-µm pad width, which is close to the width of the backscattered electron Gaussian (7 µm). The smaller pads have a larger dose latitude and the lithographic features require higher doses than optimal for the larger pads. However, more dramatically, there is a large reduction in the dose latitude with a reduction in stripe width when Fig. 4.18 (a), 4.18 (b), and 4.18 (c) are compared. For the 20-µm pads, the 500-nm stripe can be defined

with a ± 32% dose latitude, while the 100-nm stripe requires a dose accuracy of better than ± 3%. The dimensional difference between 500 and 100 nm is negligible compared to the 7 μm backscattered electron radius. A large part of the error is due to the higher precision (in ± nm) required to define a 100-nm stripe.



(left) **Fig. 4.17.** Diagram of tower pattern for studies of proximity effects in resists. [37]
(right) **Fig. 4.18.** Effect of width of unexposed stripe on process latitude. Shown in each graph is the dose range over which each stripe was defined within G10% error as a function of pad width. [36]

In nanolithography, effective spot size is close to the desired feature size. It is often thought that if one wants to write 20-nm features, a 20-nm e-beam spot size is sufficient, even if the features are densely spaced. This is not entirely true. The larger the spot size, the lower the image contrast of the feature edge and the lower the process latitude. This effect is observed even when the spot size is a factor of four or more less than the CD.

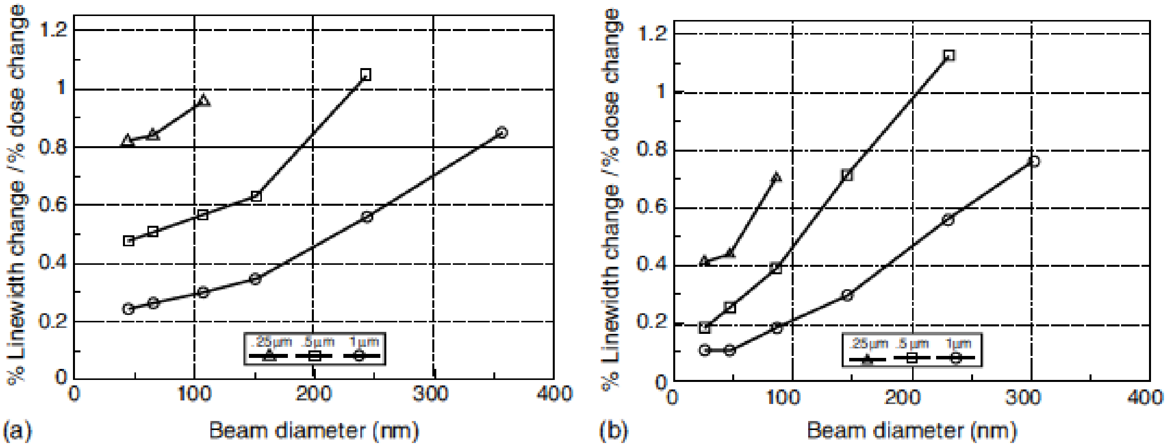


Fig. 4.19. Effect of spot size on dose latitude for isolated lines at (a) 50 kV and (b) 100 kV. [37]

McCord et al. [38] showed the process latitude, defined as percent linewidth change divided by percent dose change, decreased greatly with increasing spot size at electron probe sizes much smaller than the linewidth. McCord's results are shown in Figure 4.19, where graphs are plotted for 0.25, 0.5, and 1-μm dimensions. One can see that the spot size greatly affects the process latitude, even when the spot size is significantly smaller than the CD. One can also see that the smaller the feature, the greater the variation of size with dose. At the smaller spot sizes, the 100-kV beam offers less linewidth variation with dose than the 50-kV beam.

The effect of the width of the primary beam spot size on process latitude is demonstrated by Dobisz and Marrian [39] in Figure 4.20.

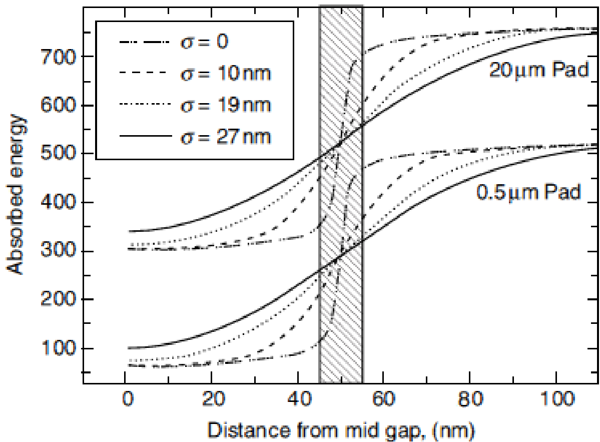


Fig. 4.20. Effect of forward Gaussian width on process latitude. Calculated absorbed energy vs. distance from center of a 100-nm unexposed stripe. The shaded region is $\pm$ 10% the coded gap dimension. Calculated from Monte Carlo code PSF with four different primary Gaussian widths. [38]

Here, an exposed rectangle with a 100-nm unexposed stripe in the center is considered. Shown in Figure 4.20 is the normalized energy deposited in a resist layer as a function of distance from the center of the unexposed stripe. The pointspread function was calculated for a 50-kV electron beam, 100-nm resist, and a Si substrate. Four cases are shown in which the Monte Carlo-code-generated PSF was convolved with Gaussian probes of three standard deviations and one without convolution. Rectangle sizes of 0.5 and 20 μm were considered. If the correct dose is applied, the clearing dose to define the stripe should occur at 50 nm from the center of the stripe. The $\pm$ 10% CD tolerance of the stripe edge is shown in the shaded region of Figure 4.20. One can see the hypothetical case of perfect point exposure with broadening only from electron scattering as calculated from a Monte Carlo code. Here, the absorbed energy changes rapidly at 50 nm and the edge has a large image contrast. If one takes into consideration a Gaussian beam of 10-nm standard deviation, the slope of the energy absorbed is reduced by a factor of 2.5. A plateau dose is not reached until 20 nm in either direction from the edge of exposure. If the finite Gaussian beam is allowed to broaden in the resist and have standard deviations of 19 and 27 nm, the slopes of the pattern edge exposure decrease further by factors of 2 and 3 from the 10-nm incident beam. These spot sizes are less than the CD, yet the process latitude turns out to be greatly affected by the spot size.

The factors that can influence the ultimate resolution of e-beam lithography are related to the performance of e-beam lithography systems, the e-beam resist processes and the nature of electron scattering in resist and substrate materials.

To achieve high-resolution e-beam lithography, the following conditions are generally preferred: *high electron energy*, *small scanning field*, *low beam current*, *low-sensitivity resist*, *thin resist layer*, *optimized resist process*, *low pattern density*, *light and conductive substrate material*, *stable environment*.

In particular:

- The advantages of higher acceleration voltages are that they enable smaller beam sizes (i.e., higher patterning resolutions) and higher beam stability (i.e., smaller deflection noise and drift). Moreover, a higher acceleration voltage reduces pattern widening due to forward-scattering electrons in resist film. These advantages are very important for precise high-resolution patterning. A higher voltage also causes longer mean free paths and thus longer ranges of electron scattering. Although this means proximity effect correction has to include a wider area, this is not disadvantagous for maintaining high resolutions, since the amount of total energy deposited in proximity areas does not significantly depend on the voltage. On the other hand, higher voltages result in a smaller interaction between electrons and resist, which leads to a reduction of resist sensitivity.
- It is not difficult to understand why low-sensitivity resists have high resolution capability: the long tail of PSF is in low energy level, which will not have significant exposure effect if the resist is low in sensitivity. In other words, the low energy electrons will only show up their impact on resist pattern profile if the resist sensitivity is very high.
- Higher throughput (larger beam current) and higher resolution (smaller beam size) are mutually exclusive. Let's examine the relationship among the beam current, beam step, and resist sensitivity:

  $$I = Dfd^2 \tag{4.2}$$

  where d is the beam step, f the deflection frequency and D the area dose on exposure (mainly decided by the resist sensitivity).

  Since the maximum deflection frequency depends on the EB apparatus, the maximum beam current is limited by the beam step and resist sensitivity. As the beam step should generally be smaller than the minimum feature size in order to obtain smooth pattern edges, the beam current may have to be reduced because of this relationship.

RESIST MATERIALS AND PROCESSES

Introduction

Electron-beam resists are most commonly liquid solutions composed of organic polymers cast in a solvent. The liquid is applied to a substrate by spin coating to form a film of uniform thickness. Following spin casting, the casting solvent is driven from the film by baking, referred to as a softbake. The softbaking leaves a durable polymer film on the substrate that is ready for electron exposure. Often resist is applied to the substrate in multiple layers or in layered combinations with inorganic materials or other organic materials that are not sensitive to electron-beam exposure. The purposes of multi-layer resist are varied and include planarization, improvement of pattern transfer, reduction of backscatter, and creation of specific feature cross-sections. The electron-beam pattern exposure deposits energy in the resist and modifies it to create the latent pattern image. Development completes the transformation of the resist coating into the three-dimensional representation of the CAD pattern, revealing the latent image. Development is most commonly performed by wet processing with organic solvent or aqueous-based chemistry. For a positive resist, the exposed regions are made more soluble and are selectively removed by the developer. For a negative resist, the exposed regions are rendered insoluble in developer and the unpatterned area (the field) is removed by the developer.

The energy of charged particles deposited in a polymer resist will have to be converted into molecular chain reaction in the resist to eventually form a lithographic pattern. A small beam size and small scattering range are not necessarily translated into a small size of resist features. On the other hand, a poor resist material and non-optimum resist process can certainly result in poor resolution and much enhanced proximity effect. Resist materials and processes play a vital role in achieving the ultimate resolution capability of lithography by charged particle beams.

5.1 Resist materials

In single-component polymeric resists, the solubility of the polymer in a solvent is a strongly varying function of its molecular weight. Exposure of the resist results in chain scission or cross-linking. The lower-molecular-weight polymer dissolves in the developer, so in the former case the resist is positive tone; in the latter case, it is negative tone. Polymeric resists are versatile and simple to use because they can be spun onto a variety of substrates.

For high-resolution e-beam lithography, high-sensitivity resists are preferred. Table 5.1 gives a list of some commonly used e-beam resists with their sensitivities and resolution capabilities.

	Resist tone	Sensitivity* (μC cm^{-2})	Resolution(nm)
PMMA	+	100	10
ZEP-520	+	30	10
ma-N 2400	–	60	80
EBR-9	+	10	200
PBS	+	1	250
COP	–	0.3	1000

*Sensitivity measured at 20-keV beam energy.

Tab. 5.1. Sensitivity and resolution of some e-beam resists.

Of course, high-sensitivity resists with a higher resolution would be very useful, but physical limits, such as pattern-edge roughness due to shot noise, could degrade the patterning. Moreover, high sensitivity and high resolution are thought to be contradictory.

5.1.1 Positive resists

5.1.1.1 PMMA

For many years, PMMA has remained the resist of choice for very high resolution electron beam lithography since its introduction into electron-beam lithography in 1968 [40]. It has extremely high resolution (below 10 nm), excellent shelf life, and does not suffer from swelling or post-exposure latency effects. Its biggest drawbacks are its poor sensitivity, poor dry etch resistance, and poor thermal stability. Historically cast in chlorobenzene, it is now commercially available cast in safer anisole solvent. It is available in molecular weights from 50 to 950K, with slightly decreasing sensitivity and higher resolution capability with increasing molecular weight. PMMA is also sensitive to exposure by deep ultraviolet light and x-rays, permitting hybrid exposure strategies. A hybrid exposure exposes the finest features of a given lithography level by electron beam with coarser features of the same lithography level exposed separately by a high-throughput optical or x-ray technique, all in the same resist and with a common development. The primary exposure mechanism for PMMA is by scission of the polymer chain (see Fig. 5.1). Scission in the exposed region reduces the molecular weight, with the lighter weight fragments selectively removed by developer.

Fig. 5.1. Mechanism of radiation-induced chain scission in PMMA.

At roughly an order of magnitude higher dose, PMMA will crosslink and exhibit negative tone, an effect commonly observed where alignment marks are scanned for registration. The poor thermal stability of PMMA is problematic for electron-beam inspection (SEM) of developed images, where the resist deforms rapidly under observation. In addition, elevated temperatures during material deposition onto patterned PMMA can result in CD variation and pattern distortion.

MIBK/IPA and water/IPA as PMMA developers: a thorough analysis

In electron beam lithography, the development of exposed resist often involves the use of binary solvent mixtures, which typically consist of a strong solvent and a moderating non-solvent. Using conventional techniques, feature sizes down to 10 nm can be fabricated with PMMA, but with low reliability and this severely restricts the complexity of devices which can be made at these dimensions. However, in recent years significant enhancements to the contrast of PMMA, through the use of new developers, have been reported. Among them, isopropanol (IPA), mixed with a small proportion of water, has been demonstrated to have superior lithographic performance [41]. Since both IPA and water separately are non-solvents for PMMA at room temperature, this mixture acts as a cosolvent system, which is defined as a system in which the polymer coil is more effectively solvated than in either of the two separate component liquids.

The basic requirements for a solvent mixture, to be used as a developer for a positive resist, can be summarized as follows. First, the mixture must have appropriate thermodynamic properties to act as a minimal solvent for the unexposed resist, so minimizing the swelling of the resist features remaining after development. Second, the mixture needs to be a good solvent for the exposed resist, so that the soluble regions are completely dissolved and washed away during the development time. The solubility of a polymer matrix after exposure depends on the molecular weight of the polymer and on the thermodynamic quality of the solvent mixture. The developer should be just capable of dissolving the low molecular weight fractions resulting from electron beam exposure whilst leaving the rest of the film with little or no swelling or distortion. The main factors that affect the rate of solubilization of a polymer are the plasticization and thermodynamic compatibility or 'solvent quality' between the resist polymer and the developer solvent mixture. Plasticization occurs when solvent penetrates the polymer matrix, increasing the free volume, and so promoting the dissolution of the resist. The rate at which resist is plasticized is related to the molecular size (molecular weight) of the polymer. Solvent quality is a measure of the closeness of the thermodynamic parameters of the solvent to those of the polymer. A high dissolution rate of the resist occurs if the thermodynamic properties are close.

The overall performance of a developer is characterised by sensitivity, contrast, exposure dose latitude, roughness and ultimate resolution, but in order to optimize developer composition, *sensitivity* and *contrast* alone can be used.

Previous attempts made to increase the sensitivity and contrast of PMMA resist have used variations in the temperature, composition and development time of standard developers such as MIBK/IPA [42], giving optimum values of 25 °C, 1:3 and 30 s, respectively. In a recent work, Yasin et al. [41] found that a high improvement in sensitivity and contrast can be achieved using isopropanol (IPA), mixed with a small proportion of water, a developer solvent that is more finely tuned to the dissolution of the degraded PMMA while leaving the unexposed regions relatively unaffected. Among the different compositions, the composition of 3:7 water/IPA has been selected as the optimum developer due to its

combination of high sensitivity and good contrast. Table 5.2 shows how water/IPA improve sensitivity by nearly 40%, and contrast by nearly 20%, compared to standard 1:3 MIBK/IPA developer.

Developer system	Sensitivity ($\mu C/cm^2$)	Contrast γ
MIBK/IPA	520	5.4
Water/IPA	380	6.4
Water/MIBK	200	3.1

Tab. 5.2. Comparison of sensitivity and contrast for different developers. [41]

Process latitude too is important in determining the usability of an exposure/development strategy, and necessary for maintaining the correct feature size across a substrate and from one substrate to another. A wide process latitude also improves the reliability of forming small features. Fig. 5.2 compares the variation of linewidth as a function of incident dose for 3:7 water/IPA and 1:3 MIBK/IPA developer using PMMA resist. As can be seen, the exposure dose latitude, with the use of 3:7 water/IPA, increases by nearly 40% compared to the use of 1:3 MIBK/IPA.

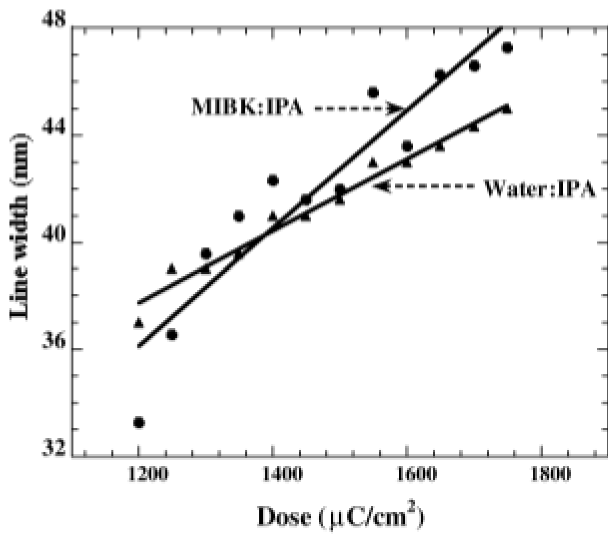


Fig. 5.2. Exposure dose latitude for 1:3 MIBK/IPA and 3:7 water/IPA using conventional dip development. The exposure dose latitude is increased by ~ 40% with water/IPA developer compared to standard MIBK/IPA developer. [41]

The origin of *roughness* in large-area *partially developed* resist features can be attributed to two-dimensional *phase separation*.

Phase separation is characteristic of development in weak solvents and for long development times. For temperatures above a critical value, all polymer-solvent mixtures are homogeneous and stable. However, development is usually performed at temperatures well below the critical value, in presence of unstable polymer-solvent mixtures which can lower the free energy of the solution by phase separation. Phase separation results in a division of the mixture into polymer-rich and polymer-poor phases. The delay in dissolution of the polymer-rich phase results in some regions being left behind during development process. These regions are indistinguishable from the unexposed resist and appear as roughness.

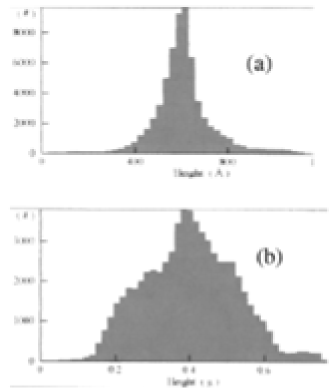


Fig. 5.3. Height histograms for PMMA developed in (a) 3:7 water/IPA; (b) 1:3 MIBK/IPA. The scale on the x-axis shows the surface roughness. [41]

Fig. 5.3 shows height histograms of underexposed and developed resist patterns using 3:7 water/IPA at 20 °C and 1:3 MIBK/IPA developer at 25 °C. The x-axis shows the size of the islands, which reflects the degree of phase separation and hence the roughness. It is evident from Fig. 5.3 that the island size reduces by an order of magnitude with the use of water/IPA developer compared to MIBK/IPA development. The decrease in the roughness with water/IPA developer indicates that the conditions necessary for phase separation are frustrated. The reason lies in the good compatibility between the water/IPA developer and the PMMA polymer, which leads to enhanced dissolution and to reduced phase separation.

5.1.1.2 ZEP

ZEP is a copolymer of chloromethacrylate and methylstyrene that provides high resolution comparable to PMMA, but with improved sensitivity and dry etch resistance. Development is accomplished with n-hexyl acetate, ZED50 (n-amyl acetate), or ZEP-RD (xylene), and is immediately followed by a rinse of isopropanol or methyl isobutyl ketone.

5.1.2 Negative resists

5.1.2.1 HSQ

As a way to improve the resolution and reduce pattern-edge roughness, the miniaturization of resist polymers/molecules has been investigated. Hydrogen silsesquioxane (HSQ) is a negative resist with small molecular weight, which exposure causes to cross-link. It can be developed in tetramethyl ammonium hydroxide (TMAH) developers. HSQ has enabled nanopatterning of 5–7 nm in size and small line edge roughness (LER) of 1–1.5 nm. A key advantage of HSQ is its etch resistance to oxygen, which makes it ideal in a bilayer resist scheme. Although PMMA has similar or slightly better resolution capability than HSQ, the LER in HSQ is much lower and the etch resistance much higher than that of PMMA. In these aspects, HSQ is a better e-beam resist for extremely high-resolution e-beam lithography.

5.2 Resist processes

5.2.1. Resist adhesion and substrate priming

Adequate adhesion of resist to a wafer surface is critical for proper process performance. Resist adhesion failure can occur not only during lithography operations, but also in subsequent etch, implant, or other masking steps. Negative resists are less prone to adhesion failure because as cross-linking results in a networked polymer that is bound to the wafer surface. Positive resists are more likely to be single-polymer chains and rely on weaker physical and chemical forces for adhesion. Etch-process undercutting can often result from inadequacies at the resist interface, resulting in loss of etch linewidth control.

The causes of resist adhesion failure are generally related to dewetting of a resist film. This can result from a large discordance between the surface tension of the wafer and that of the resist material, especially when coating over silicon oxide. Silicon dioxide is an especially difficult layer to coat because it provides a hydrophilic (water attracting) surface to a hydrophobic (water repelling) resist. Surface defects can also cause adhesion failure as surface free energy can result in dewetting.

Methods of adhesion promotion can be used for most silicon oxide layers, whether they are thermally grown, deposited, native, or glasslike. Chemical passivation of these surfaces is generally carried out using silylating priming agents that act to modify the wafer surface. Some benefit can be realized with priming of layers other than oxides if techniques promote a closer matching of material surface tension. Alkylsilane compounds are generally used to prime oxide surfaces, leading to a lowering of surface hydrophilicity. The most commonly used silane-type adhesion promoter is hexamethyldisilazane (HMDS). Reduction of substrate surface tension is carried out in two stages, as shown in Figure 5.4. The figure depicts a silicon oxide surface with adsorbed water and OH^- groups. An initial reaction of water with an alkylsilane (HMDS) produces an inert hexamethyldisiloxane and ammonia, resulting in a dehydrated surface. Further reaction with HMDS produces a trimethylsilyl-substituted hydroxyl or oxide species and unstable trimethylsilylamine. With heat, this unstable compound reacts with other surface hydroxyl groups to produce further ammonia and a trimethylsiloxy species. The process continues until steric hindrance (via the large trimethylsilyl groups) inhibits further reaction.

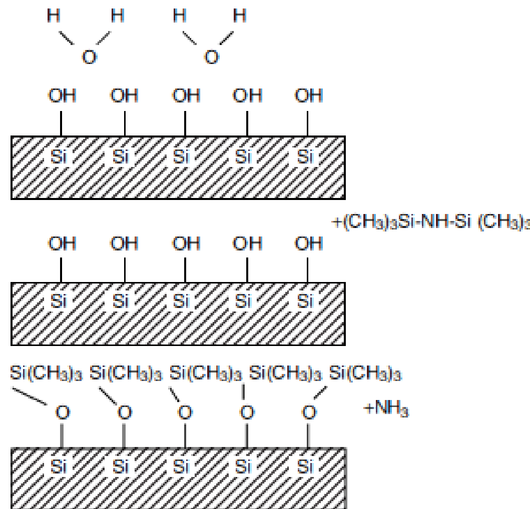


Fig. 5.4. Adhesion promotion of a silicon oxide surface with HMDS surface priming. The substrate is first dehydrated upon reaction with silane promoter. Further reaction with heat leads to a hydrophobic surface.

Elevated process temperatures (~100°C) must be reached to complete the priming reaction. Substrates should be cleaned prior to application using UV ozone, HF dip, plasma, or other “oxidative” cleaning methods.

5.2.2 Resist coating

A resist can be dispensed by several methods, including spin coating, spray coating, and dip coating. The most widely used method for coating resists onto wafer substrates are spin coating methods. During spin coating, the resist is dispensed onto a wafer substrate, accelerated to final spin speed, and cast to a desired film thickness (see Fig. 5.5).

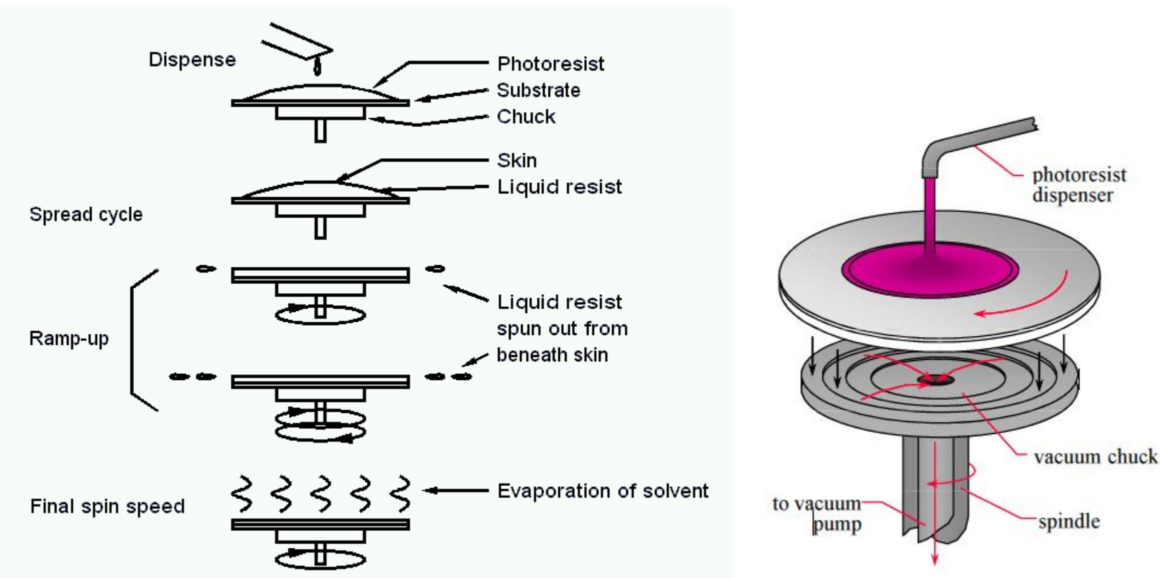


Fig. 5.5. Resist application by spin-coating.

Spin-coating processes use the dynamics of centrifugal force to disperse a polymeric resist material over the entire wafer surface. As a resist-solvent material is spin cast, the film thickness decreases uniformly at a rate dependant on the spin speed (ω), kinematic viscosity (ν), solids concentration (c), solvent evaporation rate (e), and initial film thickness, expressed by the following rate equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{-2\omega^2 h^3 c}{3\nu} \tag{5.1}$$

$$\frac{\mathrm{d}L}{\mathrm{d}t} = (1-c)\frac{2\omega^2 h^3}{3\nu} - e \tag{5.2}$$

where $\mathrm{d}S/\mathrm{d}t$ and $\mathrm{d}L/\mathrm{d}t$ are rate of change of solids (S) and solvents (L), respectively [43]. The results are shown in Fig. 5.6 for a 1-μm film, where both solids and solvent volumes are plotted against spin time.

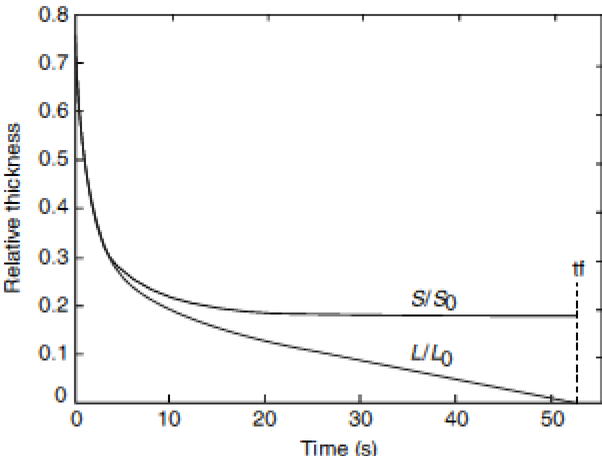


Fig. 5.6. Calculated time dependance during spin coating on the volume of solids (S) and solvent (L) per unit area normalized to initial values. When the resist thickness drops to one third of its original value, evaporation dominates and the solvent content reaches its final value. [43]

Initially, concentration changes little as resist spread dominates. When the resist thickness drops to one-third of its original value, evaporation dominates and solvent content reaches its final value. The high viscosity of the resist eliminates further flow.

The primary material factors that influence spin-coated film properties include the resist polymer molecular weight, solution viscosity, and solvent boiling point (or vapor pressure). Primary process factors include wafer spin speed, acceleration, temperature, and ambient atmosphere. The thickness of a resist film can be modified to some extent through control of the rotation speed of the substrate. Resist thickness (T) varies according to Eq. (5.3):

$$T = \frac{KC^{\beta}\eta^{\gamma}}{\omega^{\alpha}} \tag{5.3}$$

where K is a calibration constant, C the polymer concentration in grams per 100 ml solution, η the intrinsic viscosity (a measure of the molecular weight) and ω the number of rotations per minute (rpm). Once the various exponential factors (α, β and γ) have been determined, the equation can be used to predict the thickness of the film that can be spun for various molecular weights and solution concentrations (see Chapter 9, Fig. 9.9).

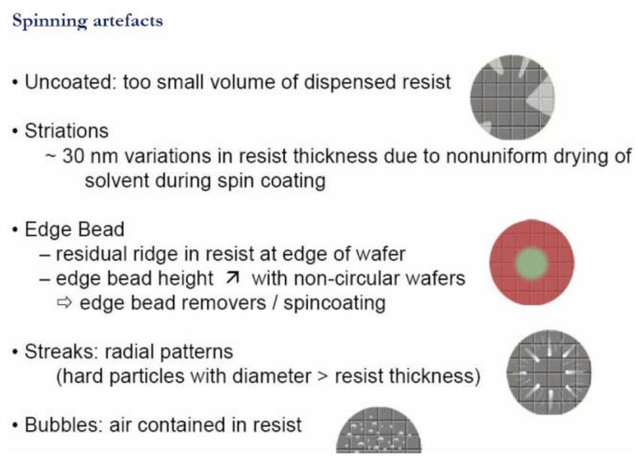


Fig. 5.7. Spinning artefacts can emerge as a result of not optimal coating.

To achieve large thickness changes, modification of the resist solution viscosity is generally required because coating at excessively low or high speeds results in poor coating uniformity. At excessively high speeds, mechanical vibration and air turbulence result in high levels of across-wafer nonuniformity. At low spin speeds, solvent loss of the resist front as it is cast over the substrate results in a situation of dynamic resist viscosity, also resulting in high levels of nonuniformity. The optimal spin range is dependent on wafer size.

Possible spinning artefacts are shown in Fig. 5.7.

5.2.3 Soft baking

After spin coating, the resist still contains up to built 15% solvent and may contain built-in stresses. The resist has therefore to be soft baked to remove solvents and stress and to promote adhesion of the resist layer to the substrate. To prevent stress arising during the cooling process, a suitable decreasing temperature ramp has to be set.

5.2.4 Exposure

The dose

The e-beam pattern consists of a matrix of points in which the beam dwells a certain amount of time to give the proper dose. The relationship among the *dose* (usually in units of micro coulomb per square centimeter), the step size (in cm), the beam current (in nano Ampère) and the dwell time (in seconds) is the following:

$$D = \frac{I\,\tau}{s^2} \times 10^{-3} \qquad (5.4)$$

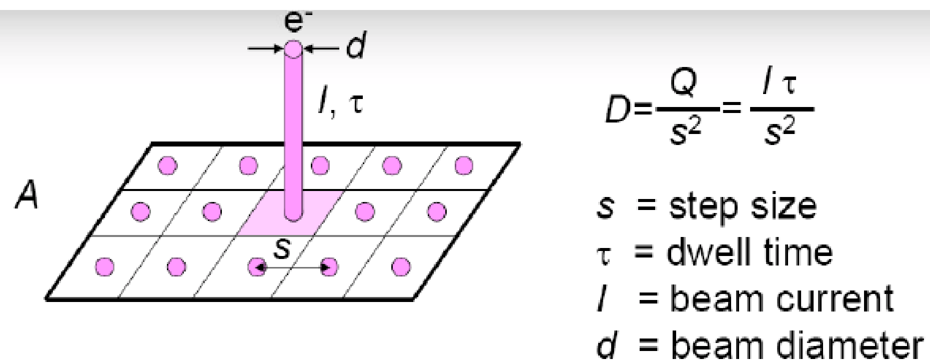


Fig. 5.8. A visual representation of Eq. (5.4).

Because the accuracy of resist pattern is so much dependent on exposure dose, it is a common practice that e-beam lithography of any new pattern designs is always started with a dose test, that is, running a series of exposure for the same pattern design at different exposure doses, from which the optimum exposure dose range is determined.

5.2.5 Sensitivity of resist materials

The most important properties of a resist are *sensitivity* and *contrast*, which are directly linked to the resolution capability of the resist. The sensitivity and contrast are derived from a resist development curve which typically looks like the plots in Fig. 5.9.

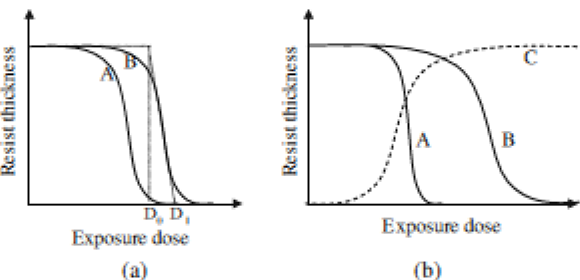


Fig. 5.9. Resist development curves: (a) resist A is of higher sensitivity than resist B; (b) resist A is of higher contrast than resist B, resist C is a negative resist.

For a positive resist, long molecule chains are broken by energized electrons into short chains, called chain secession process, so that the exposed parts of polymer become soluble in developer. For a negative resist, the initial short chain molecules are joined upon exposure to form long chains, called cross-linking process, so that the exposure parts of polymer become insoluble in developer. The development curves, also called *contrast curves*, are obtained by exposure and development of a series of large square patterns with different exposure dose (dose metrics) (see Chapter 9, Fig. 9.15). The remaining resist thickness is measured against the exposure dose, which generates the curves shown in Fig. 5.9 (see also Chapter 9, Fig. 9.10). For positive resist, there is an exposure dose at which the resist layer is fully developed to bottom. This exposure dose is defined as the *clearing dose* (see Fig. 5.10). The dose to clear is normally quoted as the sensitivity of the resist. The smaller this threshold dose, the higher the resist sensitivity.

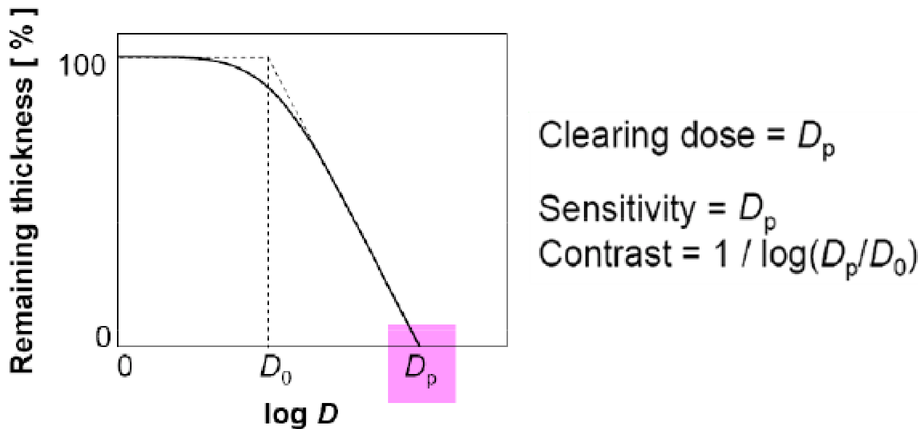


Fig. 5.10. Clearing dose, sensitivity and contrast definitions.

In Fig. 5.9 (a), resist A is more sensitive than resist B. For negative resist, the development curve looks like the plot C in Fig. 5.9 (b). With the increase of exposure dose, the degree of cross-linking increases and more resist thickness remains after development, until the whole resist layer is completely cross-linked. The sensitivity of negative resist is defined by the exposure dose which results in more than half the resist thickness remaining due to cross-linking.

One must be aware that the above definition of resist sensitivity is only for comparison purpose. It is defined by exposure of much larger pattern area, which is rarely representative in actual e-beam lithography. The actual exposure dose needed for a particular pattern exposure can be quite different. It depends on the following factors:

- *Electron energy*. Higher electron energy requires higher exposure dose, resulting in low resist sensitivity
- *Resist thickness*. The thinner the resist, the lower the resist sensitivity, because fewer electrons deposit their energy within the thin resist layer.
- *Pattern density*. Exposure of dense patterns requires less exposure dose, because electron scattering from neighboring exposure areas contribute to the overall energy deposition. Similarly, isolated patterns require high exposure dose, because a portion of electrons have scattered outside the pattern area.
- *Substrate material.* High-density substrate materials result in more backscattering of electrons into resist layer, contributing to the resist exposure. Therefore, resist coated on a high-density material appears more sensitive than on a low-density material.
- *Process conditions.* The post-exposure baking conditions, in the case of chemically amplified resists, and the strength of developer and the temperature of developer can all have influence on the resist sensitivity.

Typical sensitive parameters are resist age, developer strength, development environment (temperature and humidity), development method (dip versus puddle versus spray), and post-bake and post-exposure latency time and environment (for example, atmosphere versus vacuum).

5.2.6 Contrast of resist materials

In addition to sensitivity, the performance of a resist is also defined by its contrast. The resist contrast is defined by the slope of a development curve and is expressed in Eq. (5.5). D_p is the exposure dose at which the resist is fully exposed and D_0 is the exposure dose at which the resist starts to be exposed. They are marked in Fig. 5.9 (a) and in Fig. 5.10. The steeper the slope, the higher the resist contrast. For example, the resist A in Fig. 5.9 (b) has higher contrast than resist B. Resist contrast is an inherent property of a resist material. However, development conditions can have some influence on it.

$$\gamma = \frac{1}{log_{10}D_p - log_{10}D_0} \tag{5.5}$$

Apart from measuring the slope of resist contrast curves, another direct way of judging the resist contrast is to observe the resist profile after development. An example to compare the effect of resist contrast on resist profile is shown in Fig. 5.11, where computer-simulated resist profiles are shown for 50-keV e-beam exposure of 0.5-μm thick resist. Figure 5.11 (a) is the profile of low-contrast resist while Fig. 5.11 (b) is the profile of high-contrast resist. Although electron scattering and proximity effect are the same in both cases, high-contrast resist can produce the most desirable profile.

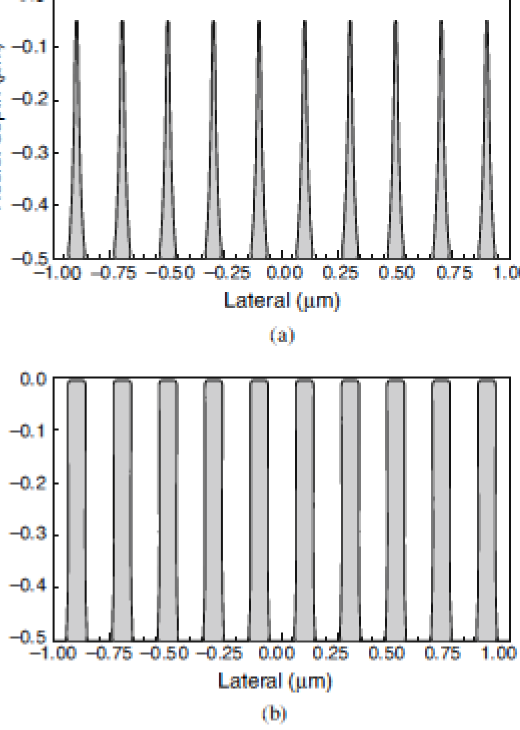


Fig. 5.11. (a) Low-contrast resist profile (b) High-contrast resist profile.

The contrast of a given resist is strongly related to the choice of developer chemicals, concentration, and temperature. Resist contrast has profound effect on lithography resolution and resist pattern usability. High contrast is always desirable because it enables resist profiles after development to have vertical sidewalls and high aspect ratios (feature height to feature width), which brings two benefits: facilitating pattern transfer by reactive-ion etching and enabling lithography of high-density patterns.

Since high resolution is easier to achieve for sparse or isolated patterns than for dense patterns with small half-pitch, one should be aware that whenever talking about the ultimate resolution of lithography, specifying at what pattern density is essential. This has been the reason why IC manufacturing industry defines the resolution in half-pitch, not in minimum feature size. In e-beam lithography, such a limit in exposure of high-density patterns is caused by the proximity effect. Proximity effect correction does not work effectively for high-density periodic features. In this case, high-contrast resists can partly offset the proximity effect, making high-density patterning possible. The other route for high-density patterning is to use high beam energy or thin resist layer (see Chapter 4 for details).

5.2.7 Resolution enhancement processes

The resist sensitivity and contrast are partly dependent on the resist material property, also partly dependent on the resist process conditions. The process conditions are also tied to pattern density and exposure condition, such as the beam energy and exposure dose. Therefore, to achieve the highest resolution for a particular resist the optimum process conditions should be investigated. In some cases, exotic methods were developed to enable very high resolution of e-beam lithography. These exotic methods include ultrasonic-assisted resist development and development at lower or higher temperature.

Ultrasonic-assisted development

The enhancement effect by ultrasonic-assisted development on e-beam lithography resolution with PMMA resist was discovered in 1993 [44]. The development is realized by short-chain polymer molecules (after chain scission action by electron exposure) dissolved in a developer solution. When the exposure region becomes very small (<10 nm), the molecules, which for PMMA have a size of around 2 nm, can be trapped in solid PMMA due to intermolecular forces which increase rapidly within the radius of 10 nm. More electron interaction with the polymer is needed to release the trapped polymer molecules, which means higher exposure dose. The exposure region then becomes wider at the increased exposure dose. With ultrasonic-assisted development, the ultrasonic agitation can provide additional energy and help to loosen up the trapped molecules to dissolve into the developer solution. The ultrasonic agitation can indeed reduce the exposure dose significantly at line feature width below 20 nm, as shown in Fig. 5.12.

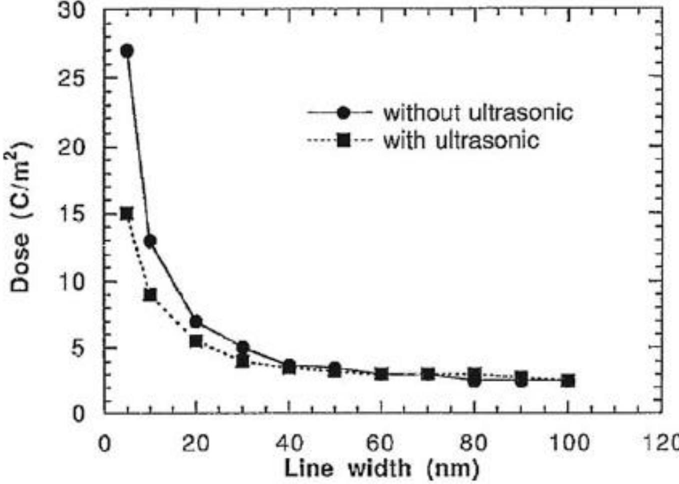


Fig. 5.12. Comparison of exposure dose required for PMMA development with and without ultrasonic assistance. [44]

It becomes no different to conventional development process for wider exposure region because the intermolecular force is no longer dominant, at above 10-nm line opening. However, it must be pointed out that the resolution enhancement by ultrasonic agitation only works for PMMA resist with high molecular weight, which demonstrates that resist development process is very much dependent on individual resist materials.

Cold development

Another method to enhance resolution is to develop the exposed PMMA in a cold developer. Sub-10 nm lines were achieved at 30-keV beam energy with MIBK+IPA(1:3) developer which was maintained at 4–8°C temperature [45]. Unlike ultrasonic agitation, which promotes the activity of PMMA molecules, cold developer actually suppresses the activity of PMMA molecules. In order to develop the exposed PMMA resist at low temperature, exposure dose much higher than development in room temperature is needed. High exposure dose is known to result in wider exposure region, as indicated by electron scattering analysis. However, energy level deposited in resist is rapidly reduced at the location away from the center of exposure region, as shown by the PSF for very thin resist layer in Fig. 4.4. In a

cold developer, those PMMA molecules exposed by low-level electron energy cannot be dissolved because of their low level of kinetic activity at low temperature. Only the very central region where the deposited electron energy is the highest can be developed, resulting in high-resolution exposure features. The similar resolution enhancement at low-temperature development was also found in e-beam lithography of ZEP 520 resist. With the developer temperature as low as –17°C, 13-nm lines in ZEP-520 exposed at 30-keV beam energy have been obtained [46]. A particular benefit of low-temperature development for ZEP-520 is the much improved line edge roughness (LER), as demonstrated in Fig. 5.13.

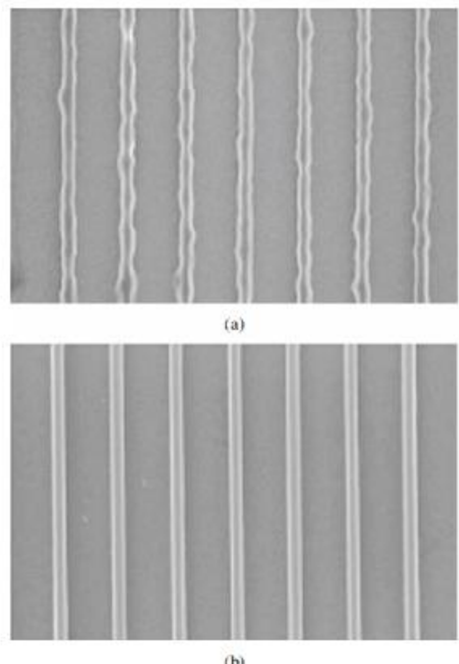


Fig. 5.13. Comparison of edge roughness of ZEP-520 resist lines (40-nm wide) developed at (a) room temperature and (b) at –4°C temperature. [46]

The ZEP-520 resist lines in Fig. 5.13 (a) were developed at room temperature, while the lines in Fig. 5.13 (b) were developed at –4°C. It clearly shows the ZEP-520 resist molecules have much lower kinetics in low-temperature developer solution. The reason that cold development works for both PMMA and ZEP-520 resists is believed to be due to the variation of molecule weight distribution before and after electron exposure. Upon electron exposure, chain scissions occur, resulting in short chains and low molecular weight molecules in the exposed region. Low-temperature developer reduces the kinetic energy of molecules. Only those molecules with very short chain can be dissolved by the developer, which effectively suppress the influence of scattering electrons because the polymer molecules exposed by low-energy scattering electrons cannot participate in the dissolving process. The same mechanism can explain why low-strength developer can help to achieve high-resolution resist patterns. Table 5.3 shows qualitatively the relation between developer concentration and resolution.

Developer concentration MIBK:IPA)	Sensitivity	Resolution
1:3	Low	Extremely high
1:2	Medium	Very high
1:1	High	High
Pure MIBK	Very high	Low

Tab. 5.3. Influence of developer concentration on resist resolution.

While low-temperature development helps to enhance resolution of PMMA and ZEP-520, it works the opposite for HSQ resist. The contrast of HSQ resist is actually enhanced at high development temperature.

5.2.8 Multiple-layer resist strategies

Multiple-layer resist strategies allow the combination of materials with different properties to be used to simultaneously achieve multiple effects, such as high resolution and planarization, high resolution and undercut, or combinations of low and high resolution for special cross sections. The strategy for all multiple-layer combinations involves the selection of materials with compatible exposure, chemistry, and process requirements. In all cases, each coated layer is baked before successive coatings to minimize intermixing. Consideration should be given to the bake temperature sequence, which ideally will progress from the highest to the lowest temperature for successive layers.

The relative sensitivity differential for PMMA of differing molecular weights can be used to achieve a natural undercut suitable for liftoff. When P(MMA-MAA) is used in a bilayer system underneath PMMA, the differential sensitivities realized with a single-step development lead to a natural undercut. This undercut is larger than that using PMMA of differing molecular weights, and is also suitable for liftoff. The fact that P(MMA-MAA) is nearly insoluble in common nonpolar solvents for PMMA makes it ideally suited for multiple-layer applications using selective developers: still larger undercut is possible using either PMMA over P(MMA-MAA) or P(MMA-MAA) over PMMA with separate selective solvents (see Fig. 5.14) [47].

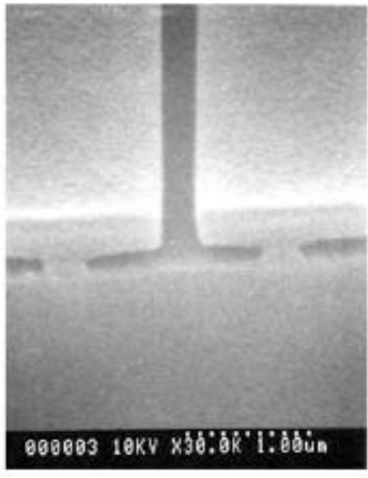


Fig. 5.14. P(MMA-MAA) over PMMA, developed selectively using ethyl-cellosolve acetate:ethanol followed by chlorobenzene. [47]

5.2.9 Pattern transfer

The resist pattern is ultimately used as a template for pattern transfer. Pattern transfer techniques are categorized as additive or subtractive. In an additive transfer, material is added to the substrate through the resist openings. In a subtractive transfer, the resist protects the underlying material while material in the resist openings is removed. In either case, the resist film is removed from the substrate following pattern transfer. Examples of additive processes are ion-implantation and liftoff.

Ion implantation requires a resist thickness sufficient to prevent penetration of the given ion species and energy into the substrate.

The *liftoff* process is shown schematically in Fig. 5.16. First, material is deposited on the patterned substrate, condensing on the resist top surface and on the substrate surface where there are resist openings. Following deposition, the resist is dissolved in a solvent, and the excess metal on top of the resist is washed away, or lifted off, leaving only the deposited material on the substrate in the patterned openings. Liftoff requires an undercut resist profile to prevent deposition on the sidewall, which can lead to ragged edges, tearing of material from the substrate, or total failure of liftoff due to complete encapsulation of the resist. Liftoff works best for deposition that is normal to the surface, such as evaporation and is more difficult for angular deposition such as with sputtering.

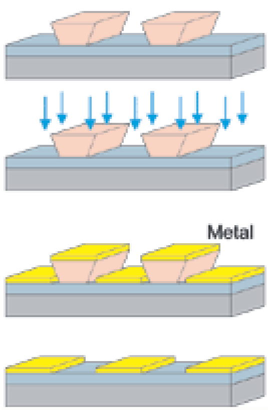


Fig. 5.16. Scheme of the liftoff process.

Another example of additive process is *electroplating*, where electrolytic metal growth takes place through the resist openings. Electroplating requires a conductive substrate and a resist thickness sufficient to prevent overgrowing on the top of the resist.

Examples of subtractive processes are wet etching and plasma etching, otherwise referred to as dry etching.

In *wet etching*, liquid solutions etch away material in the resist openings. This process places great demands on resist-to-substrate adhesion. Any adhesion failure here can lead to significant feature size variation or even total resist delamination. Removing contaminants thoroughly and possibly removing native oxides immediately before resist coating can greatly improve adhesion, depending on the substrate.

Dry etching uses energetic gas atoms to etch via physical sputtering alone or sputtering combined with chemical reaction. The primary resist concern with dry etching is selectivity, the ratio of material etch rates for resist and substrate. Low selectivity implies rapid erosion of resist during etching, leading to loss of feature size control and rough edges. Resist sidewall profiles should be vertical for isotropic dry etching to maintain critical dimension control as the top surface of the resist is eroded. The bias introduced by pattern transfer may be a resolution-limiting factor, as in the case of a narrow opening in a positive resist that is subject to undercut etching. In other cases, the bias introduced by the process may extend the resolution, as in the case of a bar of negative resist that is undercut etched to define an etched feature smaller than the resist feature.

ELECTROPLATING

Introduction

A metallic coating can only be deposited onto a substrate from an aqueous solution of a metal salt if there are sufficient electrons available, from whatever source, to neutralise the metal ions in solution, allowing the metal itself to form in the zero-valent state. In practice, there are two main sources of such electrons, each forms the basis of a technologically important process. The first of these is the so-called electroless or chemical deposition, which does not involve any external source of voltage or electron source. The second is the electrolytic deposition, where electrons are supplied from an externally applied voltage. In the following treatment, the electrolytic deposition process will be more closely examined.

6.1 Electrolytic metal deposition

The fundamental principle of the electrodeposition of a metal is shown in Fig. 6.1.

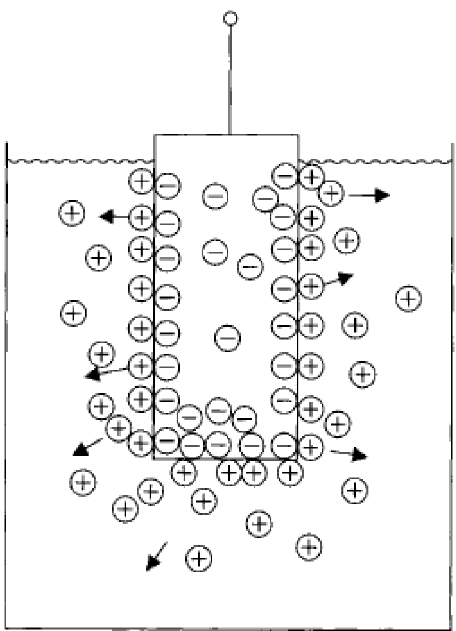


Fig. 6.1. Spontaneous establishment of a charge at a metal-solution interface for a metal immersed in aqueous solution.

From this, it is seen that when a metal is immersed in an electrolyte solution, under certain conditions, a spontaneous dissolution takes place whereby metal atoms leave the metal lattice to form positively charged ions (cations) which migrate into the electrolyte. As a result of this so-called anodic dissolution, an excess of positively charged ions is found in the immediate vicinity of the metal

electrode. Their departure leaves an equal and opposite negative charge on the metal. Electrostatic charge attraction has the opposite effect and seeks to pull the positively charged ions back to the negatively charged metal. These processes, which can be described as a dynamic equilibrium, can be represented as:

$$\mathrm{Me} \rightleftharpoons \mathrm{Me}^{z+} + z\mathrm{e}^{-} \quad (6.1)$$

The double arrow indicates that equilibrium is reached by virtue of the process going from left to right and vice versa at the same rate. The metal-solution potential difference $\Delta\varphi = \varepsilon_{\mathrm{Me/Me}^{z+}}$ has a fixed value for a given metal, immersed in a solution of standard composition, pH and temperature. The process of anodic dissolution or cathodic metal deposition implies movement of charge and thus constitutes an electric current, which is denoted as anodic or cathodic, respectively. At equilibrium, the anodic and cathodic currents are equal, and no net current then flows. However, if the metal is no longer at open circuit, but attached by means of a wire, for example, to an electron source, this equilibrium will be disturbed. More electrons will flow into the metal, which will become more negatively charged, and will attract more cations from solution. The metal is then said to be cathodic, and electrodeposition takes place. This is the basis of metal deposition or electroplating technique. Until quite recently, the process was carried out using direct current (DC).

6.1.1 Direct Current electrodeposition

DC electrolysis can be represented as in Fig. 6.2, where two electrodes, immersed in solution, are connected to the output of a DC current source [48].

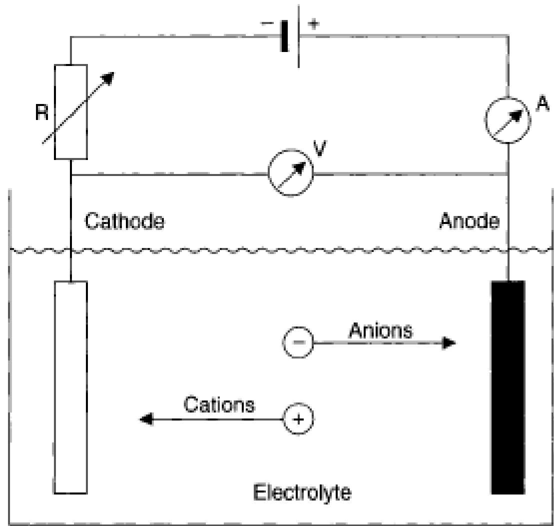


Fig. 6.2. Main components of a DC electrolysis system.

Anodes fall into two classes. Sacrificial anodes are made of the same metal that is being deposited. They will anodically dissolve, so releasing into solution metal ions to replace those which have been deposited at the cathode. The second type of anode is the so-called permanent anode, typically platinum coated titanium. Where these are used, metal ion depletion from solution is made good by

adding metal to solution in the form of a metal salt. The application of an external current source upsets the equilibrium shown in Eq. (6.1) and depending on the polarity of the connection, will tilt the equation from left to right or vice versa. The equilibrium condition shown will cease to hold, and a net electrochemical reaction will be observed. The potential driving this is known as the overpotential (η) and is defined as:

$$\eta = \left|\varepsilon - \varepsilon_{\mathrm{Me/Me^{z+}}}\right| \tag{6.2}$$

This is the energy required to overcome the reaction activation energy barrier, and the greater the overpotential, the faster will the reaction proceed. The rates of anodic metal dissolution and cathodic metal deposition can be represented by the following kinetic equations:

$$i_a = i^0 e^{(\alpha z F/RT)\eta_D} \tag{6.3}$$

$$i_c = -i^0 e^{-[(1-\alpha)zF/RT]\eta_D} \tag{6.4}$$

In these equations, α is the so-called transfer coefficient, which is usually assumed to be 0.5 but can have any value $0 < \alpha < 1$. α denotes the potential dependence of the forward charge transfer reaction, $(1-\alpha)$ denotes the potential dependence of the reverse reaction. R is the gas constant (8.135 J/mol/K), and T the absolute temperature.

6.1.1.1 Deposition of metal layers

Of critical importance in electrodeposition, is the mechanism by which metal cations are delivered to the cathode, and the means for their replenishment as they are lost to solution by deposition at the cathode. The rate at which flesh ions (and also uncharged species required for reaction) are delivered to the cathode surface from the bulk of solution, depends on the prevailing hydrodynamic conditions at and near the cathode surface. There are three main mechanisms involved in delivery of ions to the electrode surface, these being migration (under a potential gradient), diffusion (under a concentration gradient) and convection (movement of the electrolyte liquid itself).

Migration

Voltage applied across the electrodes of an electrolysis cell sets up an electrical field between anode and cathode. Assuming that the electrolytic conductivity of the electrolyte is the same at all points in solution, the potential gradient is given by the voltage across the solution divided by the distance between the electrodes. The magnitude of this potential gradient determines the rate at which ions move through solution. The term 'migration' is understood here as the movement of charged species in solution under a potential gradient. The effect operates throughout the solution, anions being electrostatically attracted to the anode, cations to the cathode. The progress of anions and cations through solution is impeded by collisions with solvent molecules and viscous drag as the ions, with their hydration sheaths, move through the liquid. The ions thereby acquire velocities which are very

low, of the order of micrometres per second. It follows that the overall contribution to the supply of ions resulting from the migration process is very small, and can generally be neglected.

Convection

In contrast to other two transport mechanisms which involve the movement of species through an electrolyte, convection can be said to be the movement of reactants, etc. with the electrolyte. The so-called 'convective mass transport' results from movement of the bulk solution. Such movement ceases to be significant in the region immediately adjacent to the electrode surface, where a liquid layer sometimes known as the 'stagnant layer' or more usually as the 'diffuse layer' is formed. Movement of ions etc across this diffuse layer takes place by diffusion (see below). Convection is important not only because it moves the solution (with dissolved species) up to the diffuse layer, but also because the thickness of the diffuse layer is determined by convective action. The stronger the agitation (pumping, stirring, air-sparging) the thinner is the diffuse layer with benefits considered below.

Diffusion

The penultimate step, before charge transfer takes place at the electrode surface, is the migration of species, both charged and uncharged, across the diffuse layer. The driving force here is the concentration gradient, more formally expressed as chemical potential. The concentration of species at the electrode surface will, under open circuit conditions, be much the same as that in bulk solution. However, once current flows, species will, by their reaction, be removed at the electrode surface and a concentration gradient will be established. The tendency of species to move from regions of high concentration to those of lower concentration, are what drive the diffusion process, and this is enshrined in Fick's laws of diffusion. The thickness of the 'diffuse layer', also known as the Nernst or Nernstian layer, is denoted by δ. Without forced convection, in a static solution, δ would be approx 0.2 mm. Under conditions of forced convection, this value will decrease, and can reach values as low as 0.001 mm. It is generally assumed that diffusion is the only significant transport mechanism operating within the Nernstian layer.

6.2 Electrodeposition electrolytes

The electrodeposition of metal coatings is usually based on aqueous electrolytes, known in the industry as electrodeposition baths or simply, baths. Their primary constituent is the metal salt of the metal to be deposited, then in most cases an acid or alkali to promote conduction. In addition, there will be additives to promote the electrodeposition process or optimise the deposit properties. In the case of electroplating, the metal deposition is brought about by current flow from an external power source.

6.2.1 Electroplating electrolytes

Over the last few decades, hundreds of different plating baths have been developed to allow or optimise the deposition of metals or alloys. A useful means of characterising such baths is in terms of their operating pH. In terms of this, one has acid, neutral and alkaline baths, where the pH values are

usually < 3, = 7 and > 9, respectively. In the following, examples of each of these three are detailed. In all three cases, it is vital to avoid significant pH changes during the electrodeposition process. For this reason, all plating electrolytes have added acid, alkali or buffers.

6.2.1.1 Acid electrolytes

Acid electrolytes are usually based on simple metal salts such as sulfates or chlorides, less commonly are phosphates or sulfamates encountered. Also present is a high concentration of the parent acid, partly to optimise electrical conductivity, partly to minimise pH changes. The metal salts will dissociate as usual, into their constituent anions and cations. A typical example is the so-called 'acid copper', typically constitutes as:

150-250 g/l copper(II) sulfate pentahydrate $CuSO_4 \cdot 5H_20$
30-75 g/l sulfuric acid (d = 1.84 g/cm^3)
30-150 mg/l chloride (Cl^- ion)

The copper salt is the source of the metal ions. From a copper sulfate solution alone, copper deposits will be mossy, coarse-crystalline and brittle. Only after the parent acid is added will a compact, fine-grained and ductile deposit be formed. In addition, by increasing electrolyte conductivity, the process can be carried out at a lower voltage, reflecting a smaller ohmic drop in solution. This type of electrolyte is usually operated in the temperature range 20-45°C and at current densities of 1-20 A/dm^2.

6.2.1.2 Neutral electrolytes

Under this heading are included systems which operate in the weakly-acid to weakly-alkaline range. Since these are poorly conducting, they are not in common use. (hydrogen cations and hydroxyl anions have far the greatest electrolytic conductivity). An important exception to this, however, are the neutral zinc plating baths, operating in the pH range 7.5-8.8.

Another example of a neutral electrolyte is that used for electrodeposition of precious metals, which incorporate added salts to increase their conductivity. So a typical composition for a neutral (pH 6.8-7.5) gold bath might be [48]:

7 g/1 gold (as the potassium dicyanoaurate(I), $K[Au(CN)_2]$)
7 g/1 copper (as potassium cyanocuprate(I) $K_2[Cu(CN)_3]$)
28 g/1 sodium dihydrogen phosphate NaH_20_4

This bath is used at 65-75°C with a deposition current density of 0.5-1 A/dm^2.

6.2.1.3 Alkaline electrolytes

This category is best sub-divided as cyanide-containing and cyanide-free types. It might be noted here that in both cases, alkaline electrolytes react slowly with carbon dioxide in air to form alkaline metal or other metal carbonates. Anything that increases electrolyte-to-air contact, such as agitation or

air-sparging, will accelerate the rate of carbonate formation. Above a certain concentration, carbonates in solution will adversely affect the metal deposition process and will have to be removed, using either a physical or chemical approach.

ELLIPSOMETRY AND SPM MICROSCOPY

7.1 Ellipsometry

Ellipsometry is a versatile and powerful technique for characterization of the optical properties of a device through the reflection of a circularly polarized monochromatic wave incident on a multilayered flat sample. Analyzing the polarization of the reflected beam, information about thickness and refractive index of each layer can be obtained. This can be considered as a non destructive, contactless, flexible and sensitive technique. The name "ellipsometry" stems from the fact that the most general state of polarization of a light wave is elliptic. It has a number of advantages compared to standard reflection intensity measurements:

- measures an intensity ratio instead of pure intensities. Therefore, it is less affected by intensity instabilities of the light source;
- both real and imaginary part of the dielectric function can be extracted without the necessity to perform a Kramers–Kronig analysis.

In the following will be described the basis of ellipsometric analysis [49].

The most general polarization of a monochromatic light wave is elliptic, where the endpoint of the electric-field vector precesses along an elliptic trajectory in a plane pependicular to the direction of propagation. The time evolution can be viewed as a superposition of two harmonic vibrations along perpendicular axes, with phase shift Δ. Assuming a wave propagation along the z-axis, the amplitude of the electric field in the xy-plane can be described in the form:

$$\mathbf{E}(t) = \begin{pmatrix} E_x(t) \\ E_y(t) \end{pmatrix} = \mathrm{Re}\left\{ \begin{pmatrix} X e^{i\Delta} \\ Y \end{pmatrix} e^{i\omega(t-t_0)} \right\} \qquad \text{(Eq. 7.1)}$$

Besides the phase shift Δ, the state of the elliptic polarization is determined by the amplitudes X and Y. More precisely, only relative amplitude X/Y is relevant in ellipsometric measurements, since multiplying both X and Y by a common constant changes merely the light intensity. Using the angle ψ defined by $tan\,\psi = X/Y$ and varying it from zero to $\pi/2$, elliptic polarization can be represented by the Jones vector:

$$\begin{pmatrix} sin\psi e^{i\Delta} \\ cos\psi \end{pmatrix} \qquad \text{(Eq. 7.2)}$$

which is determined by the two real angles ψ and Δ. Simple special cases of a general elliptic polarization are:

- Linear polarization, for $\Delta = 0$ or π
- Circular polarization, for $\psi = \pi/4$ and $\Delta = \pi/2$ (right) or $-\pi/2$ (left).

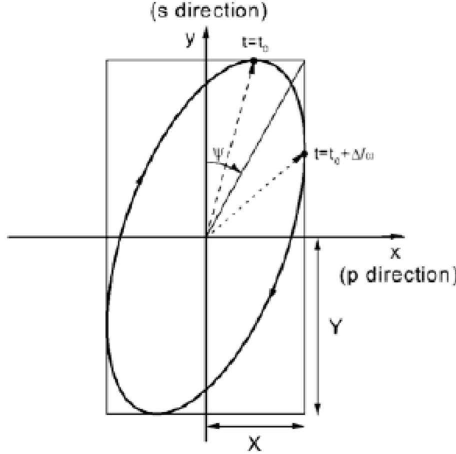


Fig. 7.1. Polarization ellipse.

Ellipsometric technique measures the change of the polarization state of the incident light upon reflection or transmission and can achieve angstrom resolution. The simplest configuration consists of a light source, a linear polarizer (P), the sample (S), a linear polarizer (called analyzer - A) and a detector (polarizer-sample-analyzer (PSA) configuration, shown in Fig. 7.2).

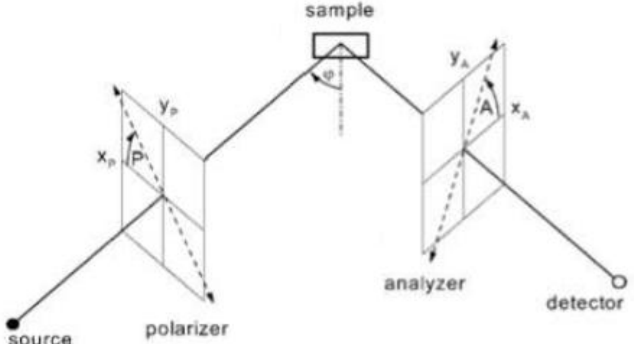


Fig. 7.2. PSA configuration.

The arm with source and polarizer prepares a known polarization state of the light incident on the sample. The arm with analyzer and detector is used to detect the change of polarization produced by the sample. A planar sample is assumed, with the angle of incidence denoted by ϕ. The beam incident on and reflected from the sample lies in the plane of incidence, which contains also the normal to the sample surface. The directions of linearly polarized light transmitted by the polarizer and the analyzer are defined by the angles P and A, respectively.

Assuming E_p the complex amplitude of the linearly polarized wave transmitted by the polarizer, the wave incident on the sample is a superposition of the component parallel and perpendicular to the plane of incidence (p and s polarizations). The reflected wave is modified by multiplying the two components by the complex reflectivities r_p and r_s. Finally, the wave transmitted by the analyzer results from the addition of incident p- and s-polarized components, projected onto the direction of the azimuth A. The field amplitude on the detector results:

$$\begin{aligned} E_A &= E_p\left(r_p \cos P \cos A + r_s \sin P \sin A\right) \\ &= E_p r_s(\rho \cos P \cos A + \sin P \sin A) \end{aligned} \quad \text{(Eq. 7.3)}$$

where ρ is the complex reflectance ratio:

$$\rho = \frac{r_p}{r_s} = \tan\psi \, e^{i\Delta} \qquad \text{(Eq. 7.4)}$$

Photometric ellipsometry is based on measurements of intensity for a number of suitably chosen settings of the optical components influencing the polarization state of light. In the PSA configuration, light intensities are measured for several properly chosen azimuths of the polarizer and analyzer. Since the analyzed state of polarization is independent of absolute intensities, one of them can be used as a reference for the measurement of relative values. Consequently, at least three independent intensities are required to determine the two real ellipsometric parameters. For fixed azimuth P, the intensity transmitted by the analyzer is

$$I(A) = I(P)|r_s|^2 \cos^2 P \,(\tan^2\psi \cos^2 A + \tan^2 P \sin^2 A + 2\tan\psi \cos\Delta \tan P \cos A \sin A) \qquad \text{(Eq. 7.5)}$$

Taking the intensity for $A = \pi/2$ as a reference, the ellipsometric angle ψ is obtained from the relative intensity measured for $A = 0$ as:

$$\tan\psi = |\tan P| \sqrt{\frac{I(0)}{I(\pi/2)}} \qquad \text{(Eq. 7.6)}$$

The third intensity can be measured for $A = \pi/4$ and provides the value of the ellipsometric angle Δ:

$$\cos\Delta = sgn(P) \frac{2I(\pi/4) - I(0) - I(\pi/2)}{2\sqrt{I(0)I(\pi/2)}} \qquad \text{(Eq. 7.7)}$$

Ellipsometry is an indirect method, i.e. in general the measured ψ and Δ cannot be converted directly into the optical constants of the sample. A model analysis must be performed. A model must be established, which considers the optical constants (refractive index or dielectric function tensor) and thickness parameters of all individual layers of the sample. The model which ψ and Δ calculated values best match the experimental data, provides the optical constants and thickness parameters of the sample.

It is possible to analyze the sample in two basics configurations: in transmission and in reflection. We have used the first to compare the experimental results with COMSOL simulations and the second to evaluate the dielectric functions and the thickness of the layers.

The ellipsometry measurements have been taken at the Veneto Nanotech research center *LaNN* (Padova) with the J.A. Woollam Co. V-VASE ellipsometer shown in Fig. 7.3.

Fig. 7.3. The J.A. Woollam Co. V-VASE ellipsometer at LaNN (Padova).

7.2 Scanning Probe Microscopy

Scanning Probe Microscopy (SPM) has enabled researchers to image surfaces at the nanometer scale. Rather than using a beam of light or electrons, SPM uses a fine probe that is scanned over a surface (or the surface is scanned under the probe). By using such a probe, researchers are no longer restrained by the wavelength of light or electrons. The resolution obtainable with this technique can resolve atoms, and true 3-D maps of surfaces are possible. Scanning Probe Microscopy is a general term, used to describe a growing number of techniques that use a sharp probe to scan over a surface and measure some property of that surface. Some examples are STM (scanning tunneling microscopy), AFM (atomic force microscopy) and NSOM (Near-Field Scanning Optical Microscopy), where the use of a tunneling current, simple force feedback or light through an aperture in the probe enables different ways to interact with the surface.

The basic idea of scanned probe techniques is illustrated in Fig. 7.4.

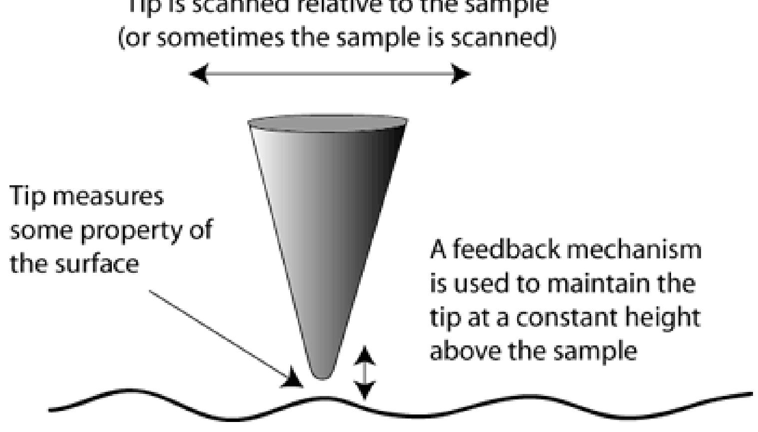


Fig. 7.4. Scheme of the SPM operation.

7.2.1 Atomic Force Microscopy

Atomic Force Microscopy (AFM) is used for surface imaging at the nanometer scale. The definition of "surface" in the microscopic world is not as trivial as in the macroscopic world. A practical definition of "surface" that can be used for AFM is: "*the locus of points traced by a "point-probe" while keeping constant the probe-sample interaction"*. The most commonly used force sensor in AFM is a micro-cantilever, with a small tip (5 nm of radius) attached to the free end, and a thin light beam, produced by a laser diode, which is reflected by the cantilever end and detected by a split photo diode (see Fig.7.5).

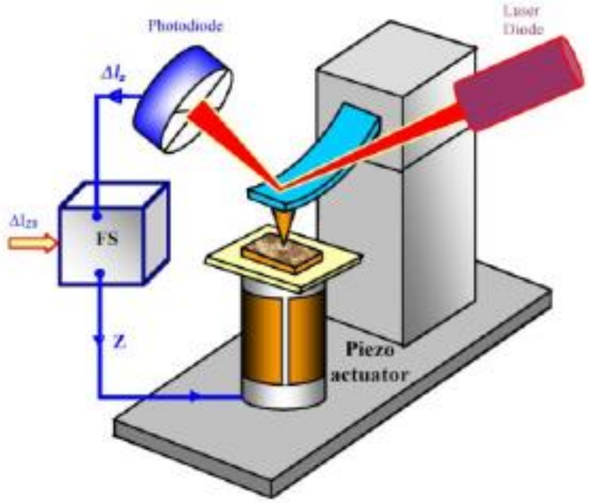


Fig. 7.5. Diagram of a AFM system.

The interactive force between the atoms of the tip and of the sample deflects the cantilever and deviates the reflected laser beam. The different illumination of the photodiode sectors produces a differential photocurrent signal that measures the interactive force.
Depending on the distance of the tip from the surface, we can say to operate in "contact" or "non contact" mode (see Fig.7.6).

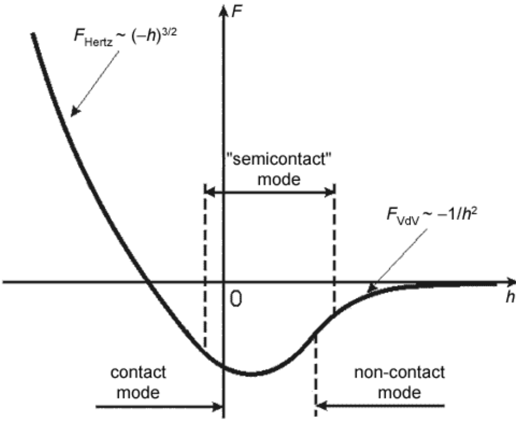


Fig. 7.6. Tip-sample interaction: short- and long-range forces. [50]

In “contact mode” we exploit the steepness of the F(z) curve slope in the repulsive region of the tip-sample interactive force (at short distances prevails the “hard sphere” repulsion due to the Pauli exclusion principle, that forbids the overlap of the electron wave functions of the outer atomic orbitals).

In “non-contact mode” (some nm from the sample surface) the Van der Waals polarization prevails and the force becomes attractive, with a much smaller gradient.

In the acquisition in “contact mode” at constant force, the operating mode used for taking our images, the tip attached to the cantilever is dragged across the surface to be probed, using the photodiode output to control (with a feedback system) the z value, thus keeping the output signal always at a constant value (i.e. keeping constant the interaction force).

The mapping of the interactive force may be obtained by sweeping in a raster scanning the sample area, and by recording in a bidimensional matrix, for each position x_iy_j of the tip, the voltage values $V_z(x_iy_j)$ fed to the scanner to keep constant the force: this matrix is a real topographic image of the scanned area.

In an AFM image, besides the lateral resolution (defined by the tip radius, ~50Å) and the vertical resolution (defined by signal/noise ratio and feedback gain), also the mapping resolution must be accounted for.

The AFM measurements have been taken at the Veneto Nanotech research center *NanoFAB (Parco Scientifico e Tecnologico VEGA, Porto Marghera* (VE)).

7.2.2 Scanning Near-field Optical Microscopy

It has been assumed for many years that imaging could only occur with classical propagating-mode solutions to Maxwell’s equations. However, imaging with the so-called “nonpropagating” exponential (evanescent) modes is also possible. This rather unconventional technique is referred to as near-field optical microscopy.

According to Abbe’s theory, the resolving capability of conventional optical microscopy is ultimately limited by diffraction.Near-field scanning optical microscopy (SNOM) is a microscopic technique for nanostructure investigation that breaks the far field resolution limit by exploiting the properties of the non-propagating fields that exist only near the surface of the object. These fields carry the high frequency spatial information about the object and have intensities that drop off exponentially with the distance from the object. Because of this, the probe must be placed very close to the surface (typically a few nanometers), much closer than the wavelength of the light. This region is the "Near-Field zone" and hence the name of the technique.

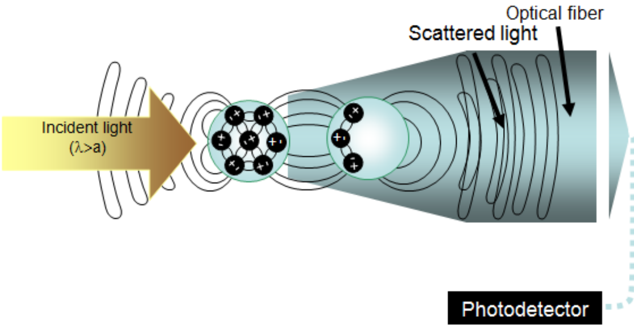


Fig. 7.7. Detection of Optical Near Field.

With this technique, the resolution of the image is limited by the size of the detector aperture and not by the wavelength of the illuminating light. SNOM offers the use of a very small light source as the imaging mechanism. By using a quasipoint light source with a diameter much smaller than the wavelength of light, one can achieve resolutions better than the diffraction limit.

Typically, laser light is fed to the aperture via an optical fiber (see Fig. 7.8). The aperture (Fig. 7.9) can be a tapered fiber coated with a metal (such as Al), a microfabricated hollow AFM probe or a tapered pipette. Normally, the size of the point light source determines the resolution obtainable.

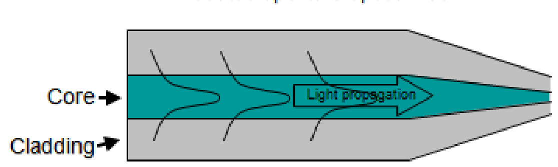


Fig. 7.8. Scheme of the aperture illumination via optical fiber.

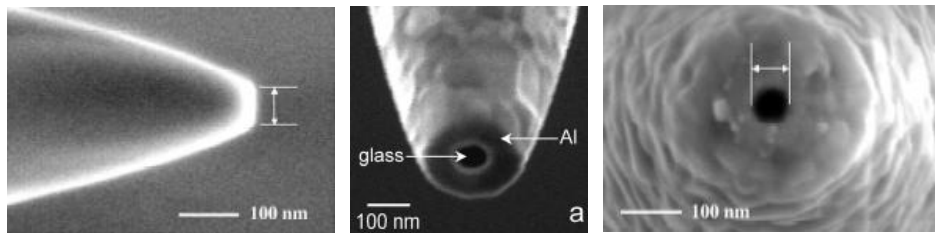


Fig. 7.9. Images of optical fiber apertures. [51]

There are two types of feedback typically used to maintain the proper working distance of the probe to the sample. One method is quite similar to how feedback works with an AFM - by using a cantilevered probe (see Fig. 7.10) the normal force is monitored, typically by using a beam-deflection setup as in most AFMs. The second method uses a tuning fork. By attaching the fiber to a tuning fork, which oscillates at its resonant frequency, we can monitor changes in the amplitude as the tip moves over the surface. The tip is moved laterally, and this techniques is normally referred to as "shear-force" feedback.

Fig. 7.10. Images of cantilever probes. [51]

Depending upon the sample being imaged, there are multiple modes of operation for SNOM:

- *Transmission*: laser light travels through the probe aperture and transmits through the sample (requires a transparent sample). This is the mode of operation we used: light comes from the back side of the sample and is collected, in transmittance, from the top of it (see Fig.7.11);

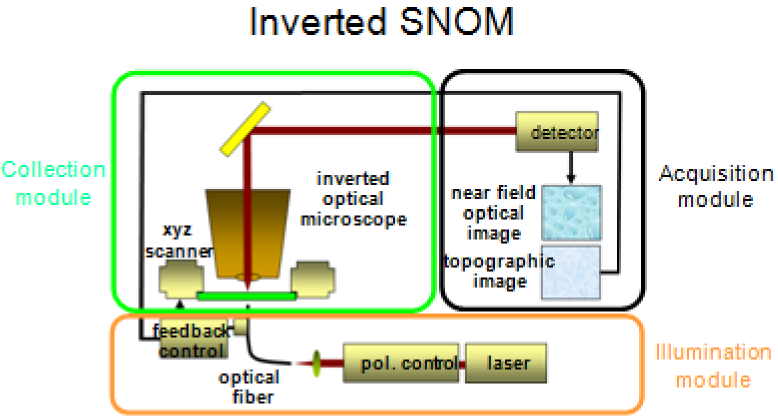


Fig. 7.11. Inverted SNOM experimental setup.

- *Reflection*: lightsource travels through the probe aperture, and reflects from the surface. Lower light intensity, and tip-dependent, but allows for opaque samples;
- *Collection*: sample is illuminated from large outside light source, and the probe collects the reflected light;
- *Illumination/Collection*: the probe both illuminates the sample and collects the reflected light.

Detection of the signal can be handled in a number of different ways: Spectrometer, APD (Avalanche Photo Diode), Photomultiplier Tube, or CCD.

All the SNOM measurements have been taken at the Veneto Nanotech research center *NanoFAB (Parco Scientifico e Tecnologico VEGA, Porto Marghera* (VE)).

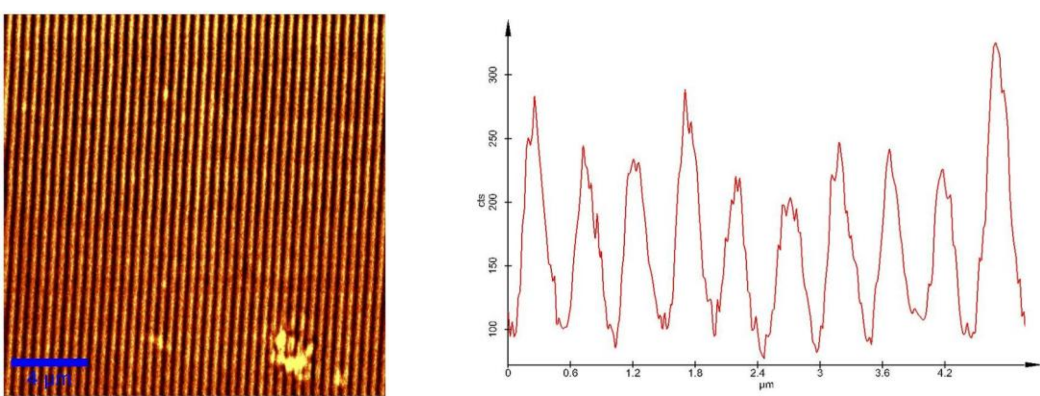


Fig. 7.12. Left: example of SNOM topographic image of a 1D gold grating. Right: SNOM light intensity profile.

DESIGN

8.1 Design of the nanodevice

Theoretical modeling provides a very powerful method for optimizing the design of the devices, i.e. tailoring their geometrical parameters to maximize optical response and efficiency.

8.1.1 The Finite Element Method

The Finite Element Method (FEM) is a numerical technique for finding approximate solutions of partial differential equations (PDE), as well as of integral equations, with proper boundary conditions. The software used for the FEM simulations is COMSOL Multiphysics, specifically the application-specific RF Module, which is a package expressly dedicated to the solutions of Maxwell's equations, and allows a variety of applications to RF devices, optics, photonics and plasmonics.

Several steps are required in order to obtain a FEM simulation. The starting point is to to define the *application mode,* i.e. the proper set of equations describing the problem. The successive step is to create a mesh, i.e. a partition of the geometry into small units of simple shapes, called mesh elements. The introduction of a mesh defines a discretization of the geometry, allowing approximations to the dependent variables that define the problem. The idea is to approximate the main functions of the problem with functions that can be described by a finite numbers of parameters, or elements. Inserting this approximations into the PDE that governs the problem, generates a set of ordinary differential equations which are then numerically integrated using standard techniques such as Euler's Method, Runge-Kutta, Lagrange's Method and others.

Numerical simulations play an essential role in the understanding of the phenomena that are object of the research, giving a quantitative description of the physics governing the process.

In the following, an explicit description of FEM simulations applied to the EOT problem is provided.

8.1.2 FEM simulations

8.1.2.1 Definition of the geometry

The first step in a FEM simulation is to define the geometry. In the present case, we are interested in studying the transmission process of light through a one-dimensional array of metallic slits, like the one represented in Fig. 8.1. By choosing simple geometries (e.g. 1D gratings), with a high degree of symmetry, it is possible to provide a better insight in the field distribution on the metal surfaces with shorter computational time.

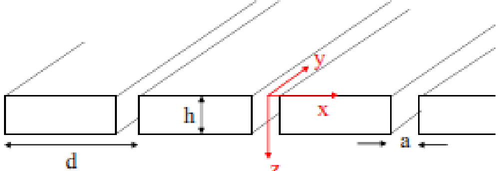


Fig. 8.1. Schematic of a 1D periodic array of slits (period d, width of the slit a) perforated on a metallic film of thickness h.

The geometrical parameters that govern the problem are: the period d of the grating in the x-direction the width a of the slits the thickness h of the grating. The grating is considered to be infinite periodic in the x-direction and the slits are considered to be infinite long. This kind of geometry allows us to restrict the problem to a one-dimensional problem, where the cross section of a single slits is represented, as shown in Fig. 8.2. For simplicity, the dielectric region above, inside and below the slits is assumed to be filled with air, with dielectric constant equal to unity. The dielectric constants of the various layers were taken from ellipsometric data, obtained directly from the samples used for the experiments.

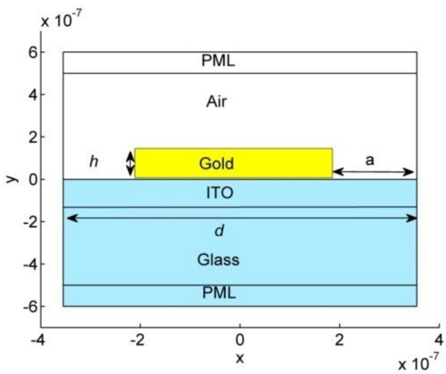


Fig. 8.2. Scheme of the simulated structure

8.1.2.2 Domain and boundary settings

After defining the geometry, the physics of the problem must be defined. As we are interested in plasmonic resonances, the incident radiation is assumed to be TM polarized, i.e. the magnetic field is directed along the major axis of the slits, and normal incidence is assumed. The PDE to be solved is then the wave equation for TM modes (see Chapter 1, Eq. 1-10c):

$$\frac{\partial^2 H_y}{\partial z^2} + \left({k_0}^2 \varepsilon - \beta^2\right) H_y = 0 \tag{8.1}$$

Since the system is periodic, on the outer edges of the grating periodic boundary conditions are imposed, while on the transmission region a scattering condition for the EM field is assumed.

8.1.2.3 Mesh

Once all geometrical and physical parameters are settled, a mesh must be generated. The mesh generation is an automatic process handled by the software, which chooses according to the geometry

of the problem where to make a more refined mesh. In the specific case, the mesh becomes particularly accurate on the edges and in the region inside the slits.

8.1.2.4 Solution

The EM field configuration shown in Fig. 8.3 for TM polarization is the result of the combination of a planar SPP on the top of the grating and SPP cavity modes excited on the vertical walls of the slits. The combination of these modes is responsible for the EOT. The SPP propagates along the vertical walls of the slits and further spreads at the bottom horizontal surface, generating a SPP mode on the other side of the grating, which further couples to the far field, generating the EOT contribution.

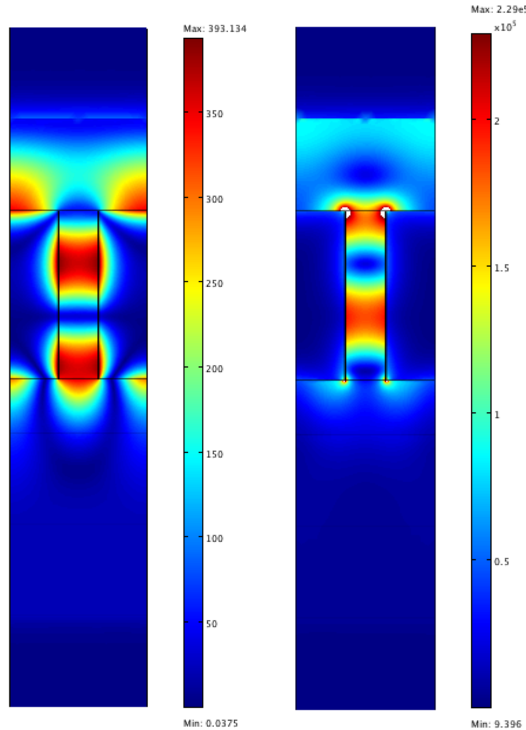


Fig. 8.3. Graphical representation of: left) norm of the magnetic field for TM polarization. Coupling of SPP on both interfaces and waveguide mode inside the slits, at the resonant wavelength, leading thus to EOT; right) norm of the electric field.

In Fig. 8.4 the EM field configuration corresponds again to Fabry–Perot resonances excited by plasmonic effects, though in this case the presence of two "hot spots" well localized inside the slits depicts an optimization in the excitation of the cavity modes. This is just the configuration we were looking for to maximize the enhancement of the electric field for our SERS experiments.

Simulations allow to determine the geometrical parameters of the grating so that the above configuration correspond to a wavelength excitation of λ = 633 nm, equivalent to the wavelength of the laser used in our SERS measurements. We will see in Chapter 12 how these conditions correspond to the possible following choice of parameters for the grating: period d=300 nm, slit width a=120nm and metal thickness h= 350 nm.

The calculation of the experimental SSEF has been carried out integrating $|E^4|$ on the gold surface of the grating, i.e. the upper surface and the walls of the slits, which corresponds to the surface which will be functionalized with Benzenethiol to measure the SERS spectra (see Chapter 10).

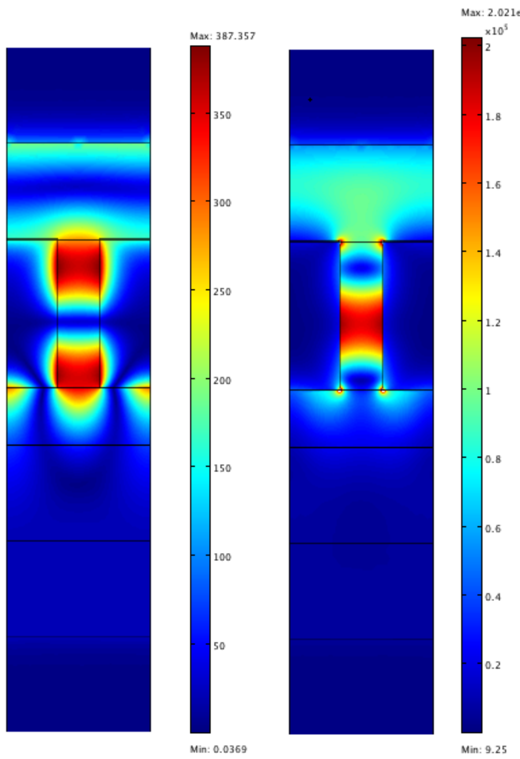


Fig. 8.4. Graphical representation of: left) norm of the magnetic field; right) norm of the electric field for TM polarization. The presence of two "hot spots" inside the slit represents an optimized configuration for SERS analysis.

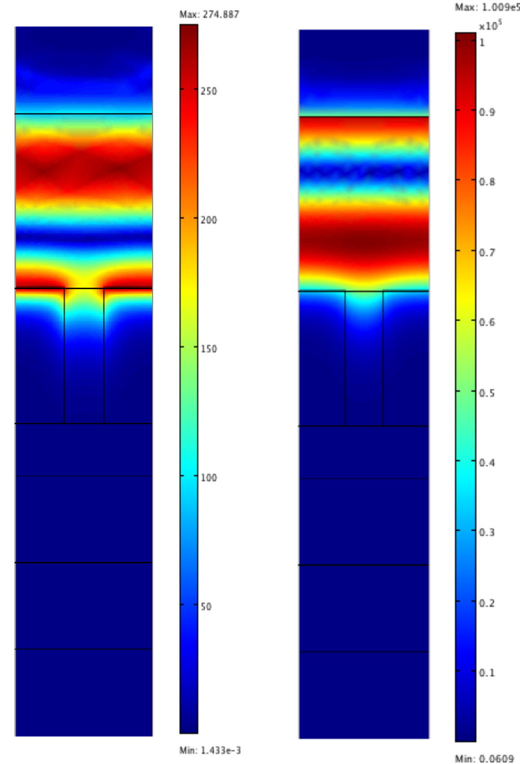


Fig. 8.5. Graphical representation of: left) norm of the magnetic field; right) norm of the electric field for TE polarization. As it is apparent, the EM field is forbidden to penetrate inside the slits.

We see in Fig. 8.4 (right) that in correspondence of the vertices of the gold blocks the electric field shows a very strong enhancement. This is due to a spurious 'tip' effect introduced by the simulation, which is most likely not present in the fabricated sample. In the calculation of the experimental SSEF we will therefore take care of excluding these regions.

For comparison, in Fig. 8.5 is shown the EM field configuration for TE polarization. As expected, since no plasmonics phenomena can take place for TE polarization, the EM field is forbidden to penetrate inside the slits.

Due to the coupling with the far field, quantities such as reflectance, transmittance and absorbance can be evaluated in order to extrapolate useful information about the EM field configuration.

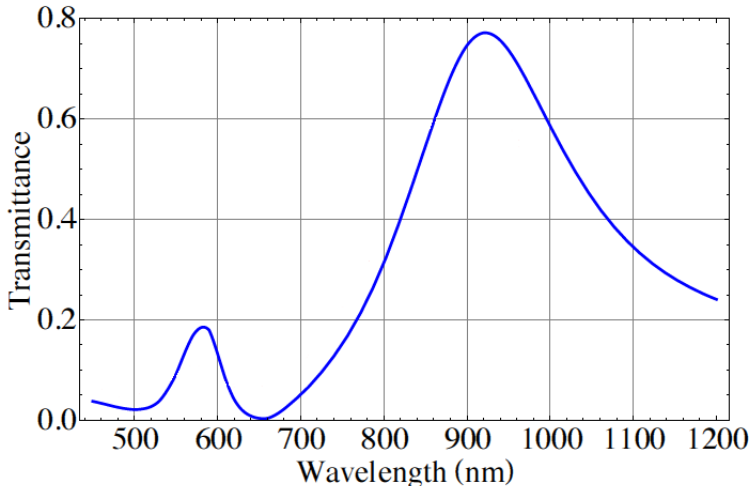


Fig. 8.6. Transmission spectrum of the simulated grating as a function of incident light wavelength, for TM polarization; The transmission peak at 570 nm is due to the hybrid mode whose configuration is shown in Fig. 8.3. The transmission dip at 633 nm correponds to the excitation of the Fabry-Perot resonances producing the 'hot spots' shown in Fig. 8.4.

The plot of the transmission spectrum for the simulated grating is shown in Fig. 8.6. The transmission spectrum is calculated dividing the time averaged energy flux (Poynting vector) emerging from the grating by the energy of the impinging radiation.

In Fig. 8.6 the transmission peak observed at 570 nm for TM polarization is due to the hybrid mode whose configuration is shown in Fig. 8.3. The transmission dip at 633 nm correponds to the excitation of the stationary waves giving rise to the 'hot spots' inside the slits shown in Fig. 8.4. Finally, around 900 nm there is a very intense cavity mode.

In addition, several simulations have been done in order to understand the effect of a variation of the geometrical parameters on the transmission spectra. Varying in turn the period of the grating (Fig. 8.7), the slit width (Fig. 8.8) and the thickness of the metal (Fig. 8.9), we can observe a variation in the shape and position of the resonance wavelength. In particular the thickness dependence can be interpreted as a sort of interference between the surface plasmons on the two sides of the gold grating: when the thickness is too low, a sort of "destructive interference" takes place between the plasmons, while as the thickness is increased, the conditions of a "constructive interference" are settled, leading to the EOT. Then, if the thickness is too great, the plasmons are mainly absorbed into the metal, and no plasmonic resonance appears.

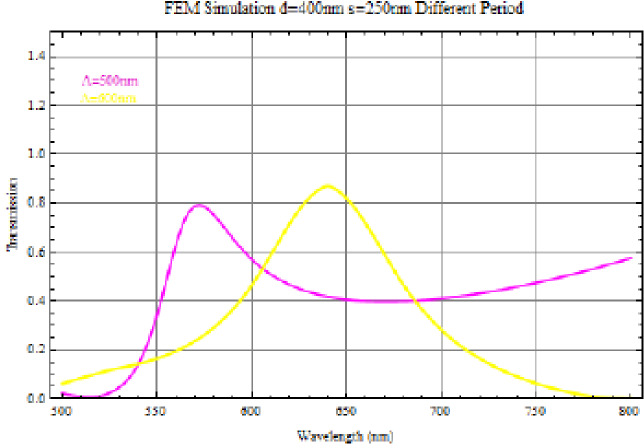


Fig. 8.7. Shift of the transmission peak due to a change in the period of the grating.

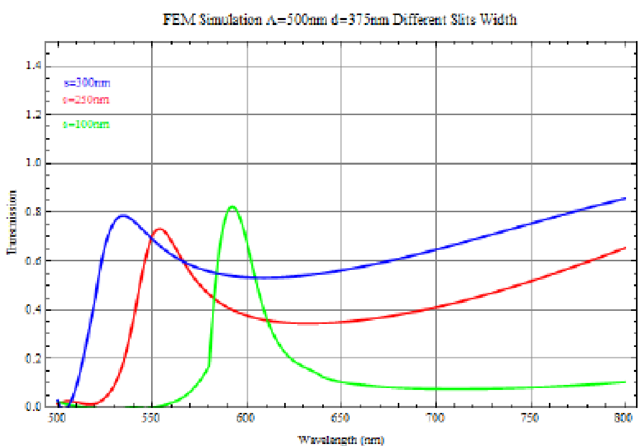


Fig. 8.8. Shift of the transmission peak due to a change in the slit width.

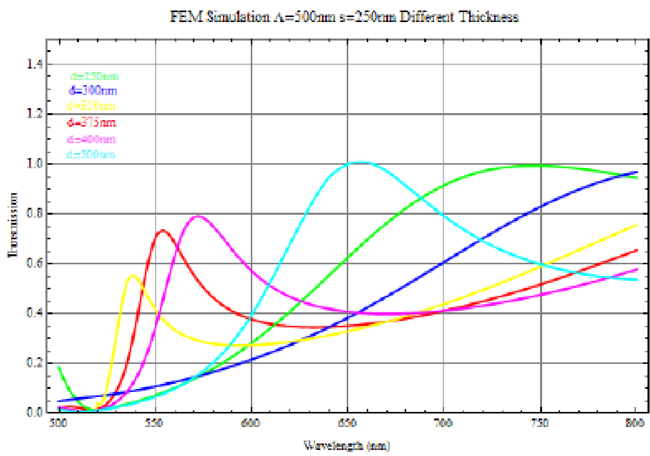


Fig. 8.9. Shift of the transmission peak due to a change in the thickness of the grating.

NANOFABRICATION

Introduction

To achieve high-resolution patterning is essential to attain a complete control on the geometry of the fabricated patterns with a resolution of few nanometers. To this aim Veneto Nanotech (VN), the hi-tech district of Veneto Region in the field of nanotechnologies applied to materials, purchased in 2009 an Electron Beam Lithography (EBL) machine (Fig. 9.1) for its new Research Center LaNN (Laboratory for the Nanofabrication of Nanodevices) located in Padova, a tool that can create structures smaller than 10 nm.

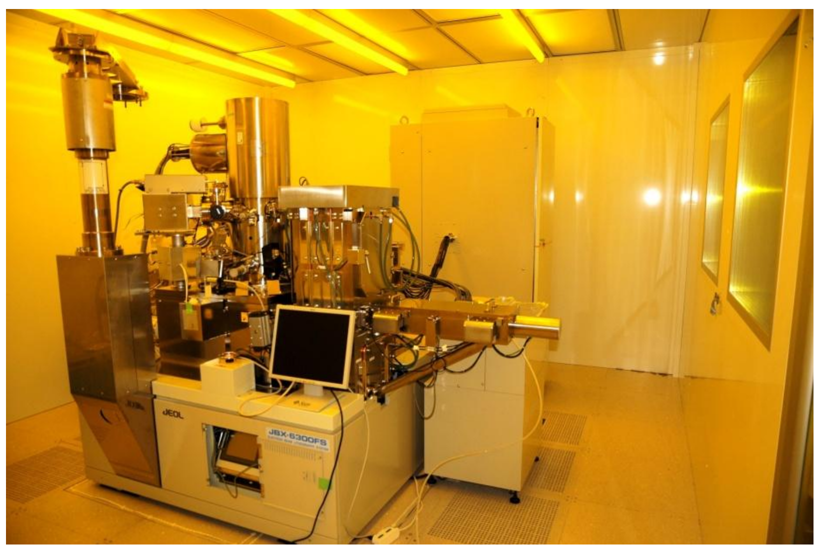


Fig. 9.1. The JEOL JBX-6300FS EBL system located at LaNN (Laboratory for Nanofabrication of Nanodevices).

Thanks to the collaboration with Veneto Nanotech, I could make use in my research work of this nanofabrication technique, allowing me to implement standardized and easily reproducible processes for the production of high resolution devices and to develop a full expertise in electron beam lithography.

9.1 The EBL system

These are the main features of the EBL JEOL JBX-6300FS system used:

- ZrO/W Thermal field-emission gun
- Spot-beam vector scanning combined with a step-and-repeat stage movement
- 4-stage electron-beam focusing system

- Accelerating voltage up to **100 kV**
- Beam scanning speeds up to **12MHz**
- Scan increments down to **0.125 nm**
- Beam diameter as low as **2 nm**
- Stage positioning resolution of λ/1024 (about **0.62 nm**)
- Deflector amplitude correction and objective-lens focus correction using the substrate height detector
- 2-stage objective lens enabling both submicron writing for small-volume production and ultra-fine processing
- Writing performance: minimum feature size at 100 kV (high-precision writing mode) **≦ 8 nm**

In order to check the writing performance of the machine, exposures were made to determine the minimum feature dimension that the EBL system can fabricate.

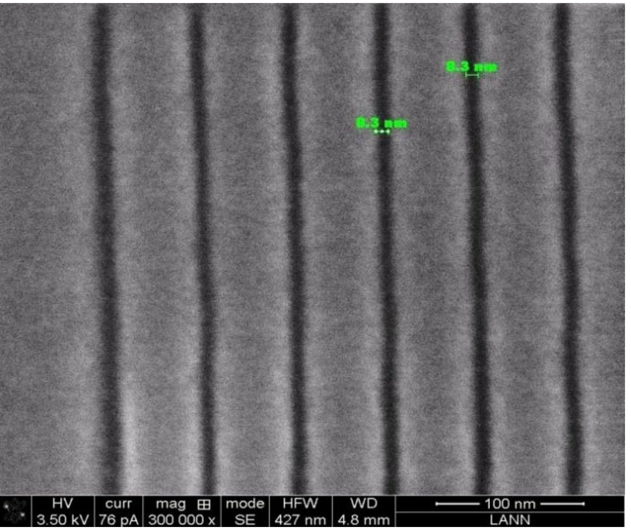


Fig. 9.2. A minimum line width of 8nm was fabricated using a 50nm thick ZEP 520A resist.

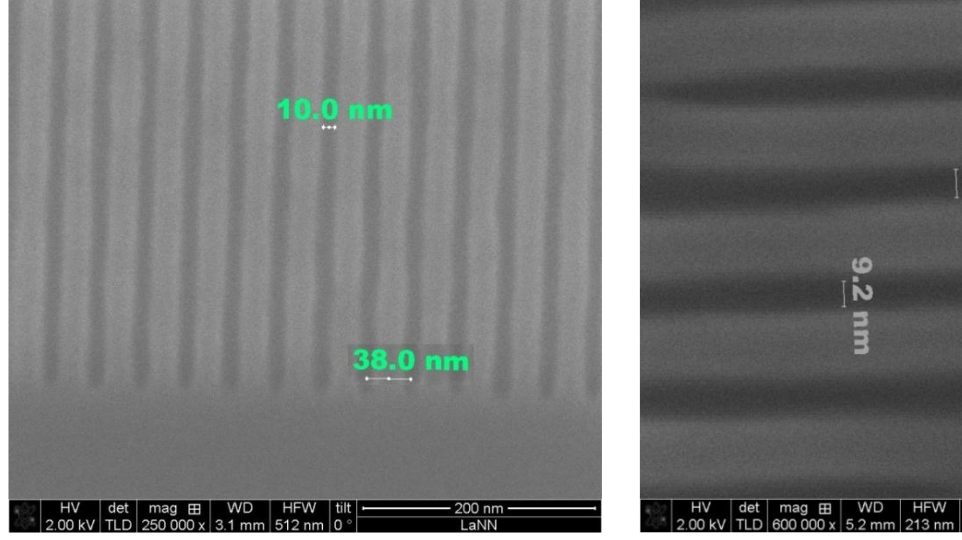


Fig. 9.3. A minimum line width of 10nm was obtained for a period as low as 38nm

As we can see in Fig. 9.2 and Fig. 9.3, using a 50nm thick ZEP 520A resist, developed in MIBK:IPA 1:1 at 7°C, we were able to fabricate slits 8nm in width (with a period of 60nm); we were moreover able to reduce the period to 38 nm, fabricating slits 10 nm in width.

In order to check the field stitching accuracy of the machine, a specific pattern is used, made up of a sort of nonius crossed in the middle by the boundary line between two fields, through which it is possible to appreciate stitching errors ≥ 8nm.

In Fig. 9.4 is shown the exposed pattern, which attests a stitching performance of the EBL machine below 8nm.

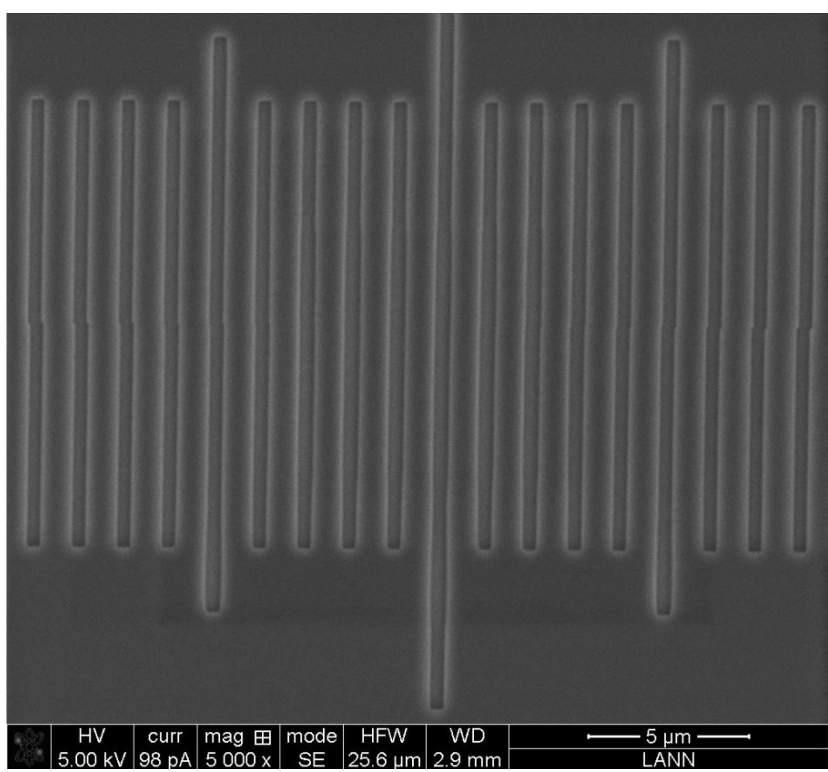


Fig. 9.4. A specific pattern was exposed to test the stitching performances of the EBL machine.

9.2 The nanofabrication process

The typical nanofabrication process we followed in our experiments is described in Fig. 9.5, and organized in the following steps:

- **1st step - Spin coating** and **soft baking**
- **2nd step** - **EBL patterning**
- **3rd step - Resist development**
- **4th step - Electrolytic growth**
- **5th step** - **Resist stripping** (+ **etching)**

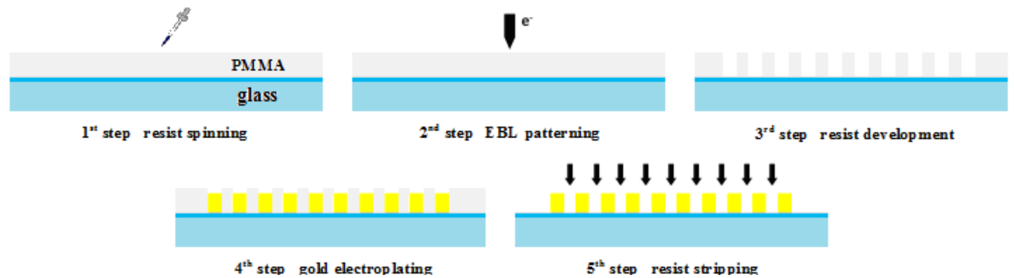


Fig.9.5. Scheme of the fabrication process.

The main concepts of the above techniques have been thoroughly described in the preceeding chapters. In the following we will go into details of the specific materials and processes used, the description of which will offer a choice to present some intermediate results of the research. In the experimental results chapters, every fabricated device will come along with its specific nanofabrication process.

9.2.1 Spin coating and soft baking

The substrate

The substrates chosen for the realization of the devices have been of two types: we started with silicon nitride membranes, very thin and transparent in the visible/near IR, but very delicate to handle, then we switched in successive works to glass substrates covered with a thin ITO layer.

In the Si_3N_4 membrane fabrication, a 2μm layer of Si_3N_4 was deposited on the back side of a silicon wafer. KOH etching of a ⟨100⟩ silicon wafer is anisotropic. KOH etches the ⟨100⟩ and ⟨110⟩ planes of single crystal silicon at a much higher rate than the ⟨111⟩ plane. This creates a V-shaped etch pit bound by the ⟨111⟩ crystal planes of silicon. This etching property of silicon has been used for the fabrication of the silicon nitride membrane (see in fig. 9.6 a schematic diagram of the fabrication process and an example of a fabricated membrane).

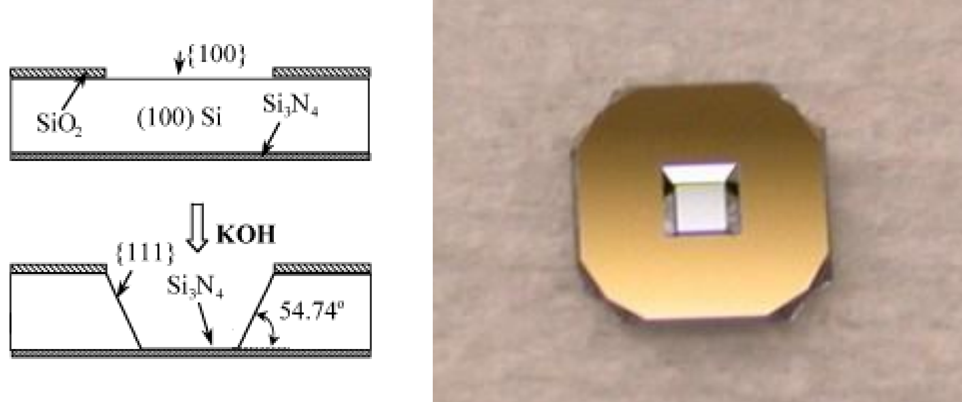


Fig. 9.6. A schematic diagram showing the microfabrication process via wet anisotropic etching of (100) single-crystal silicon for a Si_3N_4 membrane, together with an example of a fabricated membrane.

In the Si_3N_4 membranes, the optical window has been realized on a 1 mm × 1 mm, 2 µm thick Si_3N_4 self standing membrane. The evaporation of a chromium/gold bilayer (10 nm/20 nm thick) produces the base plating electrode for the successive galvanic electrolytic growth (Fig. 9.7). After growth and PMMA stripping, the Au and Cr layers have been removed respectively by an Ar dry etching and by a wet etching based on NaOH and $K_3Fe(CN)_6$ solution in H_2O.

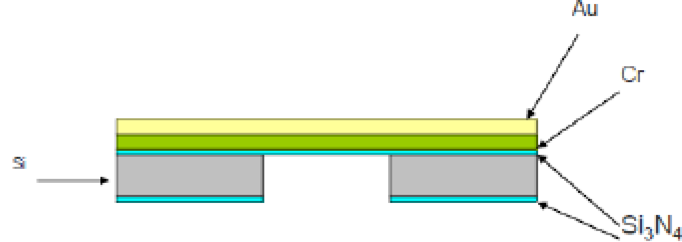


Fig. 9.7. Scheme of the membrane with the base plating electrode

In the case of the ITO/glass substrates, the slides are made with soda lime float glass with refractive index of 1.517; typical composition of the float glass is 72.6% SiO_2, 0.8% B_2O_3, 1.7% Al_2O_3, 4.6% CaO, 3.6% MgO, and 15.2% Na_2O, with a density of approximately 2.53 g/cm^3 at 20°C.
A thin layer (120-160nm) of ITO on the glass substrate acts as conductive layer, necessary both as a discharge path during exposure and for the following electrolytic growth. The surface resistivity is 8-12 Ω/sq. The transmittance of the substrate is > 83%.

A SEM image of the ITO/glass substrate is shown in Fig. 9.8.

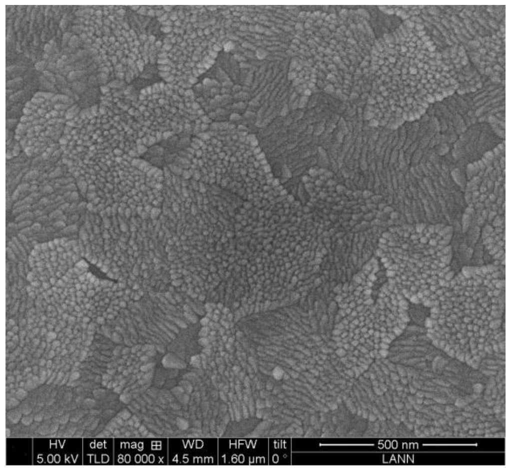


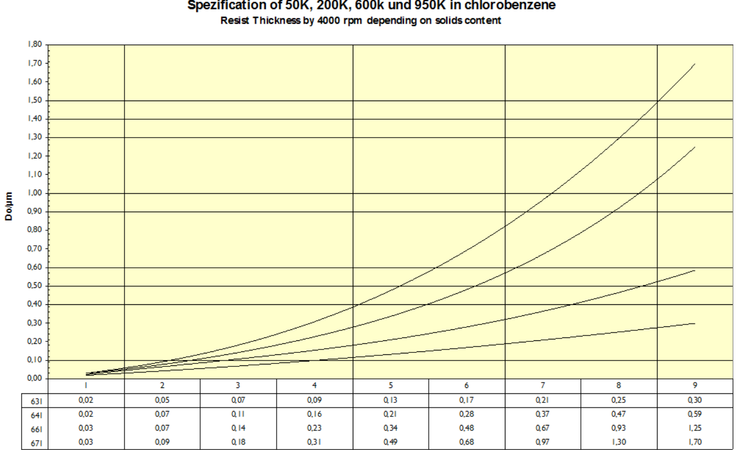
Fig. 9.8. SEM image of the ITO/glass substrate, where is clearly visible the ITO structure.

The choice of a membrane or a ITO/glass as a substrate to fabricate our devices relies on the fact we need a transparent substrate in the visible range to allow ellipsometric and SNOM transmittance measurements.

The resist

The resist used in our nanofabrication processes is the poly(meth)acrylate (PMMA) AllResist AR-P 671.05, with molecular mass 950K, viscosity 57 mPa s, solids content 5% (dissolved in

chlorobenzene). Each resist comes along with a specifications document, like the one shown in Fig.9.9, where the thickness of the resist is plotted vs the solid content, for different molecular weights and a definite coating speed.



	1	2	3	4	5	6	7	8	9
631	0,02	0,05	0,07	0,09	0,13	0,17	0,21	0,25	0,30
641	0,02	0,07	0,11	0,16	0,21	0,28	0,37	0,47	0,59
661	0,03	0,07	0,14	0,23	0,34	0,48	0,67	0,93	1,25
671	0,03	0,09	0,18	0,31	0,49	0,68	0,97	1,30	1,70

Fig. 9.9. Thickness curves of PMMA resists with different molecular weights depending on solids content (at 4000rpm). [Allresist]

A fundamental step in an electron beam lithography process is the characterization of the resist used. For this purpose the development curve, known as the contrast curve of the resist, has to be obtained by exposure and development of a series of large square patterns with different exposure doses (dose metrics). The remaining resist thickness is measured against the logarithm of the applied exposure dose. For positive resist, there will be an exposure dose at which the resist layer is fully developed to bottom. This exposure dose is defined as the *clearing dose*.

Exposing a dose metrics like the one shown in Fig. 9.15 and measuring the remaining resist thickness with a profilometer, the contrast curve generated is shown in Fig. 9.10.

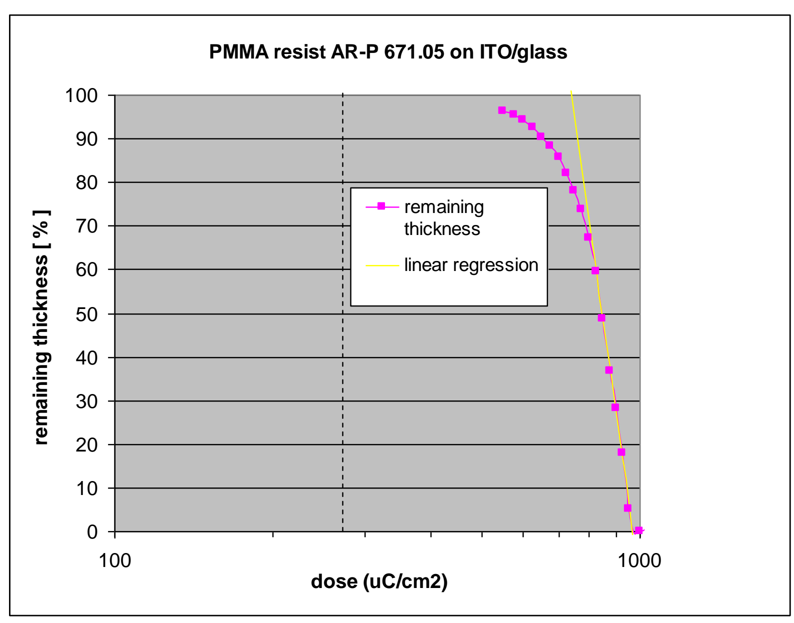


Fig. 9.10. Contrast curve of the AR-P 671.05 PMMA resist, molecular weight 950K, solids content 5% in chlorobenzene, spinned at 4000 rpm and developed in MIBK:IPA 1:3 for 30 seconds.

The analysis of the curve allows to identify a clearing dose of about 1000 μC/cm^2; the contrast of the resist has been calculated to be 5.24.

Since the contrast of a given resist is strongly related to the choice of developer chemicals, the measured contrast curve is specific of the particular developer used.

Prior to resist application, the substrate surface must be cleaned because of the inevitable contamination which occurs during storage and handling between processing steps. This is usually performed by immersion in an IPA solution and sonicate for ten minutes.

The liquid is applied to the substrate by spin coating to form a film of uniform thickness (e.g. spinning the AR-P 671.05 resist at 4000 rpm to obtain a nominal thickness of 490nm). The resist is then soft baked on a hot plate at 180°C for 10 minutes to remove solvents and stress and to promote adhesion of the resist layer to the substrate. To prevent stress arising during the cooling process, a suitable decreasing temperature ramp has been set.

Fig. 9.11. The programmable SCS G3P-8 spin coating system used to coat our samples.

9.2.2 EBL patterning

9.2.2.1 Layout editing and Proximity Effect Correction (PEC)

The next step in the fabrication process is the generation of the pattern to be exposed. This is usually accomplished by the use of CAD softwares like Layout Editor™.

To correct the pattern for the proximity effect, a commercial software has been used, GenISys Layout BEAMER™, which besides providing the tools for advanced process correction, offers a lot of complementary functions, e.g. the possibility to model the absorbed energy or showing the e-beam shot placement.

Conventional e-beam lithography processes are optimized for ‘‘binary” patterns (resist or no resist). The resist is used in a ‘‘threshold” mode, and lateral dimensions are controlled by adjusting the absorbed energies at the edge of the features to a constant level (dose to clear). The variations in the absorbed energy, due to proximity effects (scattering of electrons), must be corrected, usually using exposure dose modulation.

We are coming back here to the PEC argument introduced in Chapter 4, paragraph 4.1.2, to see its application to the case of study through the use of Layout BEAMER software.

Layout BEAMER standard PEC insures (by allocating suitable dose modulation) that the absorbed energy at features edges is uniform 0.5 (= dose to clear) at all over the layout. Thereby the development reaches all edges at the same time and positive resist is removed (negative resist will remain) in all exposed areas at absorbed energies > 0.5.

Layout BEAMER takes this 0.5 as a reference to the dose to clear. So whatever the correct results will be, the outline of the layout will receive precisely the dose to clear energy, to ensure that the resist outline on the wafer is equal to the layout design.

In Fig. 9.12 the simulated energy density of the e-beam exposure in the resist, for a non-corrected layout and a proximity effect corrected layout are compared.

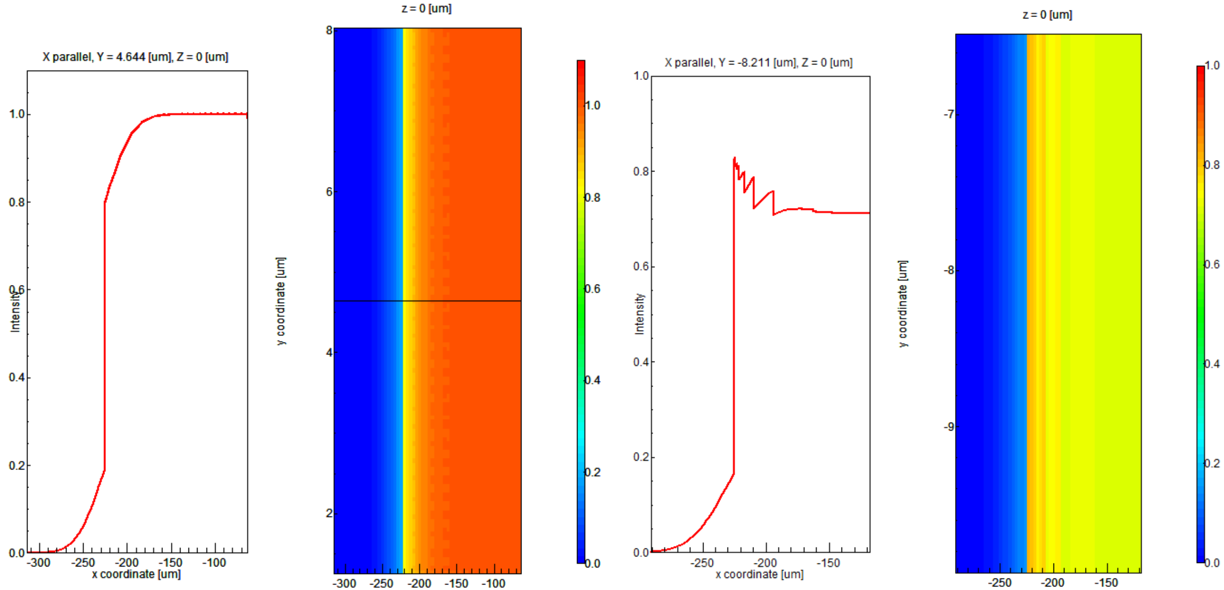


Fig. 9.12. Simulated energy density for a non-corrected (left) and a proximity effect corrected (right) layout.

The reference dose (base dose) of 1 is dependent on the resist sensitivity and needs to be experimentally calibrated for a given resist process. The base dose is independent of layout after one time calibration for a given resist process has been performed. The layout independent base dose is one of the main benefits of PEC.

When applying the proximity effect correction, choosing the appropriate correction function (the Point Spread Function (PSF)) is very important. This approach corrects the scattering effects of electrons in the material stack. The range of scattering will depend on the electron energy (acceleration voltage), resist and other layer materials and their thicknesses, including the substrate material, and it is divided into two types:

- *forward scattering / short range correction*

  The energy deposition of the forward scattering range, also called short range, is in the range of a few nanometers for 100keV machines and is designated by the letter *alpha*. In addition to the forward scattering the beam size (beam blur) needs to be considered for short range effects. The

beam blur is dependent on the beam current of the exposure and is typically in the range of a few nanometers for low beam currents (>1nA) and may become few 10nm for very high beam current. The forward scattering and beam blur can be combined to an "*effective alpha*". Depending on this effective alpha short range correction may be needed or not. For most 100keV Gaussian beam exposures short range correction will not be needed for feature >50nm.

- *back scattering / long range correction*

The deposited energy of backscattered electrons as a function of the distance from the beam can be approximated by either single or multi Gaussian function(s) with the width *beta*, *gamma1*, *gamma2*. The function strongly depends on electron energy and the stack, especially the substrate. Beta is on the order of 35µm for a 100keV, when writing on a silicon substrate. On the other hand, it is in the range of 10µm when writing on GaAs. The shape of the backscatter (long-range) PSF is simpler and can be approximated with just one Gaussian (beta) when using a 100kV machine writing on silicon. The PSF is more complex, with greater energy contributions in the mid-range (100nm-1µm), for 100keV on GaAs, requiring the approximation by multiple Gaussians or better yet by using a numerical PSF table.

There are several ways of determining the right correction functions (PSFs):

- *Monte Carlo simulation* - Monte Carlo simulation software can simulate the PSF using computational models to calculate the scattering of electrons in the material for a given stack and electron energy. The result is a table with energy as a function of distance from the beam position. This PSF function table can be used directly, or approximated with Gaussian functions.
- *Experimental methods* - A correction function can also be determined by using experimental methods (writing and measuring of test patterns) by providing a table equivalent to energy deposition as a function of distance. Such experimental methods can cover process effects, e.g., resist bake, development and etching, in addition to the electron scattering effects, and the calculated PSF is thus the most reliable.

In Fig. 9.13 is shown the PSF we have calculated by Monte Carlo simulations for 500nm PMMA on ITO/glass substrate, along with its approximation with Gaussian functions.

Fig.9.13. The PSF function calculated for the PEC correction of our samples .

In Fig. 9.14 is shown the result of the PEC correction calculated for a 500x500um^2 array of slits, with grating period of 300 nm and slit width of 120 nm, applying to the pattern the PSF shown in Fig. 9.13.

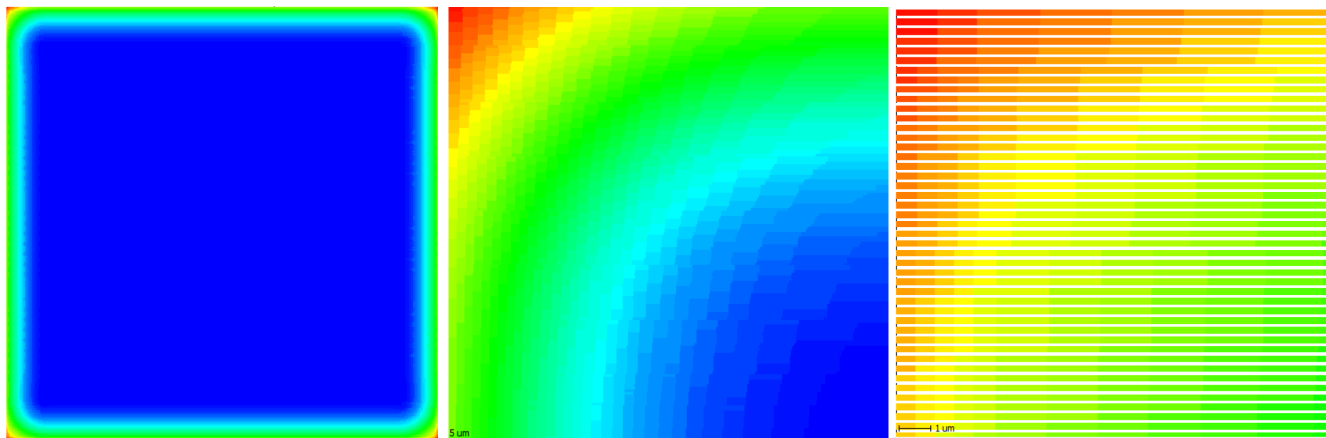


Fig. 9.14. Proximity effect compensation; detail of the dose correction at a pattern corner; further zoom (single slits and fracturing are visible).

In Fig. 9.15 is shown an optical microscopy image of a dose metrics of squares, where is apparent the effect of the proximity. Due to backscattering, at every dose the center of the square receives more energy than the borders and even more than the corners, which are in fact the last to be cleared. This is a typical example of a dose metrics used to calculate the contrast curve of a resist.

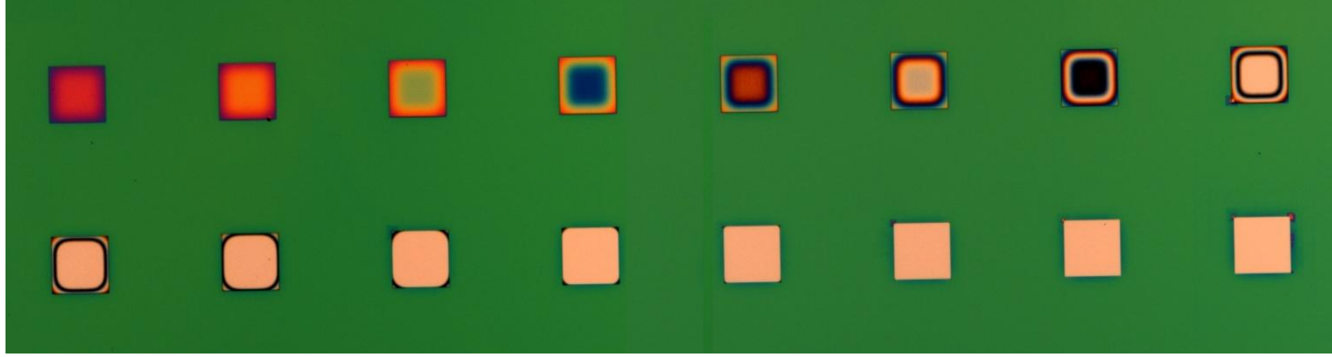


Fig. 9.15. From top left to bottom right: sequence of increasing dose squares. It is apparent the effect of the proximity.

In Fig. 9.16 we can see the comparison between a series of square patterns with increasing dose, exposed without and with proximity effect correction. In the series without PEC, measured with an optical profilometer, due to different energy densities the first square appears to be cleared in the center, but not (yet) in the corners; the second square is cleared all over, but the edges appears to bulge; this effect is even more apparent in the third square, due to the higher dose of the edges with respect to the corners. In the series with PEC it is possible to observe the fracturing generated by the PEC correction software, which allows the square to be exposed uniformly.

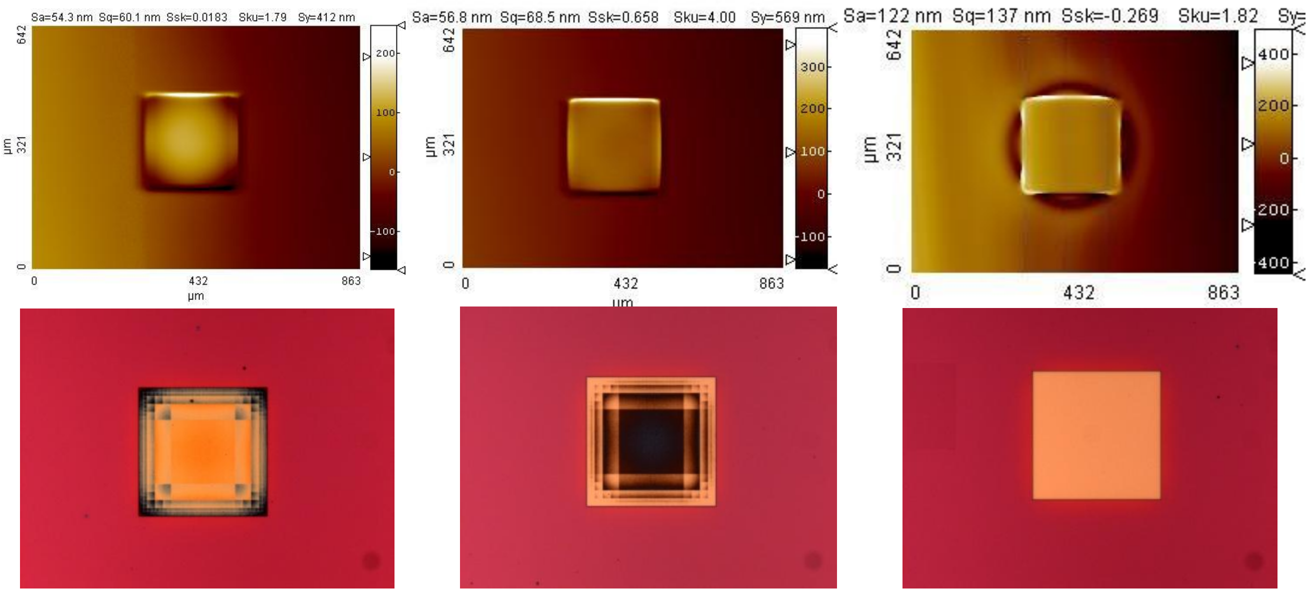


Fig. 9.16. Comparison between a dose metrics of squares exposed without (top) and with (bottom) proximity effect correction. The top series has been characterized with an optical profilometer, the bottom with an optical microscope.

In order to allow a better control of the pattern to be exposed, the Layout Beamer software offers also the possibility to show the e-beam shot placement. An example of shot simulation of a 15nm diameter beam with a shot pitch of 12 nm is shown in Fig. 9.17 for a slit of width 300nm. It can be noticed how the edges of the slit (at the top, at the left and at the bottom) are well outlined by the reticle of shots.

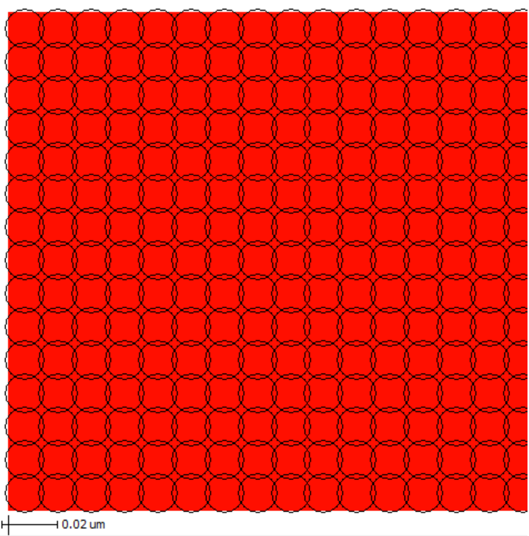


Fig. 9.17. Shot simulation of the exposure of a slit of width 300nm with a 15nm diameter beam with shot pitch of 12 nm. Note how the edges of the slit are well outlined by the array of shots.

9.2.2.2 Exposure of the pattern

The JBX-6300FS e-beam lithography system can work in a high-speed writing mode for small-volume submicron production and a nanometer-writing mode for ultra-fine processing.

In the high-speed writing mode the EBL exposure is performed typically at 100KeV with a current ≥ 2nA, a beam diameter of 6nm, a scan pitch (distance between two successive shots) of 12nm, an objective aperture of 60µm and a field size of 500x500µm^2 (EOS mode 3). In the high resolution mode the exposure is performed at 100KeV with a current of 100pA, a beam diameter of 2nm, a scan pitch of 1.5nm, an objective aperture of 25µm and a field size of 62.5x62.5µm^2 (EOS mode 6).

Dose determination

Every nanofabrication process by EBL needs a preliminary stage where the correct dose to apply to the specific pattern to expose has to be determined. A dose metrics is meant to investigate the correct dose through the exposure of an array of patterns identical in size and shape to the pattern to expose, slightly differing in dose. After the development, the best dose has to be determined through a comparative analysis of the resulting patterns.

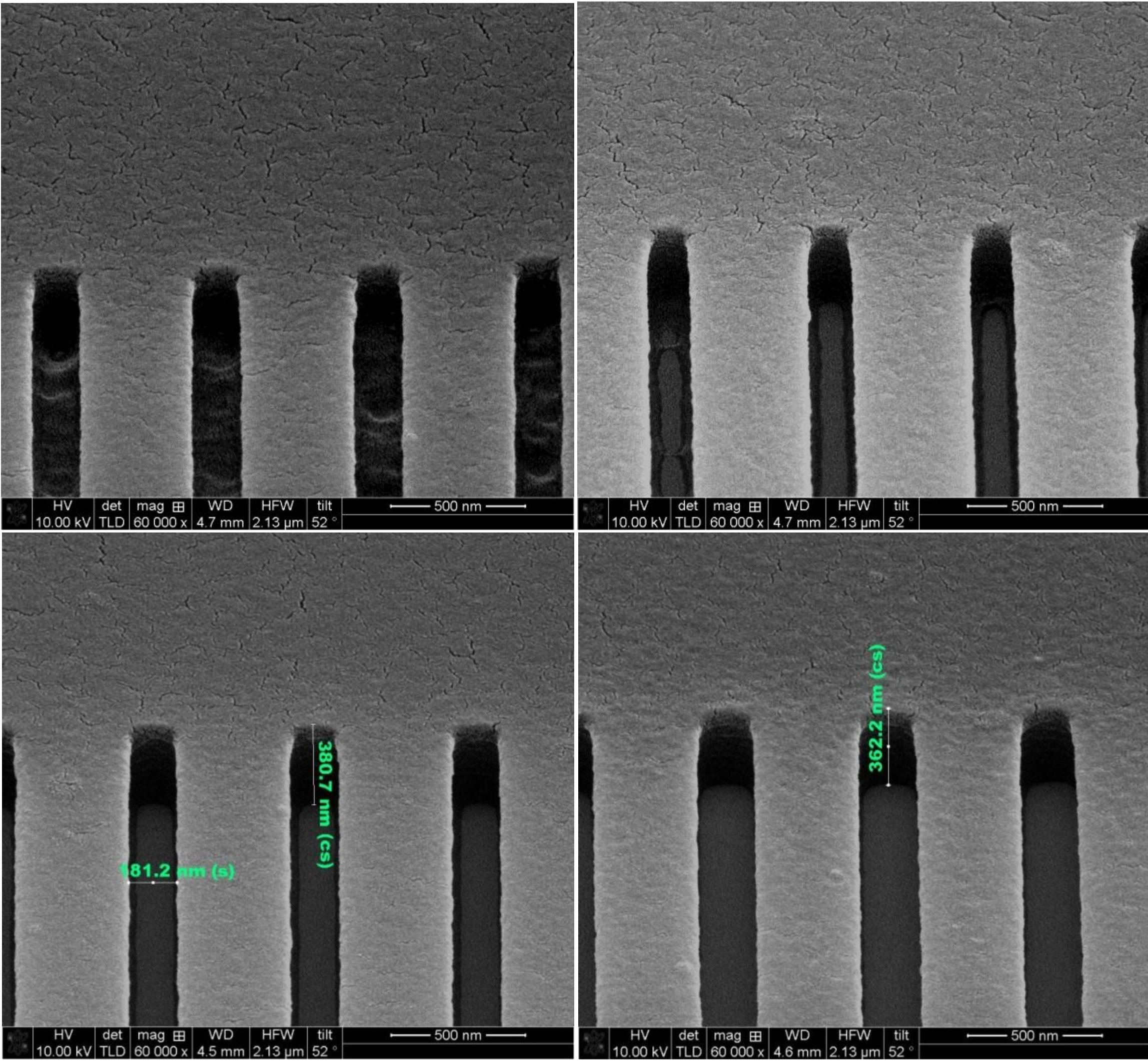


Fig. 9.18. From top left to bottom right: cross section SEM images of increasing dose patterns (slit nominal width: 180nm).

In Fig. 9.18 a series of cross section SEM images of a dose metrics for an array of slits (period 610nm, slit width 180nm) is shown, where can be observed: (a) the residual resist inside the slits in case of insufficient dose; (b) the progressive clearing of the slits at the increase of the dose; (c) the right replication of the pattern for a dose equal to the clearing dose; (d) the broadening of the slits in case of excessive dose.

To obtain the expected pattern a proximity correction must be applied to the exposure, which can be done in this case, due to the simple and highly repeated pattern, by resizing the pattern geometry. Resizing narrower (negative biasing) the slits with respect to the nominal design, we compensated the higher-than-average energy contributed from neighboring lines, thus obtaining a relatively uniform distribution of energy across the whole pattern area.

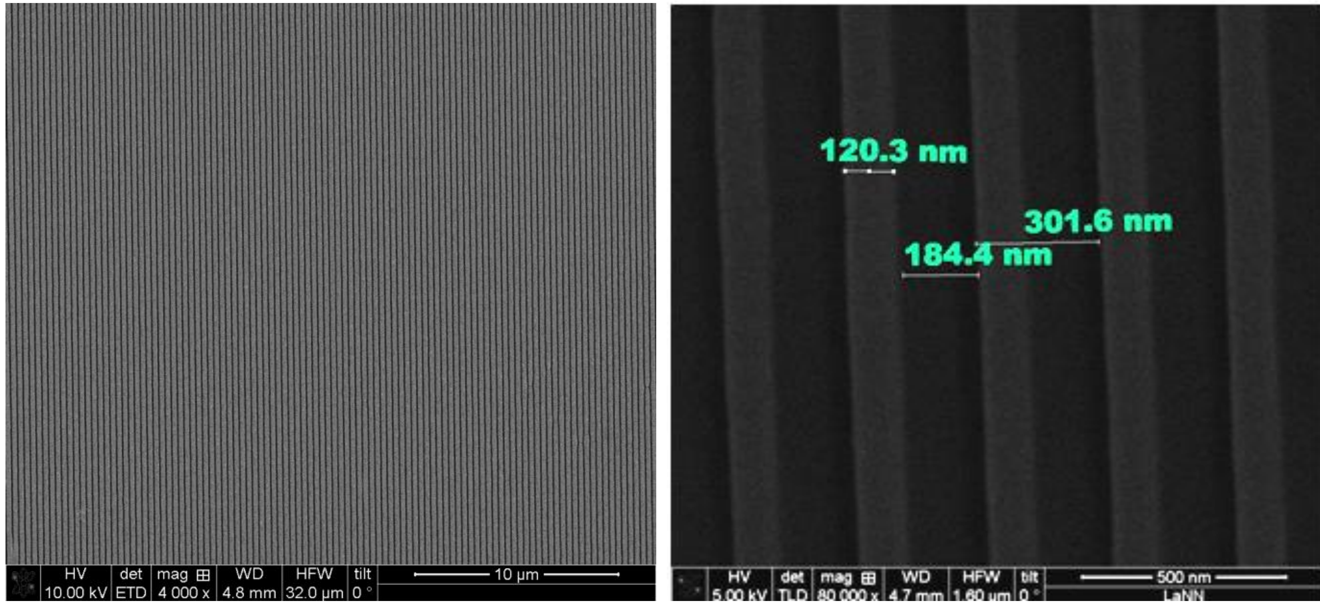


Fig. 9.19. SEM images of the optimized grating of slits (slit width: 180nm, array period 300nm)

In Fig. 9.19 two images are shown of the optimized grating pattern fabricated by EBL, where can be appreciated among other things the uniformity of the lithography, which, considered that the laser spot of the micro-Raman successively used for SERS analysis is ~1 μm^2, guarantees the reproducibility of the measures over the single chip. The overall dimension of the chips exposed is 500x500μm^2.

9.2.3 Resist development

As we have seen, a lot of different parameters concur to the final result of the lithography. Among these: the dose, the nature of the resist, the developer type, the developer composition and the development technique.

Commonly used developers for PMMA are 1:3 methyl isobutyl ketone (MIBK): isopropanol (IPA) for higher contrast and 1:1 MIBK:IPA for higher sensitivity [52]. However, recently isopropanol mixed with a small proportion of water, has been demonstrated to have superior lithographic performance [41] (see also the paragraph '*MIBK/IPA and water/IPA as PMMA developers: a thorough analysis*' in Chapter 5, within paragraph 5.1.1.1).

This has been confirmed by a paper, ‘Design, fabrication and characterization of phase masks for astronomical applications’, we published in 2011 in the ‘*Microelectronic Engineering*’ Journal on nanofabrication of 3D resist patterns [53], where we noted that the surface roughness of partially developed (in 1:3 MIBK:IPA) PMMA resist (where the resist was intentionally under-dosed in order to create the 3D features) was acceptably smooth only in the lower dose range (0 to 250 μC/cm^2); for higher doses, R_q was very high (up to hundreds of nm), due to a “phase separation effect” of the resist (see Fig. 9.20) [41,54]. Using the 3:7 deionized water:IPA developer we were able to improve the roughness R_q value of one order of magnitude (see Fig. 9.21).

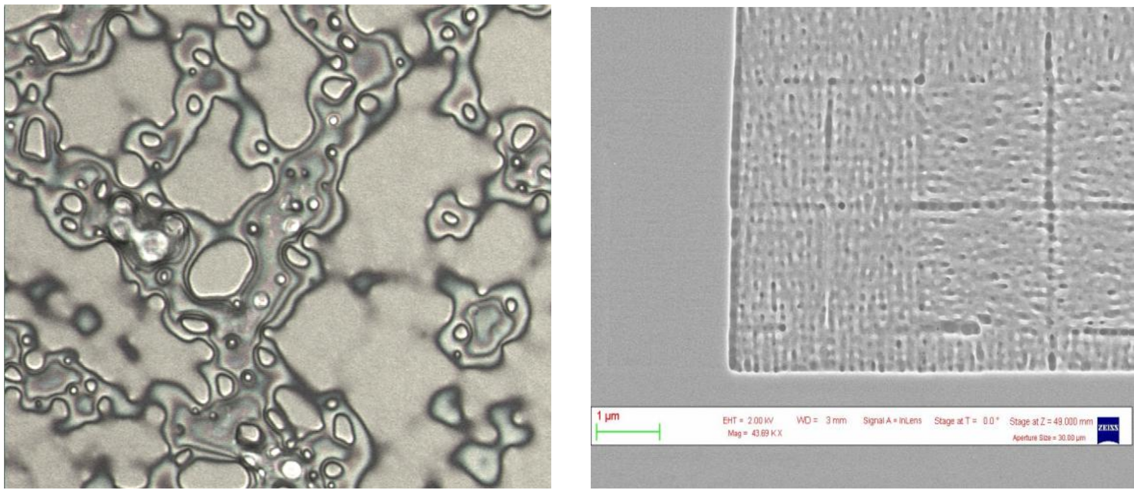


Fig. 9.20. Optical microscope (left) and SEM (right) images of the phase separation effect of PMMA partially developed in 1:3 MIBK:IPA.

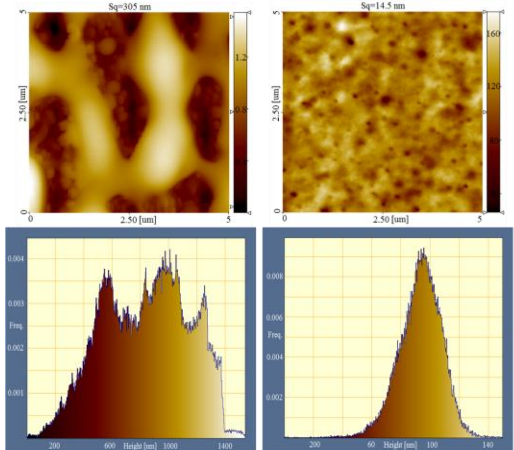


Fig. 9.21. AFM image (top) and height dispersion histogram (bottom) for PMMA partially developed in 1:3 MIBK:IPA (left) and 3:7 water:IPA (right). [53]

9.2.4 Electrolytic growth

The set-up we used for gold electroplating is a 25-liter bath solution of gold salts with two electrodes (anode and cathode) and a third electrode in between, with an active area much larger than the sample to grow. The third electrode has the function to attract the major part of the ions so that

making the electrodeposition slower and more controllable. A rotating little magnetic bar keeps in movement the solution and avoids the creation of concentration gradients.

For ITO/glass substates we used typically a voltage of 2.9V and a growth current of 100 mA DC, since previous work has demonstrated to be a good value to control the growth, with a growth rate calibrated to be from variable from 3 to 10 nm/sec. The growth was performed at IOM-CNR (Istituto per l’Officina dei Materiali) located at Basovizza (TS).

In fig. 9.22 a particular of a lithographed grating pattern is shown, before (left) and after (right) electrolytic growth.

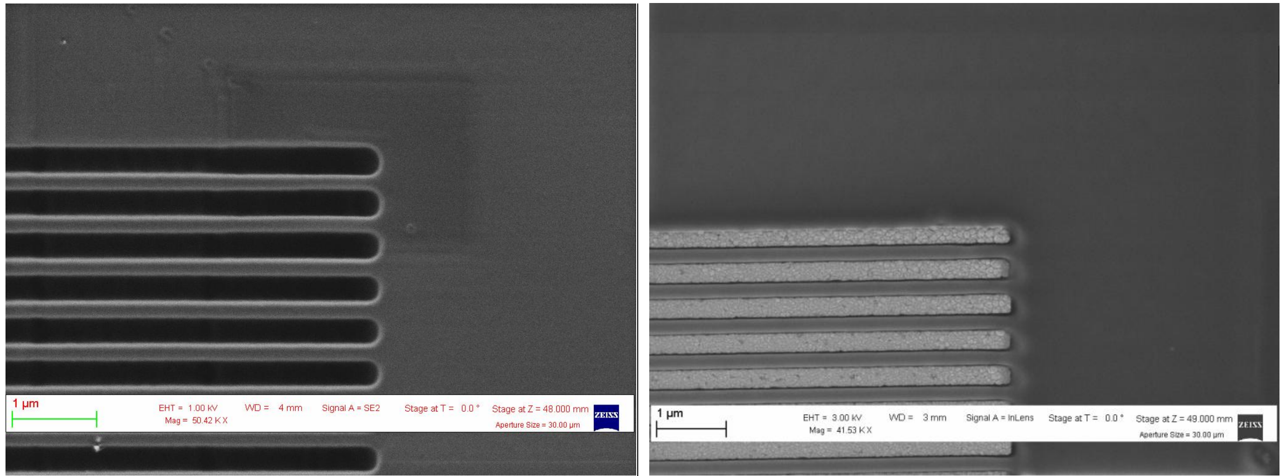


Fig. 9.22. Particular of the grating pattern before (left) and after (right) electrolytic growth.

9.2.5 Stripping

The stripping was done dipping for 3 minutes the sample into hot acetone (60°), and rinsing it in cold acetone and then in IPA.

In Fig. 9.27 is shown a grown pattern before and after stripping of the unexposed resist.

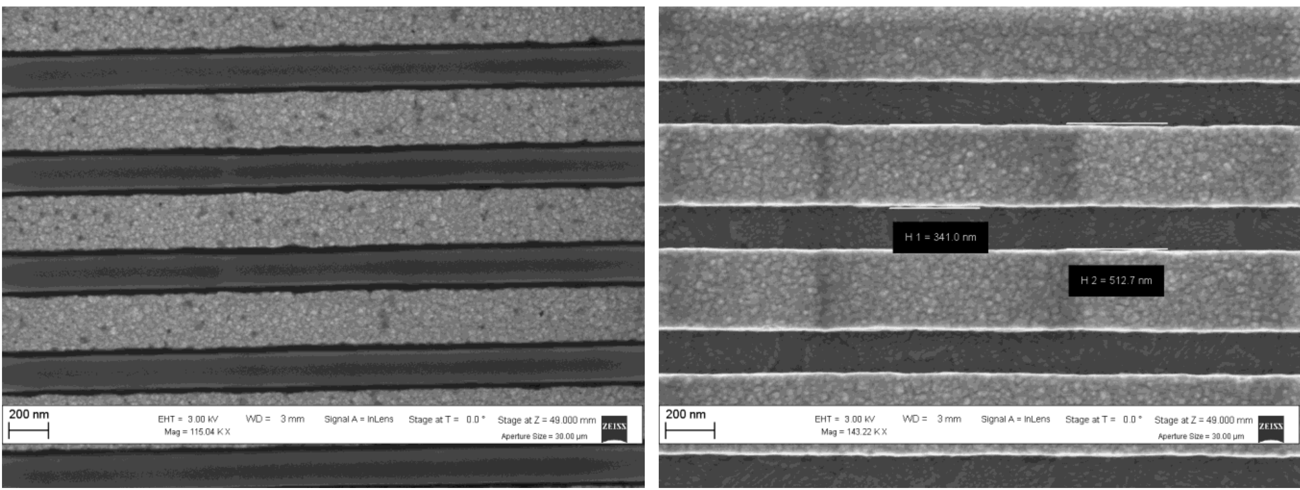


Fig. 9.27. SEM image of a grating pattern before and after the stripping of the resist.

Finally, in fig. 9.23 are shown 2D and 3D SEM images of the fabricated structures:

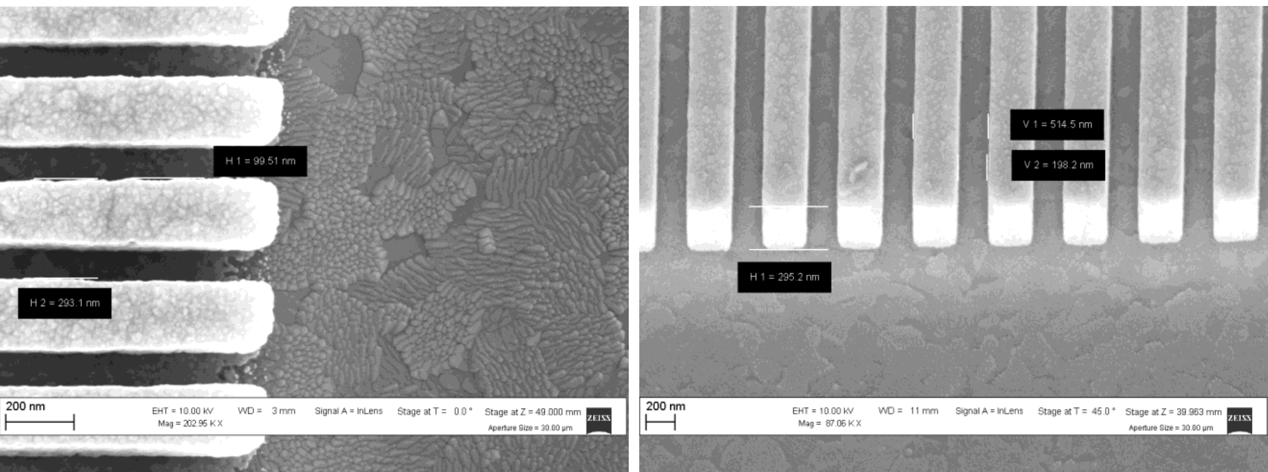


Fig. 9.23. SEM images of the nanofabricated structures.

9.2.6 Characterization of the growth

Morphology

To characterize the morphology of gold grown we analyzed the fabricated grating with the Atomic Force Microscope (AFM). In Fig. 9.24 is shown an AFM image of the grating. As can be seen, a few nm roughness is present on the surface, probably due to disuniformity in growth rate or to a replication of the ITO surface morphology. Due to the morphological variability of ITO coating, resistance will not be isotropic on the substrate, and measurable differences in growth rate can be observed based upon the orientation of the electrodes with respect to the coating.

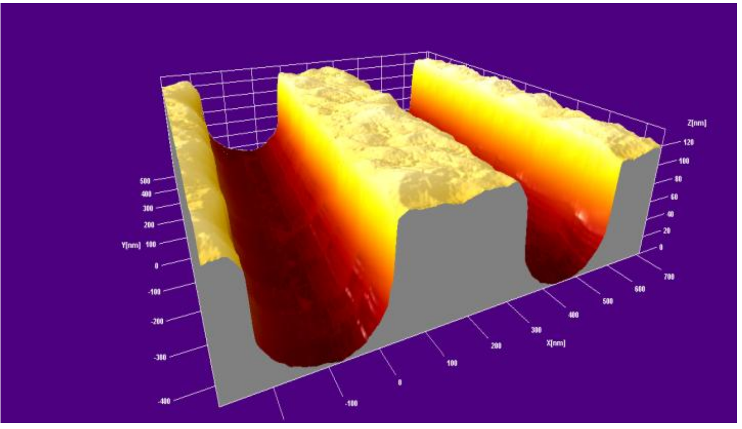


Fig. 9.24. AFM image of the surface of a nanofabricated structure.

SEM cross section

Through the use of a FIB (Focused Ion Beam) it is possible, sputtering Ga ions onto the sample, open a window inside the pattern to observe it in cross section: this allows to measure with good precision the thickness of the gold grown.

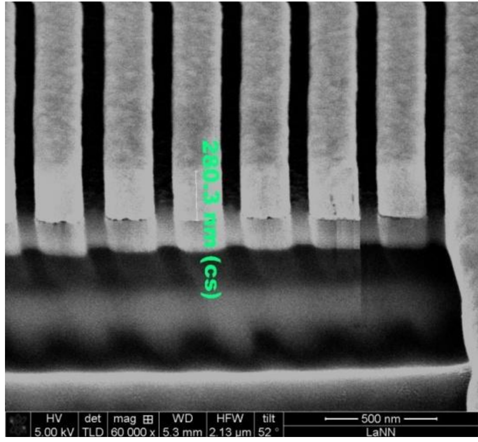


Fig. 9.25. Cross section SEM image of the grating, obtained through a FIB cut.

In Fig. 9.26 we can appreciate the uprightness of the edges of the metallic grating, due to the high energy (100KeV) used to pattern the resist (see Fig. 4.16 and paragraph "*The role of Beam Energy"* in chapter 4, within paragraph 4.1.2.1).

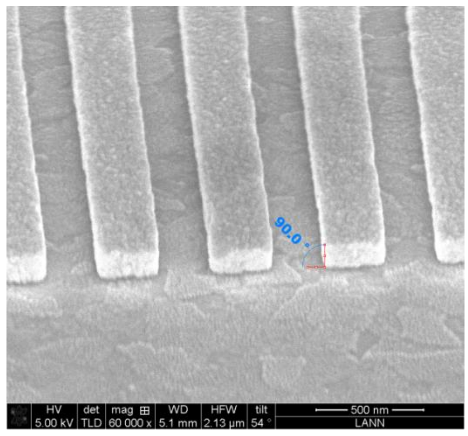


Fig. 9.26. Cross section SEM image of the grating.

SERS MEASUREMENTS

Introduction

In this chapter we will describe the functionalization of the nanofabricated device and the determination of the SERS enhancement factor through Raman measurements. We will present SERS spectra of the analyzed samples and we will discuss the way to evaluate the experimental SSEF (see Chapter 3), which comparison with theoretical estimate will be discussed in Chapter 12.

10.1 Functionalization

After the fabrication of the gold nanostructure we proceed to the functionalization of the surface with a suitable organic molecule (Benzenethiol). This molecule molecule is widely used as a test molecule since it has a large cross section, is photochemically stable and forms monolayers on metals with a known packing density. The device is functionalized by immersion (overnight) in a 10 mM solution of Benzenethiol in methanol: such concentration provides a large excess of molecules with respect to what necessary for monolayer coverage of the metal surface. The sample is then taken out, rinsed with copious amounts of methanol and dried with a nitrogen flux.

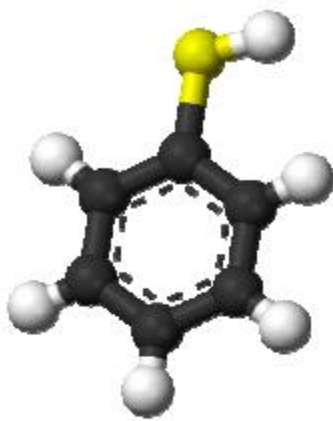


Fig. 10.1. Benzenethiol chemical structure. Black: carbon atoms, grey: hydrogen atoms, yellow: sulfur atom.

In Fig. 10.1 the atomic structure of Benzenethiol (IUPAC name thiophenol) is shown: it is an organosulfur compound with chemical formula C_6H_5SH. This foul-smelling colourless liquid is the simplest aromatic thiol. Chemical structures of thiophenols are analogous to phenols structures except the oxygen atom in the hydroxyl group (-OH) bonded to the aromatic ring is replaced by a sulfur atom. The prefix thio- implies a sulfur-containing compound and when used before a root word name of a compound which would normally contain an oxygen atom, denotes that the alcohol oxygen atom is replaced by a sulfur atom. Thiophenols have a sulfhydryl group (-SH) covalently bonded to an aromatic

ring. The dashed line inside the ring indicates the presence of delocalized bonds, typical of the benzene ring in Kekule model.

The bond between Benzenethiol and metal is realized thanks to the SH-group. The mechanism of assembling of Benzenethiol on a gold surface is quite complicated and different explanations can be found in literature [55,56,57,58,59]. The main opinion is that the bonding is reached through the loss of the hydrogen, so we can consider the sulfur atom directly bond to the metal surface. The benzenethiol is dissociatively adsorbed on metal surfaces via the S atom to form benzenethiolate (C_6H_5S). Although some cluster calculations have been reported [60,61], information about adsorption geometries is still incomplete. The bonding is generally accepted to be covalent. Thanks to the fact that benzene rings do not bond neither together (π-bond) nor with a SH, we can conclude that a monolayer of Benzenethiol is formed on the metal surface.

10.2 SERS characterization

10.2.1 Experimental set-up

Raman measurements were recorded by using a homebuilt micro-Raman setup. A Helium-Neon laser (MellesGriot) at 632.8 nm served as excitation light source. The laser beam was coupled by a beam splitter into a microscope (Olympus BX 40) and focused on the sample by a 50x objective (Olympus, NA=0.75). Backscattered Raman signal was separated from the Rayleigh scattering by a holographic Notch Filter (Kaiser Optical Systems). The Raman signal was then analyzed with a 320 mm focal length imaging spectrograph (TRIAX-320 ISA) equipped with a 1800 lines/mmdiffraction grating and a liquid nitrogen cooled CCD camera (Spectrum One, JobinYvon1024x256 pixels). Typical spot diameter was about 1μm for the 50x objective.

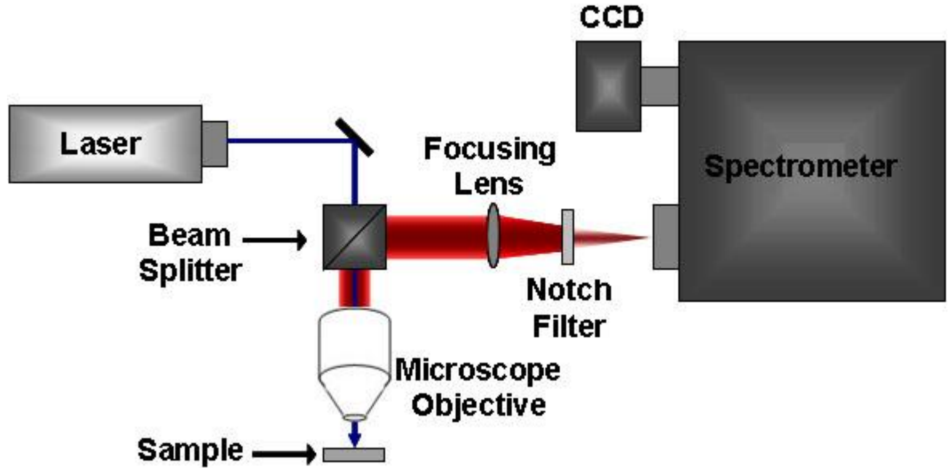


Fig. 10.2. Basic scheme of a microRaman setup.

Fig. 10.2 shows a basic sketch of the system showing the fundamental optical components. The (He-Ne) laser beam has an output power of 17 mW at 632.8 nm, which can be attenuated through a series of filters with different optical density, interchangeable via software.

10.2.2 SERS measurements varying incident light polarization

The measurements were acquired with two different integration conditions. For TM polarization (*short integration*) 50 spectra (10 seconds integration time each) were averaged together; for TE polarization (*long integration*, due to the extreme weakness of the signal for this polarization), 20 spectra (500 seconds integration time each) were averaged. The laser power was set in order to fulfill the tradeoff between integration time and possible sample decomposition due to laser exposure.

At every point measured on the sample, measurements were acquired in the following way: first a TM spectrum (*short integration*), TE spectrum (*long integration*) and again TM integration (short integration) under the same experimental conditions.

In Fig. 10.3 Raman spectra measured are shown: the first spectrum in TM polarization (black), the TE spectrum (green) and the second in TM (red); in Fig. 10.4 the blow-up of the TE spectrum is reported.

The assignment of the peaks is the following: 996cm^{-1} (a_1, β_{CCC}), 1020 cm^{-1} (a_1, β_{CH}), 1072 cm^{-1} (b_2, β_{CH}) where the a_1 and b_2 indicate the mode symmetry and β indicate the in-plain bending [62].

The following qualitative considerations can be drawn from the analysis of Fig. 10.3:

- upon laser exposure the sample does not shows substantial degradation (black and red spectra are very similar);
- correctly, in TE configuration the signal is strongly suppressed with respect to TM.

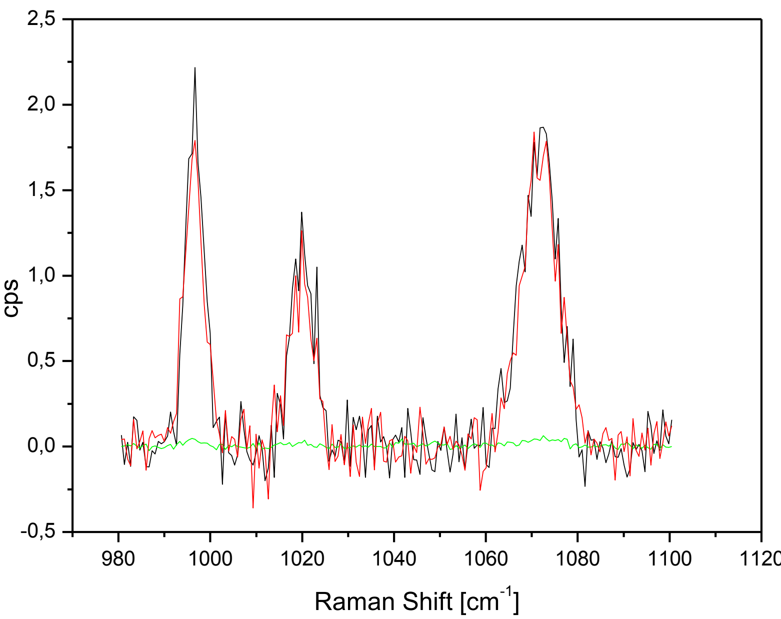


Fig. 10.3. Raman spectra measured in counts per seconds (cps): the first spectrum in TM polarization (black), the TE spectrum (green) and the second in TM (red).

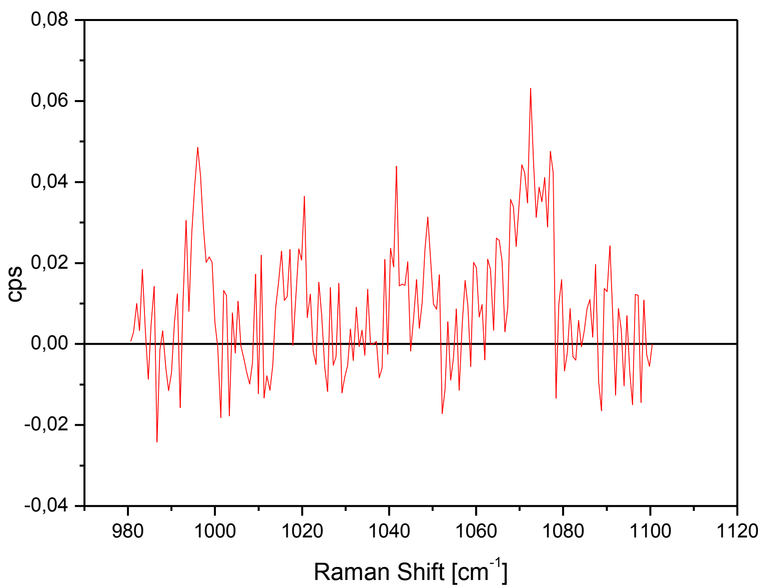


Fig. 10.4. Blow-up of the TE Raman spectrum in Fig. 10.3.

To check the reproducibility of the measures we took various acquisitions at different points of the same chip, at different chips of the same sample and for different samples, all made with the same fabrication parameters. Examples of spectra taken on different points of the same chip are shown in Fig. 10.5.

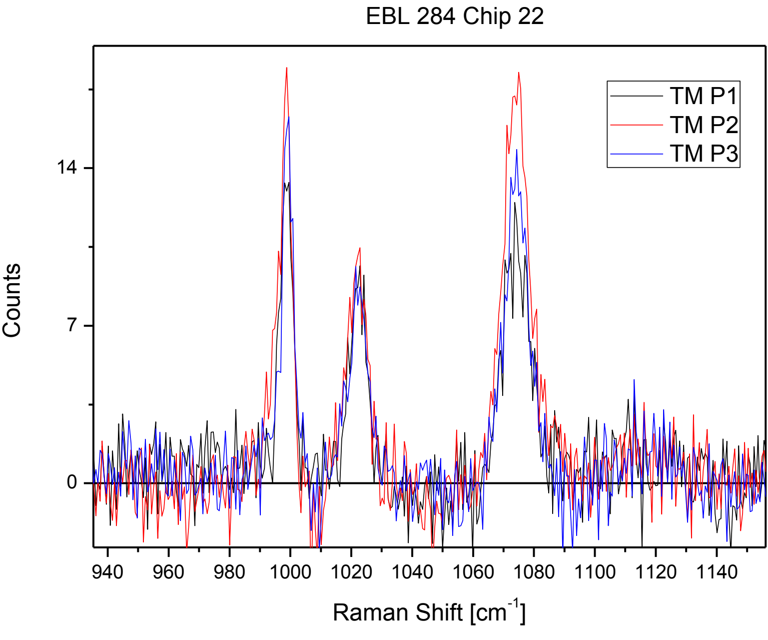


Fig. 10.5. SERS spectra (10 seconds integration time) of BT, recorded with TM polarization on different points of the same chip.

The plot summarizing the result of all measures is showed in Fig. 10.6 and will be discussed in more detail in Chapter 12.

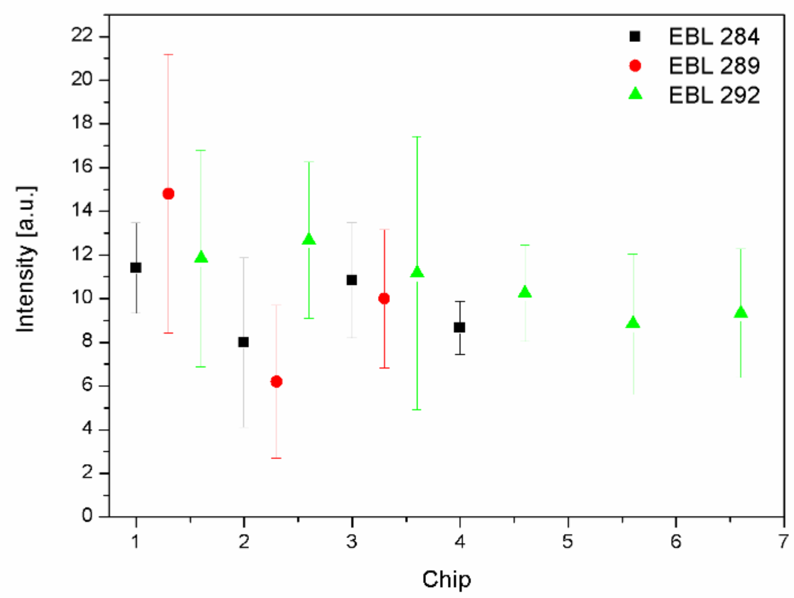


Fig. 10.6. Uniformity of the SERS intensity over different samples, different chips and different points within the same chip. The vertical bar represents the dispersion of the different measures on the surface of each chip.

10.3 Experimental SERS Enhancement Factor

In the following we will introduce the formula used to evaluate the SSEF.

In the literature EFs are often measured with respect to the 1002 cm^{-1} mode (a_1, in-plane bending of the benzenic ring), which is one of the most intense bands of the benzenthiol spectrum. Therefore we shall refer to that Raman band for the EF evaluation in the following. As a reference, BT sealed in vial was used.

In section 3.3.1 (Eq. 3.5) we defined the SSEF as:

$$\mathrm{SSEF} = \frac{I_{\mathrm{SERS}} / N_{\mathrm{Surf}}}{I_{\mathrm{RS}} / N_{\mathrm{Vol}}} \tag{10.1}$$

where I_{SERS} refers to the Raman scattering collected from the sample functionalized with BT and I_{RS} refers to the Raman signal of liquid BT. These two values are normalized by the number of BT molecules illuminated in the sample (N_{Surf}) and in the liquid (N_{Vol}).

Eq. (10.1) can be explicitly recast in the form:

$$SSEF = \frac{I_{\mathrm{SERS}}^{BT\ monolayer}}{I_{\mathrm{Raman}}^{BT\ liquid}} \frac{C_V}{C_S} H_{\mathrm{Eff}} \frac{1}{A} \tag{10.2}$$

where $C_V = 5.88 \cdot 10^{21}$ molecule/cm^3 and $C_S = 6.8 \cdot 10^{14}$ molecule/cm^2 are the volumetric density and the packing density on flat gold for BT, respectively, known from the literature. Since in SERS experiments molecules are disposed on the patterned gold surface, while the packing density is given

for a flat surface, we have to divide by a factor A, considering the ratio of the patterned and the non-patterned area.

The H_{eff} is somehow related to the depth of focus of the objective, and it can be experimentally determined by recording the intensity of the 520 cm^{-1} Raman band of Silicon as a function of the distance from the focal plane (along the propagation direction of the laser beam, z). In Fig. 10.7 is reported a typical z-scan: H_{eff} is estimated as the Full Width at Half Maxiumun (FWHM) and turned out to be 10 microns for a 50x objective.

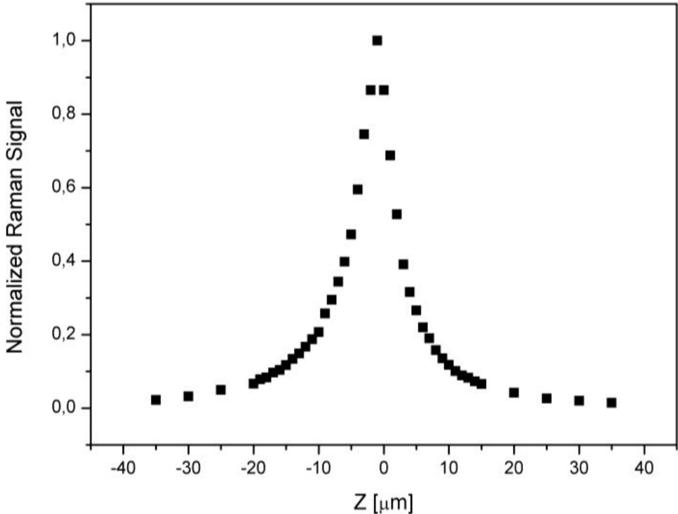


Fig. 10.7. Intensity of the 520 cm^{-1} Raman band of Silicon as a function of the distance from the focal plane.

10.4 Device homogeneity quality and accuracy

Two important concepts in device fabrication are *accuracy* and *precision*. The former is the degree of closeness of measurements of a quantity to the actual value. It depends on the possibility to measure *absolute* quantities, like $SSEF_{TM}$, and to compare them with simulated ones, without working with ratios. In absence of a theory to implement chemical effects, the achieving of high accuracy in the determination, for instance, of the actual value of $SSEF_{TM}$, appears difficult at the moment, due to possible systematic errors in simulations. The latter, also called *reproducibility* or *repeatability*, is the degree to which repeated measurements under unchanged conditions show the same results. It is related to the dispersion of experimental data around an average and it is therefore measured by the statistical error σ.

One of the project's objectives was indeed the reproducibility and uniformity of the response, fundamental for a SERS-active substrate suitable for sensing applications. In order to test the homogeneity of the signal, a number of samples has been produced using the same fabrication parameters. On every sample many chips with the same geometrical characteristics (period, duty cycle and gold thickness of the grating) have been replicated and functionalized. Using the same analytical parameters (laser power, wavelength and light polarization) many points have been measured on the surface of each chip. The resulting plot, shown in Fig. 10.6, demonstrates the uniformity of measurements and represents a proof of quality of the nanofabrication process.

Coming back to accuracy, we have already seen that in the SSEF definition appears a second factor, in addition to the electromagnetic enhancement: the chemical enhancement. Since we do not have a method to evaluate it, we cannot directly compare the simulated and the experimental SSEF. To avoid

this problem we can work with ratios, for instance between SSEF in TM polarization and SSEF in TE polarization. In this way the chemical enhancement, supposed to be polarization independent, cancels out.

In the next chapter we will evaluate the ratio between $SSEF_{TM}$ and $SSEF_{TE}$, both experimentally and numerically. Concerning the errors, all the parameters appearing in equation (10.2) will be assumed errorless, except for the peak intensity.

A good agreement between the experimental and numerical $SSEF_{TM}$ to $SSEF_{TE}$ ratio, added up to a well-tested nanofabrication process able to guarantee a good uniformity and reproducibility of the device, confirms the potentiality of nanofabrication technique in substrate engineering in order to get local fields of the desired intensity and location.

EXPERIMENTAL RESULTS - EOT

In 2011 we presented our studies on EOT in the article: 'Extraordinary optical transmission in 1D gold gratings near and far-field analysis' [63] published by the '*Applied Optics*' Journal, where extraordinary optical transmission of 1D metallic gratings is thoroughly studied in order to elucidate the different modes of light propagation through subwavelength slits.

The link between far- and near- field regimes of EM wave propagation has been investigated by means of the SNOM technique. Experimental results on gold gratings fabricated on silicon nitride membranes have been compared with numerical simulations calculated with FEM method in order to understand the nature of the resonances and provide a thorough description of the mechanisms of EOT propagation.

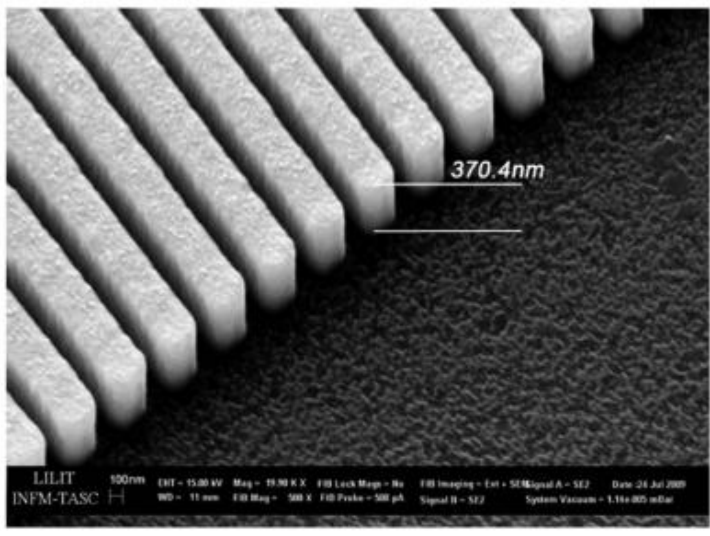


Fig. 11.1. One-dimensional gold grating fabricated on a silicon nitride membrane.

We used Electron Beam Lithography (EBL) in combination with electrolytic growth to fabricate gold gratings on thin Si_3N_4 membranes. Electroplating growth technique allows us to obtain high-quality structures with steep vertical walls regardless of the pattern aspect ratio. Silicon nitride membranes are thin and transparent in the visible/near IR: this property enables us to perform scanning near-field optical microscopy (SNOM) measurements and obtain a complete transmittance optical characterization. The optical window has been realized on a 1 mm×1 mm, 2 µm thick Si_3N_4 self standing membrane. The evaporation of a chromium/gold bilayer (10 nm/20 nm thick) produces the base plating electrode for the galvanic electrolytic growth. The nanopatterns have been realized on a 450 nm thick PMMA resist using a 30 keV accelerating voltage and a beam probe current of 140 pA. The gold electroplating growth has been performed at a current density of $1mA/cm^2$ with a growth rate of 10 nm/sec. After growth and PMMA stripping, the Au and Cr layers have been removed respectively by an Ar dry etching and by a wet etching based on NaOH and $K_3Fe(CN)_6$ solution in H_2O. An example of a fabricated grating is shown in Fig. 11.1.

The structure has a period of 500 nm with a slit aperture of 250 nm. The thickness of gold is 370 nm. Far-field transmittance measurements have been collected over the spectral range between 400 and 1100 nm at normal incidence, both for TM and TE polarization.

Experimental transmission spectra (continuous line) of the upper defined 1D gold slit array are reported in Fig. 11.2 for both TE and TM light polarization (the spectral range of TE transmission spectra has been reduced to better show the transmission peak).

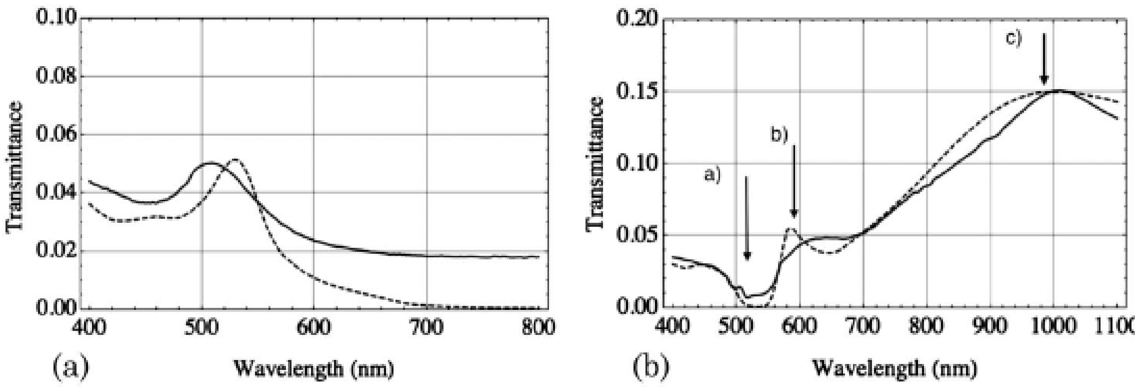


Fig. 11.2. Transmission spectra of a one-dimensional gold grating. Dashed and solid lines correspond to FEM simulation and experimental data, respectively. (a) and (b) refer to TE and TM polarization. In (b) the arrows indicate the different resonances that are directly related to the EM field distribution of Fig. 11.3.

We choose to focus our analysis on the visible/near-IR part of the electromagnetic spectrum (400–1100 nm) in order to characterize the optical response of the plasmonic gratings. Different transmission resonances are observed for the two polarizations, thus confirming that different mechanisms generating EOT are present. A transmission peak can be recognized at 510 nm for TE polarization. The TM polarization spectrum is characterized by a broad peak between 400 and 500 nm and by a strong continuous transmittance increase from 550 to 1000 nm. The depletion around 500 and 550 nm can be identified with a transmission gap.

Several numerical and theoretical studies have been performed in past years in order to explain the physics that governs the EOT phenomena for 1D grating structures. The modal analysis presented by Garcia-Vidal and co-workers [13,64] is a simplified analytical model that is able to catch the main physics effects, and it has the advantage of leading to a simple analytical expression for far-field quantities. The model was developed for an ideal infinite 1D grating, setting surface impedance boundary conditions on the interfaces, and considering only the p-polarization case. According to this model, neglecting all the evanescent modes and keeping only the fundamental propagating eigenmode inside the slits, it is possible to give an analytical expression for the transmittance [see Eq. (2.4)] which allows us to identify the main mechanisms that govern the transmission properties of 1D metallic gratings.

Using FEM simulation and modal analysis, it is possible to understand the nature of the resonances observed in the experimental spectra. FEM simulations allow us to calculate both far- and near-field quantities, i.e., the transmittance (dashed lines in Fig. 11.2) and the distribution of the EM field inside the slits (Fig. 11.3), respectively.
As can be seen in Fig. 11.2, there is a fairly good agreement between experimental and numerical transmission spectra, since all the major features of the experimental spectra appear correctly

simulated. Small discrepancies can be due to some fabrication artifacts, whereas FEM simulations consider the structure to be infinite and perfectly uniform.

For TM polarization [Fig. 11.2 (b)], the transmission dip at the wavelength of 530 nm is associated to a SPP resonance. We notice that, when the condition of Eq. (2.5) is satisfied, a minimum in the transmission coefficient [Eq. (2.4)] is observed. SPP resonances lead to transmission extinction and high electromagnetic (EM) field enhancement near the slits.

The simulation of Fig. 11.3 (a) depicts the norm of the magnetic field for TM polarization at the wavelength of 530 nm. As can be seen, the field is concentrated on the top of the gold ridges; this configuration of the field is directly related to the transmission extinction observed in the experimental spectrum. The wavelength prediction shown by the vertical arrow in Fig. 11.2 (b) [arrow a)] agrees well with the dip's spectral location, thus confirming the nature of this antiresonance. It is enough to shift the wavelength a few tens of nanometers at 580 nm in order to completely change the transmission properties of the 1D grating in TM polarization. The transmission peak observed at 580 nm for TM polarization [arrow b) in Fig. 11.2 (b)] is due to a hybrid mode whose configuration is shown in Fig. 11.3 (b). The field distribution results in the combination of a planar SPP on the top of the grating and SPP cavity modes excited on the vertical walls of the slits. The combination of these modes is responsible for the enhanced transmission. The SPP resonances on the top of the grating surface are still excited because this resonance has a quite large spectral width. However, at this wavelength SPPs can propagate along the vertical walls of the slits, exciting cavity modes and further propagate at the bottom horizontal surface, generating a symmetrical SPP mode at the interface with the dielectric silicon nitride substrate. This latter is the field that further couples to the far field, generating the EOT contribution. At higher wavelengths (up to 1000 nm), we observe an increase in the transmission value where another resonance is observed. In Fig. 11.3 (c) it can be seen that, at the resonant wavelength of 1000 nm, the norm of the magnetic field is distributed uniformly only within the slit. No SPP modes on the horizontal grating surface are generated. This is a pure cavity mode whose spectral location is well predicted by the Fabry–Perot condition for the total phase [Eq. (2.5)] in the transmission coefficient [Eq. (2.4)], which leads to a transmission maximum whose spectral position [arrow c) in Fig. 11.2 (b)] is very close to the one observed experimentally and numerically.

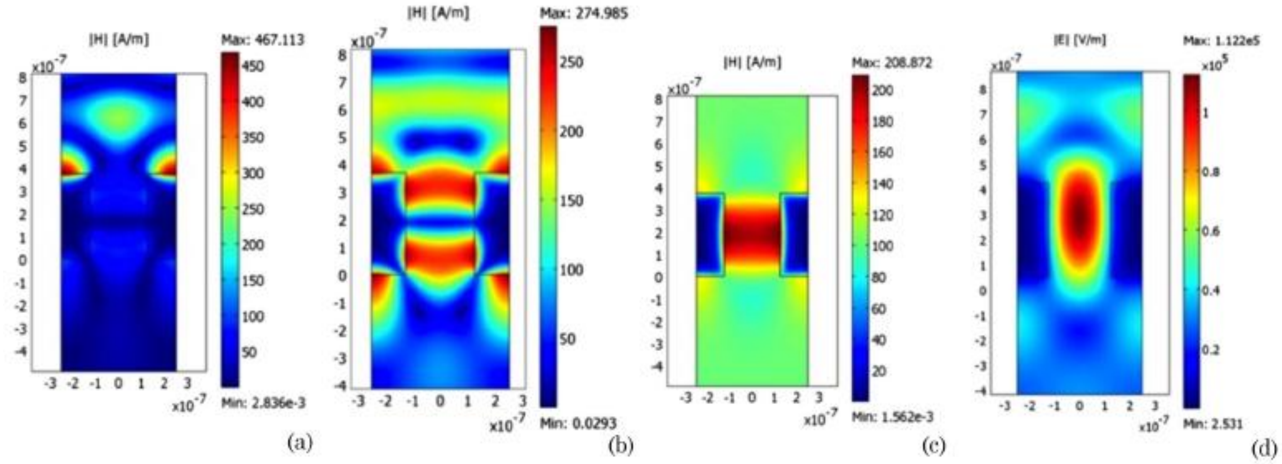


Fig. 11.3. (a) Norm of TM magnetic field at the SPP resonance (corresponding to the wavelength λ = 530 nm); (b) norm of TM magnetic field at the hybrid mode resonance (corresponding to the wavelength λ = 580 nm); (c) norm of TM magnetic field at the cavity mode resonance (corresponding to the wavelength λ = 1000 nm); (d) norm of TE electric field at the cavity mode resonance (corresponding to the wavelength λ = 530 nm).

The resonance peak at 530 nm observed in TE polarization [Fig. 11.2 (a)] is representative of a type of resonance which does not involve SPPs. In fact, for TE polarization, SPPs cannot be excited. As accounted for by different authors [65,66], this TE resonance occurs by means of a cavity mode excitation, whose EM field distribution inside the slits is however completely different from the TM case, as we can see in Fig. 11.3 (d). The distribution of the field in the TE cavity mode is concentrated in the center of the slit, with no SPP field mode located on the vertical walls. Moreover, this resonance strongly depends on the geometrical parameters of the slits, in particular width and depth, with little dependence on the grating's period.

In order to better correlate the different EM field profiles found with FEM simulations to the transmission resonances observed in Fig. 11.2, experimental near-field analyses have been performed, by means of a SNOM technique.

The SNOM (scanning near field optical microscopy) technique is a microscopy technique that is able to overcome the resolution limit of the far field, which is due to diffraction. Exploring with a probe the evanescent fields at the surface of the samples and propagating such waves to a detector in the far field, resolutions much higher than the incident light wavelength can be obtained. In the present work the grating was illuminated upside down, in a configuration known as inverted SNOM, shown in Fig. 11.4.

Fig. 11.4. Inverted SNOM experimental setup.

Fig. 11.5 (b) reports the intensity profile of the light emitted from the slits as a function of the direction perpendicular to the grating. The measurements have been collected by using collimated unpolarized light with wavelength of 514 nm at a distance of few tens of nanometers from the sample. From the far-field experimental spectra of Fig. 11.2 we see that, at 514 nm, a resonance condition is obtained for both TE and TM polarizations, since a cavity mode and an SPP resonance are excited, respectively in TE and TM spectra. In Fig. 11.5 (b) is reported the SNOM light intensity image, and two types of peaks are present, as can be better seen in the SNOM light intensity profile shown in Fig. 11.5 (c). The more intense peaks appear to be coherent to the grating's period, as can be seen by the comparison with the topographic image in Fig. 11.5 (a), and are related to the TM SPP excited on the top of the grating gold ridges. The lower peaks appearing in the near-field regime between the more intense peaks correspond to the TE cavity modes excited inside the slits, which are related to the excitation of the evanescent waveguide modes [Fig. 11.3 (d)].

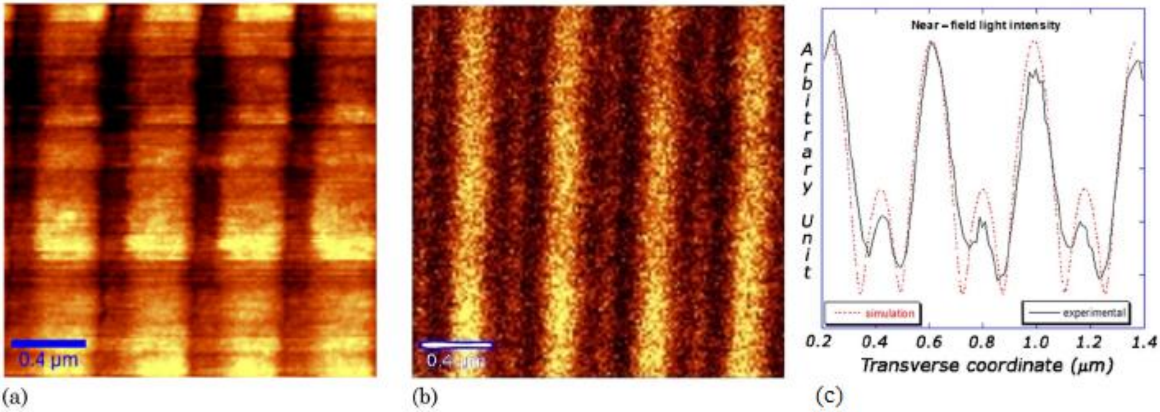


Fig. 11.5. (a) SNOM topographic image of 1D gold grating; (b) SNOM light intensity image; (c) comparison of experimental (solid line) and simulated (dotted line) SNOM light intensity profile. The FEM simulation plots the magnitude of the square of the electric field amplitude. The major peaks and the brighter areas correspond to SPP excitations on the gold ridges of the array, while the small peaks and the less bright areas correspond to the light directly scattered inside the slits.

A FEM simulation of the SNOM intensity has also been performed [the dotted red line in Fig. 11.5 (c)], which plots the magnitude of the square of the electric field amplitude vs the transverse coordinate. We can see how the alternation of peaks well fits the experimental data both in relative intensity and location. This result confirms that, using unpolarized light, both SPP and cavity mode resonances can be excited at the same time.

In 2011 we presented additional EOT studies on light polarization in the article: 'Polarization independence of extraordinary transmission trough 1D metallic gratings' [67] published by the '*Optical Society of America*' Journal, where the optical characterization of the device is focused in particular on the study of the dependence of the extraordinary transmission on the polarization of the incident light.

The aim of the paper is to show, both experimentally and numerically, that under particular conditions EOT through 1D gold grating becomes polarization independent in the optical regime. A numerical analysis performed as a function of period and wavelength allows to describe the limits of the phenomenon.

The fabrication process of the gold gratings is made by Electron Beam Lithography (EBL) in combination with electrolytic growth. The lithography is performed on a glass substrate, coated with 350 nm thick polymethylmethacrylate (PMMA) resist layer, using a 100 keV acceleration voltage and a beam current of 2 nA. A thin layer of ITO on the substrate acts as conductive layer, necessary for the electric contact during galvanic growth. A growth current of 100 mA, voltage of 2.9 V have been used, so as to obtain a gold thickness of 200nm. The optical characterization of the gratings has been made through ellipsometric measures, focusing in particular on the far-field transmission properties of the system, using the variable angle spectroscopic ellipsometer described in Chapter 7. The possibility to change wavelength, polarization and angle of incidence of the incoming light allows very flexible measurements. In our work, measurements have been performed at normal incidence, with a polarization angle varying from 0° (tangential component of the electric field perpendicular to the slit axis or TM polarization) to 90° (tangential component of the electric field parallel to the slit axis or TE polarization). The dielectric constants of the different materials have been calculated directly from ellipsometric measurements performed on the samples used for the experiment.

In Fig. 11.6 (a) are reported several transmission spectra of a 1D gold grating (period 516 nm, slit width 258 nm and metal thickness 200 nm) as a function of the incident wavelength, for different polarization angles.

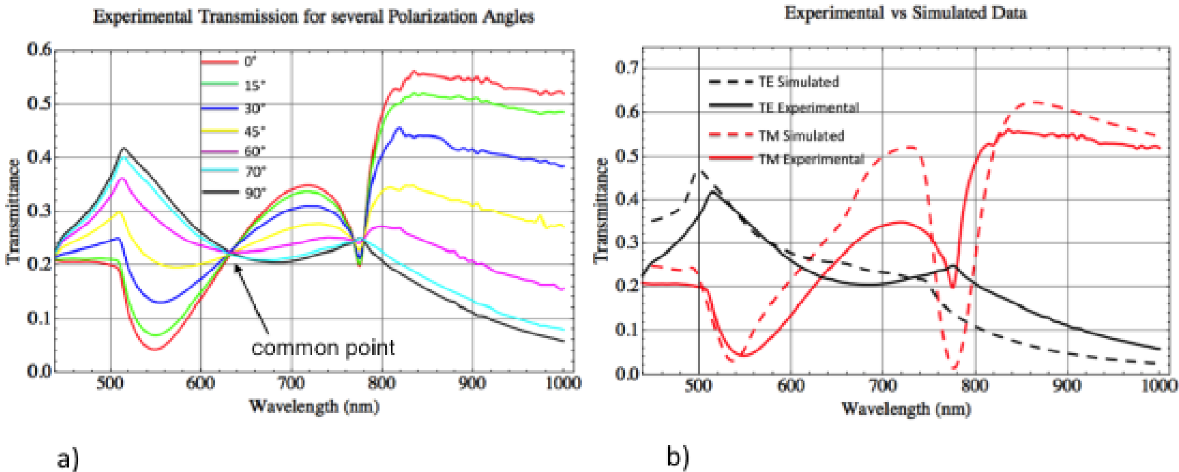


Fig. 11.6. (a) Experimental transmission spectra as a function of incident light wavelength and polarization angle, from 0° (TM) to 90° (TE); at λ = 630nm a point common to all polarization values is observed. b) Comparison between experimental (continuous lines) and simulated (dashed lines) data for TE and TM transmission spectra.

The transmission reaches a remarkable value as high as 55% for TM polarization. Different features in the transmission spectra appear for TM and TE polarization. As already seen [15], three kind of mechanisms are responsible for the peculiar transmission properties of the system: Surface Plasmons Polaritons (SPP) resonances, Cavity Modes (CM) resonances and Wood-Rayleigh (WR) anomalies. In the experimental spectrum of Fig. 11.6 (a) we observe all of these features. In Fig. 11.6 (b) is reported a comparison between an experimental spectrum and a simulated one, for both TE and TM polarization. As we can see, the spectral locations of cavity modes and transmission dips are in fairly good agreement with theoretical data, discrepancies being due to fabrication artefacts. For TM polarization we see that at a wavelength of 550 nm a transmission dip is present. The dip spectral location is very close to the one predicted by Eq. (2.6), thus confirming the plasmonic nature of this antiresonance. In fact, at this wavelength, SPPs are excited on the air-gold interface of the grating, and the electromagnetic (EM) field is confined on the top of the metallic structures, leading to transmission extinction and high near-field enhancement [65]. The peak at 720 nm is due to a hybrid SPP-CM mode formed by the combination of a planar SPP on the top of the grating and SPP cavity modes excited on the vertical walls of the slits [65]. In fact, as many authors pointed out, CMs and SPPs resonances are not independent to each other [14] and can couple in order to generate a hybrid mode that presents both CM and SPP resonant characteristics. The transmission dip at 775 nm is another SPP excitation; this time SPPs are excited on the gold-ITO interface of the grating. Finally at a wavelength of 850 nm a more intense CM mode is excited, reaching a sort of plateau in the transmission spectrum. For this type of resonance the EM field is distributed uniformly within the slits [65].

For TE polarization, SPPs cannot be excited; although the resonance observed for this polarization at a wavelength of 516 nm (equal to grating's period) is due to the excitation of a cavity mode, the EM field distribution within the slits is completely different with respect to the TM case. For this polarization, the field is located at the centre of the slits, with no SPP field mode located on the vertical walls. For TE polarization we also observe a WR anomaly, represented by a less intense transmission

peak located at 780 nm. Its spectral location is well predicted by the equation (2.7), where n_d is the refractive index of the ITO substrate (n_d=1.517+0.025i at 780 nm), thus confirming the nature of this peak.

Experimental data also clearly show that, by rotating the polarization from 0° to 90°, all the spectral features have a progressive suppression or enhancement and TM spectrum gradually transforms into TE.

A peculiar behaviour is represented by a "common point" for all the different polarization angles, located at a wavelength of about 630 nm. This point represents a symmetry of the system: at this particular wavelength, no matter the polarization angle, the far field transmission remains constant.
We have then performed multiparametric FEM simulations that allow us to explore the phase map. In Fig. 11.7(a) and Fig. 11.7 (b) transmission maps are reported as a function of incident wavelength and grating period, respectively for TM and TE polarization.

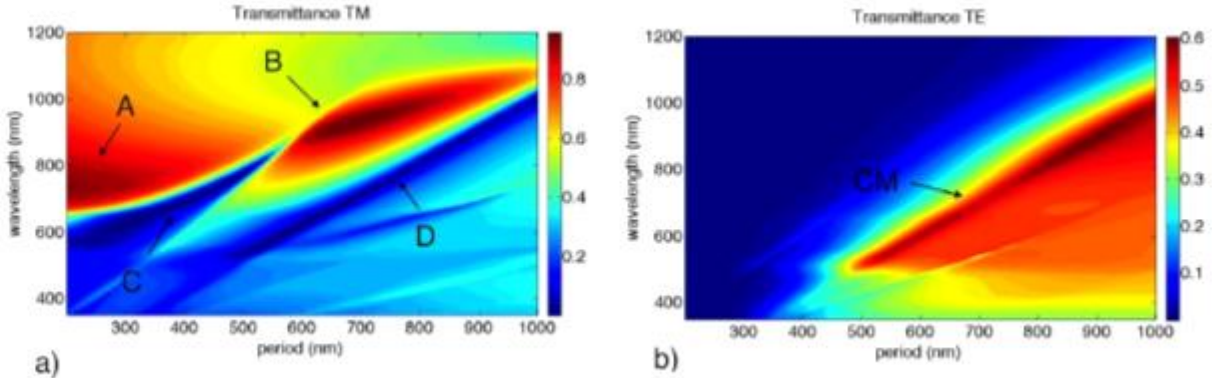


Fig. 11.7. FEM transmission map as function of grating's period and incident light wavelength. The duty cycle is kept fixed to 50% and the gold thickness is set to 200 nm. Fig. a) refers to TM case. Fig. b) refers to TE case. The letters refers to the different transmission resonances described in the main text.

For TM polarization, two distinct high-transmission areas are present, corresponding to both hybrid cavity modes-SPP and CM resonances [marked in Fig. 11.7 (a) with A and B]. In the map are also present anti-transmission bands (marked with C and D). These bands correspond to direct SPP excitation at the different interfaces (air/gold and gold/ITO), as confirmed by their spectral positions which in first approximation fit the predictions of grating-coupling formula for SPP excitation (Eq. (2.6). For TE [Fig. 11.7 (b)] we can see that only cavity modes (marked with CM on the map) are present, as expected.

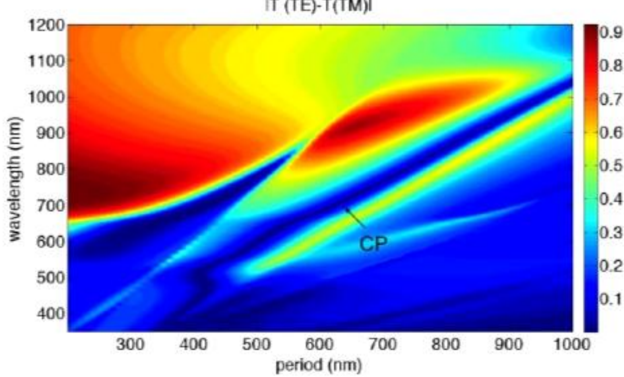


Fig. 11.8. FEM map that plots the absolute difference between TE and TM transmission values (Fig. 11.7) as a function of incident light wavelength and grating's period. Letter CP refers to the "common point" band.

In order to study the properties of the common point, we present also a map which plots the absolute value of the difference between the TE and TM transmission values (Fig. 11.8).

Since at the *common point* the transmission is the same for both polarizations, we expect to see in the map transmission minima corresponding to the common points. In fact, a *common point* band (marked with CP) starting from a wavelength of 500 nm appears in the map. As we can see, the spectral location of the common point as a function of period follows an almost linear dependence behaviour; this means that we can choose the particular wavelength at which the common point occurs just by varying the grating's period. It’s worth noting that since transmittance and reflectance are complementary to each other, this result opens the possibility to obtain also a reflectance polarization independence.

These 2D maps can be usefully exploited for bio-sensoristic applications, where the selection of the right set of parameters to obtain a strong enhancement/focusing of the EM fields is fundamental to increase the sensitivity of the device.

EXPERIMENTAL RESULTS - SERS

In 2011 we published a paper in the '*Microelectronic Engineering*' Journal, 'Design, fabrication and characterization of plasmonic gratings for SERS' [68], where we present the design and realization of a SERS-active substrate suitable for sensing applications.

The aim of the work is to theoretically optimize and fabricate a 1D gold grating to concentrate the electromagnetic (EM) radiation inside the slits. Since to our knowledge only a few papers have previously computed and measured the SERS Enhancement Factor (SSEF) [69], the SSEF of the functionalized digital gold grating will be measured and compared with the theoretical estimation.

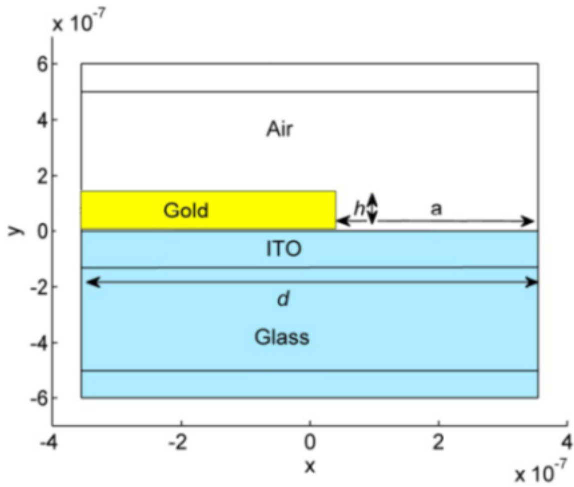


Fig. 12.1. Scheme of the grating structure.

Numerical analysis was performed using finite element method (FEM). A scheme of the simulated structure is shown in Fig. 12.1. The gold grating has been illuminated from the air side at normal incidence with TM polarization. The simulated structures were designed in order to obtain a particular EM field configuration: the presence of hot-spots inside the slits (Fig. 12.2). Since the presence of hot-spots is linked with the full excitation of plasmonic resonances, far field quantities, such as reflectance, transmittance and absorbance can be evaluated in order to extrapolate useful information about EM field configuration.

In Fig. 12.2 the optimized EM field configuration for TM (right) and TE (left) polarization is shown. It clearly appears that TE polarization is forbidden to penetrate inside the slits. On the contrary in TM polarization the EM generates a plasmonic standing waves whose intensity has been optimized changing the geometrical parameters of the grating. In fact, the EM field distribution is, in principle, dependent on all the geometric parameters: metal thickness (h), slits width (a), and period (d). The duty cycle (slit width/period ratio) has been fixed to 40% in order to have a large zone of active SERS surface.

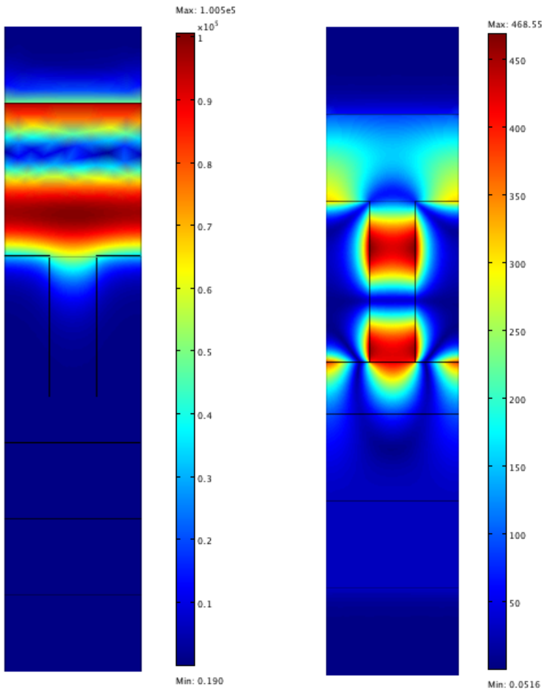


Fig. 12.2. Right: optimized field configuration for TM polarization. Left: field configuration for TE polarization. Color bar represents the norm of the magnetic field. Notice the different percentage of the incident power reflected in the two cases.

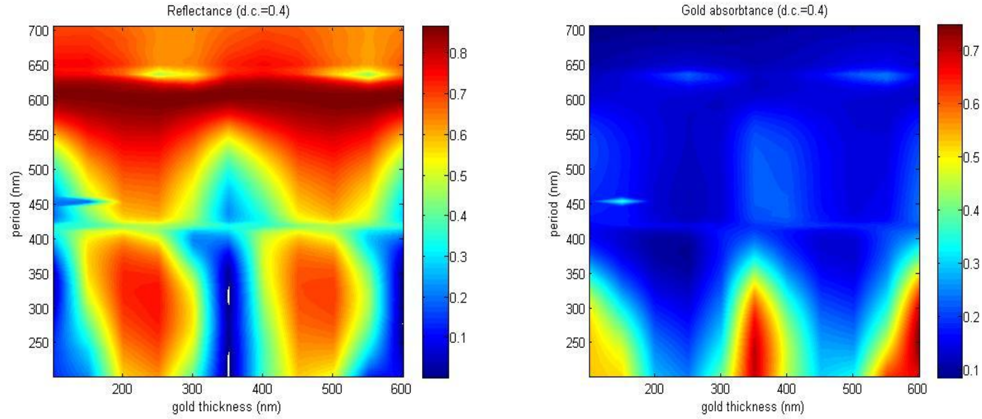


Fig. 12.3. Reflectance (left) and gold absorbance (right) maps as a function of gold grating and thickness.

Bidimensional maps of transmittance, reflectance and substrates absorbance have been computed as function of period and thickness (Fig. 12.3), keeping the incident wavelength set to 633 nm, i.e. the laser wavelength used in the experiments. As can be clearly seen, a periodic pattern is present in the maps. The areas of minimum reflectance and maximum metal absorbance correspond to Fabry–Perot resonances excited by plasmonic effects. Looking at the maps, this condition is verified in particular for a gold thickness (h) of 350 nm and up to a grating period (d) of 350 nm. From these results, a configuration with a grating period of 300 nm and a metal thickness of h = 350 nm has been chosen: it presents two hot spots well localized inside the slits. The final optimized geometry is shown in Fig. 12.2.

The fabrication process consists of two main steps. Electron beam lithography (EBL), by using a JEOL JBX-6300FS EBL system operating at 100 keV, with 2 nA current, has been performed on a ITO/glass substrate. The substrate was previously spinned with a 490 nm thick PMMA resist layer. A dose correction to compensate for the proximity effect was required to obtain good shape definition. The metallic structures were obtained through an electrolytic growth made on the ITO layer. A growth current of 100 mA, voltage of 2.9 V and growth time of 45 s have been used for the process. The nanofabrication process has been described in Chapter 9.

Fig. 12.4. Picture of a fabricated sample with five gold grating chips; SEM image of a fabricated sample; it's visible the ITO structure of the substrate underneath.

SERS measurements have been performed with a homebuilt micro-Raman setup employing a 50 x microscope objective and the 632.8 nm line of a He–Ne laser as laser source. Spectra were acquired with different laser polarization: transverse magnetic (TM) and electric (TE) modes. The determination of the SERS enhancement factor has been performed by recording the SERS signal of a benzenethiol (BT) monolayer adsorbed on the metal nanostructure. BT presents an intense Raman signal at 1002 cm^{-1}, to which we will refer for the SSEF evaluation. The SERS substrate enhancement factor can be evaluated by using Eq. 10.2:

$$\mathrm{SSEF} = \frac{I_{\mathrm{SERS}}^{BT\ monolayer}}{I_{\mathrm{Raman}}^{BT\ liquid}} \frac{C_V}{C_S} H_{\mathrm{Eff}} \frac{1}{A} \qquad (12.1)$$

where I is the Raman signal intensity of the molecule monolayer absorbed on the metal surface (SERS) and of the liquid reference (Raman), C_V and C_S are the volumetric and packing density of the reference liquid and the BT adsorbed, H_{Eff} is the collection efficiency of the experimental setup and A is the ratio between the patterned area and a flat one.

Fig. 12.5 shows SERS spectra of BT: the Raman signal recorded with TM polarization is much more intense than the one recorded with TE polarization, as foreseen by simulations.

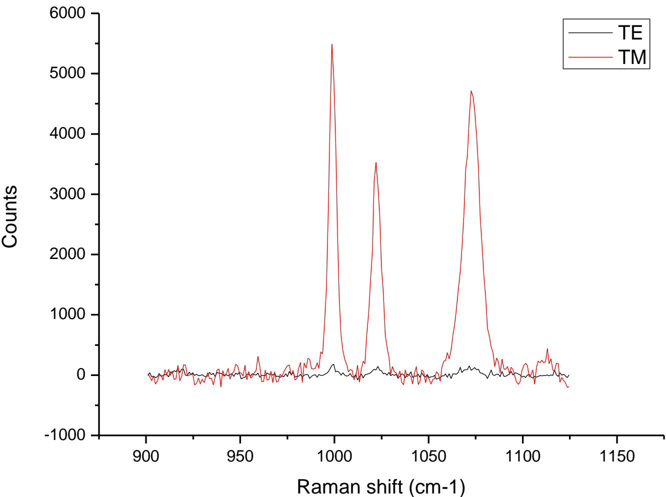


Fig. 12.5. SERS spectra (total counts) of BT, recorded with TM (red line) and TE (black line) polarization.

The ratio of the SSEF factors in Eq. (12.1) allows to identically eliminate all the other factors except for the I_{SERS} contribute of the two perpendicular polarizations, according to the formula:

$$R = \left[\frac{SSEF_{TM}}{SSEF_{TE}}\right]_{exp} = \left[\frac{I_{TM}^{SERS}}{I_{TE}^{SERS}}\right]_{exp} = \left[\frac{SSEF_{TM}}{SSEF_{TE}}\right]_{theor} \tag{12.2}$$

A similar reasoning can be performed for the simulation data [the fourth term in Eq. (12.2)]. The SSEF for polarized detection for a specific molecule functionalization on a specific substrate can in principle be computed by spatial averaging of the enhancement factor of a single molecule, averaged over the entire metallic surface (Eq. 3.58):

$$\begin{aligned} \text{SSEF}^{P} = \Big\{ & M_{\text{Loc}}(\mathbf{r},\omega_L) M_{\text{Loc}}^{\text{PW}-P}(\mathbf{r},\omega_R) \\ & \times \Big[T\Big(\hat{\alpha}_N, \mathbf{e}_{\text{Loc}}(\mathbf{r},\omega_L), \mathbf{e}_{\text{Loc}}^{\text{PW}-P}(\mathbf{r},\omega_R)\Big)\Big]\Big\} \end{aligned} \tag{12.3}$$

Rigorous definitions of the terms in Eq. (12.3) are provided in Chapter 3. The curly brackets {...} denote spatial averaging; the square brackets [...] represent averaging over allowed molecular orientations of the $T(\hat{\alpha}_N, \mathbf{e}_L, \mathbf{e}_R)$ surface selection rules factor, which couples the excitation and re-emission problems through the normalized Raman polarizability tensor (characterizing the symmetry of the Raman tensor and the molecular orientation). Substantially the total SERS enhancement factor is due to two multiplicative factors: the (dominant) electromagnetic EM factor and the chemical one [18]. The EM factor can be computed taking into account the field distribution provided by the simulation code within the $|E|^4$-approximation [18]. On the contrary, the contribute in the square brackets is independent of the EM field and depends only on the orientation of the molecules on the surface of the grating, being simplified when we compute the ratio between the theoretical estimation of SSEF in Eq. (12.2).

Theory		Experiment	
EM $SSEF_{TM}$	EM $SSEF_{TE}$	EM $SSEF_{TM}$	EM $SSEF_{TE}$
6.43	0.17	245.67±9.82	7.51±1.51
R		**R**	
37.43		32.68±6.52	

Table 12.1. Comparison of the theoretical and the experimental SSEFs. Theoretical EM $SSEF_{TM/TE}$ is evaluated in the $|E|^4$-approximation [18]. Experimental $SSEF_{TM/TE}$ is evaluated using Eq. (12.1). **R** is the ratio defined in Eq. (12.2).

Eq. (12.2) allows to directly compare the experimental and the theoretical $SSEF_{TM}$ to $SSEF_{TE}$ ratios (see Table 12.1). The difference between absolute values of theoretical and experimental SSEFs can be attributed to the contribute of the chemical enhancement and of the $T(\hat{\alpha}_N, \mathbf{e}_L, \mathbf{e}_R)$ surface selection rules factors, which most likely cancel out in the computation of the ratio R.

Although a polarization dependence is shown in Fig. 12.5 and a match between the experimental and theoretical ratio in Eq. (12.2) has been achieved (see Table 12.1), further investigation was required for a systematic confirmation of this result.

To this purpose we realized a second series of samples, with geometrical characteristics slightly different from the previous (period d = 300 nm and metal thickness h = 400 nm, optimized for a slit width a = 80nm), from the analysis of which we experimentally determined the ratios R shown in Table 12.2, which mean value can be calculated as: **R = 46 ± 5**.

I_{TM}^{SERS} @ 996 cm^{-1} [Cps*cm^{-1}]	I_{TE}^{SERS} @ 996 cm^{-1} [Cps*cm^{-1}]	**R**
4.8	0.10	48
10.1	0.24	41
9.6	0.19	51

Table 12.2. Experimental **R** determined for a further series of samples.

Due to the different grating parameters, the theoretical R ratio has been recalculated. The obtained value is **R** = **42**.

This result confirms the agreement between the experimental and theoretical ratios in Eq. (12.2) and supports the reliability both of the numerical simulations and of the nanofabrication process, thus confirming the validity of the $|E|^4$-approximation in the evaluation of the Enhancement Factor and the potentiality of the nanofabrication technique in the engineering of plasmonic substrates.

In order to experimentally characterize the relation between the SSEF factor and the polarization angle (ϕ) of the incident light, i.e. the dependence of SSEF on a generic hybrid polarization, we must sum together the contributions of $SSEF_{TM}$ and $SSEF_{TE}$ at the different polarization angles. To estimate it theoretically, we must simply calculate the percentage of the incident power polarized in the TM and TE modes. The Malus's law says that when a polarizer is placed in a plane polarized beam of light, the intensity, I, of the light that passes through is given by:

$$I = I_0 cos^2(\phi) \quad (12.4)$$

where I_0 is the initial intensity and ϕ is the angle between the light's initial polarization direction and the axis of the polarizer. Therefore, $I_{TM} = I_0 cos^2(\phi)$ and $I_{TE} = I_0 sin^2(\phi)$. Since SSEF is proportional to $|E|^4$, the expected dependence of the SSEF factor on the polarization angle will be approximately of the type $Acos^4(\phi) + Bsin^4(\phi)$, where, due to the absence of plasmonic effects, the contribution of the TE mode will be in general very small. Measuring experimentally the variation of the SERS signal intensity to the changing angle ϕ, we obtained the plot showed in Fig. 12.6, which appears to be properly fitted by a curve of the type $Acos^4(\phi) + Bsin^4(\phi)$, thus confirming the theoretical predictions and the reliability of the SERS measurements.

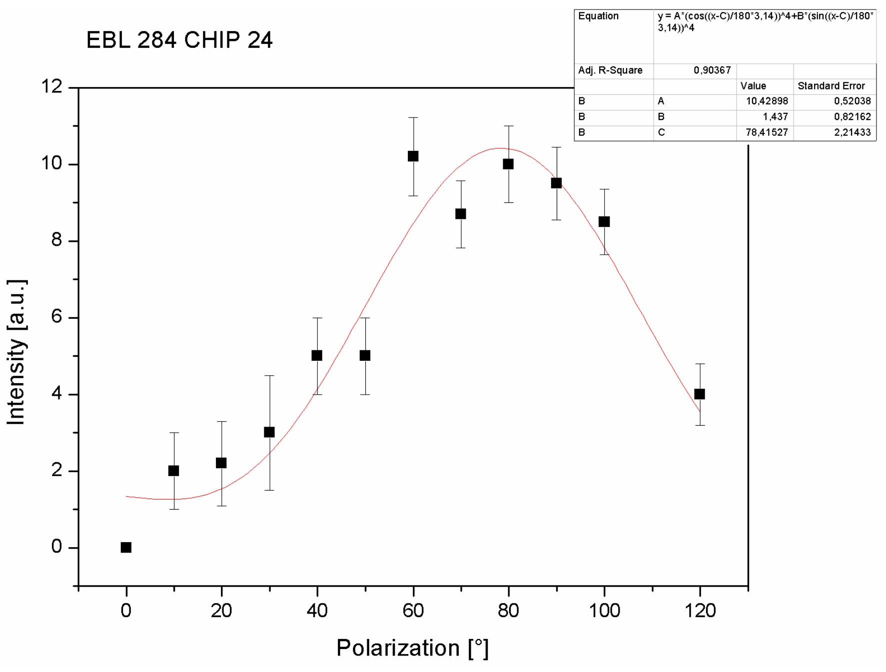


Fig. 12.6. SERS signal intensity dependence on the incident polarization angle. The error bars have been estimated from the signal to noise ratio of the single measures.

The optical characterization of the device has been made through ellipsometric measures. In Fig. 12.7 are shown the experimental transmission spectrum (red line) and the simulated spectrum (blue line).

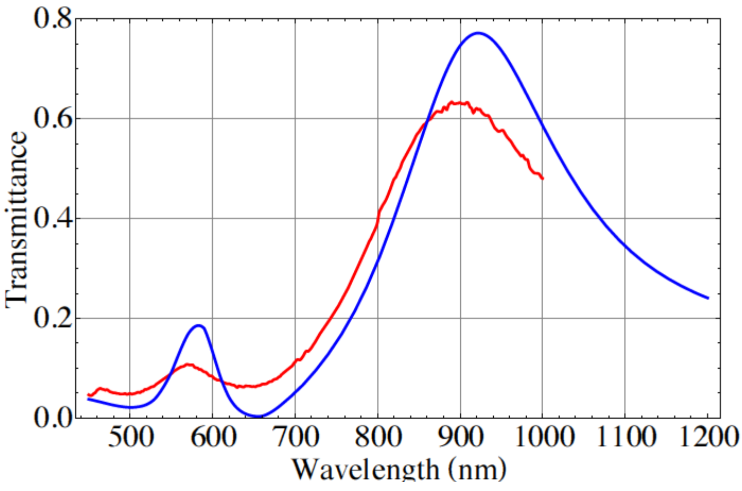


Fig. 12.7. Experimental (red line) and simulated (blue line) transmission spectra.

As can be seen, the agreement between theory and experiment is very good. Since the light spot impinging on the device is quite broad (hundreds of μm of diameter), the ellipsometer delivers a global information of the sample analyzed: the good agreement between the simulated and experimental spectra is therefore indicative of the quality of the device.

A plasmonic peak is observed at 570 nm, followed by a dip at 633 nm, which is exactly the wavelength of the SERS laser beam used.

Note also the scattering of the cavity mode on the ITO substrate, generating a SPP on the gold-ITO interface. Around 900 nm there is a very intense cavity mode, with a transmission experimental value of 63%. The difference in intensity and position between theory and experiment can be due to the uncertainty on gold thickness and on the determination of the dielectric constant of the glass substrate.

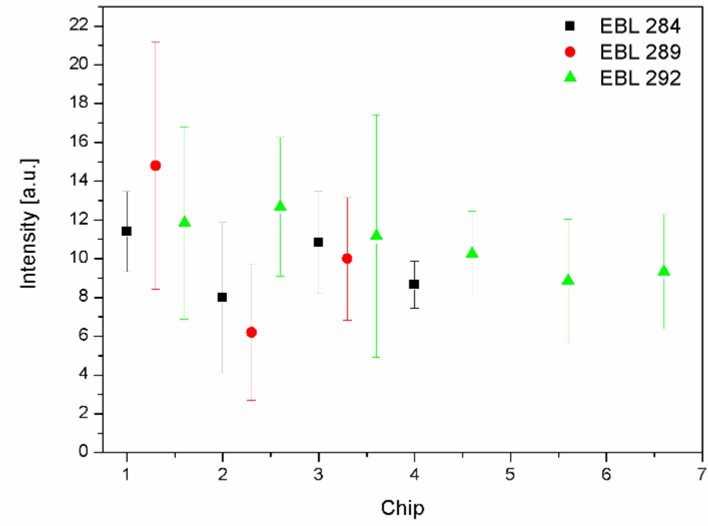


Fig. 12.8. Uniformity of the SERS intensity over different samples, different chips and different points within the same chip.

In order to develop a well-tested nanofabrication process for the realization of an easily manufacturable and reproducible device, characterized by a good uniformity of the response, a number of samples has been made using the same fabrication parameters. On every sample many chips with the same geometrical characteristics (period, duty cycle, gold thickness of the grating) have been replicated and functionalized. Using the same analytical parameters (power of the laser, wavelength, polarization of light) many points have been measured on the surface of each chip. The plot summarizing the result of the measures is showed in Fig. 12.8.

As we can see, there is a good uniformity of the Raman counts (in the plot the vertical bar represents the dispersion of the different measures on the surface of each chip). This is a proof of the quality of the nanofabrication process, which was one of the goals to achieve to fabricate SERS-active substrates suitable for sensing applications.

In conclusion, a 1D digital gold grating has been simulated varying its geometrical parameters in order to find an EM field configuration showing hot-spots inside the slits. The Enhancement Factor of the optimized structure has been evaluated in the $|E|^4$-approximation. A device with the optimized geometry has been fabricated by Electron Beam Lithography and electrolytic growth developing an easy and reliable nanofabrication process, and Raman performances have been measured with a 632.8 nm laser line. This result is a good starting point for tailoring the properties (like field distribution) of more complex nanostructures, which can be optimized and exploited for example in SERS and micro-fluidics applications.

METAMATERIALS

Introduction

Photonic metamaterials are man-made structures composed of tailored micro- or nanostructured metallodielectric subwavelength building blocks. This apparently simple yet powerful concept allows the realization of many new and unusual optical properties, such as magnetism at optical frequencies, negative refractive index, large positive refractive index, zero reflection through impedance matching, perfect absorption, giant circular dichroism and enhanced nonlinear optical properties. Possible applications of metamaterials include ultrahigh-resolution imaging systems, compact polarization optics and invisible cloaking devices.

Introduced around a decade ago, metamaterials represent a broad class of densely packed micro- or nanostructured building blocks that are much smaller than the wavelength of light. The Mie-like resonances of these building blocks open the door to a world of new possibilities. In particular, magnetic dipole resonances can be produced by shrinking the size of macroscopic metallic electromagnets such that their resonances enter the optical regime.

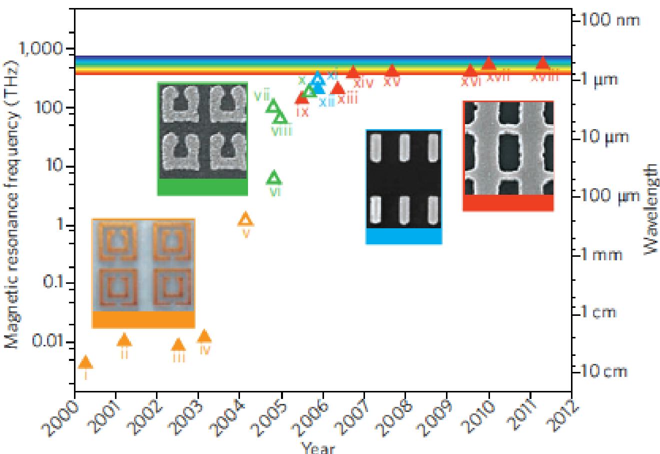


Fig. 13.1. Progress in metamaterial operating frequency over the past decade. The operating frequency of metamaterials with negative magnetic permeability μ (empty triangles) and negative index of refraction n (solid triangles) is shown on a logarithmic scale ranging from microwave to visible wavelengths. Orange: structures based on double SRRs; green: U-shaped SRRs; blue: metallic cut-wire pairs; red: negative-index double-fishnet structures. The four insets show optical or electron micrographs of the four types of structure. [70]

A paradigm example of this is the Split-Ring Resonator (SRR), in which an incident light field induces a circulating and oscillating electric current that generates a magnetic dipole moment normal to the ring. Naively, the SRR can be viewed as a half-wave antenna rolled into an almost closed circle. This picture gives a resonance wavelength that is 2π times larger than the ring's diameter - a value not far from more detailed calculations that also account for plasmonic effects occurring when the operating frequency approaches the metal's plasma frequency, which gradually turns the antenna resonance into a Mie resonance. The magnetic dipole resonance of the SRR, together with its numerous

variations, reminds us that obtaining complete control over an electromagnetic wave inside a material requires independent control over both its electric and the magnetic field.

In Fig. 13.1 is reviewed how the operating frequency of magnetic and/or negative-index metamaterials has developed over the years. The tremendous increase in operating frequency has been made possible by miniaturizing and redesigning magnetic SRRs.

13.1 Metamaterials as sensors

Beyond the main research work achieved in this thesis, concerning the design, fabrication and characterization of plasmonic gratings for EOT and SERS, other two parallel projects, both devoted to the realization of plasmonic nano-optics to be used as biosensors, have been performed. Work on these topics is currently in progress.

To explore the new possibilities offered by plasmonic nano-optics realized through MetaMaterials (MMs), we designed and fabricated two different MetaMaterials (MMs), both formed by arrays of Split Ring Resonators (SRRs). The goal is to create plasmonic crystals able to concentrate the light within so-called 'hot spots', thus resulting in an amplification of the Raman signal, coupling the biosensing properties to the peculiar properties of MetaMaterials.

Split-ring resonators, which have been used to generate negative refractive index media, [71,72] posses geometric and plasmonic qualities which lend them to coupled plasmonic applications, and have recently been demonstrated as effective Surface Enhanced Resonance Raman Spectroscopy (SERRS) substrates for multi-analyte sensing assays. [73]

13.1.1 Design, fabrication and characterization of SRRs using X-Ray Lithography

Split Ring Resonators (SRRs) are one of the most common elements used to fabricate metamaterials and many different fabrication approaches have been investigated during last years by several groups [72,74,75].

We have presented at "*MNE 2011, 37th International Conference on Micro & Nano Engineering*", Berlin, 19-23 september 2011, a SRR fabrication process based on X-Ray Lithography (XRL), a technique able to obtain high aspect ratio structures and large area samples at the same time.

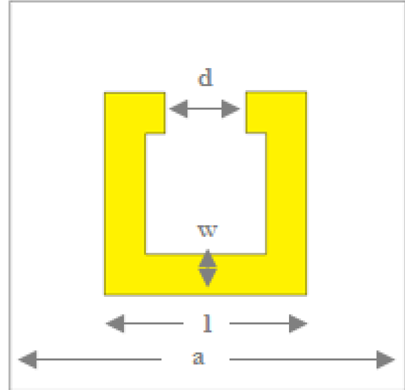


Fig. 13.2. Scheme of the structure of a fabricated SRR (length l=500nm, width w=100nm, gap d=200nm, period a=1µm).

It has been shown that gold SRRs exhibit plasmonic resonances, which can be exploited for sensing small quantities of biological and chemical substances [76] and that the sensitivity of devices based on SRRs can be improved by increasing their height [77]. In order to use the SRRs for sensing it is then equally important to fabricate large area samples. To this purpose, a fabrication process developed combining EBL and XRL tecniques has been used. Although the resolution of XRL is lower than Electron Beam Lithography (EBL), the advantage of this technique stays in the possibility of getting higher aspect ratio structures and rapid lithography of larger areas, up to few square centimeters. In XRL, high energy X-radiation is used to transfer a pattern from a mask to a photoresist; after the exposure and the development, a three-dimensional resist structure mold is produced. Subsequently metal deposition fills the mold and the resist is removed.

Design

We designed the SRR pattern (see Fig. 13.2) in order to obtain a plasmonic response at 2-3 μm. In order to understand the importance of the thickness of the SRR array, we have simulated the transmittance of the structure with the COMSOL software in TE polarization for several heights (see Fig. 13.3).

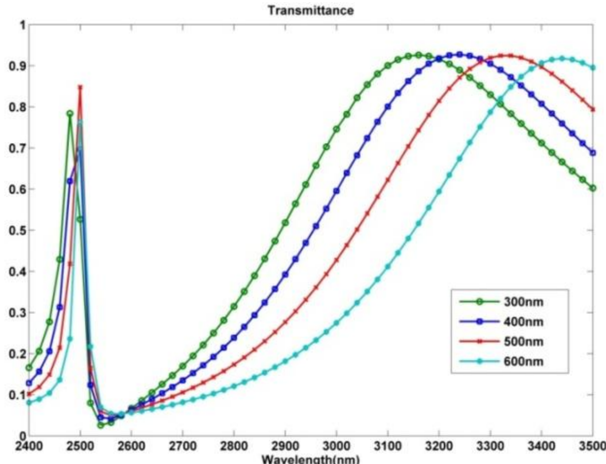


Fig. 13.3. Simulated transmittance in TE configuration for different aspect ratios.

In Fig. 13.3 we show that the shift of the LC resonance of the SRR (see Fig. 13.4) is affected by its aspect ratio, which can be tuned in order to achieve a more suitable operative wavelenght for sensing. The plasmonic external resonance at 2.5 μm is also shown. This resonance appears not to be affected by a variation of the aspect ratio.

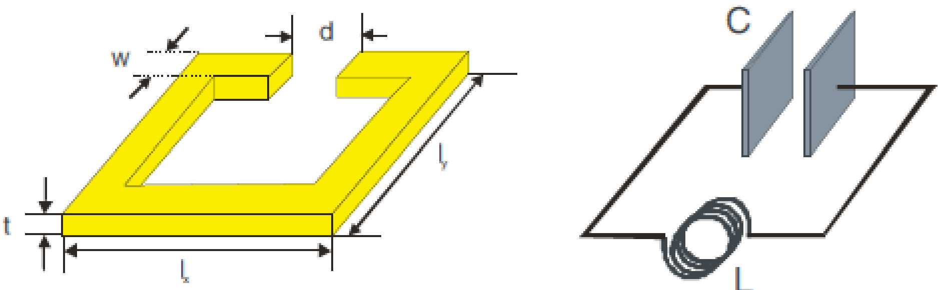
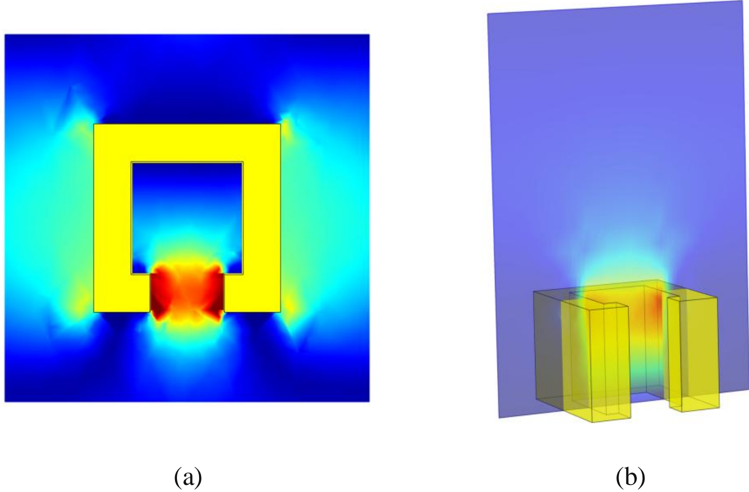


Fig. 13.4. The split-ring resonator is a metallic structure mimicking an LC-circuit consisting of a capacitor, C, and a magnetic coil with inductance L with a resonant frequency at $\omega_0 = (LC)^{-1/2}$.

In TE configuration, a clear plasmonic current is induced by the incident electric field, as appears in Fig. 13.5 (a). The plasmonic current is due to the resonance of the oscillating electric charges among the capacitive gap of the SRR. Moreover, the magnetic field induced by the circular currents counteracts with the z projection of the incident magnetic field, thus producing a magnetic resonance within the SRR arms [Fig. 13.5 (b)]. No magnetic resonance appears in the case of normal incidence.



(a) (b)

Fig. 13.5. (a) Plasmonic current induced by an incident electric field in TE configuration. (b) Magnetic field resonance induced by the interaction of the incident magnetic field with the field produced by the circular currents.

Fabrication and results

The mask substrate is a 2 mm silicon nitride membrane, fabricated at IOM-CNR, and the sample substrate is silicon: membrane and silicon has been covered in Cr/Au to promote gold growth. The mask has been fabricated on a 400 nm thick PMMA resist layer using the JEOL JBX-6300FS EBL machine operating at LaNN (Laboratory for Nanofabrication of Nanodevices) in Padova, using a 100keV acceleration voltage and a beam current of 2nA. Successively the mask has been grown in a gold electroplating bath at IOM-CNR in Trieste and used for XRL lithography at LILIT (Laboratory for Interdisciplinary LITography) beamline at ELETTRA sinchrotron, in soft X-ray regime (beam energies between 1 and 2 keV). XRL has been performed on both PMMA (positive resist) and SAL 601 (negative resist). Exposed PMMA has been developed in 3:7 deionized water:IPA mixture after both EBL and XRL, SAL has been developed in MF319. Electrochemical growth provided the final sample. The nanofabrication process is described in Fig. 13.6.

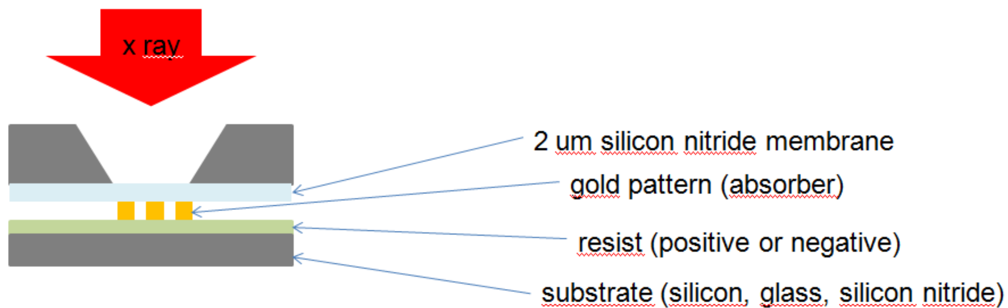


Fig. 13.6. Scheme of the fabrication process.

In the following SEM images is summarized the whole process for the fabrication of the final SRR structure. The process uses a positive resist (PMMA); using a negative resist (e.g. SAL601) it is possible to obtain a direct copy of the EBL fabricated mask.

The process is made up of the following steps:

1. a mask for XRL is fabricated by EBL on a membrane coated with PMMA and electroplated with gold (Fig. 13.7);
2. a XRL of the mask is made on a membrane coated with PMMA [Fig. 13.8 (left)];
3. the electrolytic grown gold structure constitutes a mask for the successive XRL exposure on PMMA [Fig. 13.8 (right)];
4. the gold grown final SRRs structure (on silicon substrate) (Fig. 13.9).

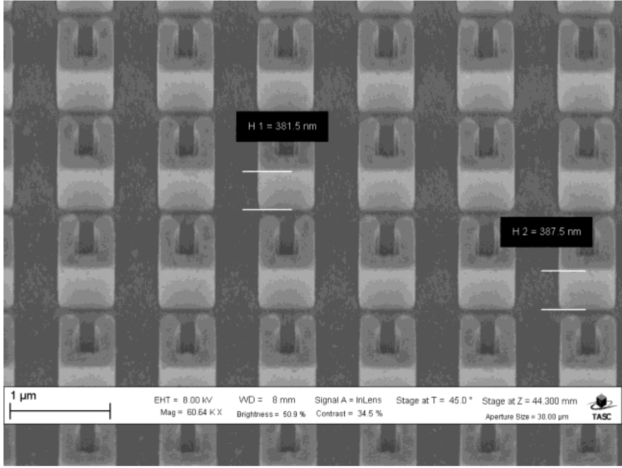
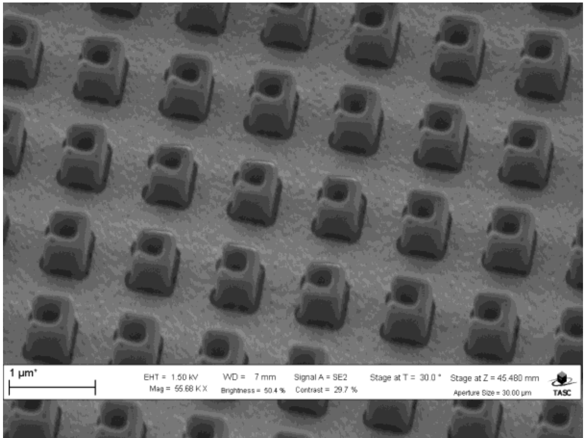


Fig. 13.7. SEM image of a (EBL fabricated, electroplated) XRL mask (thickness=385nm). The mask itself constitutes a gold SRR array.



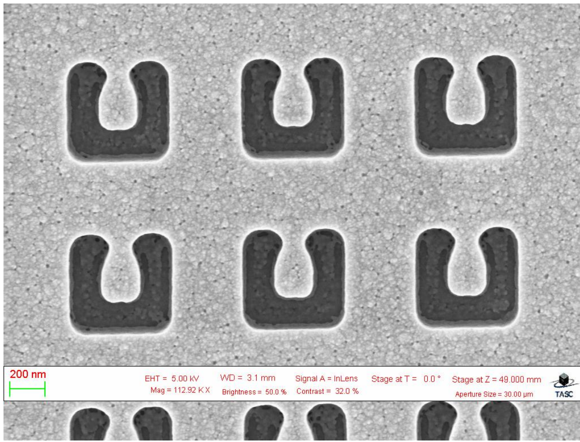


Fig. 13.8. SEM image of XRL on PMMA. On the left: exposed PMMA. On the right: gold grown structure. This structure, grown on a membrane, constitutes a mask to obtain the final SRRs structure.

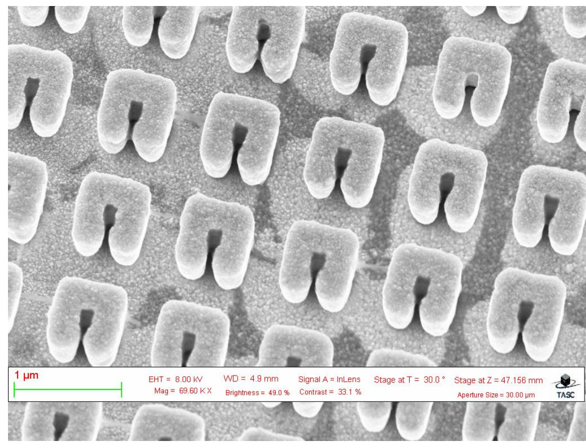


Fig. 13.9. SEM image of the gold pattern growth, resulting in the final SRRs structure.

Fig. 13.10 a) shows the result of the simulation of the SRR reflectance in TE configuration (electric field along the SRR gap and orthogonal to the incidence plane) at an incidence angle of 70 degrees, where is noticeable the plasmonic resonance at a wavelenght of 1.97 µm. The unit cell consists of a gold SRR built upon 20 nm of gold, 10 nm of chrome and a semi infinite layer of silicon nitride. Fig. 13.10 b) shows the measured reflectance in s-polarization (TE configuration) at the same angle of incidence. It's present a plasmonic resonance at 1.85 µm, almost in agreement with the value found with simulated data.

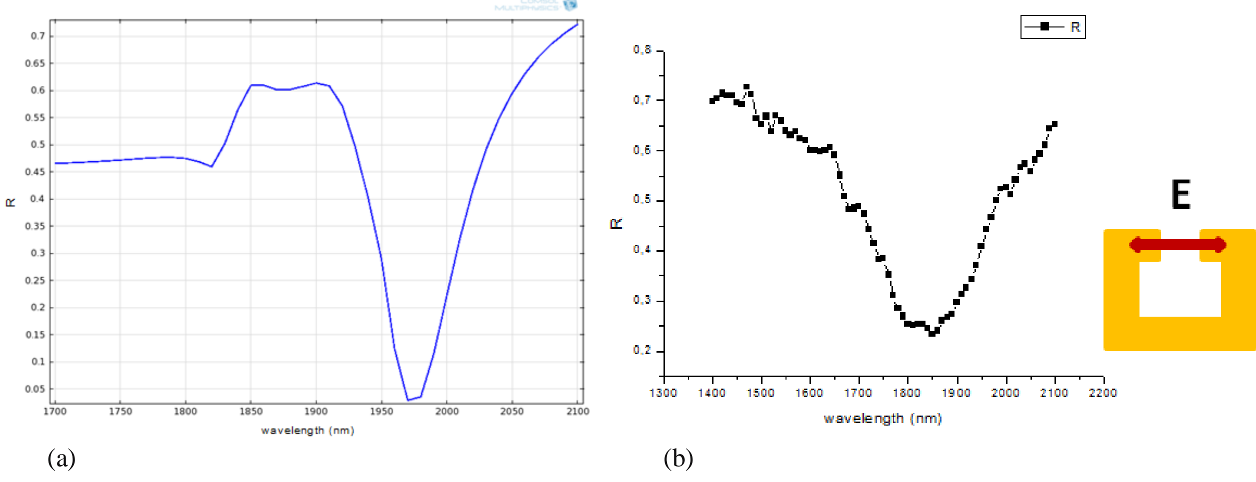


Fig. 13.10. a) Simulation and b) experimental measure of the SRR reflectance in TE polarization for an angle of incidence of 70°.

We have also simulated the SRR reflectance in TE configuration at an incidence angle of 0° and 45° [Fig. 13.11 a)]. The resulting plot shows two resonances: the first, around 1 µm, is plasmonic in nature and is commonly predicted in grating theories in presence of a metal when the magnetic field lies in the plane of the metal; the latter, around 1.75 µm, is electric in nature and is due to the presence of the electric field resonating at the SRR gap. The experimental data [Fig. 13.11 b)] are in agreement with the simulation. The measurements have been performed at LaNN, using the variable angle spectroscopic ellipsometer described in Chapter 7.

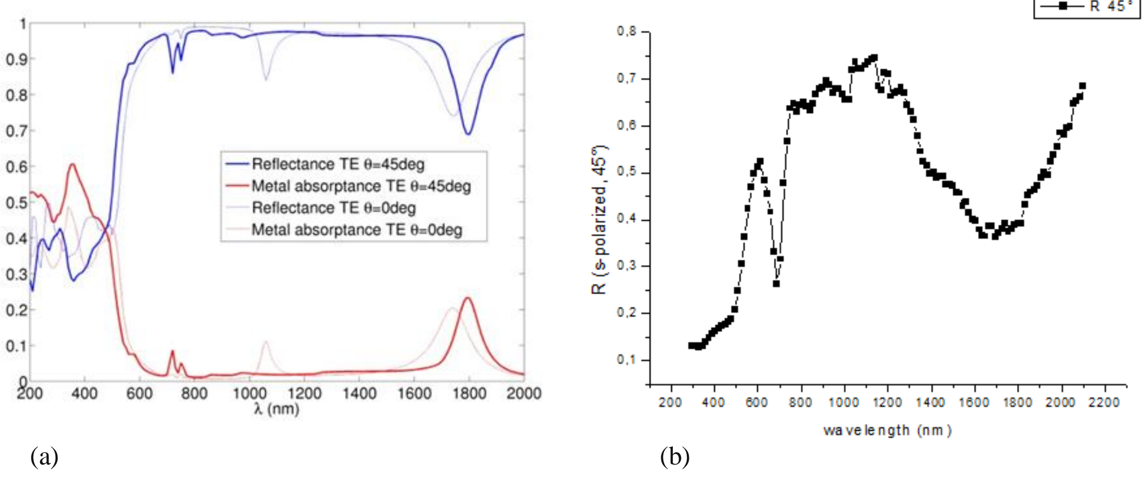


Fig. 13.11. a) Simulation and b) experimental measure of the SRR reflectance in TE polarization for an angle of incidence of 0° (only simulation) and 45°.

Conclusions

SRR fabrication using a combination of EBL and XRL techniques offers the possibility to obtain large area sample with high aspect ratio and a good lithographic quality. Tilted exposure is also achievable with XRL tecnique: this feature opens the opportunity to obtain oblique structures, a new geometry which will be explored as a next step of the research.

13.1.2 Nanogap Ring Antennae as Plasmonically coupled SERRS substrates

Electron beam lithography has enabled a new generation of optically active nanostructures with defined geometries and plasmonic characteristics to be created. Among the more interesting of these new structures, Clark and Cooper highlight multiple-split-ring resonators (MSRRs) [78,79]. MSRRs are nanoring antennae of circular geometry which take advantage of the interparticle plasmonic-coupling effect, concentrating and amplifying the optical field in the splits between the ring segments.

They represent promising structures for enhanced Raman-based sensors, as control over the number and dimension of the splits would allow resonance frequency, volume, and intensity control of the plasmon confined therein. Further, by fabricating rings with small radii, it is possible not only to tune the fundamental resonance frequency to any wavelength within the visible spectrum, but also to maximize the density of the localized "hot spots" (regions of high electromagnetic intensity required for sensing) by the subsequent reduction in array periodicity.

Coupling between ring segments takes place when the external electric field is polarised across the split. In such cases the surface charges at the tips of neighbouring segments are opposed, leading to hybridisation of each segment's plasmon, a phenomenon which acts to increase the magnitude of the field in those areas. When the field is polarised parallel to the split surface charges on either side are the same, so no coupling can take place and the subsequent repulsive force prevents field enhancement. This phenomenon is most immediately apparent in Figure 13.12 [75,76].

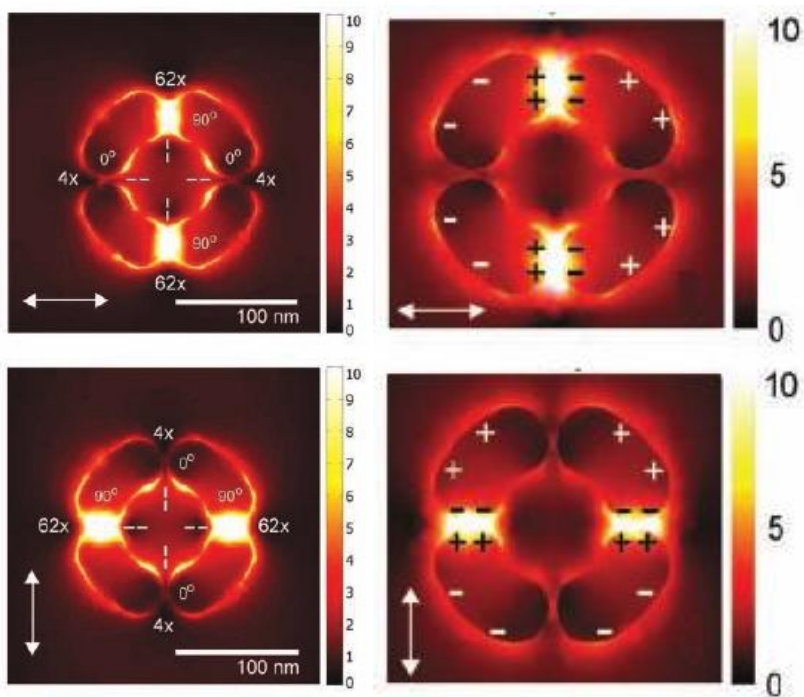


Fig. 13.12. Using finite-element analysis, the plasmonic characteristics of the nanogap MSRR structure have been explored at two different polarizations, describing how variations in the surface charge distribution affect intersegmental coupling at different polarization angles. The hot-spot maxima are located at the top of the vertical faces of each segment at the narrowest point of the split, decaying exponentially in all directions away from this point. Left: local field enhancements with respect to the excitation field have been calculated for gaps equal to 6nm. Right: + and - signs indicate the charge distribution over the ring surface. The direction of the electric component of the driving field is denoted by the arrows in the bottom left of each graphic. The color bar represents the field-enhancement factor. Computed wavelengths is 560 nm. [75,76]

By using Electron-Beam Lithography, MSRRs made up of four 28nm-dimension splits have been engineered into silver nanophotonic ring structures to create concentrated areas of localized field coupling, which can be exploited for enhanced plasmonic applications. The geometry of these multiple-split rings leads to unique plasmon hybridization between the segments of a single structure. This phenomenon is demonstrated to be applicable to extreme Raman sensitivity and may also find use in metamaterial applications.

Fabrication and results

The MSRRs have been fabricated with the JEOL JBX-6300FS EBL machine operating at LaNN (Laboratory for Nanofabrication of Nanodevices) in Padova, working in HR mode with a 100keV acceleration voltage and a beam current of 100pA. The pattern was written on a 130 nm thick PMMA resist layer, spinned on a ITO/glass substrate and developed in a 3:7 deionized water:IPA mixture; we used thermal evaporation of metal followed by acetone lift-off to fabricate the array of Ag split-ring structures. A titanium layer of 3nm was used to ensure good adhesion of the 20 nm Ag layer to the substrate. The evaporation was carried out at IOM-CNR in Trieste. In fig. 13.13 is shown a scheme of the nanofabrication process. In Fig. 13.14 evidence of the Proximity Effect Correction (PEC) applied to the pattern to optimize exposure is shown.

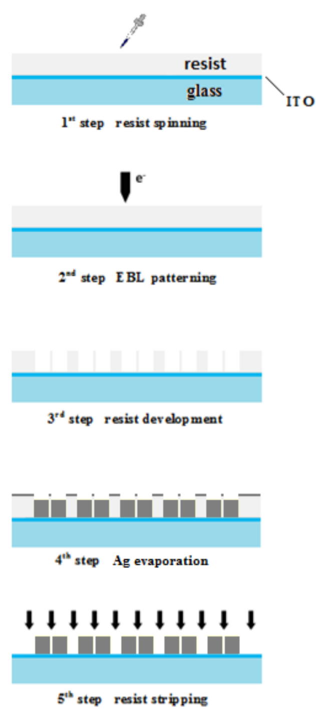


Fig. 13.13. Scheme of the nanofabrication process.

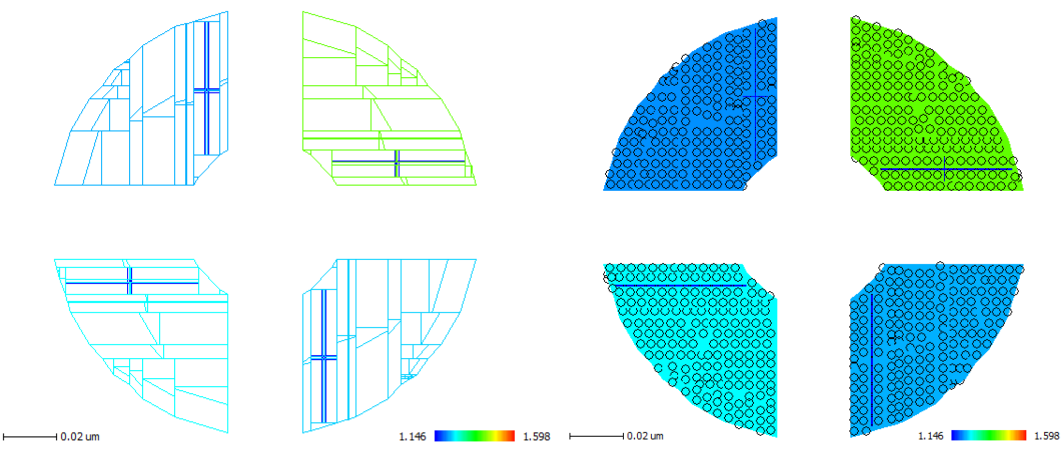


Fig. 13.14. PEC of a MSRR to optimize exposure. Left: fracturing of the pattern is visible. Right: evidence of shot beam arrangement (spot diameter 2nm).

The following SEM images (Fig. 13.15) show the good quality of the high resolution MSRR nanostructures fabricated. The geometrical parameters of the structures are: period 230nm, radius 80nm, width 50nm, gap 28nm. The chips produced have a size of 500x500µm.

Fig. 13.15. SEM images at increasing magnification of the multiple-split ring resonator array fabricated.

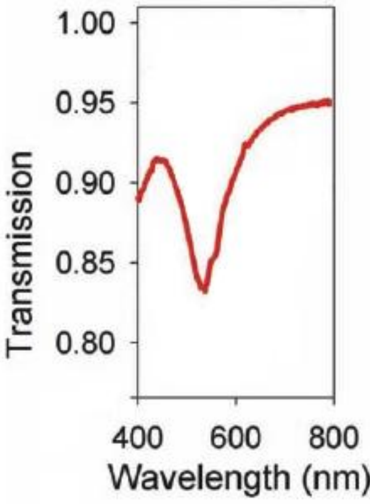


Fig. 13.16. Transmission spectrum of a MSRR array of period 230nm, radius 80nm, width 50nm and gap 6nm, showing the peak plasmonic-resonance of the structures. The resonance wavelength turns out to be independent on the field polarization. [76]

Conclusions

We have been able to fabricate circular multiple-split ring antennas with feature sizes significantly small (radius 80nm, width 50nm, gap 28nm). Optical characterization are in progress to collect ellipsometric transmission spectra to compare with simulations.

Gaining precise control over nanometal's morphology will enable us to spectrally tune and confine the localised plasmon resonances (see Fig. 13.16) and their associated field enhancements to specific areas of the ring. Such control allows to manipulate location, density, volume and electromagnetic intensity of field hot spots available for sensing. Achieving very small gaps allow to confine the plasmon to nanometric volumes, generating the enhancements required for extreme Raman sensitivity. As such, the MSRRs sensors fabricated in this research represent a promising step toward reproducible single-molecule detection using complex engineered plasmonic substrates.

Conclusions

The object of the PhD research was the design, modeling, fabrication and experimental characterization of plasmonic nanostructures to be used as biosensors.

In order to fabricate SERS-active substrates suitable for sensing applications, a well-tested nanofabrication process for the realization of an easily manufacturable and reproducible device was developed, exploring the different mechanisms of light transmission through 1D digital metallic gratings in order to optimize the optical response and efficiency of the device.

A further development of the project has concerned the realization of novel types of sensing nanodevices, coupling the biosensing properties to MetaMaterials to explore the peculiar properties of such kind of structures in combination with plasmonic nano-optics.

The strategic innovation of the project has consisted in the combination of theoretical tools (the finite elements method COMSOL Multiphysics) to foresee and tailor the features of arrays of periodic nanostructures to be used as sensors, nanofabrication techniques (Electron Beam Lithography, X-Ray Lithography) to develop and realize the devices, and nano-optical characterization techniques such as Surface Enhanced Raman Spectroscopy (SERS), Scanning Near-Field Microscopy (SNOM) and Ellipsometry to analyze and measure the response of the nanostructures and compare it with the theoretical simulations.

In the first part of the project Extraordinary Optical Transmission of 1D metallic gratings was studied. Numerical simulations have enabled us to identify the different resonance mechanisms that generate EOT and to give quantitative results in terms of far-field transmission intensity and near-field EM field configurations. In particular, the TM polarization has been demonstrate to support both SPP and cavity modes in hybrid configurations providing the formation of so called ‘hot spots’ inside the slits, characterized by a strong increase of the local electromagnetic field.

The aim of the second part of the project was to theoretically optimize and fabricate 1D digital gold grating devices, able to concentrate the electromagnetic radiation inside the slits. Such plasmonic nanostructures can generate strong light concentration in such a way that Raman signal can be increased by several orders of magnitude. The SERS Enhancement Factor (SSEF) of the realized nanostructures and its dependence on a generic hybrid polarization have been measured and compared with the theoretical estimations.

In order to fabricate SERS-active substrates suitable for sensing applications, reproducibility and efficiency issues must be tackled. The project aimed at developing a well-tested nanofabrication process for the realization of sensing devices easily manufacturable and reproducible and characterized by a good uniformity of the response. To test the reliability of the process a number of samples has been made using the same fabrication parameters and characterized under the same experimental conditions.

A good agreement between the experimental and numerical $SSEF_{TM}$ to $SSEF_{TE}$ ratio, added up to a well-tested nanofabrication process able to guarantee a good uniformity and reproducibility of the device, have confirmed the potentiality of the nanofabrication technique in substrate engineering to provide local fields of the desired intensity and location.

This result is a good starting point for tailoring the properties (like field distribution) of more complex nanostructures, which can be optimized and exploited for example in SERS and micro-fluidics applications.

A further part of the project, still in progress, have concerned the realization of plasmonic nano-optics through MetaMaterials (MMs).

MetaMaterials are artificial materials engineered to have properties, which usually gain from structure rather than composition, that may not be found in nature. The research in this area is interdisciplinary and involves fields such as optoelectronics, material sciences and nanoscience. Potential applications of MetaMaterials are diverse and include, among the others, sensor detection.

In particular gold Split Ring Resonators (SRRs), a MetaMaterial exhibiting plasmonic properties, can be exploited for sensing small quantities of biological and chemical substances.

The aim of this part of the project was to design and fabricate two different kind of Split Ring Resonators. The goal was to create optically active nanostructures able to concentrate the field within 'hot spots', thus resulting in an amplification of the Raman signal, coupling the biosensing properties to peculiar properties of MetaMaterials.

The aim of the first MetaMaterial's project was to exploit the increase in sensitivity produced by a high aspect ratio of the SRR structure. A fabrication process developed combining EBL and XRL tecniques has enabled us to obtain large area samples with high aspect ratio and good lithographic quality.

In the second MetaMaterial's project, by using electron beam lithography circular multiple-split ring antennas (MSRRs) with very small feature sizes have been engineered into silver nanophotonic ring structures, to create concentrated areas of localized field coupling which can be exploited for enhanced plasmonic applications.

Achieving very small gaps allows to confine the plasmon to nanometric volumes, generating the enhancements required for extreme Raman sensitivity. As such, the MSRRs sensors fabricated in this research represent a promising step toward future reproducible single-molecule detection using engineered plasmonic substrates.

Projects and Publications

Research projects:

- Progetto di Eccellenza 2008-2009 della fondazione CARIPARO: "Surface PLasmonics for Enhanced Nano Detectors and Innovative Devices (SPLENDID)";

- Progetto Strategico di Ateneo 2008: "PLAsmonic nano-Textured materials and architectures FOR enhanced Molecular Sensing (PLATFORMS)";

- PRIN(Programma di ricerca scientifica di Rilevante Interesse Nazionale): "Design and fabrication of nano-based plasmon optics designed for interaction with structured light with orbital angular momentum".

Publications:

- "Extraordinary optical transmission in one-dimensional gold gratings: near- and far-field analysis" [63], Applied Optics **50**, 4529 (2011);
- "Polarization independence of extraordinary transmission trough 1D metallic gratings" [67], Optics Express **19**, 9426 (2011);
- "Design, fabrication and characterization of plasmonic gratings for SERS" [68], Microelectronic Engineering **88**, 2717 (2011);
- "Fabrication of metamaterials in the optical spectral range" [80], Microelectronic Engineering **88**, 1951 (2011);
- "Fabrication and testing of phase masks for optical vortex coronagraph to observe extrasolar planets" [81], Proceedings of SPIE-The International Society for Optical Engineering, 7735 (2010);
- "Design, fabrication and characterization of phase masks for astronomical applications" [53], Microelectronic Engineering **88**, 2675 (2011).

Acknowledgments

First of all, I would like to express my gratitude to all the PhD students and reseachers who helped me all along this thesis work: a special thanks goes to Tommaso Ongarello for the optical simulations and the ellipsometric measures, Valentina Giorgis for the electrolytic growths and X-ray lithography, Giuseppe Parisi for MetaMaterial simulations, Marta Carli for AFM characterizations, Gianluca Ruffato and Denis Garoli for ellipsometric measures, Marco Natali for evaporations, Pierfrancesco Zilio for COMSOL assistance and Tommaso Ongarello, Giuseppe Pirruccio, Valentina Giorgis and Marta Carli for sharing part of the work.

I am really grateful to Roberto Pilot for the SERS measurements and useful discussions and for the unvaluable assistance during the reviewing phase.

I am also grateful to Piero Schiavuta, Francesco Marinello and Roberto Pierobon for their precious advice and support during my stay at NanoFAB and for helping me with AFM, SNOM and optical profilometer measurements.

A special thank goes to my supervisor, Prof. Filippo Romanato, for his precious advice and support during these years and for giving me the opportunity to take part of the 'LaNN' adventure.

I express my gratitude also to Prof. Paolo Mazzoldi, for having always supported me.

A great thanks also to Davide Rurale, Sandro Vestini and Kazunori Kato (Kato-san) for their unvaluable advice and assistance with the EBL machine.

I would like also to thank Mauro Prasciolu for his helpfulness and Claudio De Lazzari and Simone Brusa for their patience and problem solving skill.

Thanks to my wife Raffaella for never precluding me any road, standing by my side and believing in me and for being a mother to the world.

Thanks to my parents for having always been there when I needed their help.

Thanks to my sister, to my friends and to all who I have forgotten here, but they deserve to.

Finally, a huge kiss to my kids, Gabriele and Emanuele, for whom I did all of this and I love every day more.

Bibliography

[1] Wood R. W., *Philosophical Magazine* **4**, 396 (1902).
[2] Fano U., *J Opt Soc Am* **31**, 231 (1941).
[3] Turbadar T., *Proc Phys Soc* **73**, 40 (1959).
[4] Otto A., *Zeits Phys* **216**, 398 (1968).
[5] Kretschmann E., Raether H., *Z Naturforsch* **A 23**, 2135 (1968).
[6] Homola J. (Ed.), "*Surface Plasmon Resonance Based Sensors*", Springer (2006).
[7] Maier S. A., "*Plasmonics: Fundamentals and Applications*", Springer (2007).
[8] Offerhaus H. L. et al., *Nano Letters* **5**, 2144 (2005).
[9] Ebbesen T. W. et al., *Nature* **391**, 667 (1998).
[10] Garcia-Vidal F. J. et al., *Reviews of Modern Physics* **82**, 729 (2010).
[11] Yang F., Sambles J. R., *Phys. Rev. Lett.* **89**, 63901 (2002).
[12] Schroter U., Heitmann D., *Phys. Rev. B* **58**, 15419 (1998).
[13] Porto J. A. et al., *Phys. Rev. Lett.* **83**, 2845 (1999).
[14] Marquier F. et al., *Opt. Express* **13**, 70 (2005).
[15] Zilio P. et al., *Optics Express* **18**, 19558 (2010).
[16] Barbara A. et al., *Phys. Rev. B* **66**, 161403 (2002).
[17] Pang Y. et al., *Opt. Commun.* **280**, 10 (2007).
[18] Le Ru E., Etchegoin P., "*Principle of Surface Enhanced Raman Spectroscopy*", Elsevier (2009).
[19] Aroca R., "*Surface-enhanced vibrational spectroscopy*", John Wiley & Sons (2006).
[20] Lombardi J. R. et al., *J. Chem. Phys.* **84**, 4174 (1986).
[21] Kerker M. et al., *Appl. Opt.* **19**, 4159 (1980).
[22] Rojas R., Claro F., *J. Chem. Phys.* **98**, 998 (1993).
[23] Dulkeith E. et al., *Phys. Rev. Lett.* **89**, 203002 (2002).
[24] Novotny L., Hecht B., "*Principles of nano-optics*", Cambridge University Press (2006).
[25] Jackson J. D., "*Classical electrodynamics*", Wiley (1998).
[26] Barnett S. M. et al., *Phys. Rev. Lett.* **68**, 3698 (1992).
[27] Broers A., *J. Electrochem. Soc.: Solid State Sci. & Technol.* **128**, 166 (1981).
[28] Liddle J. et al., *Microlithography World* **6**, 15 (1997).
[29] Murata K. et al., *Jpn. J. Appl. Phys.* **10**, 678 (1971).
[30] Parikh M., Kyser D.F., *J. Appl. Phys* **50**, 1104 (1979).
[31] Cui Z., "*Nanofabrication*", Springer (2008).
[32] Chang T. et al., *IBM J. Res. Develop.* **32**, 464 (1988).
[33] Broers A. N., *IEEE Electron. Dev.* **ED-28**, 1268 (1981).
[34] Owen G., Rissman P., *J. Appl. Phys.* **54**, 3573 (1983).
[35] Dobisz E. et al., *Journal of Vacuum Science and Technology B* **7**, 1507 (1989).
[36] Hinsberg W. et al., *Journal of Vacuum Science and Technology B* **16**, 3689 (1998).
[37] Dobisz E. A., Marrian C. R. K., *Proceedings of SPIE* **2723**, 383 (1996).
[38] McCord M. A. et al., *Journal of Vacuum Science and Technology B* **11**, 2958 (1993).
[39] Dobisz E. A., Marrian C. R. K., *Journal of Vacuum Science and Technology B* **15**, 2327 (1997).
[40] Haller I. et al., *IBM J. Res. Develop.* **12**, 251 (1968).
[41] Yasin S. et al., *Microelectronic Engineering* **61**, 745 (2002).

[42] Yasin S., "*Nanotechnology using Electron Beam Lithography and Ultrasonically Assisted Development in Organic resists*", *Ph.D. Thesis*, University of Cambridge (2001).
[43] Meyerhofer D., *J. Appl. phys.* **49**, 3993 (1978).
[44] Chen W., Ahmed H., *Appl. Phys. Lett.* **62**, 1499 (1993).
[45] Hu W. et al., *J. Vac. Sci. Technol.* **B22**, 1711 (2004).
[46] Ocola L.E., Stein A., *J. Vac. Sci. Technol.* **B24**, 3061 (2006).
[47] Hatzakis M., *J. Vac. Sci. Technol.* **16**, 1984 (1980).
[48] Wendt H., "*Electrolysis in: Ullman's Encyclopedia of Industrial Chemistry*", WILEY-VCH (1973).
[49] Tompkins H. G., Irene E. A. (Ed.), "*Handbook of Ellipsometry*", Springer (2005).
[50] Torzo G., "*Introduction to Scanning Probe Microscopy*", IMN Interuniversity Master in Nanotechnologies (2007).
[51] Schiavuta P., "*Near-Field Optics and Scanning Near-Field Optical Microscopy (SNOM)*", IMN Interuniversity Master in Nanotechnologies (2007).
[52] Bernstein G. H. et al., *Journal of Applied Physics* **71**, 4066 (1992).
[53] Massari M. et al., *Microelectron. Eng.* **88**, 2675 (2011).
[54] Yasin S. et al., *J. Vac. Sci. Technol.* **B22**, 574 (2004).
[55] Shun-Chi Chang et al., *J. Am. Chem. Soc.* **116**, 6792 (1994).
[56] Nara J. et al., *J. Phys. Chem. C* **120**, 6705 (2004).
[57] Li-Jun Wan et al., *J. Phys. Chem. B* **104**, 3563 (2000).
[58] Yu-Tai Tao et al., *Langmuir* **13**, 4018 (1997).
[59] Whelan C. M. et al., *Langmuir* **15**, 116 (1999).
[60] Johansson A., Stafstroem S., *Chem. Phys. Lett.* **322**, 301 (2000).
[61] Bauschlicher Jr. C.W., Ricca A., *Chem. Phys. Lett.* **367**, 90 (2003).
[62] Joo T. H. et al., *J. Raman Spectr.* **18**, 57 (1987).
[63] Romanato F. et al., *Applied Optics* **50**, 4529 (2011).
[64] Garcia-Vidal F. J. et al., *Phys. Rev. B* **66**, 155412 (2002).
[65] Crouse D., Keshavareddy P., *Opt. Express* **13**, 7760 (2005).
[66] Crouse D., Keshavareddy P., *Opt. Express* 15, 1415 (2007).
[67] Ongarello T. et al., *Opt. Express* **19**, 9426 (2011).
[68] Romanato F. et al., *Microelectron. Eng.* **88**, 2717 (2011).
[69] Reilly T.H. et al., *J. Phys. Chem. C* **111**, 1689 (2007).
[70] Soukoulis Costas M., Wegener M., *Nature Photonics* **5**, 523 (2011).
[71] Shelby R. A. et al., *Science* **292**, 77 (2001).
[72] Linden S. et al., *Science* **306**, 1351 (2004).
[73] Clark A. W. et al., *J. Am. Chem. Soc.* **131**, 17615 (2009).
[74] Gorelick S. et al., *Nanotechnology* **21**, 295303 (2010).
[75] Peralta X.G. et al., *Appl. Phys. Lett.* **94**, 161113 (2009).
[76] Al-Naib I.A.I. et al., *Appl. Phys. Lett.* **93**, 083507 (2008).
[77] Chiam S. et al., *Appl. Phys. Lett.* **94**, 064102 (2009).
[78] Clark A.W., Cooper J.M., *Adv. Mater.* **22**, 4025 (2010).
[79] Clark A.W., Cooper J.M., *small* **7**, 119 (2011).
[80] Garoli D. et al., *Microelectron. Eng.* **88**, 1951 (2011).
[81] Mari E. et al., *Proceedings of SPIE-The International Society for Optical Engineering*, 7735 (2010).