
E	·	T	H	L
P		F	L	
S	F	I	T	L



Faculty STI – Section of Microtechnology – Institute of Applied Optics
October 2002 to February 2003

Design and Evaluation of subwavelength Apertures for real time Fluorescence Correlation Spectroscopy

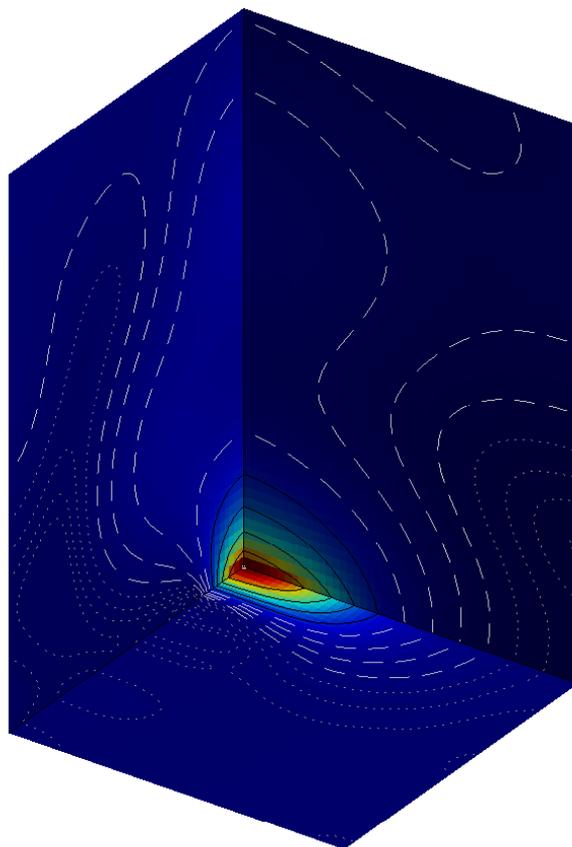
Diploma work

21 February 2003

Marcel Leutenegger Adrian von Mühlengen

monitored by

Prof. Theo Lasser Dr. Tiemo Anhut



This work could not be finished without the help and generous support of many people. In particular, we are indebted to Prof. Theo Lasser, Prof. Patrick Hoffmann, Dr. Ivo Utke, Dr. Tiemo Anhut, Kai Hassler and Alexandre Perentes at EPFL. Also, we would like to thank Prof. Olivier Martin at ETHZ and Prof. Rudolf Rigler at Karolinska Institutet.

We are grateful to everybody else who has supported this work.

Lausanne, 21 February 2003

Marcel Leutenegger

Adrian von Mühlengen

Many chemical reactions and biological processes were investigated by fluorescence correlation spectroscopy *FCS*. Often, a confocal microscope is used to locally excite a molecule group containing very few fluorophores. The *FCS* method yields a high signal to noise ratio *SNR* if very few - at best one or less - fluorophores are excited. Hence, *FCS* provides an elegant method to measure single molecule behaviour and parameters.

This paper describes a new approach to confine the excitation light simultaneously in a multitude of light fields. In contrast to the classical approach based on confocal microscopy and limited by diffraction, we use a thin metal film containing periodic nanoholes. This approach paves the way to parallel *FCS* based on a simple microoptical system and promising light confinement beyond the limits of confocal optics. We developed a computation method and designed a sample mask by simulation.

Contents

1	Introduction	3
2	Theory	4
2.1	The model	4
2.2	Maxwell equations	4
2.3	Boundary conditions	5
2.4	Surface impedance	7
2.5	Permittivity	9
2.6	Resonance	16
2.6.1	Surface plasmon	17
3	Slit model	22
3.1	Surface impedance approximation	22
3.2	Discussion of the model	24
4	Basic concepts	27
4.1	Interface	27
4.2	Layers	29
4.3	Diffraction	30
4.4	Propagation through apertures	31
4.5	Evanescence waves	34
4.6	Electromagnetic intensity	35
5	Metallic slits	37
5.1	Field representation	37
5.2	Evaluation	39
5.3	Slit array	41
5.4	Rayleigh expansion	42
5.5	Performance trade-off	44
5.6	Example	45
5.7	Conceptual characteristics	46
6	Metallic holes	47
6.1	Geometry	47
6.2	Field representation	47
6.3	Evaluation	49
6.4	Example	51

7	Calculations	52
7.1	Surface plasmons	52
7.2	Light fields	54
8	Applications	55
8.1	FCS setup	55
8.2	FCS	56
8.3	Cross-correlation FCS	57
9	Conclusions	58
A	Dielectric constant of metals	61
B	ATR	64
B.1	Kretschmann-Raether configuration	64
B.2	Otto configuration	65
C	Implementation	66
C.1	Metallic slits	66
C.2	Metallic holes	70
C.3	Fresnel coefficients	76
C.4	Status bar	77
C.5	User interface	78
C.6	Dielectric constant of metals	83

1 Introduction

Today, fluorescence correlation spectroscopy *FCS* is widely used to observe chemical reactions and detect single molecules. The technology advances towards high-speed DNA sequencing and real time investigation of living cells. In this paper, we investigate a novel FCS setup to reach these goals.

Chapters 2 and 3 show the theoretical background; chapters 5 and 6 present the preliminary design of the core part for the FCS setup outlined in chapter 8.

Challenges:

1. The investigation of living cells asks for FCS at fluorophore concentrations $C_f \approx 10\mu\text{M}$. For optical excitation, live cell FCS needs a **smaller spot volume** $V_e \approx 2 \cdot 10^{-19}\text{l}$. This is about three orders of magnitude smaller than the actual confocal spot volume.
2. In a FCS experiment, molecules have to drift through the excitation spot. The relatively slow motion imposes an observation time $\Delta t \approx 10..100\text{ms}$ while maintaining a sampling frequency $f \gtrsim 20\text{MHz}$. This means that a confocal microscope is able to capture $\approx 10\text{FCS}$ measures per second. In contrast, the **parallelisation of the FCS process** will be able to speed up DNA sequencing far beyond the current values.

Trying to answer the mentioned challenges, this paper presents a novel FCS method. We design and evaluate a mask of nanoholes in a thin metal film. The mask is backlit by a laser beam and confines it into a multitude of small light fields. Particular nanoholes are believed to approach the characteristics of a point light source. Note that live cell FCS needs a spot radius $r \lesssim 60\text{nm}$. We can control its size adjusting the dimensions of the nanoholes.

The sample is put in contact with the mask and observed by means of a standard microscope. We propose to use a novel array detector¹ to measure the fluorescence light. Only the number of detectable nanoholes limits the parallelisation. Nevertheless, we may not be able to investigate a 'thick' sample.

¹A standard CCD array is not able to provide the necessary sampling frequency. But we may profit from the current development by the group of Prof. Radivoje Popovic at the Microsystems Laboratory at the EPFL.

2 Theory

2.1 The model

Our model describes a metallic layer with slits or square holes smaller than the wavelength of the incident light and allowing a periodicity as small as $1 \dots 2$ wavelengths. We separate the space in four different regions, characterised by their material constants, the permittivity ϵ and the permeability μ . A linear polarised plane wave incident from region I (free space or dielectric medium) interacts at the interface $I \rightarrow II$, $I \rightarrow III$ with the metallic layer containing openings in a periodic configuration. Region II and III are on top of region IV . Region II is characterised by a metallic material whereas III contains a transparent optical material.

The plane wave incident from region I excites waveguide modes in region III . These waveguide modes propagate in region III and couple again to plane waves in region IV .

Each linear polarised plane wave can be decomposed in a *transverse electric* wave (TE = s-polarised) and a *transverse magnetic* wave (TE = p-polarised) as outlined in figures 1(a) and 1(b).

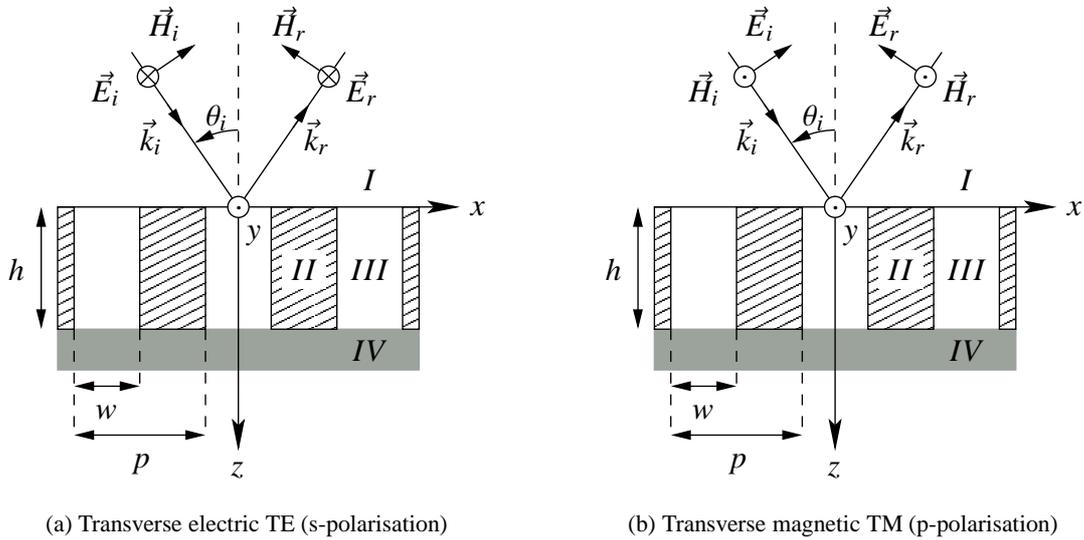


Figure 1: Model outline for s- and p-polarisation.

2.2 Maxwell equations

The electric field $\vec{\mathcal{E}}_{(\vec{x},t)}$ and the magnetic field $\vec{\mathcal{H}}_{(\vec{x},t)}$ are vector functions in position and time and obey the Maxwell equations

$$\vec{\nabla} \times \vec{\mathcal{E}} = \mu \frac{\partial \vec{\mathcal{H}}}{\partial t} \quad (1a) \quad \vec{\nabla} \times \vec{\mathcal{H}} = \vec{\mathcal{J}} + \frac{\partial \vec{\mathcal{D}}}{\partial t} \quad (1c)$$

$$\vec{\nabla} \cdot \vec{\mathcal{D}} = \rho \quad (1b) \quad \vec{\nabla} \cdot \vec{\mathcal{B}} = 0 \quad (1d)$$

The electromagnetic field-matter interaction involves the electric flux density $\vec{\mathcal{D}}$, the current density $\vec{\mathcal{J}}$ and the magnetic flux density (or magnetic induction) $\vec{\mathcal{B}}$

$$\vec{\mathcal{B}} = \mu_r \mu_0 \vec{\mathcal{H}} = \mu \vec{\mathcal{H}} \quad (2a) \quad \vec{\mathcal{J}} = \sigma \vec{\mathcal{E}} \quad (2b) \quad \vec{\mathcal{D}} = \epsilon_0 \epsilon_r \vec{\mathcal{E}} = \epsilon \vec{\mathcal{E}} \quad (2c)$$

Combining the Maxwell equations (1a) and (1c) and (2a) results in *the vectorial wave equation*²

$$\begin{aligned}\vec{\nabla} \times (\vec{\nabla} \times \vec{\mathcal{E}}) &= \mu \nabla \times \frac{\partial \mathcal{H}}{\partial t} = \mu \left(\frac{\partial \mathcal{J}}{\partial t} + \epsilon \frac{\partial^2 \vec{\mathcal{D}}}{\partial t^2} \right) \\ &= \mu \left(\sigma \frac{\partial \vec{\mathcal{E}}}{\partial t} + \epsilon \frac{\partial^2 \vec{\mathcal{E}}}{\partial t^2} \right)\end{aligned}\quad (3)$$

For a plane wave $\vec{\mathcal{E}}_{(\vec{r},t)} = \vec{E}_{(\vec{r})} \exp\{-i\omega t\}$ we can write $\frac{\partial}{\partial t} \vec{\mathcal{E}} = -i\omega \vec{\mathcal{E}}$ and $\frac{\partial^2}{\partial t^2} \vec{\mathcal{E}} = -\omega^2 \vec{\mathcal{E}}$. In addition the vector relation $\vec{\nabla} \times (\vec{\nabla} \times \vec{a}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{a}) - \nabla^2 \vec{a}$ ³ allows us to write the time independent wave equation as

$$\begin{aligned}\vec{\nabla} \underbrace{(\vec{\nabla} \cdot \vec{E})}_{=0 \text{ if } \rho=0} - \nabla^2 \vec{E} &= \mu_r \mu_0 (-i\omega\sigma - \omega^2 \epsilon) \vec{E} \\ \nabla^2 \vec{E} - \omega^2 \epsilon_0 \mu_0 \mu_r \left(\epsilon_r + i \frac{\sigma}{\omega \epsilon_0} \right) \vec{E} &= 0\end{aligned}\quad (4)$$

The dielectric constant is complex as we will use metallic as well as dielectric materials: $\underline{\epsilon} = \epsilon' + i\epsilon''$. For the light velocity we have $c_0 = (\sqrt{\epsilon_0 \mu_0})^{-1}$

$$\begin{aligned}\vec{\nabla}^2 \vec{E} + \frac{\omega^2}{c_0^2} \mu_r \left(\epsilon_r + i \frac{\sigma}{\omega \epsilon_0} \right) \vec{E} &= 0 \\ \vec{\nabla}^2 \vec{E} - \frac{\omega^2}{c_0^2} \mu_r (\epsilon' + i\epsilon'') \vec{E} &= 0 \\ \vec{\nabla}^2 \vec{E} - \frac{\omega^2}{c_0^2} \mu_r \underline{\epsilon} \vec{E} &= 0\end{aligned}$$

Using $k_0^2 = \frac{\omega^2}{c_0^2}$ and $n^2 = \mu_r \underline{\epsilon}$ we find the *Helmholtz equation*

$$\vec{\nabla}^2 \vec{E} - k^2 n^2 \vec{E} = 0 \quad (5)$$

or the scalar wave equation

$$\vec{\nabla}^2 \phi - k^2 n^2 \phi = 0 \quad (6)$$

2.3 Boundary conditions

At any interface, the electromagnetic field has to fulfil the boundary conditions that will be derived using the Maxwell equations.

Let ΔV be a finite cylindrical volume element containing the boundary (see figure 2). Using the Maxwell equation (1b) we can write for the electromagnetic field in the volume ΔV using the *Gauss law*

$$\iiint_{\Delta V} \vec{\nabla} \cdot \vec{\mathcal{D}} \, dx dy dz = \oiint_{\Delta S} \vec{\mathcal{D}} \cdot \vec{n} \, dS = \iiint_{\Delta V} \rho \, dx dy dz \quad (7)$$

²Here, the permittivity μ was assumed to be constant

³The Laplacian is applied to each component of the vector \vec{a} :
 $\vec{\nabla} \cdot \vec{\nabla} \vec{a} = \vec{\nabla}^2 \vec{a} = \{(\nabla^2 a_{1,(x,y,z)}, \nabla^2 a_{2,(x,y,z)}, \nabla^2 a_{3,(x,y,z)})\}$

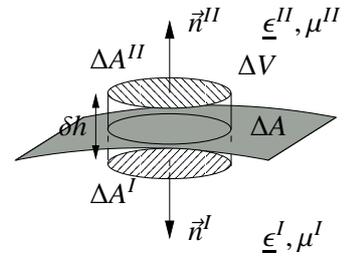


Figure 2: The Gauss law.

If the surfaces ΔA^I and ΔA^{II} are sufficiently small, the field at ΔA is constant: $\vec{\mathcal{D}}^I, \vec{\mathcal{D}}^{II}$. This results in

$$\vec{\mathcal{D}}^I \cdot \vec{n}^I \Delta A^I + \vec{\mathcal{D}}^{II} \cdot \vec{n}^{II} \Delta A^{II} + \text{wall contribution} = \iiint_{\Delta V} \rho \, dx dy dz \quad (8)$$

The wall contribution vanishes if we shrink the cylinder height to zero. Any charge inside the volume element ΔV is contained in the surface charge density ρ_s and for the unit vectors pointing outwards $\vec{n}^{I-II} = -\vec{n}^I = \vec{n}^{II}$, we find

$$\begin{aligned} (\vec{\mathcal{D}}^I \cdot \vec{n}^I + \vec{\mathcal{D}}^{II} \cdot \vec{n}^{II}) \Delta A &= \lim_{h \rightarrow 0} \iiint_{\Delta V} \rho \, dx dy dz \\ \vec{n}^{I-II} \cdot (\vec{\mathcal{D}}^{II} - \vec{\mathcal{D}}^I) \Delta A &= \iint_{\Delta A} \rho_s \, dA = \rho_s \Delta A \\ \vec{n}^{I-II} \cdot (\vec{\mathcal{D}}^{II} - \vec{\mathcal{D}}^I) &= \rho_s \end{aligned} \quad (9)$$

In presence of a layer with surface charge density ρ_s , the normal component of $\vec{\mathcal{D}}$ to the surface changes across the interface.

Using analogue the Maxwell equation (1d) the Gauss law for the magnetic field can be deduced

$$\iiint_{\Delta V} \vec{\nabla} \cdot \vec{\mathcal{B}} \, dx dy dz = \oiint_{\Delta S} \vec{\mathcal{B}} \cdot \vec{n} \, dS = 0 \quad (10)$$

Thus no magnetic monopole can exist.

Similar to (8) we find

$$\vec{n}^{I-II} \cdot (\vec{\mathcal{B}}^{II} - \vec{\mathcal{B}}^I) = 0 \quad (11)$$

The normal component of the magnetic field to the surface is continuous across the surface separating the materials.

According to figure 3, let $\Delta A = \square(\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{D})$ be a rectangular loop containing the boundary. The vector \vec{b} is the unit normal vector of the surface enclosed by the ΔA . The path elements $\overline{\mathcal{A}\mathcal{B}} = \delta l^I$ and $\overline{\mathcal{C}\mathcal{D}} = \delta l^{II}$ are parallel to the boundary. Using the Maxwell equation (1a) and the Stokes theorem we find

$$\iint_{\Delta A} \vec{\nabla} \times \vec{\mathcal{E}} \cdot \vec{b} \, dS = \oint_C \vec{\mathcal{E}} \cdot d\vec{s} = \mu \iint_{\Delta A} \frac{\partial \vec{\mathcal{H}}}{\partial t} \cdot \vec{b} \, dS$$

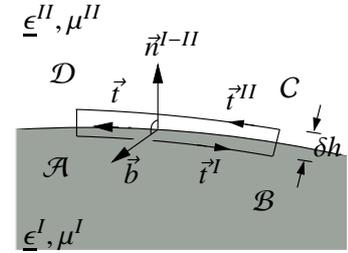


Figure 3: The Stokes law.

If the surface ΔA is sufficiently small, the electric $\vec{\mathcal{E}}^I, \vec{\mathcal{E}}^{II}$ and the magnetic field $\vec{\mathcal{B}}$ are constant on ΔA . The cross contribution as well as the contribution from the magnetic field vanish if we shrink the height δh to zero:

$$\begin{aligned} \vec{\mathcal{E}}^I \vec{t}^I \delta l^I + \vec{\mathcal{E}}^{II} \vec{t}^{II} \delta l^{II} + \text{cross contribution} &= \vec{\mathcal{B}} \cdot \vec{b} \delta h \delta l \\ (\vec{\mathcal{E}}^I \vec{t}^I + \vec{\mathcal{E}}^{II} \vec{t}^{II}) \delta l &= 0 \end{aligned}$$

$\vec{t} = \vec{b} \times \vec{n}^{I-II}$ is the unit tangential vector to the surface. Therefore $\vec{t}^I = -\vec{t}$ and $\vec{t}^{II} = \vec{t}$.

$$\begin{aligned} \vec{b} \cdot (\vec{n}^{I-II} \times (\vec{\mathcal{E}}^{II} - \vec{\mathcal{E}}^I)) &= 0 \\ \vec{n}^{I-II} \times (\vec{\mathcal{E}}^{II} - \vec{\mathcal{E}}^I) &= 0 \end{aligned} \quad (12)$$

The components of the electric field tangential to the interface are continuous across the surface.

Similar we can deduce using the Maxwell equation (1c) and being aware of the surface current density \mathcal{J}_s

$$\begin{aligned} \mathcal{H}^I \hat{t}^I \delta l^I + \mathcal{H}^{II} \hat{t}^{II} \delta l^{II} + \text{cross contribution} &= \vec{\mathcal{J}} \cdot \vec{b} \delta l + \vec{\mathcal{J}} \cdot \vec{b} \delta h \delta l \\ \vec{b} \cdot (\vec{n}^{I-II} \times (\mathcal{H}^{II} - \mathcal{H}^I)) &= \vec{\mathcal{J}}_s \cdot \vec{b} \\ \vec{n}^{I-II} \times (\mathcal{H}^{II} - \mathcal{H}^I) &= \vec{\mathcal{J}}_s \end{aligned} \quad (13)$$

The tangential components of the magnetic field to the interface are discontinuous across the boundary surface.

Summary:

$$\vec{\mathcal{D}}_{\perp}^{II} - \vec{\mathcal{D}}_{\perp}^I = \rho_s \quad (9) \quad \vec{\mathcal{E}}_{\parallel}^{II} - \vec{\mathcal{E}}_{\parallel}^I = 0 \quad (12)$$

$$\vec{\mathcal{B}}_{\perp}^{II} - \vec{\mathcal{B}}_{\perp}^I = 0 \quad (11) \quad \mathcal{H}_{\parallel}^{II} - \mathcal{H}_{\parallel}^I = \vec{\mathcal{J}}_s \quad (13)$$

2.4 Surface impedance

In chapter 2.3 we have shown that tangential components of the electric field must be continuous.

The boundary condition (13) is only valid for a perfect conductor (characterised by infinite conductivity). The surface current $\vec{K} = \vec{n} \times \vec{\mathcal{H}}$ is the response due to the time varying tangential magnetic field (which leads to zero magnetic field inside the perfect conductor). The charges of a conductor with a finite conductivity have finite mobility and therefore the response is not instantly any more. Hence, the magnetic field penetrating in the conductor is exponentially attenuated on a characteristic length: the skin depth $\delta = \sqrt{2/(\mu\omega\sigma)}$. Therefore, the tangential magnetic field at the boundary of a non-perfect conductor has to be continuous ⁴ and equation (13) is modified to

$$\vec{n}^{I-II} \times (\mathcal{H}^{II} - \mathcal{H}^I) = 0 \quad (14)$$

Landau [19] postulated that these parallel fields are proportional and are related by the ‘‘surface impedance’’ $Z = \sqrt{\mu\epsilon^{-1}}$ ⁵ if $|\frac{\epsilon'}{\epsilon''}| \gg 1$. ⁶

$$\vec{E}_{\parallel} = Z \vec{n} \times \vec{H}_{\parallel} \quad (15)$$

The equation (15) describes the light - matter interaction through the physical constants permittivity ϵ and the permeability μ . Consequently, the surface impedance is an adequate representation of the boundary condition. Therefore we investigate the two particular cases of a linear polarised plane wave, the TE (s-polarisation) and the TM (p-polarisation) wave:

Transverse electric waves:

The plane of incidence lies in the (x,z)-plane and the interface is in the (x,y)-plane (see figure 1(a)). The tangential field of the incident wave \vec{E}_{\parallel}^I has only a y-component E_y^I . Using (12) and (15) we find

$$\vec{E}_{\parallel}^I - \vec{E}_{\parallel}^{II} = E_y^I - E_y^{II} = 0$$

⁴For a detailed discussion of the fields at the surface of a conductor, refer [14].

⁵The impedance of free space is $Z_0 = \sqrt{\mu_0/\epsilon_0} = 376, 73\Omega$

⁶Remember the complex notation of the dielectric constant: $\underline{\epsilon} = \epsilon' + i\epsilon''$

With equations (15) and (14), we can express the fields in only one half space

$$E_y^I \vec{e}_y - Z^{II} \vec{n} \times \vec{H}_{\parallel}^{II} = E_y^I \vec{e}_y - Z^{II} \vec{n} \times \vec{H}_{\parallel}^I = 0$$

$$E_y^I \vec{e}_y - Z^{II} \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \times \begin{pmatrix} H_{\parallel,x}^I \\ H_{\parallel,y}^I \\ H_{\parallel,z}^I \end{pmatrix} = E_y^I \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - Z^{II} \begin{pmatrix} H_{\parallel,y}^I \\ -H_{\parallel,x}^I \\ 0 \end{pmatrix} = 0 \quad (16)$$

Using the Maxwell equation (1a)

$$\begin{pmatrix} \frac{-\partial E_y^I}{\partial z} \\ 0 \\ \frac{\partial E_y^I}{\partial x} \end{pmatrix} = -i\mu^I \omega \begin{pmatrix} H_x^I \\ H_y^I \\ H_z^I \end{pmatrix}$$

$$\Rightarrow H_{\parallel}^I = H_x^I = \frac{-1}{i\mu^I \omega} \frac{-\partial E_y^I}{\partial z}$$

$$= \frac{1}{ik_0 \mu_r^I \sqrt{\mu_0}} \frac{\partial E_y^I}{\partial z}$$

$$= \frac{1}{ik_0 n^I} \frac{1}{Z^I} \frac{\partial E_y^I}{\partial z}$$

Introducing this expression in (16) leads to

$$E_y^I + \frac{1}{ik^I} \frac{Z^{II}}{Z^I} \frac{\partial E_y^I}{\partial z} = 0 \quad (17)$$

The equation (17) is the boundary condition at an interface between a 'good' conductor and a dielectric TE wave. Only the fields in the dielectric half space have to be taken into account.

Transverse magnetic waves:

For the same plane of incidence and the same boundary as above (see figure 1(b)), we recognise that the tangential component of the incident magnetic field \vec{H}^I has only a y-component H_y^I .

Using the equations (12) and (15)

$$\vec{E}_{\parallel}^I - \vec{E}_{\parallel}^{II} = Z^I \vec{n} \times H_y^I \vec{e}_y - \vec{E}_{\parallel}^{II}$$

$$= Z^I H_y^I \vec{e}_x - \vec{E}_{\parallel}^{II} = 0 \quad (18)$$

Applying the Maxwell equation (1c)

$$\begin{pmatrix} \frac{-\partial H_y^{II}}{\partial z} \\ 0 \\ \frac{\partial H_y^{II}}{\partial x} \end{pmatrix} = -i\omega \epsilon_0 \left(\epsilon^{II} + \frac{i\sigma^{II}}{\omega} \right) \begin{pmatrix} E_x^{II} \\ E_y^{II} \\ E_z^{II} \end{pmatrix}$$

$$\Rightarrow E_{\parallel}^{II} = E_x^{II} = \frac{1}{i\omega \underline{\epsilon}^{II}} \frac{\partial H_y^{II}}{\partial z}$$

$$= \frac{1}{ik_0 \sqrt{\epsilon_0} \underline{\epsilon}^{II}} \frac{\partial H_y^{II}}{\partial z}$$

$$= \frac{1}{ik_0 n^{II}} \frac{\partial H_y^{II}}{\partial z} Z^{II}$$

Introducing this in (18)

$$Z^I H_y^I - \frac{1}{ik_0 n^{II}} \frac{\partial H_y^{II}}{\partial z} Z^{II} = 0$$

and applying the boundary condition (13), it follows

$$H_y^I - \frac{1}{ik^{II}} \frac{Z^{II}}{Z^I} \frac{\partial H_y^I}{\partial z} = 0 \quad (19)$$

The equation (19) is the boundary condition for a TM wave at an interface separating a good conductor and a dielectric. Only the fields in the half space of the dielectric have to be taken into account.

2.5 Permittivity

The permittivity describes the ability of the electrons to deflect of their position of rest. This concept reflects the light-matter interaction which evocates the excitation of electrons the permittivity varies with frequency.

Dielectric constant ϵ_0 :

The dielectric constant is the force of 2 electric charges of charge 10^{-4}C separated by 1m: $1F = f \cdot 10^{-8}\text{C}^2\text{m}^{-2}$. Therefore $f = 8,9875 \cdot 10^9\text{N}^2$. We rewrite the constant factor $f = 1/(4\pi\epsilon_0)$. Thus the dielectric constant is defined as

$$\epsilon_0 = 8.854 \cdot 10^{-12} \text{A}^2\text{s}^4\text{kg}^{-1}\text{m}^{-3} \quad [\text{AsV}^{-1}\text{m}^{-1}] \quad (20a)$$

This is also important for the speed of light c_0

$$c_0 = \frac{1}{\sqrt{\epsilon_0\mu_0}}$$

where $c_0 = 299792458\text{ms}^{-1}$ and $\mu_0 = 4\pi \cdot 10^{-7}\text{kgmA}^{-2}\text{s}^{-2}$ is the magnetic permeability of vacuum. Therefore

$$\epsilon_0 = \frac{1}{c_0^2\mu_0} = \frac{1\text{A}^2\text{s}^2}{c_0^2 4\pi \cdot 10^{-7}\text{kgm}} = 8.8542 \cdot 10^{-12} \text{AsV}^{-1}\text{m}^{-1} \quad (20b)$$

Dispersion:

The refraction index of the medium depends on the frequency. For the analysis we introduce the simplified model of a dispersing medium due to H. A. Lorentz [4]: A molecule consists of heavy particles (the atomic nuclei) surrounded by light particles (the electrons). For the sake of simplicity, we exclude polar molecules, in other words, the centre of gravity of the negative charges coincides with the centre of gravity of the positive charges residing in the atomic nucleus. Thus, the resulting momentum is zero.

Averaging the total field over a region derives the properties of the material. As the dimension of the averaging area is larger than the dimension of the molecules, their properties can be described by electric and magnetic dipoles. We write for the polarisation $\vec{P}_{(\vec{r},t)} = \epsilon_0\chi\vec{E}_{(\vec{r},t)}$ and for the magnetisation $\vec{M}_{(\vec{r},t)} = \chi\vec{H}_{(\vec{r},t)}$. Using the material equations (2a) and (2b) allows to write

	MKS ⁷	Gauss	
electric permittivity	$\epsilon_r = 1 + \chi$	$\epsilon_r = 1 + 4\pi\chi$	(21a)
magnetic permeability	$\mu_r = 1 + \chi$	$\mu_r = 1 + 4\pi\chi$	(21b)

Further we must distinguish between the *effective* $\vec{\mathcal{E}}'$ and the *observed* or *mean field* $\vec{\mathcal{E}}$. The first is the field acting on the atoms, the second is obtained by averaging over a region containing a great number of atoms or molecules.

To approximate the difference $\vec{\mathcal{E}}' - \vec{\mathcal{E}}$ consider a molecule to be in a small sphere which is surrounded by a homogeneously polarised medium. The molecular structure outside the sphere is neglected due to the great number of molecules. The medium is treated as continuous. In determining the effects inside the sphere, we assume the molecules to be distributed in a random manner and hence not to produce any resulting field at the central molecule.

We have to describe the potential Φ of that sphere. In other words, the discontinuity at the spherical boundary at which the polarisation changes from the exterior continuous value to the interior value. Considering the “opposite” situation; the potential of a sphere filled with particles embedded in vacuum, $\tilde{\Phi}$. By superposition of both configurations we have a homogeneously polarised material without any boundary. Thus

$$\begin{aligned} \Phi + \tilde{\Phi} &= 0 \\ \text{where} \quad \tilde{\Phi} &= -\vec{P}_{(\vec{r},t)} \cdot \int \vec{\nabla}' R^{-1} dV' \Big|_{(x',y',z')} \\ \text{and} \quad R &= \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2} \\ \text{therefore} \quad \Phi &= -\vec{P}_{(\vec{r},t)} \cdot \underbrace{\vec{\nabla} \left(- \int \frac{dV'}{R} \right)}_{=\Phi_0} \\ \Phi &= -\vec{P}_{(\vec{r},t)} \cdot \vec{\nabla} \Phi_0 \end{aligned} \quad (22)$$

The potential Φ_0 can be interpreted as the potential of a uniformly charged sphere of $\rho = -1\text{Cm}^{-3}$ and follows the Poisson equation ⁸

$$\vec{\nabla}^2 \Phi_0 = 1\text{Vm}^{-2} \quad (23)$$

The field strengths associated with the potential are

$$-\frac{\partial \Phi_0}{\partial x} = P_x \frac{\partial^2 \Phi_0}{\partial x^2} + P_y \frac{\partial^2 \Phi_0}{\partial x \partial y} + P_z \frac{\partial^2 \Phi_0}{\partial x \partial z} \quad (24)$$

At the centre of the sphere we have by symmetry

$$\begin{aligned} \frac{\partial^2 \Phi_0}{\partial x \partial y} &= \frac{\partial^2 \Phi_0}{\partial y \partial z} = \frac{\partial^2 \Phi_0}{\partial z \partial x} = 0 \\ \frac{\partial^2 \Phi_0}{\partial x^2} &= \frac{\partial^2 \Phi_0}{\partial y^2} = \frac{\partial^2 \Phi_0}{\partial z^2} \end{aligned} \quad (25)$$

so that we find by use of (22), (23), (24) and (25)

$$-\vec{\nabla} \Phi = \frac{1}{3} \vec{P}_{(\vec{r},t)} \quad (26)$$

⁷In electrodynamics often the Gaussian units are used instead of the MKS (SI) units. The Gaussian units define ϵ_0 and μ_0 both equal to unity whereas in the MKS units they are defined through the speed of light $c_0 = (\epsilon_0 \mu_0)^{-1/2}$ and hence $\epsilon_0 = 1/4 \cdot 10^7 \pi^{-1} c_0^{-2} \text{A}^2 \text{s}^4 \text{kg}^{-1} \text{m}^{-3}$ and $\mu_0 = 4\pi \cdot 10^{-7} \text{kgmA}^{-2} \text{s}^{-2}$.

⁸From the Maxwell equations (1a) and (1b) follows for a static field the Poisson equation $\vec{\nabla}^2 \Phi = \rho/\epsilon$.

The total effective field is then the sum of all fields acting on the central molecule in the sphere

$$\vec{\mathcal{E}}'_{(\vec{r},t)} = \vec{\mathcal{E}}_{(\vec{r},t)} + \frac{1}{3}\vec{P}_{(\vec{r},t)} \quad (27)$$

Knowing the effective field, the displacement of the charge of its equilibrium position is determined via the Coulomb law. The electron behaves as if a restoring force to its equilibrium position binds it. In other words, the electron is bound to the atomic nucleus in order to respect the equilibrium of charge. The restoring force is

$$\vec{F} = -q\vec{r}$$

The equation of motion of the electron with the mass m_e and the charge $-e$ becomes then

$$m_e \frac{\partial^2 \vec{r}}{\partial t^2} - q\vec{r} = -e\vec{\mathcal{E}}'_{(\vec{r},t)} \quad (28)$$

Assuming a time harmonic incident field $\vec{\mathcal{E}}'_{(\vec{r},t)} = \vec{\mathcal{E}}'_{0,(\vec{r})} e^{-i\omega t}$ where ω is the angular frequency, the homogeneous solution of (28) is

$$\vec{r} = \vec{r}_0 e^{-i\omega t}$$

what leads to the stationary solution

$$\vec{r} = \frac{-e\vec{\mathcal{E}}'_{(\vec{r},t)}}{m_e(\omega_0^2 - \omega^2)} \quad (29)$$

where ω_0 is the *resonance frequency* or the *absorption frequency*⁹

$$\omega_0 = \sqrt{q/m_e} \quad (30)$$

Each electron contributes to the polarisation with the momentum $\vec{p} = -e\vec{r}$. We neglect the contribution from the atomic nuclei since their masses are heavy in comparison with the mass of the electron. The total polarisation $\vec{P}_{(\vec{r},t)}$ for N molecules per unit volume having one effective electron with the resonance frequency ω_0 is

$$\begin{aligned} \vec{P}_{(\vec{r},t)} &= N\vec{p} \\ &= N \frac{e^2}{m_e} \frac{\vec{\mathcal{E}}'_{(\vec{r},t)}}{\omega_0^2 - \omega^2} \end{aligned} \quad (31)$$

Using the mean polarisability α and the consequential density of polarisation $N\alpha$

$$\vec{P}_{(\vec{r},t)} = \alpha N \vec{\mathcal{E}}'_{(\vec{r},t)} \quad (32)$$

Introducing the effective field (27) and the permittivity (21a) leads to an expression for the mean polarisability

$$\alpha = \frac{3}{N} \frac{\epsilon - 1}{\epsilon + 2} \quad (33)$$

⁹The absorption frequency represents the eigenfrequency of electrons and ions in a solid

The comparison of (31) and (32) leads to the dispersion formula

$$\frac{\epsilon(\omega) - 1}{\epsilon(\omega) + 2} = \frac{1}{3} \frac{Ne^2}{m_e(\omega_0^2 - \omega^2)} \quad (34)$$

for a gas n is approximate 1, we then assume $\epsilon + 2 \cong 3$ and through $\epsilon(\omega) = n_\omega^2$ we write

$$n_{(\omega)}^2 - 1 = \frac{Ne^2}{m_e(\omega_0^2 - \omega^2)} \quad (35)$$

We see, n is an increasing function of the frequency. For $\omega > \omega_0$ n approaches unity with increasing ω (see figure 2.5). The dispersion is then said to be normal. At the resonance frequency the permittivity tends to infinity. The singularity arises only formally as we neglected the damping effect. The damping involves momentum transfer between the electron and the lattice vibrations, lattice imperfections and impurities. Therefore we introduce a resisting force $\gamma \frac{\partial \vec{r}}{\partial t}$. Then equation (28) becomes

$$m_e \frac{\partial^2 \vec{r}}{\partial t^2} + m_e \gamma \frac{\partial \vec{r}}{\partial t} + m_e \omega_0 \vec{r} = e \vec{E}'_{(\vec{r},t)}$$

and its solution is of the form

$$\vec{r} = \frac{-e \vec{E}'_{(\vec{r},t)}}{m_e(\omega_0^2 - \omega^2) - i\gamma\omega}$$

Sellmeier's dispersion formula:

Until now, we assumed the system to have only one resonance frequency. Let us introduce a more general expression for the dispersion equation (34)

$$\frac{n_{(\nu)}^2 - 1}{n_{(\nu)}^2 + 2} = \frac{1}{3} N \alpha = \frac{1}{3} N \frac{e^2}{m_e} \sum_k \frac{f_k}{\omega_k^2 - \omega^2} = \sum_i \frac{\rho_k}{v_k^2 - \nu} \quad (36)$$

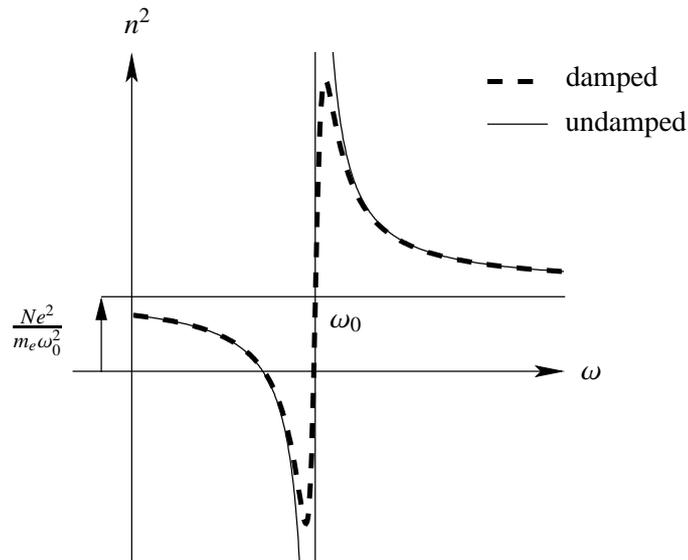


Figure 4: Dispersion curves.

Nf_k density of electrons corresponding to the resonance frequency ω_k

$$\begin{aligned}\rho_k &= N \frac{e^2}{4\pi^2 m_e} f_k \\ \nu_k &= \frac{\omega_k}{2\pi} \\ \nu &= \frac{\omega}{2\pi}\end{aligned}$$

The absorption is small compared with the resonant frequency. In the visible range, an electron in a dielectric does not see significant absorption: the medium is transparent.

We rewrite (36) as

$$n_{(\nu)}^2 - 1 = \frac{3N\alpha}{3 - N\alpha} \quad (37)$$

It will be sufficient to take into account a finite number of absorption frequencies, therefore we do a partial decomposition of (37). Hence we search the zeros of the denominator of (37)

$$3 - N\alpha = 3 - \sum_k \frac{\rho_k}{\nu_k^2 - \nu^2} = 0$$

If $\bar{\nu}_k$ are the zeros then we find the *Sellmeier's dispersion formula*

$$n_{(\nu)}^2 - 1 = \sum_k \frac{\bar{\rho}_k}{\bar{\nu}_k^2 - \nu^2} \quad (38)$$

Notes:

Sellmeier's dispersion formulas is a tool to approximate the refraction index of a material, but it is restricted to a given frequency domain because of the simplifications we made. At very long wavelengths (infra-red) the nuclei can follow the field and their motion have to be taken into account. At very short wavelengths (UV) the mean time between collisions of the free electron shrinks with their increased collision rate among themselves due to their increased energy.

A special case represents the metals because of their unbound electrons in the "electron gas cloud". They are moving almost undisturbed through the lattice. Further, in contrast to dielectrics the metals are opaque (in the visible frequency spectrum). The incident field strength shrinks to $1/e$ within a tenth of the wavelength, the so called *skin depth* δ . Therefore we introduce another model that is more suitable to describe an electron in metal.

Permittivity of metals: the Drude model

As mentioned, the valence electrons of a metal are moving almost unimpeded through the lattice. They are said to be *free electrons*. In absence of an external electromagnetic field the electrons move in a random manner. The damping effect comes from collisions involving momentum transfer between the electron and the lattice vibrations, lattice imperfections and impurities and they are represented via the damping constant γ .

If an external electromagnetic field is applied, the electrons acquire an additional velocity. Their motions become more orderly although they collide with the essentially stationary atoms. This more orderly motion of the electrons gives rise of a polarisation in the medium.

The equation of motion for a model electron with charge $-e$ and mass m_e is

$$m_e \frac{\partial^2 \vec{r}}{\partial t^2} + m_e \gamma \frac{\partial \vec{r}}{\partial t} = -e \vec{E}_{(\vec{r},t)} \quad (39)$$

In order to understand the meaning of the damping constant γ in (39), we first consider the case where no external field is present. The homogeneous solution is

$$\frac{\partial \vec{r}}{\partial t} = \vec{v} = \vec{v}_0 e^{-\gamma t}$$

The model electron has an initial velocity v_0 and slows down in an exponential way, with γ as decay constant. The decay time (or relaxation time) $\tau = 1/\gamma$ is typically of the order of 10^{-14} s.

Assuming the external field to be time harmonic $\vec{\mathcal{E}}_{(\vec{r},t)} = \vec{\mathcal{E}}_{(\vec{r})}' e^{-i\omega t}$, the solution of (39) becomes the sum of a decaying motion and a periodic motion

$$\vec{r} = \frac{-e}{m_e (\omega^2 + i\gamma\omega)} \vec{\mathcal{E}}_{(\vec{r},t)}' \quad (40)$$

This periodic motion gives rise to a dipole moment in the medium

$$\vec{p} = -e\vec{r} = \frac{e^2}{m_e (\omega^2 + i\gamma\omega)} \vec{\mathcal{E}}_{(\vec{r},t)}' \quad (41)$$

The total polarisation $\vec{P}_{(\vec{r},t)}$ for N free electrons per unit volume is similar (31)

$$\vec{P}_{(\vec{r},t)} = \frac{e^2 N}{m_e (\omega^2 + i\gamma\omega)} \vec{\mathcal{E}}_{(\vec{r},t)}' = \epsilon_0 \chi \vec{\mathcal{E}}_{(\vec{r},t)}' \quad (42)$$

Using equation (21a) and the definition of the plasma frequency

$$\omega_p^2 = \frac{Ne^2}{\epsilon_0 m_e} \quad (43)$$

we find the complex notation of the frequency dependent permittivity

$$\underline{\epsilon}_{(\omega)} = 1 - \frac{\omega_p^2}{\omega^2 + \gamma^2} + i \frac{\gamma \omega_p^2}{\omega (\omega^2 + \gamma^2)} \quad (44)$$

Comparing to Sellmeier's diffusion formula (38) the permittivity calculated by the Drude model takes into account the damping effect which is represented through the imaginary part of the $\epsilon_{(\omega)}$ -function.

The *Langmuir frequency* ω_L at which the sign of the real part of the permittivity changes is of particular interest. This determines the domain of attenuation $\Re\{\epsilon_{(\omega)}\} < 0$ and the domain of propagation $\Re\{\epsilon_{(\omega)}\} > 0$. We determine ω_L through (44)

$$\omega_L = \sqrt{\omega_p^2 - \gamma^2} \cong \omega_p \quad \text{if } \gamma \ll \omega_p \quad (45)$$

Also we can express the conductivity as a function of the angular frequency ω . We know the current density $\vec{\mathcal{J}} = -eN \frac{\partial \vec{r}}{\partial t}$. Applying Ohm's Law (2b) allows to write

$$\sigma_{(\omega)} = \frac{Ne^2}{m_e (\gamma - i\omega)} \quad (46)$$

Validity of the Drude model:

We will compare the permittivity approximated by use of the Drude model over a frequency domain to a reported set of optical properties of Gold and Silver [23]. The necessary parameters for the Drude model are reported in table 1. As it can be seen in figure 5 the fitted ¹⁰ Drude approximation holds

	N [m^{-3}]	ω_p [rads^{-1}]	γ_{reported} [s^{-1}]	γ_{fitted} [s^{-1}]
Gold	$8.7 \cdot 10^{28}$	$1.6640 \cdot 10^{16}$	$1.0417 \cdot 10^{14}$	$6.0 \cdot 10^{14}$
Silver	$6.9 \cdot 10^{28}$	$1.4819 \cdot 10^{16}$	$3.2258 \cdot 10^{13}$	$3.6 \cdot 10^{14}$

Table 1: Free electron density N at 300K (source [16]), the plasma frequency ω_p computed using equation (43) and the damping constant $\gamma = \tau^{-1}$ from [23] and fitted by the least square method.

for the visible frequency range. But in the IR domain, the Drude model fails completely, after all the imaginary part does not correspond to the values from literature.

It can be said, that the Drude model represents a adequate approximation in the visible spectral range. But in the IR and UV region, the simple model does not represent satisfyingly the physics. In these cases, only quantum mechanics can provide a passable theory.

Permittivity of gold and silver:

We have seen a model for dielectric and metallic media to estimate their permittivity. As we use reported

¹⁰The fit was done by the least square method. Further, the curve of the real part for silver has been shifted up by 5 in order to get the match. For the gold curve, no shift was adequate to match the curve with the reported values

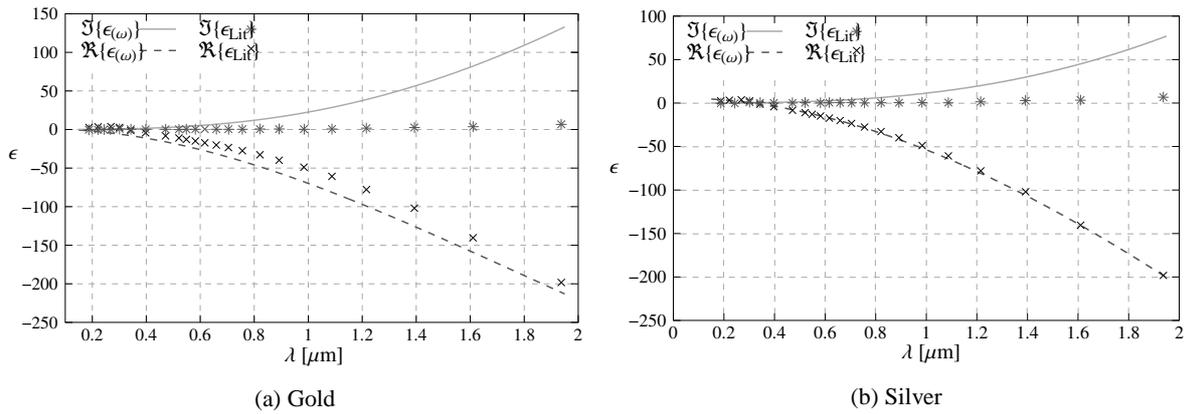


Figure 5: Comparison between the Drude model and reported data.

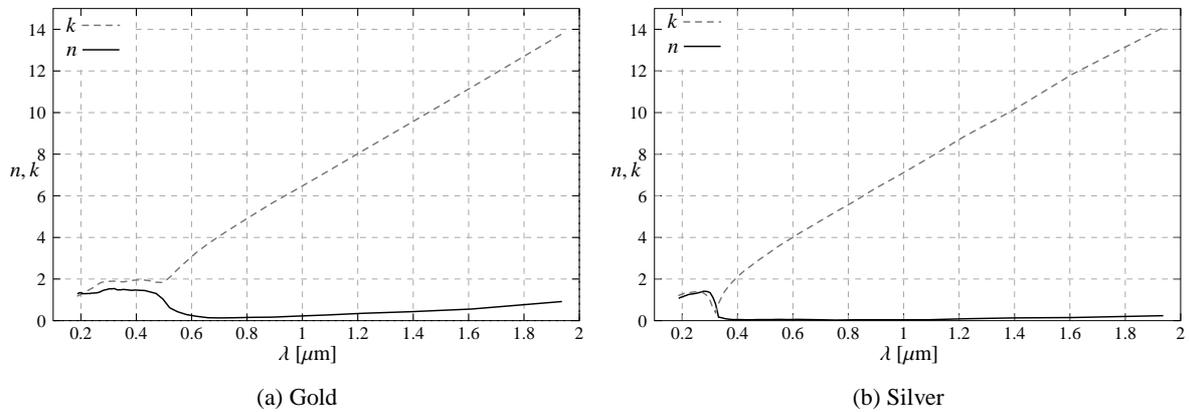


Figure 6: Complex representation of the refractive index.

values for the calculation we give the plot of the permittivity in function of the wavelength for Gold and Silver in figure 6.

For the interpretation we introduce the complex notation of the refraction

$$\underline{n} = n + ik \quad (47)$$

using the complex notation of the permittivity resulting from (6)

$$\underline{n}^2 = \left(\epsilon_r + i \frac{\sigma}{\omega \epsilon_0} \right)$$

we can express n^2 and k^2

$$n^2 = \frac{1}{2} \sqrt{\mu^2 \epsilon_r^2 - \frac{\sigma^2 \mu^2}{\omega^2 \epsilon_0^2}} + \epsilon_r \mu \quad (48)$$

$$k^2 = \frac{1}{2} \sqrt{\mu^2 \epsilon_r^2 - \frac{\sigma^2 \mu^2}{\omega^2 \epsilon_0^2}} \quad (49)$$

It can be seen that the absorption of visible light in Gold and Silver grows with increasing wavelength. As consequence the light does penetrate less deep into the metal. Does the light then experience an increased reflectivity?

2.6 Resonance phenomena in the reflection spectrum

Let us take a 3 layer model sketched in figure 15 to answer that question. Assuming a substrate (region *III*) coated with a film of Gold or Silver of height h (region *II*). We illuminate the probe with a unit amplitude plane wave coming from region *I* (air) under different angles. As we will show in chapter 4.2 the total reflection coefficient R_{13} and the total transmission coefficient T_{13} can be deduced by the Fresnel equations 86d and 86c

$$T_{13} = \frac{T_{12} T_{23} e^{ik_0 \beta_2 h_2}}{1 + R_{12} R_{23} e^{2ik_0 \beta_2 h_2}}$$

$$R_{13} = \frac{R_{12} + R_{23} e^{2ik_0 \beta_2 h_2}}{1 + R_{12} R_{23} e^{2ik_0 \beta_2 h_2}}$$

where

β_2 is the propagation constant in region 2

h_2 is the height of the metallic film

R_{ij} is the reflection coefficients at the interface ij

T_{ij} is the transmission coefficients at the interface ij

The plots of the reflection spectrum of gold and silver in figure 7 show a astonishing behaviour. At the incident angle θ_{SP} a sharp fall of slope appears and the reflected intensity drops to zero whereas the transmission is enhanced. The explanation of these anomalous absorption is given by a phenomenon called *surface plasmon* or short *SPs*.

2.6.1 Surface plasmon

From the plasma concept follows that longitudinal density fluctuations, plasma oscillations, will propagate through the volume of the metal.

At the interface of a metal film longitudinal electron density fluctuations are allowed. These oscillations appear at a different frequency than the plasma frequency ω_p and are strictly confined to the surface. The periodic surface charge fluctuation is source of a macroscopic electric field, schematically represented in figure 8 and calculated in figure 21. The SP wave vector is derived by use of the Maxwell equations (1) and the boundary conditions for an semi finite metal with the permittivity $\epsilon_2 = \epsilon'_2 + i\epsilon''_2$ adjacent to a dielectric medium ϵ_1 and leads to the dispersion relation

$$k_x^{SP} = \frac{\omega}{c} \sqrt{\frac{\epsilon_1 \epsilon_2}{\epsilon_1 + \epsilon_2}} \quad (50)$$

As ω and ϵ_1 are real k_x^{SP} is complex and we may write

$$k_x^{SP} = \underbrace{\frac{\omega}{c} \sqrt{\frac{\epsilon_1 \epsilon'_2}{\epsilon_1 + \epsilon'_2}}}_{k_x^{SP'}} + i \underbrace{\frac{\omega}{c} \sqrt{\frac{\epsilon_1 \epsilon'_2}{\epsilon_1 + \epsilon'_2} \frac{\epsilon''_2}{2(\epsilon'_2)^2}}}_{k_x^{SP''}} \quad (51)$$

Real metals fulfils the conditions $\epsilon'_2 < 0$ and $\Re\{\epsilon''_2\} \ll |\epsilon'_2|$. Therefore $k_x^{SP'}$ is real and represent the wave vector of the SP and $k_x^{SP''}$ the internal absorption.

As already mentioned, the charge density variations are confined to the surface. Perpendicular to the boundary no propagation is allowed and the field decays exponentially, hence $k_{z,i}$ has to be purely

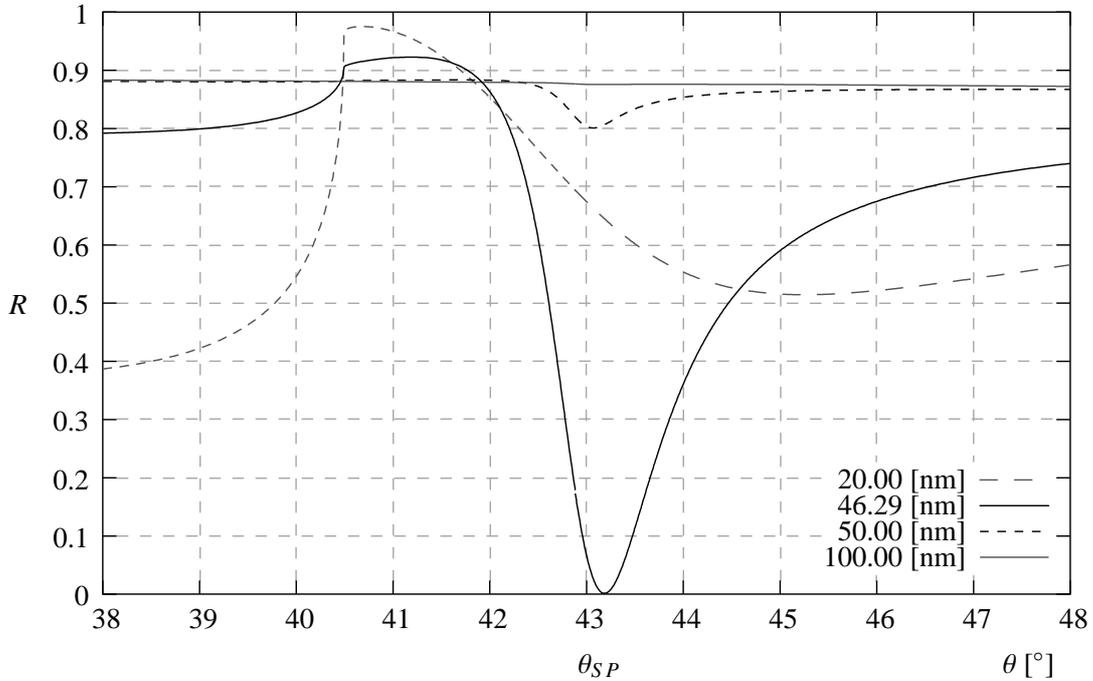


Figure 7: Reflected intensity at a thin gold film in function of the thickness h . Incident light at $\lambda_0 = 617\text{nm}$, $\epsilon_1 = 1.54^2$ (glass), $\epsilon_2 = -10.662 + 1.374i$ (gold), $\epsilon_3 = 1$ (air). Plasmon excitation for $h_{min} = 46.29\text{nm}$ at $\theta_{SP} = 43.22^\circ$.

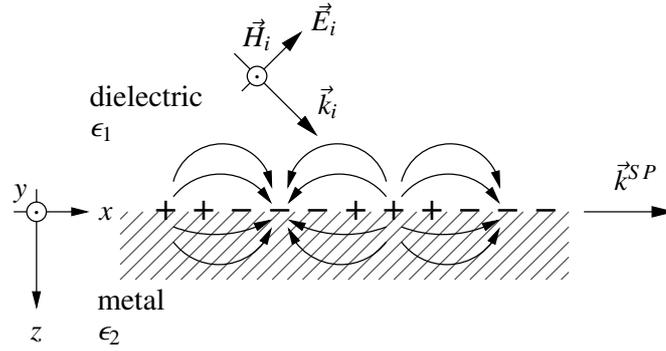


Figure 8: Longitudinal fluctuation of the surface charge density.

Electromagnetic field of SPs propagating on a smooth surface. \vec{E}_i , \vec{H}_i and \vec{k}_i represent the incident TM wave, \vec{k}^{SP} is the SP wave vector.

imaginary

$$k_{(z,i)} = \sqrt{\epsilon_i \left(\frac{\omega}{c} \right)^2 - k_x^2}$$

using equation (50), we find

$$k_{z,1}^{SP} = \frac{\omega}{c} \sqrt{\frac{\epsilon_1^2}{\epsilon_1 + \epsilon_2'}} \quad k_{z,2}^{SP} = \frac{\omega}{c} \sqrt{\frac{\epsilon_2'^2}{\epsilon_1 + \epsilon_2'}} \quad (52)$$

Through the wave vector we can describe the localisation of the surface plasmon. The *penetration depth* or *skin depth* is given by

$$\delta^{SP} = \frac{1}{2|k_z^{SP}|} \quad (53)$$

and the propagation length on the boundary is given by the absorption, the imaginary part of k_x^{SP}

$$l_x = \frac{1}{|2k_x^{SP''}|} \quad (54)$$

Coupling light to plasmons:

The dispersion relation (51) points out that at the same frequency ω for real metals the wave vector of the SP remains always larger than a wave vector in an adjacent dielectric medium

$$\begin{aligned} \frac{\omega}{c} \sqrt{\frac{\epsilon_1 \epsilon_2'}{\epsilon_1 + \epsilon_2'}} &> \frac{\omega}{c} \sqrt{\epsilon_1} \sin(\theta_i) \\ \sqrt{\epsilon_1} \underbrace{\sqrt{\frac{\epsilon_2'}{\epsilon_1 + \epsilon_2'}}}_{>1} &> \sqrt{\epsilon_1} \\ k_x^{SP'} &> k_{1,x} \end{aligned} \quad (55)$$

Therefore, on smooth surfaces, SPs are 'non-radiative'. They are not directly accessible to an incident plane wave. However, it is possible to increase the incident wave vector by Δk through a transformation

such that it matches the wave vector of the SP. This approach is called *momentum matching*.

Another important point of the coupling is the state of polarisation of the incident light. The electric field of the SPs is created by a longitudinal charge density variation in x -direction (see figure 8). Therefore the electric field component of the incident light has to lie in the plane of incidence (x, z) indicating that the incident light has to be TM polarised to allow maximal coupling. Hence SPs are not accessible to TE polarised light.

Attenuated total reflection ATR method:

The prism coupling use the total internal reflection as transformation to increase the wave vector of the incident light by Δk . Remember the experiment with the 3 layer system in chapter 2.6. The reflection spectrum showed an astonishing abrupt drop. This 3 layer stack, Air, Glass, Gold is the Kretschmann-Raether configuration sketched in figure 9. The minimum in the reflection spectrum of Gold (figure 7) represents the efficient excitation of a SP mode at the dielectric air interface.

According to figure 9, the projection of the incident wave vector on the x -axis is

$$k_x = \sqrt{\epsilon_3} \sin(\theta_i) k_0$$

Under the condition that θ_i corresponds to the angle at which occurs the minimum in the reflection spectrum, k_x matches with k_x^{SP} and the SP mode at the air - metal becomes excited.

In the dispersion curve in figure 11, the presence of a dielectric medium lowers the slope of the light line in air, ck . The phase velocity decreases due to the dispersion in the dielectric medium. As soon as the light line overlaps the dispersion relation, light can excite SPs of frequencies below the crossing point P . That equals the transformation to increase the wave vector k_x by Δk .

In other words, the evanescent wave propagates at the dielectric - metal interface with a phase velocity less than the speed of light:

$$v = \frac{\omega}{k_x} = \frac{c}{\sqrt{\epsilon_3} \sin(\theta_i)} \quad (56)$$

The resonance condition of the SPs follows out of its dispersion equation (51)

$$\frac{\omega^{SP}}{k_x^{SP}} = \sqrt{\frac{\epsilon'_2 \epsilon_3}{\epsilon'_2 + \epsilon_3}} \quad (57)$$

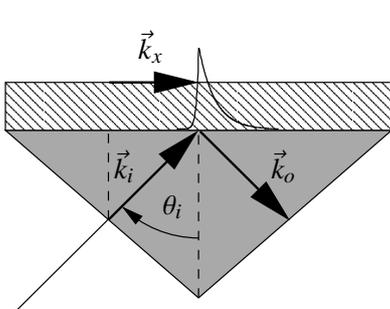


Figure 9: Kretschmann-Raether configuration. The evanescent wave created at the dielectric ϵ_3 - metal ϵ_2 interface couples into SP.

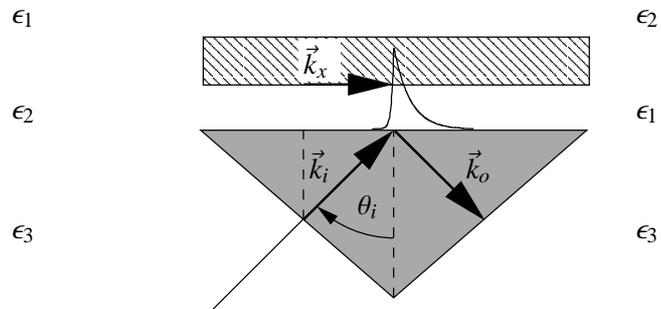


Figure 10: Otto configuration. The evanescent wave created at the dielectric ϵ_3 - air ϵ_1 interface couples into SPs on the metal ϵ_2 - air interface.

The light - SP coupling arises if the right hand side of the equations (56) and (57) are equal

$$\frac{c}{\sqrt{\epsilon_3} \sin(\theta_i)} = \sqrt{\frac{\epsilon'_2 \epsilon_3}{\epsilon'_2 + \epsilon_3}} \quad (58)$$

In contrast to the Kretschmann-Raether configuration the Otto configuration (sketched in figure 10) permits an SP excitation without that the metal film is in contact with the prism. It results also another reflection spectrum. The complete set of reflection and transmission curves are reported in appendix B.

Both configurations can be used to determine the permittivity of thin conductive films. Therefore is illuminated under different angles in order to find the minimum in the reflection spectrum that corresponds to the SP momentum k_x^{SP} as shown in figure 7. The parameters of the SP excitation are the angle of incident and the film thickness. As it is outlined in the reflection spectrum, the excitation is very sensitive to both of those parameters. The corresponding setup for the determination of the ϵ -function is given in appendix A.

The grating coupler:

At a grating the incident light is diffracted according to the grating equation

$$\begin{aligned} k_{x,m} &= k_{x,i} \pm mK & m \in \mathbb{N} \\ k_{y,m} &= k_{y,i} \end{aligned} \quad (59)$$

where $k_{x,i} = \sqrt{\epsilon_1} \omega/c \sin \theta_i$ is the projection of the incident wave \vec{k}_i from medium 1 on the surface of the grating. The grating momentum is $K = 2\pi/\Lambda$ where Λ is the period of the grooves. The reflected or transmitted diffraction orders are described via $k_{x,m} = k_m \sin(\theta_m)$. The zero order passes the grating without refraction following the Snellian law. The grating has to be conceived such that the diffracted

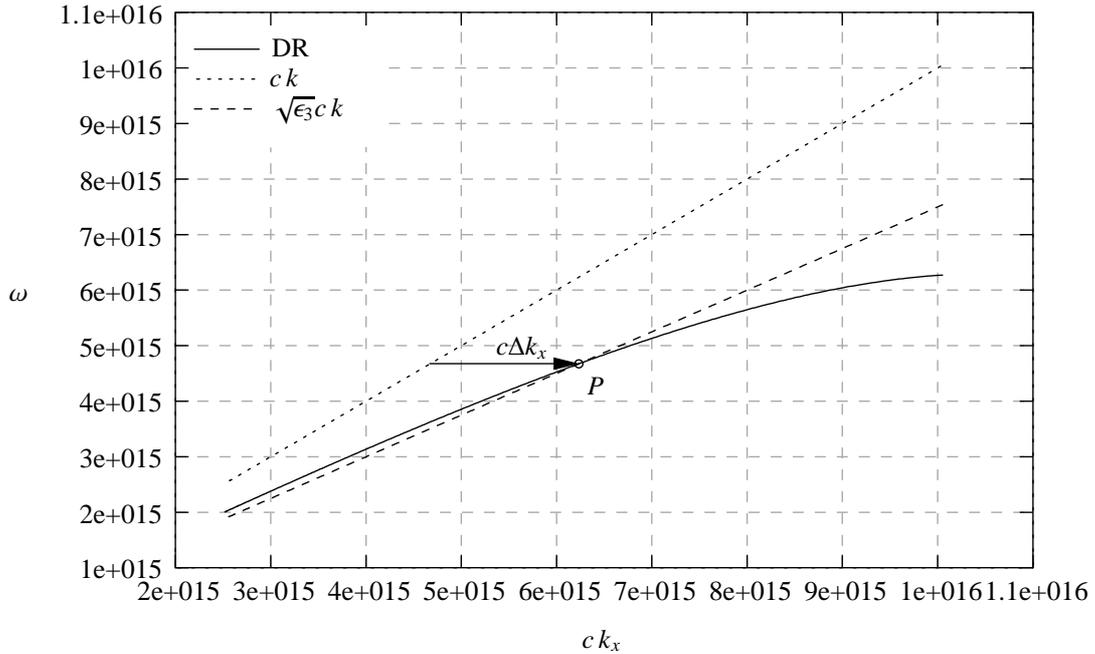


Figure 11: Dispersion relation (DR) of SPs at the gold-glass interface and the light lines in air and in the dielectric medium. Incidence at $\theta_i = 60^\circ$.

orders match the SP momentum so that the SP resonance condition following the dispersion equation (50) will be fulfilled

$$\begin{aligned} k_x^{SP} &= \frac{\omega}{c} \sqrt{\frac{\epsilon_1 \epsilon_2'}{\epsilon_1 + \epsilon_2'}} = \sqrt{\epsilon_1} \frac{\omega}{c} \sin(\theta_i) \pm \Delta k_x \\ &= k_{x,i} \pm mK \end{aligned} \quad (60)$$

The reverse process is also possible: SP propagating along a rough surface such as the corrugated surface of a grating can reduce their wave vector by Δk_{xd} and decay into a photon.

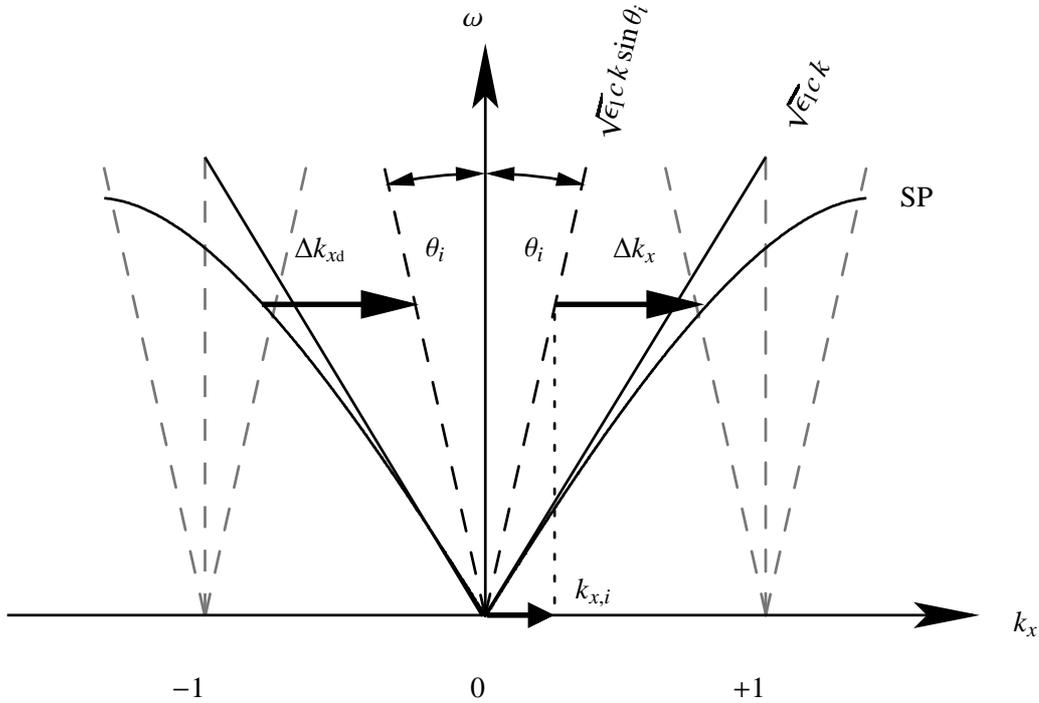


Figure 12: Momentum match on a grating.

The 'non-radiative' dispersion relation SP, right the light line $\sqrt{\epsilon_1} c k$ and the light line in medium 1 $\sqrt{\epsilon_1} c k \sin \theta_i$. The incident light $k_{x,i}$ is transformed into SP by taking up Δk_x . On the left the reverse process is sketched out: a SP decays radiatively via Δk_{xd} . At normal incidence it is possible that two SP are excited at the same time.

3 Slit model

This chapters intends to introduce the slit model that uses the surface impedance as boundary condition essentially due to A. Barbara.

3.1 Application of the surface impedance approximation

We have seen in chapter 2.3 the tangential fields at any boundary have to be continuous. Further we outlined in chapter 2.4 that the surface impedance can be applied as boundary condition so that the fields are considered exclusive in one half space. Applying this to the 2D model sketched in figure 1(b) for an incident linear polarised TM plane wave. The TM polarisation is preferred because it allows to excite SPs. We have seen, that the electric component of the incident field has to lie in the plane of incidence for best excitation of the surface charge density fluctuations that are the source of the SP. For the sake of simplicity, the magnetic component is considered for the calculations instead of the commonly used electric component because the magnetic field p-polarised light has only one y-component.

Field representation:

A Bloch wave is incident on the slits coming from region *I* excites waveguide modes in region *III*. These waveguide modes couple again in a Bloch wave in region *IV* after travelling through region *III*. The Bloch wave can be represented in a Fourier expansion. As this model uses the surface impedance boundary condition, it does consider only the fields outlying the metal. For the representation of the waveguide modes, the modal expansion is used. The fields in the three regions are

Region I

$$H_{(x,z;t)}^{(I)} = \left(e^{ik_0(\gamma_0 x - \beta_0 z)} + \sum_{n=-\infty}^{\infty} R_n e^{ik_0(\gamma_n x + \beta_n z)} \right) e^{-i\omega t} \quad (61)$$

where

$$\begin{aligned} k_0 &= 2\pi/\lambda \\ \gamma_0 &= \sin(\theta_i) &= k_{1x}^{(0)}/k_0 \\ \gamma_n &= \sin(\theta_i) + 2n\lambda/p &= k_{1x}^{(n)}/k_0 \\ \beta_0 &= \sqrt{1 - \gamma_0^2} &= k_{1z}^0/k_0 \\ \beta_n &= \sqrt{1 - \gamma_n^2} &= k_{1z}^{(n)}/k_0 \\ R_2 & & \text{is the reflected amplitude} \end{aligned}$$

Region III

$$H_{(x,z;t)}^{(III)} = \sum_{m=1}^{\infty} \cos \left[\frac{m\pi}{w} \left(x + \frac{w}{2} \right) \right] \left(A_m e^{ik_0 \mu_m z} + B_m e^{-ik_0 \mu_m z} \right) e^{-i\omega t} \quad (62)$$

where

$$\begin{aligned} \mu_m &= \sqrt{k_0^2 - (m\pi/w)^2} &= k_{2z}^{(m)}/k_0 \\ A_m & & \text{amplitude of the up propagating wave in the slit} \\ B_m & & \text{amplitude of the down propagating wave in the slit} \end{aligned}$$

Region IV

$$H_{(x,z;t)}^{(IV)} = \sum_{n=-\infty}^{\infty} T_n e^{ik_0(\gamma_n x - \beta_{n,t}(z+h))} e^{-i\omega t} \quad (63)$$

where

$$\begin{aligned} \beta_{n,t} &= \sqrt{1 - \gamma_n^2/n_4^2} &= k_{3z}^{(n)}/k_0 \\ n_4^2 &= \sqrt{\epsilon_4} &\text{is the complex dielectric constant in region IV} \\ T_n & &\text{is the transmitted amplitude} \end{aligned}$$

Boundary condition:

The boundary conditions were taken from the work of Wirgin et al. [18]. As mentioned, the surface impedance method serves as boundary condition as simplification of the calculation: Only the fields outlying the metal were matched.

over a period the boundary conditions are

$$\left(\frac{\partial H_y^{(I)}}{\partial z} + ik_0 \frac{Z_3}{Z_1} H_y^{(I)} \right)_{z=0^+} = \left(\frac{\partial H_y^{(III)}}{\partial z} + ik_0 \frac{Z_3}{Z_1} H_y^{(III)} \right)_{z=0^-} \quad \forall |x| < p/2 \quad (64a)$$

$$\left(\frac{\partial H_y^{(III)}}{\partial z} + in_4 k_0 \frac{Z_3}{Z_3} H_y^{(III)} \right)_{z=-h^+} = \left(\frac{\partial H_y^{(IV)}}{\partial z} + in_4 k_0 \frac{Z_3}{Z_3} H_y^{(IV)} \right)_{z=-h^-} \quad \forall |x| < p/2 \quad (64b)$$

in the slits

$$H_z^{(I)} \Big|_{z=0^+} = H_z^{(III)} \Big|_{z=0^-} \quad \forall |x| < W/2 \quad (64c)$$

$$H_z^{(III)} \Big|_{z=-h^+} = H_z^{(IV)} \Big|_{z=-h^-} \quad \forall |x| < W/2 \quad (64d)$$

where

$$Z_i = \sqrt{\mu_0 \mu_i / (\epsilon_0 \epsilon_i)} \quad \text{is the surface impedance of medium } i$$

Determination of the coefficients:

First, the modal expansion is truncated to the zero order term. Under this assumption, all coefficients necessary to determine the field at any point in the space can be determined

$$A_0 = \frac{\left(\frac{2\beta_0 S_0}{\beta_0 + Z_3/Z_1} \right) (1 - D_4^+)}{(1 - D_1^+)(1 - D_4^+) - e^{-2ik_0 h} (1 + D_1^-)(1 + D_4^-)} \quad (65a)$$

$$B_0 = \frac{\left(\frac{-2\beta_0 S_0}{\beta_0 + Z_3/Z_1} \right) e^{-2ik_0 h} (1 + D_4^-)}{(1 - D_1^+)(1 - D_4^+) - e^{-2ik_0 h} (1 + D_1^-)(1 + D_4^-)} \quad (65b)$$

with

$$D_1^\pm = (1 \pm Z_3/Z_1) \Gamma \sum_{n=-\infty}^{\infty} \frac{S_n^2}{\beta_n + Z_3/Z_1} \quad (66a)$$

$$D_4^\pm = (1 \pm Z_3/Z_1) \sqrt{\epsilon_4} \Gamma \sum_{n=-\infty}^{\infty} \frac{S_n^2}{\beta_{n,t} + Z_3/Z_4} \quad (66b)$$

where

$$\Gamma = \frac{w}{p} \quad \text{is the geometrical factor}$$

$$S_0 = \text{sinc}(\gamma_n k_0 w/2) \quad \text{is the overlap function}$$

Reflection and transmission coefficients:

With the coefficients A_0 and B_0 , the reflection and transmission coefficient of any order can be deduced

$$R_m = \frac{\beta_0 - Z_3/Z_1}{\beta_0 + Z_3/Z_1} \delta_{m,0} + \frac{\Gamma}{\beta_m + Z_3/Z_1} S_m [A_0(1 + Z_3/Z_1) - B_0(1 - Z_3/Z_1)] \quad (67)$$

$$T_m = n_4 \frac{\Gamma}{\beta_{m,t} + Z_3/Z_4} S_m e^{ik_0 h} [-A_0 e^{-ik_0 h} (1 - Z_3/Z_1) + B_0 e^{ik_0 h} (1 + Z_3/Z_1)] \quad (68)$$

3.2 Discussion of the model

The first point to discuss is the transmission spectrum presented in figure 13, the zero order transmittance. The implementation of the equations (67) and (68) gives the results showed in figure 14. We recognise also absorption in the reflection spectrum at similar wave numbers, but we have found a more or less flat curve with very sharp absorptions in contrast to the strong varying transmittance curve of Barbara. May be our permittivity approached by the Drude model is a source of error, but we found the same shape by using values for the dielectric constant reported in the "Handbook of Optical constants".

The second point is the definition of the boundary conditions used in Barbara's work. Referring on [18] we were not able to reproduce the equations (64a) and (64b). Therefore we propose the following boundary conditions.

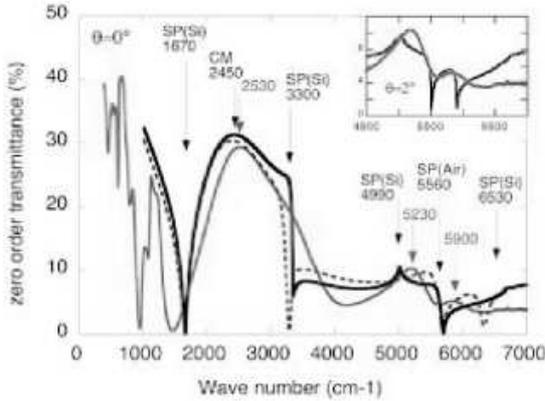


Figure 13: Zero order transmittance by [3]. Measured (dark thin line), calculated using one mode (dotted line) and calculated using four modes (dark strong line) transmission spectrum. The setup is given in figure 1(b). A thin gold film of thickness $h = 0.96\mu\text{m}$ with slits at the period $p = 1.75\mu\text{m}$ and a groove width $w = 0.61\mu\text{m}$ on a Si substrate.

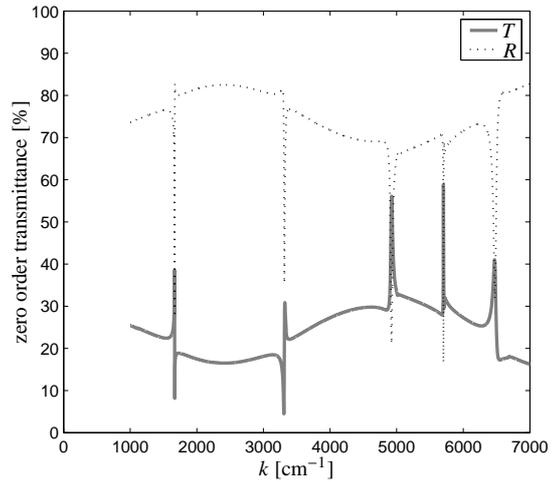


Figure 14: Computation of the total reflected and transmitted intensity by the Barbara's model. The parameters are given in figure 13. The dielectric constant was previously computed by means of the Drude model.

The walls in the slit are assumed to have no resistance

$$H_{y(\vec{x},t)}^{III} \Big|_{x=\pm w/2} = 0 \quad 0 \leq z \leq h \quad (69)$$

Interface I → II / I → III:

$$H_{y(\vec{x})}^I \Big|_{z=0_-} + \frac{1}{ik_1} \frac{Z_3}{Z_1} \frac{\partial H_{y(\vec{x})}^I}{\partial z} \Big|_{z=0_-} = 0 \quad \frac{w}{2} \leq |x| \leq \frac{d}{2} \quad (70a)$$

$$H_{y(\vec{x})}^{III} \Big|_{z=0_+} + \frac{1}{ik_3} \frac{Z_3}{Z_1} \frac{\partial H_{y(\vec{x})}^{III}}{\partial z} \Big|_{z=0_+} = 0 \quad |x| \leq \frac{w}{2} \quad (70b)$$

$$H_{y(\vec{x})}^I \Big|_{z=0_-} - H_{y(\vec{x})}^{III} \Big|_{z=0_+} = 0 \quad |x| \leq \frac{w}{2} \quad (70c)$$

$$\frac{\partial H_{y(\vec{x})}^I}{\partial z} \Big|_{z=0_-} - \frac{\partial H_{y(\vec{x})}^{III}}{\partial z} \Big|_{z=0_+} = 0 \quad |x| \leq \frac{w}{2} \quad (70d)$$

Interface III → IV:

$$H_{y(\vec{x})}^{III} \Big|_{z=h_-} + \frac{1}{ik_3} \frac{Z_3}{Z_4} \frac{\partial H_{y(\vec{x})}^{III}}{\partial z} \Big|_{z=h_-} = 0 \quad |x| \leq \frac{w}{2} \quad (71a)$$

$$H_{y(\vec{x})}^{IV} \Big|_{z=h_+} + \frac{1}{ik_4} \frac{Z_3}{Z_4} \frac{\partial H_{y(\vec{x})}^{IV}}{\partial z} \Big|_{z=h_+} = 0 \quad \frac{w}{2} \leq |x| \leq \frac{d}{2} \quad (71b)$$

$$H_{y(\vec{x})}^{III} \Big|_{z=h_-} - H_{y(\vec{x})}^{IV} \Big|_{z=h_+} = 0 \quad |x| \leq \frac{w}{2} \quad (71c)$$

$$\frac{\partial H_{y(\vec{x})}^{III}}{\partial z} \Big|_{z=h_-} - \frac{\partial H_{y(\vec{x})}^{IV}}{\partial z} \Big|_{z=h_+} = 0 \quad |x| \leq \frac{w}{2} \quad (71d)$$

Field representation:

By the boundary conditions and the wave equation the field is defined at any point in space-time. Respecting the Floquet theorem and using the Rayleigh expansion is leading to the same fields as in the Barbara model.

Application of the boundary conditions to the field:

The boundary conditions are applied to the field by projection on a set of base vectors. The scalar product describes the field component accepted on each side of the interface. Projecting (70a) and (70b) on $e^{ik_j x}$

$$\frac{1}{d} \int_{-d/2}^{d/2} \left(H_{(x,z;t)}^I + \frac{1}{ik} \frac{Z_3}{Z_1} \frac{\partial H_{(x,z;t)}^I}{\partial z} \Big|_{z=0^-} e^{-ik_j x} dx \right) = \frac{1}{w} \int_{-w/2}^{w/2} \left(H_{(x,z;t)}^{III} + \frac{1}{ik_3} \frac{Z_3}{Z_1} \frac{\partial H_{(x,z;t)}^{III}}{\partial z} \Big|_{z=0^-} e^{-ik_j x} dx \right) \quad (72)$$

After a few calculations the right hand side becomes

$$\underbrace{\frac{1}{d} \int_{-d/2}^{d/2} \left(e^{ik_0 \gamma_0 x} (1 + \beta_0 Z_{31}) \right) e^{-ik_0 \gamma_j x} dx}_{\frac{1}{d} \delta_{0,j} (\beta_0 Z_{31} + 1)} + \underbrace{\frac{1}{d} \int_{-d/2}^{d/2} \left(\sum_{n=-\infty}^{\infty} R_n e^{ik_0 \gamma_n x} (1 - \beta_0 Z_{31}) \right) e^{-ik_0 \gamma_j x} dx}_{\frac{1}{d} \delta_{n,j} R_n (\beta_n Z_{31} - 1)} \quad (73)$$

and the left hand side

$$\sum_{m=1}^{\infty} B_m (e^{2ik_0\gamma_m n_3 h} - 1) (\gamma_m Z_{31} - 1) \cdot \underbrace{\frac{1}{w} \int_{-w/2}^{w/2} \sin \left[\frac{m\pi}{w} \left(x + \frac{w}{2} \right) \right] e^{-ik_0 n_3 \gamma_j x} dx}_{S_{jm}} \quad (74)$$

Hence, we find a expression for the reflection coefficient

$$R_j = \frac{\gamma_m Z_{31} + 1}{\gamma_m Z_{31} - 1} \delta_{0,j} - \sum_{m=1}^{\infty} B_m (e^{2ik_0\gamma_m n_3 h} - 1) \frac{\gamma_m Z_{31} - 1}{\gamma_j Z_{31} - 1} S_{j,m} \frac{d}{w} \quad (75)$$

Projecting (70d) on the set of base vectors given by the waveguide mode $\sin \left[\frac{l\pi}{w} \left(x + \frac{w}{2} \right) \right]$

$$\frac{1}{w} \int_{-w/2}^{w/2} \frac{\partial H^I(x,z,t)}{\partial z} \Big|_{z=0^+} \sin \left[\frac{l\pi}{w} \left(x + \frac{w}{2} \right) \right] dx = \frac{1}{w} \int_{-w/2}^{w/2} \frac{\partial H^{III}(x,z,t)}{\partial z} \Big|_{z=0^-} \sin \left[\frac{l\pi}{w} \left(x + \frac{w}{2} \right) \right] dx \quad (76)$$

and after some calculation we find

$$\begin{aligned} & \frac{1}{w} \int_{-w/2}^{w/2} \sum_{m=1}^{\infty} \sin \left[\frac{m\pi}{w} \left(x + \frac{w}{2} \right) \right] B_m \left(e^{2ik_0\gamma_m n_3 h} \frac{\gamma_m Z_{31} - 1}{\gamma_m Z_{31} + 1} - 1 \right) \gamma_m n_3 \sin \left[\frac{l\pi}{w} \left(x + \frac{w}{2} \right) \right] dx \\ & \quad \underbrace{\hspace{15em}}_{B_l \left(e^{2ik_0\gamma_l n_3 h} \frac{\beta_l Z_{31} - 1}{\beta_l Z_{31} + 1} - 1 \right) \gamma_l n_3} \\ & = \underbrace{\frac{1}{w} \int_{-w/2}^{w/2} \sin \left[\frac{l\pi}{w} \left(x + \frac{w}{2} \right) \right] n_1 \beta_0 e^{ik_0 \gamma_0 x} dx}_{S_{0,l} n_1 \beta_0} - \underbrace{\frac{1}{w} \int_{-w/2}^{w/2} \sin \left[\frac{l\pi}{w} \left(x + \frac{w}{2} \right) \right] \sum_{n=-\infty}^{\infty} R_n \beta_n n_1 e^{-ik_0 \gamma_n x} dx}_{\sum_{n=-\infty}^{\infty} S_{n,l} \beta_n n_1} \end{aligned} \quad (77)$$

We get for the amplitude coefficient B_l

$$B_l = \frac{1}{\left(e^{2ik_0\gamma_l n_3 h} \frac{\beta_l Z_{31} - 1}{\beta_l Z_{31} + 1} - 1 \right) \gamma_l n_3} \left\{ S_{0,l} - \sum_{n=-\infty}^{\infty} S_{n,l} R_n \beta_n \right\} \quad (78)$$

The reflection coefficient differs from the one given by Barbara due to the other interpretation of the surface impedance as boundary condition.

4 Basic concepts

This chapter is intended to present all formulas in a consistent way and to propose a reference for the following numerical analysis.

First, we introduce the main concepts about reflection, refraction and diffraction of light. Next, we generalise the concept used for the rigorous analysis. We finish with a summary on evanescent waves and intensity in view of an experimental verification.

4.1 Reflection at and transmission through an interface

Combining the Maxwell equations (1) and the time independent wave equation (6), we derive that the field vectors \vec{E} and \vec{H} and the wavevector \vec{k} form an orthogonal right-hand system. Therefore, a general plane wave is a linear combination of two orthogonally polarised components - a *transverse electric* (TE = s-polarised) and a *transverse magnetic* (TM = p-polarised) component. Figure 15 details the situation for a plane TE wave incident on the interface $I \rightarrow II$. The transverse field vectors reduce to a single non-null component E_y for a TE wave, respectively H_y for a TM wave.

$$\vec{E} = \begin{pmatrix} 0 \\ E_y \\ 0 \end{pmatrix} e^{ik_0(\gamma x + \beta z)} \quad \vec{H} = \begin{pmatrix} 0 \\ H_y \\ 0 \end{pmatrix} e^{ik_0(\gamma x + \beta z)} \quad (79)$$

where

$$\begin{aligned} \beta &= n \cos \theta &= \pm k_y/k_0 \\ \gamma &= n \sin \theta &= k_x/k_0 \\ n &= \sqrt{\epsilon} &\text{is the complex refraction index} \\ \theta &= \angle(\perp, \vec{k}) &\text{is the angle from the surface normal to the propagation direction} \end{aligned}$$

Shown in figure 15, the tangential components E_y and H_y have to be continuous across the interface as we have seen in subchapter 2.3.

$$E_y^I = E_y^{II} \quad H_y^I = H_y^{II} \quad (80a)$$

Of course, this rule applies also for the tangential components of the associated magnetic and electric field.

$$H_x^I = H_x^{II} \quad E_x^I = E_x^{II} \quad (80b)$$

Using the Maxwell equation (1a) and (1c), we derive ¹¹

$$\begin{aligned} H_x &= -\frac{ik_0\beta}{i\omega\mu_0} E_y & E_x &= \frac{ik_0\beta}{i\omega\epsilon_0\epsilon} H_y \\ &= -\beta \sqrt{\frac{\epsilon_0}{\mu_0}} E_y & &= \frac{\beta}{n} \sqrt{\frac{\mu_0}{\epsilon_0}} H_y \end{aligned}$$

At the interface $z = 0$, the boundary conditions (80) yield for a TE wave

$$E_y^I = E_0 \left(e^{ik_0\gamma_i x} + R_{12} e^{ik_0\gamma_r x} \right) = E_y^{II} = E_0 T_{12} e^{ik_0\gamma_t x} \quad (81a)$$

$$H_x^I = -H_0 \left(\beta_i e^{ik_0\gamma_i x} - R_{12} \beta_r e^{ik_0\gamma_r x} \right) = H_x^{II} = -H_0 T_{12} \beta_t e^{ik_0\gamma_t x} \quad (81b)$$

Where R_{12} denotes the reflection coefficient at and T_{12} the transmission coefficient through the interface $I \rightarrow II$. Also, $H_0 = \sqrt{\frac{\epsilon_0}{\mu_0}} E_0 = Z_0^{-1} E_0$, where Z_0 is the electromagnetic impedance in free space.

Equations (81) must be fulfilled for any point on the surface. This means that the phases $ik_0\gamma_r x$, $ik_0\gamma_t x$ and $ik_0\gamma_i x$ must be equal. Therefore, if we set¹²

$$\gamma_0 = \gamma_i = \gamma_r = \gamma_t \quad (82a)$$

$$\beta_1 = \beta_i = \beta_r \quad (82b)$$

$$\beta_2 = \beta_t \quad (82c)$$

equations (81) simplify to

$$1 + R_{12} = T_{12}$$

$$\beta_1(1 - R_{12}) = \beta_2 T_{12}$$

and we get the *Fresnel coefficients* for a TE wave

$$R_{12} = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \quad (83a)$$

$$R_{21} = -R_{12} \quad (83b)$$

$$T_{12} = 1 + R_{12} \quad (83c)$$

$$T_{21} = 1 - R_{12} \quad (83d)$$

Respectively, for TM waves, equations (82) and (83) remain valid¹³, but equation (83a) transforms to

$$R_{12} = \frac{\epsilon_2\beta_1 - \epsilon_1\beta_2}{\epsilon_2\beta_1 + \epsilon_1\beta_2} \quad (84)$$

¹²Note that the refraction index is contained in γ and β .

¹³Though they apply now to the magnetic field.

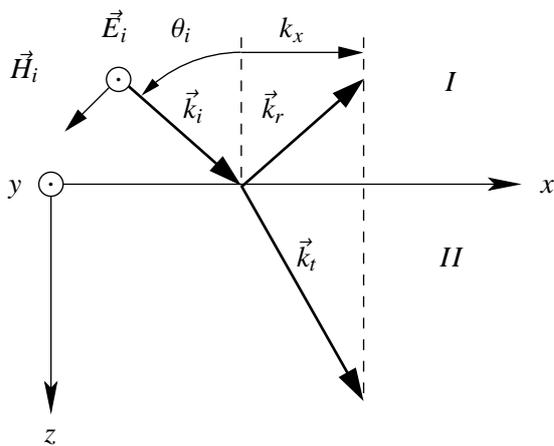


Figure 15: Reflection at and transmission through an interface $I \rightarrow II$ at $z = 0$ for a plane TE wave. The interface lies in the xy -plane, whereas xz is the incidence plane.

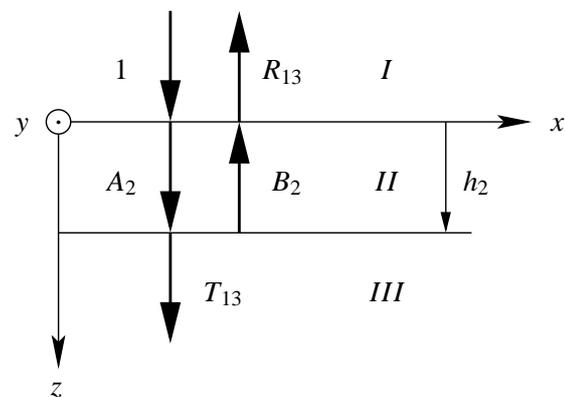


Figure 16: Reflection at and transmission through a layer II between $z = 0$ and $z = h_2$. As before, xz is the incidence plane and the interfaces are parallel to the xy -plane.

4.2 Reflection at and transmission through layers

Here, we consider the reflection and transmission of light incident onto a stack of material layers. Figure 16 outlines the simplest example made of a single layer.

First, we express the transverse electromagnetic field U_y , where \vec{U} denotes either the electric field \vec{E} for TE or the magnetic field \vec{H} for TM polarisation. The incident wave is set to unit amplitude. Hence, the field writes as

$$\begin{aligned} U_y^I &= e^{ik_0(\gamma_0 x + \beta_1 z)} + R_{13} e^{ik_0(\gamma_0 x - \beta_1 z)} && \text{incident and reflected wave} \\ U_y^{II} &= A_2 e^{ik_0(\gamma_0 x + \beta_2 z)} + B_2 e^{ik_0(\gamma_0 x - \beta_2 z)} && \text{propagating and withdrawing wave} \\ U_y^{III} &= T_{13} e^{ik_0(\gamma_0 x + \beta_3(z-h_2))} && \text{transmitted wave} \end{aligned}$$

Knowing the Fresnel coefficients at the layer interfaces, we obtain

$$R_{13} = R_{12} + B_2 T_{21} \quad (85a)$$

$$A_2 = T_{12} + B_2 R_{21} \quad (85b)$$

$$B_2 = A_2 R_{23} e^{2ik_0 \beta_2 h_2} \quad (85c)$$

$$T_{13} = A_2 T_{23} e^{ik_0 \beta_2 h_2} \quad (85d)$$

Combining equation (85b) and (85c), we find

$$\begin{aligned} A_2 &= T_{12} + A_2 R_{21} R_{23} e^{2ik_0 \beta_2 h_2} \\ &= T_{12} - A_2 R_{12} R_{23} e^{2ik_0 \beta_2 h_2} \\ &= \frac{T_{12}}{1 + R_{12} R_{23} e^{2ik_0 \beta_2 h_2}} \end{aligned} \quad (86a)$$

Hence, the other equations yield

$$B_2 = \frac{T_{12} R_{23} e^{2ik_0 \beta_2 h_2}}{1 + R_{12} R_{23} e^{2ik_0 \beta_2 h_2}} \quad (86b)$$

$$T_{13} = \frac{T_{12} T_{23} e^{ik_0 \beta_2 h_2}}{1 + R_{12} R_{23} e^{2ik_0 \beta_2 h_2}} \quad (86c)$$

$$\begin{aligned} R_{13} &= R_{12} + \frac{T_{12} T_{21} R_{23} e^{2ik_0 \beta_2 h_2}}{1 + R_{12} R_{23} e^{2ik_0 \beta_2 h_2}} \\ &= \frac{R_{12} + R_{23} e^{2ik_0 \beta_2 h_2}}{1 + R_{12} R_{23} e^{2ik_0 \beta_2 h_2}} \end{aligned} \quad (86d)$$

In the presence of a stack of material layers, we can sequentially step through the layers. Beginning with the last one, we use equations (86) and add iteratively the previous layers.

Example:

- In the presence of three surfaces, we compute the coefficients R_{24} , A_3 , B_3 and T_{24} first.
- Then, we use R_{24} and T_{24} instead of R_{23} and T_{23} to compute R_{14} , A_2 , B_2 and T_{14} .
- Finally, we account for the modified incidence on surface $II \rightarrow III$ and multiply A_3 and B_3 by $A_2 e^{ik_0 \beta_2 h_2}$.

Caution:

Care must be taken when computing the fields in a layer stack because the z -coordinate is always referred to the corresponding surface. If the first surface is at $z = 0$, the transmitted field writes as

$$U_{y(x,z>h_2+h_3)}^{IV} = T_{14} e^{ik_0(\gamma_0 x + \beta_4(z-h_2-h_3))}$$

4.3 Diffraction versus refraction

Subchapter 4.1 showed that the refraction of light conserves the tangential wavevector components at any interface. Review figure 15 where the tangential component $\vec{k}_{\parallel} = k_x$ is conserved in all regions. Hence, we find a major difference between refraction and diffraction:

- At an interface, refraction exclusively modifies the component \vec{k}_{\perp} normal to the interface. The modification is proportional to the change of the refraction index.
- Diffraction mainly modifies the component \vec{k}_{\parallel} tangential to the interface respectively normal to the propagation direction. As a side effect, changing \vec{k}_{\parallel} influences \vec{k}_{\perp} to keep $k = |\vec{k}_{\perp} + \vec{k}_{\parallel}|$ proportional to the refraction index. We should also note that *diffraction occurs everywhere in space* regardless of the modification of the refraction index.

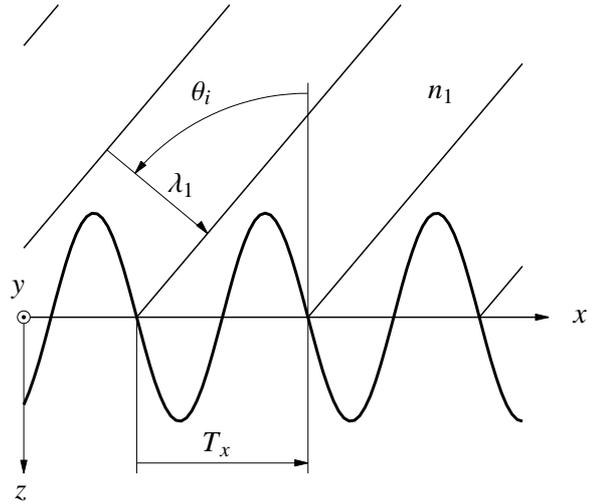


Figure 17: A plane wave falls under an angle θ_i onto the xy -plane. Its amplitude on the x -axis is a moving sine with period T_x .

As figure 17 shows, an incident plane wave creates a sinusoidal field on the surface which is the amplitude and phase of the transverse field component U_y . This field depends on the incidence angle θ_i and the free space wavelength λ_0 .

$$U_{y(x,z=0)} = U_i e^{ik_x x}$$

with

$$k_x = k_i \sin \theta_i = n_1 k_0 \sin \theta_i = \frac{2\pi}{T_x} \quad \text{is the projection of } \vec{k} \text{ onto the surface}$$

$$\frac{k_x}{k_0} = n_1 \sin \theta_i = \frac{2\pi}{T_x} \frac{\lambda_0}{2\pi} = \frac{\lambda_0}{T_x} \quad \text{where } T_x \text{ is the spatial period in } x \quad (87)$$

Application of the Gauss theorem:

If we know the wavelength of an electromagnetic wave and the field it creates on a closed surface, we definitively know how it propagates in space. The sense of propagation is determined when we know it at a single point.

We limit on a plane $z = z_0$ separating the space into two half-spaces. If we know the wavelength of the wave and its field on the plane, then we know the total field at any point in space. The propagation direction θ_i is uniquely fixed by equation (87) and its sense is known when we fix it for some point.

A plane wave has infinite lateral extent. However electromagnetic waves are always of limited in space. This limitation causes diffraction.

Any signal of finite energy can be presented as well in real time-space domain as in temporal-spatial frequency domain. By means of the Fourier transformation, we can represent the fields in the frequency domain as well as in the real space. Hence, we compute the spatial spectrum of the wave pattern at $z = z_0$ and use the inverse Fourier transform to compute the electromagnetic field at any point in space.

$$F(k_x) = \int_{-\infty}^{+\infty} U(x, z=z_0) e^{-ik_x x} dx \quad \text{plane wave decomposition} \quad (88)$$

$$U(x, z) = \int_{-\infty}^{+\infty} F(k_x) e^{ik_x x + ik_z(z-z_0)} dk_x \quad \text{three-dimensional reconstruction} \quad (89)$$

This is the *Fourier expansion* of the signal $U(x, z)$. $U(x, z)$ is a weighted sum of plane waves, where the weights are the Fourier coefficients $F(k_x)$. See figure 18 and 19 for an example.

4.4 Propagation through a pierced metal layer

Underlying principle:

Figure 20 outlines the model geometry treated in this subchapter. Region *II* adjacent to region *I* contains the metallic film. Region *III* is an aperture through the film to region *IV*.

In region *I*, a plane wave \vec{U}_i is incident on the boundary $I \rightarrow II$. \vec{U}_i is either TE or TM polarised. A part \vec{U}_r is reflected back into region *I* whereas \vec{U}_t is transmitted through the film into region *IV*. We model the electromagnetic field in five steps:

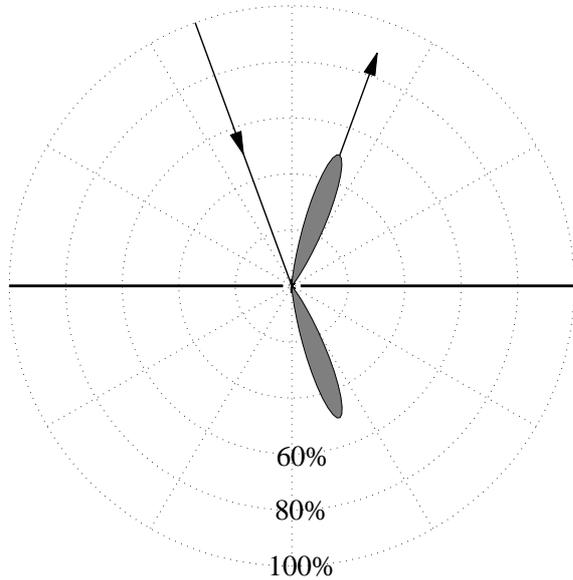


Figure 18: Spatial intensity spectrum of the incident, reflected and transmitted light on a 10nm thick metal sheet with no impedance. The aperture is $2\mu\text{m}$ wide and $100\mu\text{m}$ long (not shown). The plane wave arrives at $\theta_i = 20^\circ$ in TM polarisation. Its wavelength is $\lambda_0 = 633\text{nm}$. Note that the incident and the direct reflected wave are at 0.7% of their nominal values.

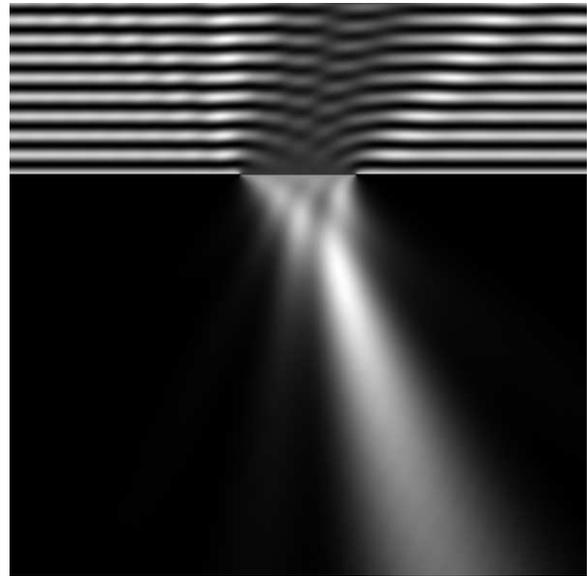


Figure 19: Fourier expansion of the spatial spectrum shown at the left. The transmitted intensity has been doubled for better contrast.

1. Find the field $U_{c(\vec{r})}$ for a metal film without any aperture.
2. Determine the waveguide modes inside the aperture. For each eigenmode, compute the reflection and transmission through the aperture.
3. Expand the incident wave into a linear combination of these eigenmodes.
4. At the interfaces $I \rightarrow III$ and $III \rightarrow IV$, determine the differences $\Delta\vec{U}_r$ and $\Delta\vec{U}_t$ between the effective amplitudes and the fields obtained in step 1.
5. Expand $\Delta\vec{U}_r$ and $\Delta\vec{U}_t$ into plane waves and propagate them in the regions I and IV .

Caution:

For the following discussion, we replace \vec{U} by its scalar form U .

Mathematical formulation:

1. The solution in region I , II and IV is given in subchapter 4.2 by equations (86). Therefore, we start with the ansatz

$$U_{c(x,y,z<0)}^I = U_{i(\vec{r})} + U_{r(\vec{r})} = U_0 e^{ik_0(\gamma_x x + \gamma_y y)} \left(e^{ik_0\beta_1 z} + R_2 e^{-ik_0\beta_1 z} \right) \quad (90a)$$

for the incident and reflected wave,

$$U_{(x,y,z \in [0,h])}^{II} = U_{a(\vec{r})} + U_{b(\vec{r})} = U_0 e^{ik_0(\gamma_x x + \gamma_y y)} \left(A_2 e^{ik_0\beta_2 z} + B_2 e^{-ik_0\beta_2 z} \right) \quad (90b)$$

for the forward and backward propagating wave, and

$$U_{c(x,y,z>h)}^{IV} = U_{t(\vec{r})} = U_0 e^{ik_0(\gamma_x x + \gamma_y y)} T_2 e^{ik_0\beta_4(z-h)} \quad (90c)$$

for the transmitted wave.

2. The propagation in region III is described by the *eigenmodes* $U_{m(x,y)} e^{ik_0\beta_m z}$.

$$\begin{aligned} U_{(x,y,z \in [0,h])}^{III} &= U_p(\vec{r}) + U_w(\vec{r}) \quad \forall (x,y) \in [\text{aperture}] \\ &= U_0 \sum_m U_{m(x,y)} \left(A_m e^{ik_0\beta_3 z} + B_m e^{-ik_0\beta_3 z} \right) \end{aligned} \quad (91)$$

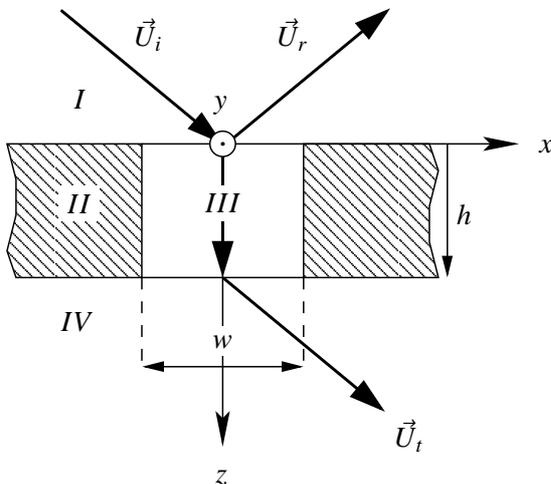


Figure 20: Reflection at and transmission through a metal film of thickness h . The film has a single aperture at $x \in [-\frac{w}{2}, \frac{w}{2}]$.

In equation (91), every eigenmode $U_{m(x,y)}$ is scaled by its coefficient A_m and B_m .

We first determine the eigenmodes and then the Fresnel coefficients R_m , A_m , B_m and T_m for a plane wave at the interfaces $I \rightarrow III$ and $III \rightarrow IV$. The Fresnel coefficients are computed again by equations (86).

3. Next, we decompose the incident field in eigenmode contributions within the aperture. Matching the amplitude of the incident wave to aperture modes is similar to a Fourier series expansion of a periodic signal. Or, to be exact, the Fourier series expansion must take into account the signal period T_0 by containing the harmonics of the ground frequency $f_0 = T_0^{-1}$.

Although the eigenmodes $U_{m(x,y)}$ need not be harmonics of $U_{i(x,y,z=0)}$, the amplitude closely matches to a linear combination of a large number of modes because we need not represent it over a large region.

$$U_{i(x,y,z=0)} = U_0 \sum_m C_m U_{m(x,y)} \quad \forall (x,y) \in [\text{aperture}] \quad (92)$$

with the *coupling coefficients*

$$C_m = \frac{1}{(\text{aperture})U_0} \iint_{\text{aperture}} U_{i(x,y,z=0)} \cdot U_{m(x,y)}^* dx dy \quad (93)$$

4. Based on the coupling coefficients C_m , we calculate the *effective fields* at the interfaces. Hence, $\forall (x,y) \in [\text{aperture}]$

$$\begin{aligned} U_{e(x,y,z=0)}^I &= U_0 \sum_m C_m (1 + R_m) U_{m(x,y)} && \text{field at the interface } I \rightarrow III && (94a) \\ &= U_{i(x,y,z=0)} + U_0 \sum_m C_m R_m U_{m(x,y)} \end{aligned}$$

$$U_{e(x,y,z=h)}^{IV} = U_0 \sum_m C_m T_m U_{m(x,y)} \quad \text{field at the interface } III \rightarrow IV \quad (94b)$$

and we obtain the *perturbations* (or *difference fields*)

$$\Delta U_{r(x,y,z=0)} = U_{e(x,y,z=0)}^I - U_{c(x,y,z=0)}^I \quad \text{at the interface } I \rightarrow III \quad (95a)$$

$$\Delta U_{t(x,y,z=h)} = U_{e(x,y,z=h)}^{IV} - U_{c(x,y,z=h)}^{IV} \quad \text{at the interface } III \rightarrow IV \quad (95b)$$

5. Finally, we have to propagate the perturbations into region I and IV . Applying the Fourier transformation, we get their spatial spectra.

$$\begin{aligned} F_{r(\gamma_{rx}, \gamma_{ry})} &= \iint_{\mathbb{R}^2} \Delta U_{r(x,y,z=0)} e^{-ik_0(\gamma_{rx}x + \gamma_{ry}y)} dx dy \\ &= \iint_{\text{aperture}} \Delta U_{r(x,y,z=0)} e^{-ik_0(\gamma_{rx}x + \gamma_{ry}y)} dx dy \end{aligned} \quad (96a)$$

$$F_{t(\gamma_{tx}, \gamma_{ty})} = \iint_{\text{aperture}} \Delta U_{t(x,y,z=h)} e^{-ik_0(\gamma_{tx}x + \gamma_{ty}y)} dx dy \quad (96b)$$

The expansion into region I yields

$$\begin{aligned} U_{y(x,y,z<0)}^I &= U_0 e^{ik_0(\gamma_x x + \gamma_y y)} \left(e^{ik_0 \beta_1 z} + R_2 e^{-ik_0 \beta_1 z} \right) \\ &\quad + \iint_{\mathbb{R}^2} F_{r(\gamma_{rx}, \gamma_{ry})} e^{ik_0(\gamma_{rx}x + \gamma_{ry}y - \beta_1(\gamma_{rx}, \gamma_{ry})z)} d\gamma_{rx} d\gamma_{ry} \end{aligned} \quad (97a)$$

The propagation in region IV is described in an analogue manner.

$$U_{y(x,y,z>h)}^{IV} = U_0 T_2 e^{ik_0(\gamma_x x + \gamma_y y + \beta_4(z-h))} + \iint_{\mathbb{R}^2} F_t(\gamma_{tx}, \gamma_{ty}) e^{ik_0(\gamma_{tx} x + \gamma_{ty} y + \beta_4(\gamma_{tx}, \gamma_{ty})(z-h))} d\gamma_{tx} d\gamma_{ty} \quad (97b)$$

4.5 Evanescent waves

The results in subchapter 4.4 are a little more *complex* than it seems at first glance. The wavevector \vec{k} may have complex components corresponding to a *damping* or an *amplification* of the field.

For the investigated setups, amplification cannot occur because they are built of purely passive optical elements. Hence, the imaginary part of the wavevector corresponds to damping in that direction. This damping is related to the directional coefficients of the wavevector. Indeed, we have

$$\vec{k}_l = k_0 \begin{pmatrix} \gamma_x \\ \gamma_y \\ \beta_l \end{pmatrix} \quad \text{where} \quad k_l^2 = k_0^2 n_l^2 = k_0^2 \epsilon_l \quad (98a)$$

With the projection factors

$$\epsilon_l = \gamma_x^2 + \gamma_y^2 + \beta_l^2 \quad \text{and} \quad \beta_l = \pm \sqrt{\epsilon_l - \gamma_x^2 - \gamma_y^2} \quad (98b)$$

Obviously, whenever $\gamma_x^2 + \gamma_y^2$ exceeds ϵ_l , the propagation constant $k_0 \beta_l$ is complex. For the sign, we always take β_l positive and fix the sign in the corresponding equations to propagate in the correct sense. As we fix the sign of the real part, we fix it again for the imaginary part to guarantee damping instead of amplification.

- When a wave can not pass through a dielectric interface, it is totally reflected. However, its field enters in the second medium decaying exponentially with increasing distance to the interface. Hence, we have an *evanescent wave* in the second medium with an imaginary propagation constant $k_0 \beta_2$. In this case, the field is damped in the second medium but **not** absorbed.

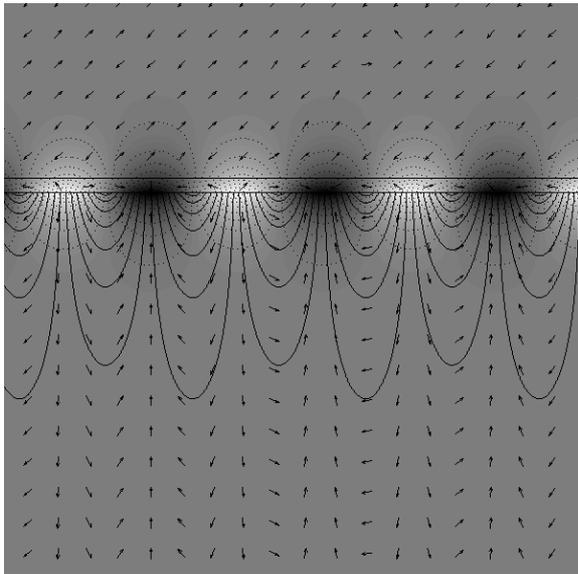


Figure 21: A plane TM HeNe beam is incident through a BK7 glass prism at $\theta_i = 44.5^\circ$ onto a 51.5nm thick gold film. Although the transmission is zero due to total reflection at the opposite gold-air interface, the reflectance is zero too. The beam excites a surface plasmon at the gold-air interface that dissipates the energy in the film and boosts the electric field to 4.23 times the incident amplitude.

Straight lines indicate the interfaces. The shading and dotted lines show the instantaneous electric potential. Solid lines are electric field lines and the vectors point in the direction of the electric field.

- Every metal has a dielectric constant ϵ_m with negative real part. In the case of a good conductor, we have $\Re(\epsilon_m) \ll -1$. Hence, a metal always damps electromagnetic waves.
- If a wave is propagating along the interface and is **evanescent in both medias**, it is a *surface wave* tightly bound to the interface by its nature. In the case of a dielectric-metal or an intermetal interface, surface waves are called *surface plasmons* and can achieve extremely high field amplitudes. The term 'plasmon' stands for the *collective oscillation* of the electron gaz in the metal. See figures 8 and 21 for an illustration.

4.6 Electromagnetic intensity

The *Poynting vector* $\vec{\mathcal{J}}_{(\vec{r},t)}$ corresponds to the energy flux.

$$\vec{\mathcal{J}}_{(\vec{r},t)} = \vec{\mathcal{E}}_{(\vec{r},t)} \times \vec{\mathcal{H}}_{(\vec{r},t)}$$

The instantaneous intensity is given by its absolute value $|\vec{\mathcal{J}}_{(\vec{r},t)}|$ whereas the time-averaged intensity is

$$I_{(\vec{r})} = |\vec{S}_{(\vec{r})}| = \left| \langle \vec{\mathcal{J}}_{(\vec{r},t)} \rangle_t \right| = \frac{1}{2} \Re(\vec{E}_{(\vec{r})} \times \vec{H}_{(\vec{r})}^*) \quad (99a)$$

$$= \frac{1}{2} \Re \left(\sqrt{\frac{\epsilon}{\mu}} \vec{E}_{(\vec{r})} \cdot \vec{E}_{(\vec{r})}^* \right) \quad (99b)$$

$$= \frac{1}{2} \Re \left(\sqrt{\frac{\mu}{\epsilon}} \vec{H}_{(\vec{r})} \cdot \vec{H}_{(\vec{r})}^* \right) \quad (99c)$$

where

$$\begin{aligned} \epsilon &= \epsilon_0 \epsilon_r & \epsilon_r & \text{is the complex dielectric constant} \\ \mu &= \mu_0 \mu_r & \mu_r & \text{is the complex permeability}^{14} \end{aligned}$$

The instantaneous intensity of visible light can not be measured due to the light frequency of about 500THz. In contrast, the time-averaged intensity is easily accessible by many detector types.

- The time-averaged power flowing through an *area* is given by

$$P_{area} = \iint_{area} I_{(\vec{r})} \frac{\vec{S}_{(\vec{r})} \cdot \vec{n}_{(\vec{r})}}{S_{(\vec{r})}} dx dy$$

where

$$\vec{n}_{(\vec{r})} \text{ is the unit normal vector } \perp \text{ area}$$

When a plane wave with wavevector \vec{k} passes through a plane *area*, the power flow simplifies to

$$P_{area} = I_{(area)} \frac{\vec{k} \cdot \vec{n}}{k} area = I_{(area)} \cos(\angle(\vec{k}, \vec{n})) (area)$$

- The imaginary part in equations (99a) describe the mechanical vibrations of charges - for example oscillating electrons in a metal. Hence, the *apparent intensity* is

$$I_{(\vec{r})} = \frac{1}{2} \vec{E}_{(\vec{r})} \times \vec{H}_{(\vec{r})}^* = I_{r(\vec{r})} + iI_{m(\vec{r})}$$

where

$I_r(\vec{r})$ is the radiative power density passing by the point \vec{r}

$I_m(\vec{r})$ counts for the mechanical energy stored at the point \vec{r}

Note that the mechanical energy is bound to its location and does not contribute to the transported energy. In general, intensity stays for the radiative power density only.

5 Numerical model – metallic slit

This chapter details our rigorous computation of the electromagnetic field through a metallic slit. The geometry was already outlined in figure 20, but the incidence plane is now fixed parallel to the xz -plane.

5.1 Ansatz for the electromagnetic field

The incident plane wave may have any polarisation. However, we can always decompose it into its TE and TM parts. Hence, the computation is restricted to the E_y component of the electric field \vec{E} for TE polarisation, respectively H_y of the magnetic field \vec{H} for TM polarisation.

Region I:

$$\begin{aligned} U_{y(x,z)}^I &= U_0 (U_{i(x,z)} + U_{r(x,z)}) \\ &= U_0 \left(e^{ik_0(\gamma_0 x + \beta_1 z)} + R_2 e^{ik_0(\gamma_0 x - \beta_1 z)} + \int_{-\infty}^{+\infty} F_{r(\gamma_r)} e^{ik_0(\gamma_r x - \beta_r z)} d\gamma_r \right) \end{aligned} \quad (100)$$

where

$$\begin{aligned} U_{i(x,z)} & \text{ is the incident plane wave} \\ U_{r(x,z)} & \text{ counts for the reflection} \\ \gamma_0 = n_1 \sin \theta_i & = k_{1x}/k_0 \\ \beta_1 = \sqrt{\epsilon_1 - \gamma_0^2} & = k_{1z}/k_0 \\ \beta_r = \sqrt{\epsilon_1 - \gamma_r^2} & = k_{rz}/k_0 \\ \epsilon_1 = n_1^2 & \text{ is the complex dielectric constant in region I} \\ R_2 & \text{ is the reflection coefficient at the metal film II} \end{aligned}$$

$$F_{r(\gamma_r)} = \int_{-\infty}^{+\infty} \Delta U_{r(x,z=0)} e^{-ik_0 \gamma_r x} dx \quad (101)$$

are the plane wave expansion coefficients of the perturbation

Region II: Metal film

$$\begin{aligned} U_{y(x,z)}^{II} &= U_0 (U_{p(x,z)} + U_{w(x,z)}) \\ &= U_0 (A_2 e^{ik_0(\gamma_0 x + \beta_2 z)} + B_2 e^{ik_0(\gamma_0 x - \beta_2 z)}) \end{aligned} \quad (102)$$

where

$$\begin{aligned} U_{p(x,z)} & \text{ is the forward propagating wave} \\ U_{w(x,z)} & \text{ is the backward propagating wave} \\ \beta_2 = \sqrt{\epsilon_2 - \gamma_0^2} & = k_2/k_0 \\ \epsilon_2 = n_2^2 & \text{ is the complex dielectric constant in region II} \\ A_2 & \text{ is the internal transmission coefficient in II} \\ B_2 & \text{ is the internal reflection coefficient in II} \end{aligned}$$

Region III: Slit

For simplicity, the vertical walls of the slit are assumed to have no impedance. Therefore, the slit is considered as a hollow, ideally metallic waveguide. Its transverse eigenmodes must not have an electric field E_{\parallel} on the walls. Hence

$$U_{m(x)}^{III} = E_{m(x)}^{III} = E_m \sin\left(m\frac{\pi}{w}\left(x + \frac{w}{2}\right)\right) \quad \forall m \in \mathbb{N} \quad (103a)$$

for TE polarisation as in figure 22 and

$$U_{m(x)}^{III} = H_{m(x)}^{III} = H_m \cos\left(m\frac{\pi}{w}\left(x + \frac{w}{2}\right)\right) \quad \forall m \in \mathbb{N}_0 \quad (103b)$$

for TM polarisation. The modal expansion yields

$$E_{y(x,z)}^{III} = E_0 \sum_{m=1}^{+\infty} C_m(\gamma_m) \sin\left(m\frac{\pi}{w}\left(x + \frac{w}{2}\right)\right) (A_{3(\gamma_m)} e^{ik_0\beta_m z} + B_{3(\gamma_m)} e^{-ik_0\beta_m z}) \quad (104a)$$

respectively

$$H_{y(x,z)}^{III} = H_0 \sum_{m=0}^{+\infty} C_m(\gamma_m) \cos\left(m\frac{\pi}{w}\left(x + \frac{w}{2}\right)\right) (A_{3(\gamma_m)} e^{ik_0\beta_m z} + B_{3(\gamma_m)} e^{-ik_0\beta_m z}) \quad (104b)$$

where

$$\gamma_m = \frac{m\pi}{wk_0} = k_{mx}/k_0 \quad (105a)$$

$$\beta_m = \sqrt{\epsilon_3 - \gamma_m^2} = k_{mz}/k_0 \quad (105b)$$

$$\epsilon_3 = n_3^2$$

is the complex dielectric constant in the slit III

$$A_3$$

is the internal transmission coefficient in III

$$B_3$$

is the internal reflection coefficient in III

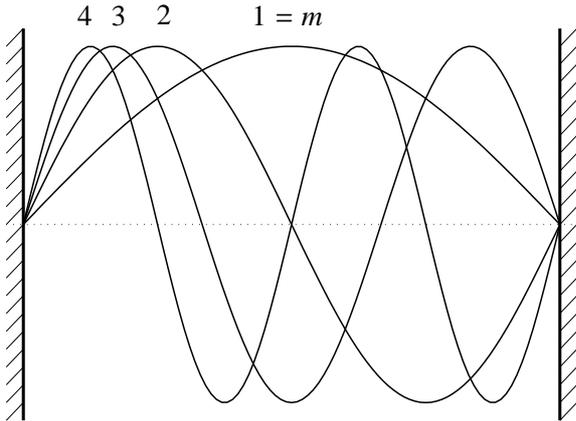


Figure 22: Electric field of the four lowest TE modes in a hollow, ideally metallic slit. Modes are standing waves, where the mode number m yields the phase increase $2m\pi$.

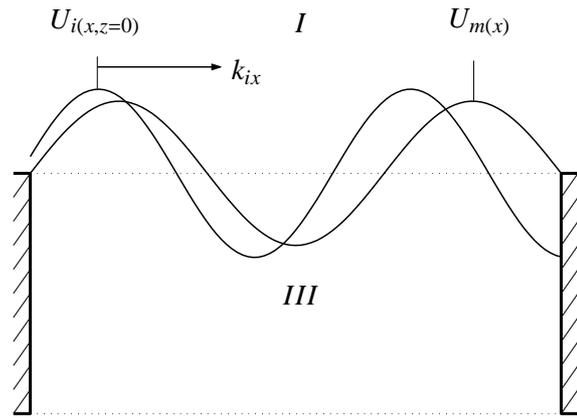


Figure 23: Field $U_{i(x,z=0)}$ of the incident wave entering in the aperture. The mode $U_{m(x)}$ illustrates the eigenmode $m = 3$. Whereas the mode field does not move, the incident field laterally shifts with a speed that is proportional to k_{ix} . The coupling coefficient C_m describes the overlap between the incident field and the m^{th} mode.

$$C_{m(\gamma_m)} = \frac{2}{w} \int_{-w/2}^{+w/2} E_{i(x,z=0)} \sin\left(m\frac{\pi}{w}\left(x + \frac{w}{2}\right)\right) dx \quad (106a)$$

respectively

$$C_{m(\gamma_m)} = \frac{2 - \delta(m)}{w} \int_{-w/2}^{+w/2} H_{i(x,z=0)} \cos\left(m\frac{\pi}{w}\left(x + \frac{w}{2}\right)\right) dx \quad (106b)$$

are the coupling coefficients from the incident wave to the slit modes

See figure 23 for an illustration.

Region IV:

$$\begin{aligned} U_{y(x,z)}^{IV} &= U_0 (U_{t(x,z)}) \\ &= U_0 \left(T_2 e^{ik_0(\gamma_0 x + \beta_4 z)} + \int_{-\infty}^{+\infty} F_{t(\gamma_t)} e^{ik_0(\gamma_t x + \beta_t z)} d\gamma_t \right) \end{aligned} \quad (107)$$

where

$$\begin{aligned} U_{t(x,z)} & \text{stands for the transmitted waves} \\ \beta_4 &= \sqrt{\epsilon_4 - \gamma_0^2} = k_{4z}/k_0 \\ \beta_t &= \sqrt{\epsilon_4 - \gamma_t^2} = k_{tz}/k_0 \\ \epsilon_4 &= n_4^2 \text{ is the complex dielectric constant in region IV} \\ T_2 & \text{is the transmission coefficient through the metal film II} \end{aligned}$$

$$F_{t(\gamma_t)} = \int_{-\infty}^{+\infty} \Delta U_{t(x,z=h)} e^{-ik_0 \gamma_t x} dx \quad (108)$$

are the plane wave expansion coefficients of the perturbation

5.2 Discretisation and numerical evaluation

We are going to present the numerical analysis for the TE polarisation. The results for the TM polarisation can be derived in an analogue manner.

Region III: Slit

We start computing the coupling coefficients C_m from the incident wave to the slit modes. Based on figure 23, we discretise the slit segment $x \in [-\frac{w}{2}, \frac{w}{2}]$ into $2N$ segments Δx . The integral in equation (106a) is Hence

$$C_{m(\gamma_m)} = \frac{2}{w} \sum_{n=-N}^N E_{i(n\Delta x, z=0)} \sin\left(m\frac{\pi}{w}\left(n\Delta x + \frac{w}{2}\right)\right) \Delta x$$

Taking $\Delta x = \frac{w}{2N}$, we obtain

$$C_{m(\gamma_m)} = \frac{1}{N} \sum_{n=-N}^N e^{ik_0 \gamma_0 n \Delta x} \sin\left(m\frac{\pi}{w}\left(n\Delta x + \frac{w}{2}\right)\right) \quad (109)$$

We consider only the first M slit modes. To get the coefficients C_1 to C_M with good accuracy, we set

$$N_{(M)} \geq 5M \quad \text{to ensure at least 20 sampling points for the eigenmodes.}$$

Introducing equations (105),

$$\beta_{1m} = \sqrt{\epsilon_1 - \gamma_m^2} \quad \text{and} \quad \beta_{4m} = \sqrt{\epsilon_4 - \gamma_m^2}$$

in equations (83), we compute the reflection and transmission coefficients through the slit by means of equations (86)¹⁵. We determine the effective fields

$$E_{e(x,z=0)}^I = E_0 \sum_{m=1}^M C_m E_m(x) (1 + R_m) \quad \text{at the interface } I \rightarrow III$$

$$E_{e(x,z=h)}^{IV} = E_0 \sum_{m=1}^M C_m E_m(x) T_m \quad \text{at the interface } III \rightarrow IV$$

and get the perturbations

$$\Delta E_{r(x,z=0)} = E_{e(x,z=0)}^I - E_0 e^{ik_0 \gamma_0 x} \quad \text{in region } I \quad (110a)$$

$$\Delta E_{t(x,z=h)} = E_{e(x,z=h)}^{IV} - E_0 T_2 e^{ik_0 \gamma_0 x} \quad \text{in region } IV \quad (110b)$$

Figure 24 shows the typical differences between the incident field and its modal expansion.

Now, we computed the complete solution in the metallic slit III . Next, we propagate the perturbations into region I and IV .

Region I and IV :

Introducing a spatial increment Δx and a frequential increment $\Delta \gamma$ in equation (96a), we get the discrete spectrum

$$F_r(\gamma_r) = \Delta x \sum_s \Delta E_{r(x,z=0)} e^{-ik_0 \gamma_r x} \quad (111)$$

¹⁵The eigenmodes are a superposition of plane waves in region I and IV with $\gamma_{1m} = \gamma_{4m} = \pm \gamma_m$. So, we compute the Fresnel coefficients R_{13} , T_{13} , R_{34} and T_{34} as for plane waves.

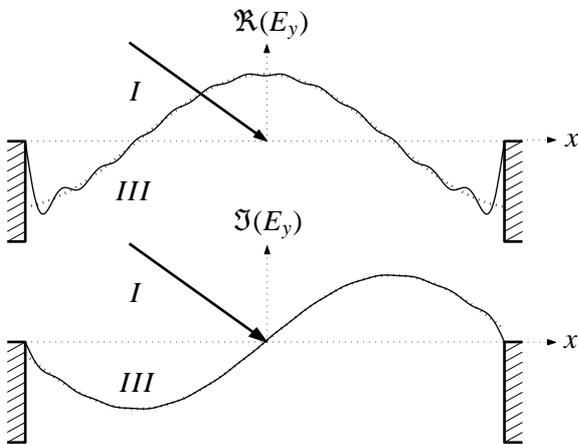


Figure 24: This example illustrates a plane wave incident under $\theta_i = 48^\circ$ onto a freely suspended gold film. The wavelength is $\lambda_0 = 633\text{nm}$ and the slit width is $w = 800\text{nm}$.

The thick dashed lines show E_i and the thin solid lines its modal expansion $E_0 \sum_{m=1}^M C_m E_m$ computed with $M = 26$ slit modes. Note the good accuracy except near the walls of the slit due to the assumption of ideally conductive walls.

where

$$\gamma_r = r\Delta\gamma \quad \forall r \in \mathbb{Z} \quad \wedge \quad |\gamma_r| < \frac{\lambda_0}{2\Delta x} \quad \text{are the sampled spatial frequencies} \quad (112a)$$

$$x = s\Delta x \quad \forall s \in \mathbb{Z} \quad \wedge \quad |x| < \frac{w}{2} \quad \text{are the sampled positions in the slit} \quad (112b)$$

The plane wave expansion (100) rewrites as

$$E_{y(x,z<0)}^I = E_0 e^{ik_0\gamma_0 x} \left(e^{ik_0\beta_1 z} + R_2 e^{-ik_0\beta_1 z} \right) + \sum_r F_{r(\gamma_r)} e^{ik_0(\gamma_r x - \beta_r z)} \quad (113)$$

In an analogue manner, applying equation (96b) and (107) yields the result for region *IV*.

Discussion:

- Equation (111) is nothing else than the discrete Fourier transformation *DFT* of the perturbation in region *I*. Or, as a major difference to the continuous Fourier transformation, the DFT has a **periodic spectrum** and its inverse transformation creates a **periodic signal**. Indeed, the spatial (frequential) sampling forces the spectrum (signal) to be periodic where

$$\Gamma = \frac{2\pi}{k_0\Delta x} = \frac{\lambda_0}{\Delta x} \quad \text{is the spectral period} \quad (114a)$$

and

$$T = \frac{2\pi}{k_0\Delta\gamma} = \frac{\lambda_0}{\Delta\gamma} \quad \text{is the spatial period.} \quad (114b)$$

Therefore, we limit computation to the first period

$$|\gamma_r| < \frac{\Gamma}{2} \quad \text{in spectrum}$$

and

$$|x| < \frac{T}{2} \quad \text{in space}$$

and get the full spectrum respectively field by simple repetition. Whereas the full spectrum has no particular interest ¹⁶, the spatial period significantly changes the field characteristics. Hence, *we always get a periodic field when using a numerical plane wave expansion* in region *I* and *IV*.

- We could evaluate (111) for many γ_r to get a pseudo-continuous spectrum and a very large field period *T*. Unfortunately, the computation time increases with the number of considered spatial frequencies. Subchapter 5.6 estimates the calculation time by means of an example.

Next, we modify the metal film *II* and introduce a periodic array of slits instead of the single one we just calculated. The following subchapter 5.3 outlines the appropriate modifications to evaluate an array of slits. Subchapter 5.5 gives hints about a reasonable accuracy-performance trade-off.

5.3 Evaluation of a periodic slit array

In the metal film *II*, we introduce a periodic array of slits, where *p* is the distance between the centres of two neighbouring slits ¹⁷.

¹⁶The higher orders are present because of aliasing. Review Shannon's sampling theorem for further information.

¹⁷See figure 20.

Region III: Slits

The eigenmodes of the slits are conserved according to equations (103). Here, $U_{y(x)}^{III}$ is the field in all slits and $U_{y(x)}^{III}$ in a single slit as in (104). In principle, we can simply repeat $U_{y(x)}^{III} \forall |x| < \frac{w}{2}$ at every position $x_n = np$. Taking into account the spatial phase shift $\Delta\phi = k_0\gamma_0 x_n$ per slit position, we get for the slit array

$$U_{y(x)}^{III} = \sum_{n=-\infty}^{+\infty} U_{y(x-np)}^{III} e^{ik_0\gamma_0 np} \quad (115)$$

Region I and IV:

We use again the discrete Fourier transform (111). Inserting the spatial period $p = T$ in equation (114b), the frequential increment

$$\Delta\gamma = \frac{\lambda_0}{p} \quad (116)$$

is now fixed by the sample geometry. Nevertheless, the sampling step Δx remains a free parameter. To take advantage of the discrete fast Fourier transformation *DFFT*, 2^l sampling points a period p must be chosen¹⁸ - where $l \in \mathbb{N}$.

Discussion:

- We passed from a single slit to a periodic array of slits to take advantage of the DFT. The spectral increment $\Delta\gamma$ guarantees a phase increment $\Delta\phi_0 = 2\pi$ over the period p of the slit array. Every sampling frequency of the discrete spectrum is an integer multiple of $\Delta\gamma$ and gives rise to $\Delta\phi_r = 2r\pi$ per period. Therefore, we can effectively limit the plane wave reconstruction on a period p and repeat the result to get the field at any point in region *I* and *IV*.
- But what is about the phase increment $\Delta\phi_i = k_0\gamma_0 p$ imposed by the incident plane wave? In general, $\Delta\phi_i \notin \{\Delta\phi_r\}$ and the reconstructed field is **not** periodic on p . Do we need to rebuild the concept from the start accounting for this very likely case? No, but we have to modify the plane wave expansion method instead!

5.4 Rayleigh plane wave expansion

We computed the eigenmodes in region *III* and expressed U^I as linear combination of them. In analogy, we determine the 'correct' eigenmodes in region *I* and *IV* and express U^I and U^{IV} in terms of these modes.

The eigenmodes in region *I* and *IV* are all plane waves with different propagation directions¹⁹. The discrete Fourier transformation selects a particular subset of plane waves given by equation (116) and (112a). We have to modify this subset to account for the incident plane wave U_i and do not change the sampling step $\Delta\gamma$. However, we shift the sampling frequencies by an arbitrary value. Therefore, we select

$$\gamma_r = \gamma_0 + r\Delta\gamma = \gamma_0 + r\frac{\lambda_0}{p} \quad \forall r \in \mathbb{Z} \quad \wedge \quad |\gamma_r| < \frac{\Gamma}{2} = \frac{\lambda_0}{2\Delta x} \quad (117a)$$

and similar

$$\gamma_t = \gamma_0 + t\Delta\gamma = \gamma_0 + t\frac{\lambda_0}{p} \quad \forall t \in \mathbb{Z} \quad \wedge \quad |\gamma_t| < \frac{\Gamma}{2} = \frac{\lambda_0}{2\Delta x} \quad (117b)$$

¹⁸More sampling points give access to higher frequencies as we have seen in equation (114a).

¹⁹Review figure 17.

to conserve the phase increment $\Delta\phi_i \in \{\Delta\phi_r\}$ and $\Delta\phi_i \in \{\Delta\phi_t\}$. This modified Fourier expansion is called *Rayleigh expansion*.

Discussion:

- Remind that the Fraunhofer diffraction formula and the Fourier transformation are very similar. The Fraunhofer formula describes the diffraction pattern in the very far field behind an object. In the case of a periodic object, only particular waves contribute to the far field diffraction pattern because most of the waves mutually erase each other by destructive interference. The waves we just selected are such particular waves as illustrated in figure 25.
- Note that equations (117) analytically describe the *Ewald sphere* shown in figure 26. We applied the Ewald sphere by means of equations (98) to our setup²⁰.

²⁰The concept of the Ewald sphere is widely applied in crystallography to describe diffraction experiments.

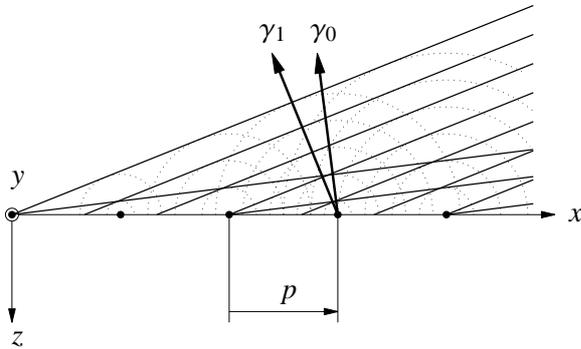


Figure 25: Rayleigh expansion and far field pattern. The incident wave imposes a phase shift $\Delta\phi_i = k_0\gamma_0 p$ between successive slits. This phase shift creates the plane wave propagating at γ_0 . The slit period p fixes $\Delta\gamma$ such that the next plane wave propagates at γ_1 .

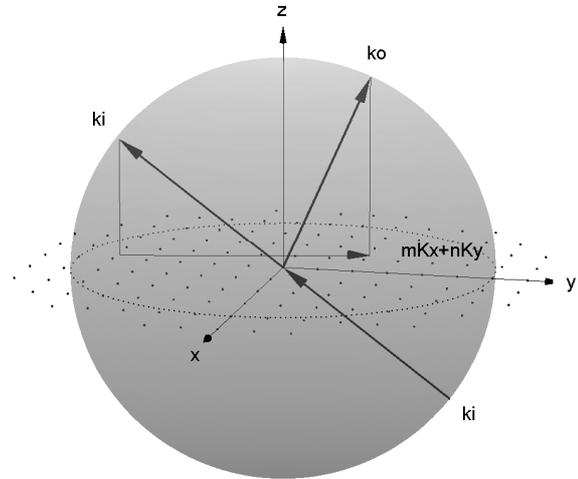


Figure 26: Ewald sphere illustrating the Rayleigh diffraction at a periodic object in the xy -plane. The object has a period p_x in x -direction and p_y in y -direction represented as grating $\vec{K}_x = \frac{2\pi}{p_x}\vec{e}_x$ and $\vec{K}_y = \frac{2\pi}{p_y}\vec{e}_y$. The object modifies the tangential wavevector component where $\vec{k}_{o\parallel} - \vec{k}_{i\parallel} = m\vec{K}_x + n\vec{K}_y$ is always the sum of integer multiples of the grating vectors \vec{K}_x and \vec{K}_y . Propagating waves have the same wavelength and their wavevectors \vec{k}_o finish on the surface of the Ewald sphere defined by \vec{k}_i . But evanescent waves have always wavevectors parallel to the xy -plane and $k_o > k_i$.

5.5 Accuracy of numerical results

This subchapter outlines three major parameters fixing the accuracy and performance of numerical analysis:

1. The number M of eigenmodes in the aperture ²¹.
2. The number of plane waves in the Rayleigh expansion.
3. The number of necessary sampling points to calculate the field.

1. Eigenmodes:

The number of eigenmodes M in the slits is determined by

- setting a criterion on the difference of the incident field and its modal expansion. For example, setting M to fulfil

$$\frac{\langle |U_i - U_0 \sum_{m=1}^M C_m U_m|^2 \rangle}{\langle |U_i|^2 \rangle} < P_{tot} = \text{maximum relative intensity mismatch} \quad (118)$$

allows an accurate analysis in all regions.

- setting a transmission criterion. Choosing M to achieve $\forall n > M$

$$\frac{|C_n T_n|^2}{|T_m|^2} < P_{mod} = \text{contribution to the transmitted intensity of the } n^{\text{th}} \text{ mode} \quad (119)$$

where T_m is the zero-order transmission with

$$m = \begin{cases} 0 & \text{for TM polarisation} \\ 1 & \text{for TE polarisation} \end{cases}$$

This criterion allows an accurate analysis of the transmitted field in region *IV*. Introducing $|C_n| \leq 1$ and $|T_n| < |e^{ik_0 \beta_n h}| = e^{\Re(ik_0 \beta_n h)}$, we get an explicit formula.

$$-\frac{\log(P_{mod}|T_m|^2)}{2k_0 h} \leq \Im(\beta_M) \quad \text{where} \quad \beta_M = \sqrt{\epsilon_3 - \gamma_M^2} \quad (120a)$$

If region *III* is free space or a dielectric, we get

$$\Im(\beta_M) = \sqrt{\gamma_M^2 - \epsilon_3} \quad \text{where} \quad \gamma_M = M \frac{\pi}{w} > \sqrt{\epsilon_3} \quad (120b)$$

2. Plane waves:

Here, we set a criterion on the sampling points to compute the perturbations. This is equivalent to set a limit Γ_{max} .

- The most rigorous criterion fulfils Heisenberg's uncertainty relation

$$\Delta W \cdot \Delta t \gtrsim \hbar \quad \text{where} \quad W = \text{the photon energy}$$

²¹Remind that the aperture is described as a hollow, metallic waveguide.

which sets a lower limit for the accuracy of energy and time measures. Introducing

$$\Delta W = \hbar\omega = \hbar kc \quad \text{and} \quad \Delta t = \frac{\Delta x}{c}$$

we get the *optical criterion*

$$\Delta\gamma \cdot \Delta x \lesssim \frac{1}{k} = \frac{1}{nk_0} \quad \text{and finally} \quad \Delta x \lesssim \frac{1}{nk_0\Delta\gamma} \quad (121)$$

because the frequential sampling $\Delta\gamma$ is fixed by the period p of the slit array.

- Besides the optical criterion just outlined, we may choose another sampling step Δx ²² and explicitly limit the field reconstruction at $|\gamma_r|, |\gamma_t| < \Gamma_{max}$. For example, if we are not interested in the surface field up to d from the interfaces, setting

$$\Gamma_{max} \approx \sqrt{\epsilon + \left(\frac{l}{k_0 d}\right)^2}$$

assures that the field of the omitted surface waves falls off to e^{-l} of their maximal values at the interface. For example, $l = 2$ limits the intensity error for $|z - z_{surface}| > d$ to less than 2%.

3. Sampling points:

Note that the reconstruction of the fields in region *I* and *IV* is the slowest task. This is especially true when we are going to compute the three-dimensional field as explained in chapter 6.

Hence, we should carefully choose the area (volume) where we are interested in. But we may also select some particular points and accommodate the sampling grid²³.

5.6 Calculation example

We assume a gold film with thickness $h = 150\text{nm}$. The slits have a width $w = 200\text{nm}$ and they are repeated at $p = 900\text{nm}$. Region *I* is a BK7 glass²⁴ whereas water fills region *IV* and the slits *III*. A HeNe laser beam ($\lambda_0 = 633\text{nm}$) is incident at $\theta_i = 0^\circ$ under TE polarisation. At this wavelength, we have $\epsilon_1 = 2.296$, $\epsilon_2 \approx -9.386 + 1.147i$ and $\epsilon_3 = \epsilon_4 = 1.774$.

We compute the electromagnetic field in all regions over two periods p along the x -axis and $1.8\mu\text{m}$ along the z -axis in total.

1. We use the criterion (118) and obtain for $P_{tot} = 1\%$ the number of TE modes $M = 51$.
2. Applying the optical criterion with $\Delta\gamma = 0.7033$, $n_1 = 1.515$ and $n_4 = 1.332$, we get $\Delta x_1 \lesssim 95\text{nm}$ and $\Delta x_4 \lesssim 108\text{nm}$. To resolve all M slit modes, we choose $\Delta x_1 = \Delta x_4 = \Delta x = 0.4\text{nm}$ according to equation (109). Now, $\Gamma_{max} = 10 \gtrsim \frac{\lambda_0}{95\text{nm}} = 6.7$ better matches the optical criterion. Maybe, the near field will be inaccurate up to about $d_{(l=2)} \approx \frac{l}{k_0\Gamma_{max}} = 20\text{nm}$ from the gold film *II*.
3. We choose a 300×300 point grid to obtain a uniform resolution of 6nm for the field image.

The calculation on a 2GHz Pentium *IV* machine with MatLab 6.0 took 3seconds. For $\Gamma_{max} = \frac{\lambda_0}{\Delta x} \approx 1600$, it took 35seconds. The overall intensity difference is 0.011%. In the near field up to 20nm , the mean difference is 0.45%. See figure 27 for an illustration of the result.

²²We advise to take the same Δx in equations (109) and (110).

²³As a consequence of our field representation in subchapter 5.1, we get the field accurately at any sampling point. Hence, we can arbitrarily choose a sampling grid without any constraint other than pure imaging considerations.

²⁴For example a standard microscopy cover slip, $170\mu\text{m}$ thick.

5.7 Characteristics of the numerical model

+ Our numerical model is straight forward in the sense that it is **explicit**. It does not contain any implicit equation that needs an iterative numerical resolution. The calculation is fast – even for a high number of aperture modes – at the exception of the plane wave expansion. Unfortunately, the plane wave expansion is slow due to the huge number of performed operations. It cannot become much faster because it is definitively not complexity that slows down.

If we could define the spatial spectrum of the fields in region *I* and *IV* instead of the fields themselves, we could suppress the plane wave expansion. Then, the design process would get the potential for automatisation instead of the actual try-and-error procedure.

- The major weakness is the fact that we do not model any **back-coupling** through the metal film *II*. Light that has passed through the apertures into region *IV* never propagates back through the metal film. This consideration does not apply to the propagating waves, but every surface wave may propagate back into region *I* – namely if it is a surface plasmon.
- We modelled the apertures as hollow, ideally metallic waveguides neglecting the **resistive loss** on the vertical walls. This is a minor weakness because it can be solved without significantly increasing the complexity of the calculation. Instead of the ideal waveguide modes, we should consider the real eigenmodes that extend laterally into region *II*.

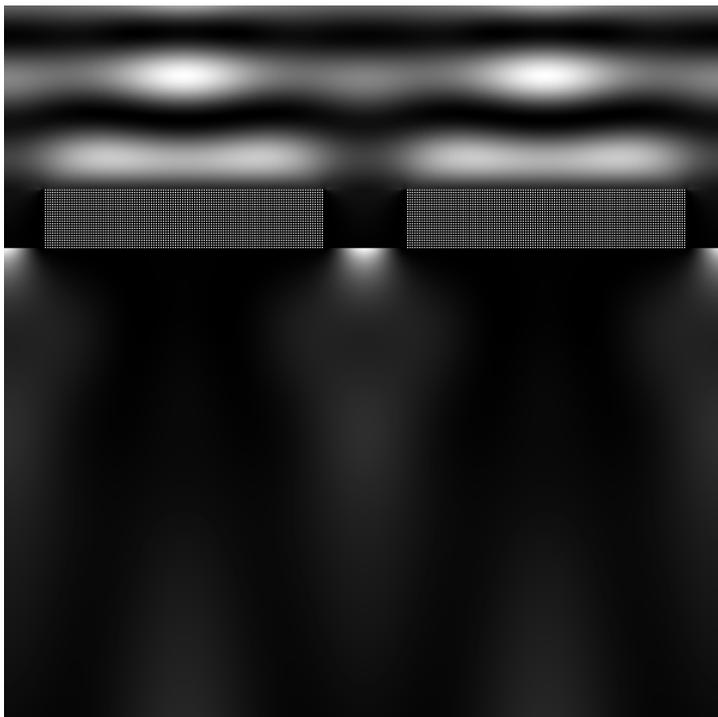


Figure 27: The computed intensity distribution in the xz -plane of the outlined example.

The transmitted intensity in region *IV* was normalised to the intensity in region *I* for better contrast and that the gold film was manually outlined after calculation.

6 Numerical model – metallic holes

The previous chapter 5 was just preliminary work - now, we are going to investigate the behaviour of nanoholes. Hence, we expand our geometry and the computation method to the third dimension.

6.1 Geometry of an array of rectangular holes

Shown in figure 28, we investigate a metallic layer with rectangular holes smaller than the wavelength λ_0 and allowing a period $p \gtrsim \lambda_0$ ²⁵. Again, we separate the space in four different regions characterised by their permittivity ϵ . A linearly polarised plane wave is incident in region *I* onto the metal film *II* and the nanoholes *III*.

6.2 Ansatz for the electromagnetic field

In this subchapter, we consider only plane TE waves because their modal expansion in the holes *III* is simpler and very similar to the modal expansion we have seen before. We focus on the most important equations.

The computation is based on the dominant component E_y of the electric field. Nevertheless, as a result of diffraction along the y -axis, the residual components E_x and E_z must be taken into account in region *I* and *IV*.

Region *I*:

$$\vec{E}_{(\vec{r})}^I = E_0 \begin{pmatrix} 0 \\ e^{ik_0(\gamma_0 x + \beta_1 z)} + R_2 e^{ik_0(\gamma_0 x - \beta_1 z)} \\ 0 \end{pmatrix} + E_0 \iint_{\mathbb{R}^2} F_{r(\gamma_{rx}, \gamma_{ry})} \vec{G}_{r(\gamma_{rx}, \gamma_{ry})} e^{ik_0(\gamma_{rx} x + \gamma_{ry} y - \beta_r z)} d\gamma_{rx} d\gamma_{ry} \quad (122)$$

where

$$\beta_r = \sqrt{\epsilon_1 - \gamma_{rx}^2 - \gamma_{ry}^2} \quad (123a)$$

$$F_{r(\gamma_{rx}, \gamma_{ry})} = \iint_{\mathbb{R}^2} \Delta E_{ry(x, y, z=0)} e^{-ik_0(\gamma_{rx} x + \gamma_{ry} y)} dx dy \quad (123b)$$

are the plane wave expansion coefficients of the perturbation

$$\vec{G}_{r(\gamma_{rx}, \gamma_{ry})} \quad \text{are the geometrical projection coefficients} \quad (123c)$$

²⁵The nanoholes are considered as independent, hollow, ideally metallic waveguides. In reality, the metallic walls between the holes should not be thinner than about 100nm.

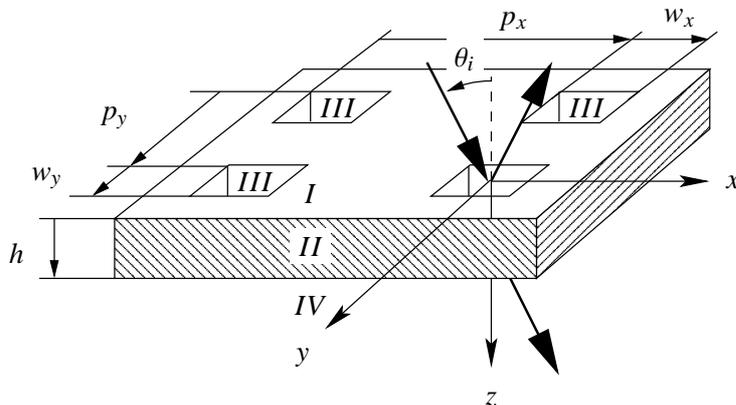


Figure 28: Investigated geometry. The metal layer is h thick and contains rectangular holes $w_x \times w_y$ in size. The holes are repeated in a rectangular array with periods p_x along the x -axis and p_y along the y -axis. The structure is lit by a plane wave incident in the xz -plane under an angle θ_i .

Region II: Metal film

The result is equal to equation (102). Therefore

$$\vec{E}_{(\vec{r})}^{II} = E_0 \begin{pmatrix} 0 \\ A_2 e^{ik_0(\gamma_0 x + \beta_2 z)} + B_2 e^{ik_0(\gamma_0 x - \beta_2 z)} \\ 0 \end{pmatrix} \quad (124)$$

Region III: Holes

The holes are modelled as hollow, ideally metallic waveguides. On the walls, the electric field $E_{\parallel} = 0$, this means that the resistance is identically to zero.

We rewrite equations (6.19) and (6.20) in [21] as

$$\vec{E}_{mn(x,y)} = E_{mn} \begin{pmatrix} -\gamma_{ny} \cos\left(k_0 \gamma_{mx} \left(x + \frac{w_x}{2}\right)\right) \sin\left(k_0 \gamma_{ny} \left(y + \frac{w_y}{2}\right)\right) \\ + \gamma_{mx} \sin\left(k_0 \gamma_{mx} \left(x + \frac{w_x}{2}\right)\right) \cos\left(k_0 \gamma_{ny} \left(y + \frac{w_y}{2}\right)\right) \\ 0 \end{pmatrix} \quad \forall \quad |x| < \frac{w_x}{2} \quad \wedge \quad |y| < \frac{w_y}{2} \quad (125)$$

where

$$E_{mn} \quad \text{are the amplitudes of the eigenmodes } (m, n) \text{ in the holes}$$

$$\gamma_{mx} = \frac{m\pi}{w_x k_0} = k_{mx}/k_0 \quad \forall \quad m \in \mathbb{N}_0 \quad \wedge \quad m \neq n = 0 \quad (126a)$$

$$\gamma_{ny} = \frac{n\pi}{w_y k_0} = k_{ny}/k_0 \quad \forall \quad n \in \mathbb{N}_0 \quad (126b)$$

The incident wave \vec{E}_i has no E_x field. Hence, we simplify equation (125) by setting $\gamma_{ny} = 0$ and $\gamma_m = \gamma_{mx}$ and obtain

$$\vec{E}_{m(x,y)} = E_m \begin{pmatrix} 0 \\ \sin\left(k_0 \gamma_m \left(x + \frac{w_x}{2}\right)\right) \\ 0 \end{pmatrix} \quad \forall \quad |x| < \frac{w_x}{2} \quad \wedge \quad |y| < \frac{w_y}{2} \quad (127)$$

The modal expansion

$$\vec{E}_{(\vec{r})}^{III} = E_0 \begin{pmatrix} 0 \\ \sum_{m=1}^{+\infty} C_m(\gamma_m) \sin\left(m \frac{\pi}{w_x} \left(x + \frac{w_x}{2}\right)\right) \left(A_{3(\gamma_m)} e^{ik_0 \beta_m z} + B_{3(\gamma_m)} e^{-ik_0 \beta_m z}\right) \\ 0 \end{pmatrix} \quad (128)$$

is identical to equation (104a) where the notation was outlined in equations (105) and the coupling coefficients C_m in equation (106a).

Region IV:

$$\vec{E}_{(\vec{r})}^{IV} = E_0 \begin{pmatrix} 0 \\ T_2 e^{ik_0(\gamma_0 x + \beta_4 z)} \\ 0 \end{pmatrix} + E_0 \iint_{\mathbb{R}^2} F_{t(\gamma_{tx}, \gamma_{ty})} \vec{G}_{t(\gamma_{tx}, \gamma_{ty})} e^{ik_0(\gamma_{tx} x + \gamma_{ty} y + \beta_t z)} d\gamma_{tx} d\gamma_{ty} \quad (129)$$

where

$$\beta_t = \sqrt{\epsilon_4 - \gamma_{tx}^2 - \gamma_{ty}^2} \quad (130a)$$

$$F_{t(\gamma_{tx}, \gamma_{ty})} = \iint_{\mathbb{R}^2} \Delta E_{ty(x,y,z=h)} e^{-ik_0(\gamma_{tx} x + \gamma_{ty} y)} dx dy \quad (130b)$$

are the plane wave expansion coefficients of the perturbation

$$\vec{G}_{t(\gamma_{tx}, \gamma_{ty})} \quad \text{are the geometrical projection coefficients} \quad (130c)$$

6.3 Evaluation modifications

This subchapter analytically presents the expansions transforming the two-dimensional evaluation into the desired three-dimensional analysis. Note that the numerical evaluation detailed in subchapters 5.2 and 5.3 and the outlined hints in subchapter 5.5 still hold.

Diffraction:

As a result of the bidirectional periodicity in the xy -plane, we have to account for diffraction at the entire xy -plane. Nevertheless, this task is considerably simplified if we note that diffraction along x is independent from diffraction along y . Hence, we separate the x - from the y -direction and the perturbation $\Delta\vec{E}_{r(x,y,z=0)}$ yields

$$\Delta\vec{E}_{r(x,y,z=0)} = \begin{pmatrix} 0 \\ \Delta E_{r(x,z=0)} \\ 0 \end{pmatrix} \sum_{n \in \mathbb{Z}} \text{rect}\left(\frac{y - np_y}{w_y}\right) \quad (131)$$

where

$$\text{rect}\left(\frac{y}{w_y}\right) = \begin{cases} 1 & \forall |y| < \frac{w_y}{2} \\ 0 & \text{else} \end{cases} \quad (132)$$

So, we can still apply the results shown in subchapter 5.3 along x but we have to include the effects of diffraction along y .

Because the incidence is parallel to the xz -plane²⁶, the Fourier expansion and the Rayleigh expansion are identical along y . The partial expansion coefficients along y are samples of the spectrum of equation (132). Hence

$$F_{ry(\gamma_{ry})} = \frac{w_y}{p_y} \text{sinc}\left(u \frac{w_y}{p_y}\right) \quad (133)$$

Combined with equation (111), the expansion coefficients write as

$$\begin{aligned} F_{r(\gamma_{rx}, \gamma_{ry})} &= F_{rx(\gamma_{rx})} \cdot F_{ry(\gamma_{ry})} \\ &= \Delta x \sum_s \Delta E_{ry(x,y=0,z=0)} e^{-ik_0 \gamma_{rx} x} \cdot \frac{w_y}{p_y} \text{sinc}\left(u \frac{w_y}{p_y}\right) \end{aligned} \quad (134)$$

where γ_{rx} is given in equation (117a), the sampling points x in equation (112b) and

$$\gamma_{ry} = u \frac{w_y}{k_0 p_y} \quad \forall u \in \mathbb{Z} \quad \wedge \quad |\gamma_{ry}| < \frac{\Gamma}{2} \quad (135)$$

Of course, the expansion coefficients $F_{l(\gamma_{rx}, \gamma_{ry})}$ are expressed analogously.

Projection coefficients:

The Rayleigh expansion now contains plane waves propagating at different angles to the xz -plane. A plane wave parallel to the xz -plane can perfectly match the perturbation²⁷, but this does not hold in general. As figure 29 shows, we have to find the best match between the perturbation $\Delta\vec{E}_r$ and the plane wave's field vector $\vec{E} \perp \vec{k}_r$. Here, best match means that the projection of $\Delta\vec{E}_r$ onto \vec{E} achieves its maximum. Obviously, the geometrical angle $\phi = \angle(\Delta\vec{E}_r, \vec{E})$ has to be least.

²⁶This means that $\gamma_{y0} = 0$ whereas $\gamma_{x0} \neq 0$.

²⁷Without explicit notation, we already used this fact excessively in chapter 5.

Inspecting figure 29, we note that $\Delta\vec{E}_r$, \vec{E} and \vec{k}_r are coplanar. To find the direction of \vec{E} , we first compute the unit vector $\vec{a} \perp \Delta\vec{E}_r, \vec{k}_r$

$$\vec{a} = \frac{\Delta\vec{E}_r \times \vec{k}_r}{|\Delta\vec{E}_r \times \vec{k}_r|} = \frac{\Delta\vec{E}_r}{\Delta E_r} \times \frac{\vec{k}_r}{k_r}$$

to get the unit vector $\vec{b} \parallel \vec{E} \perp \vec{a}, \vec{k}_r$

$$\vec{b} = \frac{\vec{k}_r \times \vec{a}}{|\vec{k}_r|} = \frac{\vec{k}_r}{k_r} \times \left(\frac{\Delta\vec{E}_r}{\Delta E_r} \times \frac{\vec{k}_r}{k_r} \right)$$

\vec{E} is now the projection of $\Delta\vec{E}_r$ onto \vec{b}

$$\vec{E} = |\Delta\vec{E}_r| \cos(\langle \Delta\vec{E}_r, \vec{b} \rangle) \vec{b} = (\Delta\vec{E}_r \cdot \vec{b}) \vec{b}$$

Remind that $\Delta\vec{E}_r$ has only a non-null y -coefficient. Hence

$$\vec{E} = \Delta E_{ry} \cdot b_y \vec{b} = \Delta E_{ry} \cdot \vec{G}_{r(\gamma_{rx}, \gamma_{ry})} \quad (136a)$$

where

$$\vec{b} = \frac{\vec{k}_r}{k_r} \times \left(\left(\begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) \times \frac{\vec{k}_r}{k_r} \right) \quad (136b)$$

Applying equations (98) and keeping only the geometrical direction²⁸, we get

$$\vec{k}_r = k_0 \Re \begin{pmatrix} \gamma_{rx} \\ \gamma_{ry} \\ \beta_r \end{pmatrix} = k_0 \begin{pmatrix} \gamma_{rx} \\ \gamma_{ry} \\ \Re \sqrt{\Re^2(n_1) - \gamma_{rx}^2 - \gamma_{ry}^2} \end{pmatrix}$$

$$k_r = k_0 \Re(n_1)$$

and get

$$\vec{G}_{r(\gamma_{rx}, \gamma_{ry})} = \frac{\Re^2(n_1) - \gamma_{ry}^2}{\Re^4(n_1)} \begin{pmatrix} -\gamma_{rx}\gamma_{ry} \\ \Re^2(n_1) - \gamma_{ry}^2 \\ -\gamma_{ry} \Re \sqrt{\Re^2(n_1) - \gamma_{rx}^2 - \gamma_{ry}^2} \end{pmatrix} \quad (137)$$

Note that $\vec{G}_{r(\gamma_{rx}, \gamma_{ry}=0)}$ is effectively the unit vector \vec{e}_y as used in chapter 5.

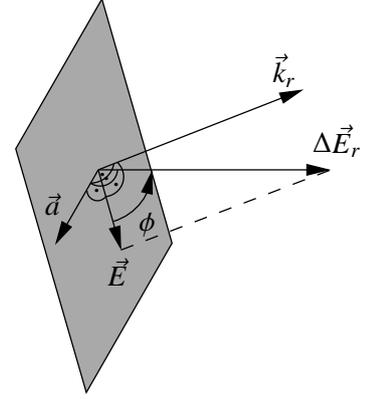


Figure 29: Projection of the difference field $\Delta\vec{E}_r$ to get the field $\vec{E} \perp \vec{k}_r$.

²⁸Here, \vec{k}_r accounts only for the propagation direction and excludes any amplitude effect. Indeed, we must not consider damping to get the geometric projection of $\Delta\vec{E}_r$. By definition, γ_{rx} and γ_{ry} are always real in this paper.

6.4 Calculation example

We keep all parameters of the example given in subchapter 5.6. To obtain a square array of square holes, the periods are set to $p_x = p_y = p = 900\text{nm}$ and the hole dimensions are $w_x = w_y = w = 200\text{nm}$. The field image has a size of $p_x \times p_y \times 1.8\mu\text{m}$ along the x -, y - and z -axis and is sampled at $100 \times 100 \times 200$ grid points. Hence, we obtain a uniform image resolution of 9nm . Again, we compute for $M = 51$ aperture modes, $\Delta x = 0.4\text{nm}$ sampling step and $\Gamma_{max} = 10$ frequency limit. Now, the calculation on a 2GHz Pentium *IV* machine with MatLab 6.0 took 13minutes and yield 54MBytes of data. Figure 30 illustrates the result.

The number of field points was multiplied by about 21 and the number of diffraction orders by about Γ_{max} because of the second diffraction direction. So, t was estimated to $21 \cdot 10 \cdot 3\text{seconds} \approx 11\text{minutes}$. Similar, we estimate $t \approx 14\text{days}$ for $\Gamma_{max} \approx 1600$. Here, it is definitively not worth to wait 2weeks on a slightly 'better' result.

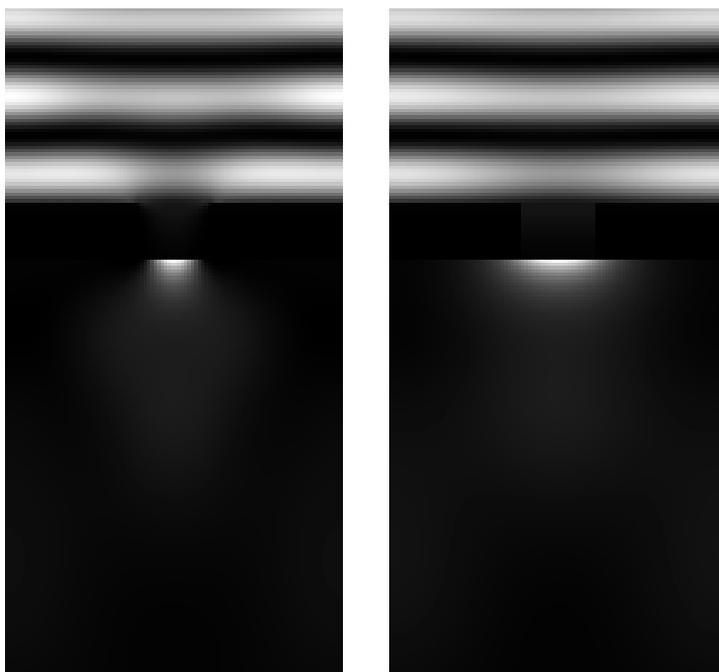


Figure 30: The computed intensity distribution in the xz -plane and the yz -plane.

The transmitted intensity in region *IV* was normalised to the intensity in region *I* for better contrast.

7 Computation example

This section present the performance of the actual MatLab implementation outlined in appendix C. We will show how to use the different modules for computation and data representation.

7.1 Surface plasmons at slit edges

We are going to set up a gold mask *II* on a BK7 glass prism *I*. The mask contains an array of slits *III* that will be backlit by a HeNe laser at TM polarisation. In view of the target application FCS, we assume water in region *IV*.

1. We define the parameter structure.

```
r.i=75.23;           % incidence angle
r.l=633e-9;         % wavelength [m]
r.h=47e-9;          % dimensions [m]
r.wx=250e-9;
r.px=432e-9;
r.wy=250e-9;        % for 3D functions
r.py=432e-9;
r.e1=2.2955;        % BK7 glass
r.e2=-9.3863+1.1470i; % gold film
r.e3=2.2955;        % BK7 glass
r.e4=1.7737;        % water
```

We choosed the film thickness h and the incidence angle θ_i to excite surface plasmons at the gold-water interface $II \rightarrow IV$.

2. We compute the field over two periods p_x for 600×600 points. The function applies the default parameter $\Gamma_{max} = 10$.

```
s=slitFieldTM(r,600,600,2);
```

We get the message "Total relative intensity mismatch $P_{tot} = 0.12\%$ " and the field is displayed after a couple of seconds. The intensity in region *I* was boosted to the level of the surface plasmons in region *IV*. Here, we would like to see the intensity enhancement and launch another display.

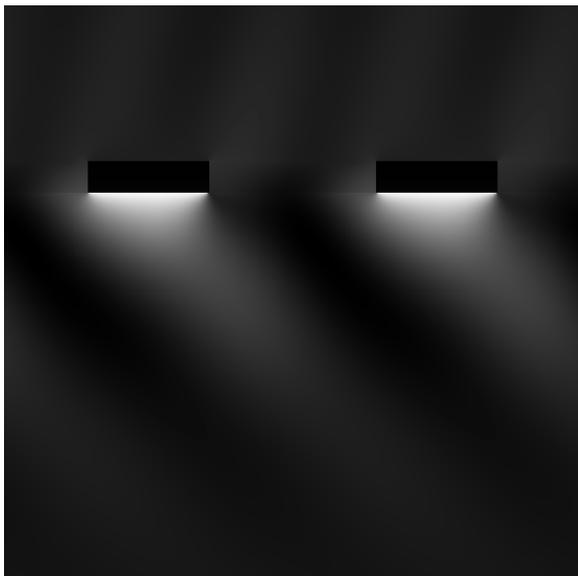


Figure 31: Intensity through the slit mask.



Figure 32: Magnetic field component H_y through the slit mask.

```
display3D(s);
```

Now, we click twice on "Field" to see how the electromagnetic field propagates through the structure. A click on "Phase" switches to the (wrapped) phase display.

- At the moment, the surface plasmons move to the right. For FCS, we should better get a standing wave. Hence, we will add a wave incident at $-\theta_i$ such that it has a phase shift $\Delta\phi \approx \pi$ over the apertures. This should suppress most of the residual transmission through the slits and boost the contrast in region IV. We could recalculate the field but we prefer to flip it along the x -axis and add it manually.

```
t=s; % new container
f=double(t.Hy); % add with phase shift
t.Hy=single(f+exp(0.5i*t.k0*t.g0*t.px)*f(t.W(1):-1:1,,:));
f=double(t.Ex);
t.Ex=single(f+exp(0.5i*t.k0*t.g0*t.px)*f(t.W(1):-1:1,,:));
f=double(t.Ez);
t.Ez=single(f+exp(0.5i*t.k0*t.g0*t.px)*f(t.W(1):-1:1,,:));
t=magneticIntensity(t); % recompute the
display3D(t);
```

Now, we get a promising set of light lines along the y -axis. If we review the field, we note that it is now a standing wave on both faces of the metal film.

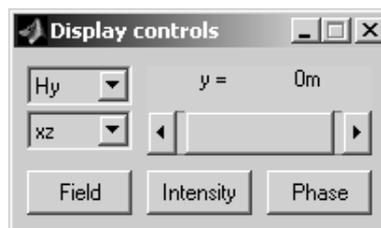


Figure 33: 3D display controls.

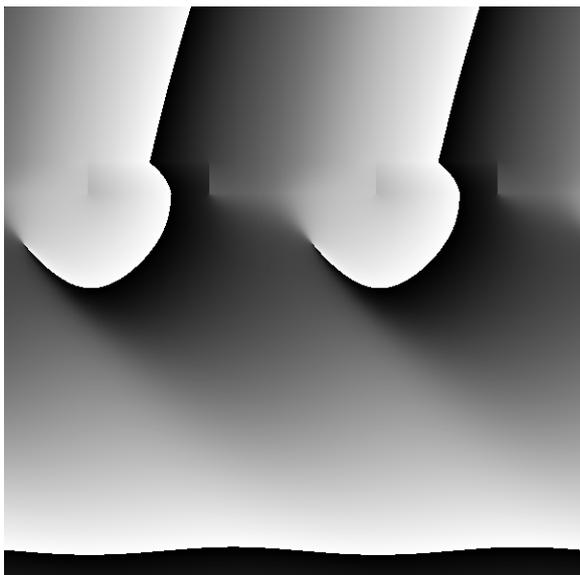


Figure 34: Phase of the magnetic field component H_y through the slit mask.

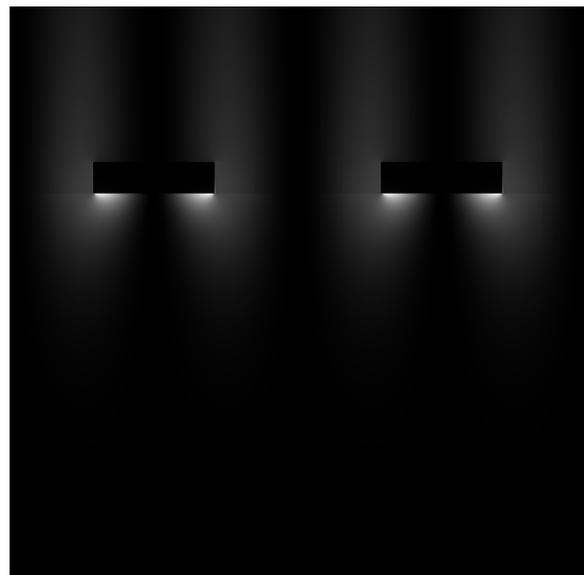


Figure 35: Intensity through the slit mask under symmetric illumination.

7.2 Light fields through holes

4. We will reuse the previous geometry setting $\theta_i = 0$ and launch a calculation in 3D for metallic holes. Here, we will focus on a single xy -period that we will sample at $100 \times 100 \times 200$ points. Because we switch to TE polarisation, we manually increase the number of considered hole modes but truncate the plane wave expansion at $\Gamma_{max} = 10$ as before.

```
r.i=0;  
u=holeFieldTE(holeTransmissionTE(r,80),[],[100 100 200],10);
```

We get the message "Total relative intensity mismatch $P_{tot} = 0.79\%$ " indicating the error of the eigenmode expansion over the holes. After about 2minutes, we can now browse through the tridimensional field at xy -, xz - and yz -slices.

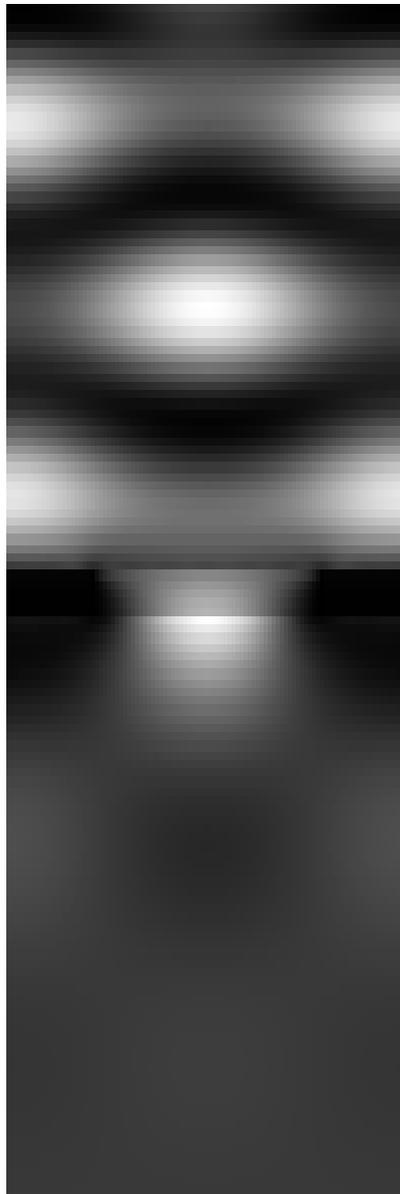


Figure 36: Intensity through the hole mask.

8 Applications for FCS

This chapter presents two applications for fluorescence auto- and cross-correlation spectroscopy. In general, FCS is made by means of a confocal microscope implying a serial scanning to gather information from a manifold of $(x; y; z)$ -positions in the sample. As detailed in the introduction 1, we follow two main objectives.

1. Speed up FCS through parallelisation of the detection - explained in subchapter 8.1.
2. Reduce the volume of the excitation spots to enable FCS on living cells - illustrated in the sub-chapters 8.2 and 8.3.

8.1 Setup for parallel FCS

Figure 37 outlines a setup for parallel FCS. Here, parallel FCS means that the intensity is detected by a sensitive array detector ²⁹ rather than an avalanche diode.

A plane collimated laser beam backlights the sample that is imaged by a standard microscope onto two FCS array detectors. The left detector receives only the background light ³⁰ whereas the upper detector only gets the fluorescence. The background has the same wavelength λ_e as the incident laser beam whereas the fluorescence contains longer wavelengths λ_f . Hence, with a dichroic mirror, the background is separated from the fluorescence and directed to its array detector.

The setup relies on a particular sample that creates itself a multitude of confined light fields replacing the single spot of a confocal microscope. The spots should have a volume $V \ll \lambda_e^3$ and they should be resolved individually by the array detectors.

Advantages:

- The setup is robust because it needs no particular adjustment. Therefore, cross-correlation at several wavelengths sets no particular problem due to the fixed spot positions.
- The fluorescence is detected at several thousand individual spots. This allows to get simultaneously the auto- and cross-correlation for each individual spot or to compute the spatial cross-correlation over several spots.
- The left detector gets the background and resolves the light fields as long as the sample is in focus. This detector controls the sample position and provides a real time position reference to the second array detector.

²⁹As for example the latest development by the group of Prof. Radivoje Popovic at the Microsystems Laboratory at the EPFL.

³⁰The transmitted excitation light.

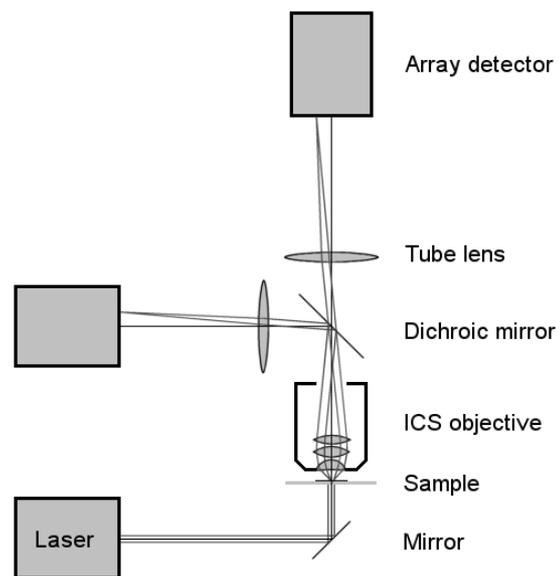


Figure 37: Proposed setup for parallel FCS.

- The setup may be considerable smaller than a standard confocal microscope because we do not need a voluminous scanning stage.

Disadvantages:

- The sample has to confine the incident light into individual, small light fields.
- The fluorophores are excited near the sample surface. This means that our approach is limited to two dimensions and that quenching may become a problem.

8.2 Fluorescence correlation spectroscopy

Figure 38 shows the intensity profile created by the underlying sample. Centred on a hole, the intensity profile is drawn for the first quadrant of a xy -period. The intensity is outlined by shading and isolines at $I/I_{max} = e^{-n/2}$ where $n \in \mathbb{N}$. The isolines are solid for $n \leq 4$, dashed for $5 \leq n \leq 8$ and dotted otherwise. Hence, the last solid isoline indicates the $I/I_{max} = e^{-2}$ surface.

The sample is a $h = 150\text{nm}$ thick gold mask of square holes with $w_x = w_y = 200\text{nm}$ in a square array of $p_x = p_y = 940\text{nm}$. The gold mask is deposited on top of a $170\mu\text{m}$ thick BK7 cover slip. Its holes are filled with BK7 glass to inhibit contamination by fluorophores. In this example, a HeNe laser beam is incident at $\theta_i = 0^\circ$ under TE polarisation. The mask transmits a fraction of the incoming energy to create individual bright spots at the top of the holes.

A drop of an aqueous fluorophore solution is put on the mask. The fluorophores are excited at λ_e and emit at $\lambda_f > \lambda_e$.

Results:

The spots have an $I/I_{max} = e^{-2}$ surface similar to a semi-ellipsoid with half-axes

$$r_x \approx 70\text{nm} \quad r_y \approx 190\text{nm} \quad r_z \approx 120\text{nm}$$

Hence, the spot volume is

$$V_{e(\text{HeNe})} \approx 2.5 \cdot 10^{-18}\text{l} \quad (138)$$

Assuming an average of $N_f = 2$ fluorophores floating simultaneously in the same spot, V_e allows FCS up to a fluorophore concentration

$$C_f = \frac{N_f}{V_e \cdot 6.02 \cdot 10^{23}\text{mol}^{-1}} \quad \text{that is} \quad C_{f(\text{HeNe})} \approx 1.3\mu\text{M} \quad (139)$$

The polarisation $\vec{E} \parallel y$ -axis shows up in $r_y > r_x$. Even for $w_y \rightarrow 0$, $r_y \gtrsim 120\text{nm}$. For square holes, we obtained $r_y \approx r_x + 120\text{nm}$.

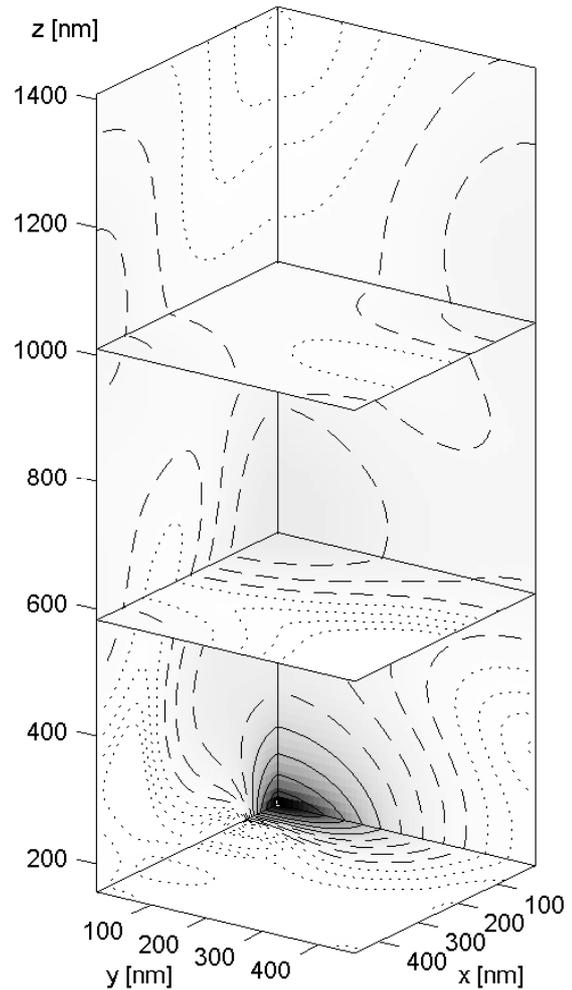


Figure 38: Intensity for $\lambda_e = 633\text{nm}$

8.3 Fluorescence cross-correlation spectroscopy

We reuse the sample of the previous subchapter 8.2. For cross-correlation spectroscopy, we combine an Argon laser $\lambda_{Ar} = 488\text{nm}$ with the HeNe laser $\lambda_{HeNe} = 633\text{nm}$ ³¹ to get a simultaneous excitation at $\lambda_e = \{\lambda_{Ar}, \lambda_{HeNe}\}$. Figure 39 depicts the intensity profile at the Argon wavelength.

Results:

Again, the spots have an $I/I_{max} = e^{-2}$ surface similar to a semi-ellipsoid. But the half-axis along z is significantly longer.

$$r_x \approx 90\text{nm} \quad r_y \approx 180\text{nm} \quad r_z \approx 250\text{nm}$$

The spot volume increases to

$$V_{e(Ar)} \approx 6.4 \cdot 10^{-18}\text{l} \quad (140)$$

and, at $N_f = 2$, the fluorophore concentration drops to

$$C_{f(Ar)} \approx 0.52\mu\text{M} \quad (141)$$

Comparison:

P. Schwille [2] worked with a confocal microscope by Carl Zeiss. She used an apochrome ICS objective 40x1.2 for the excitation at λ_{Ar} and λ_{HeNe} as well as for the fluorescent light. With a standard FCS test sample, she measured the transversal radii of the confocal spots

$$r_{t(Ar)} \approx 180\text{nm} \quad \text{and} \quad r_{t(HeNe)} \approx 250\text{nm}$$

From table 1 in [13], we approximate the axial radii as

$$r_a \approx \frac{0.88\lambda_e}{n - \sqrt{n^2 - NA^2}} \quad \text{where} \quad n = \sqrt{\epsilon_{H_2O}}$$

Hence, we deduce the confocal spot volumes

$$V_{e(Ar)} \approx 5.8 \cdot 10^{-17}\text{l} \quad \text{and} \quad V_{s(HeNe)} \approx 1.5 \cdot 10^{-16}\text{l} \quad (142)$$

and get for $N_f = 2$ the fluorophore concentrations

$$C_{f(Ar)} \approx 57\text{nM} \quad \text{respectively} \quad C_{f(HeNe)} \approx 23\text{nM} \quad (143)$$

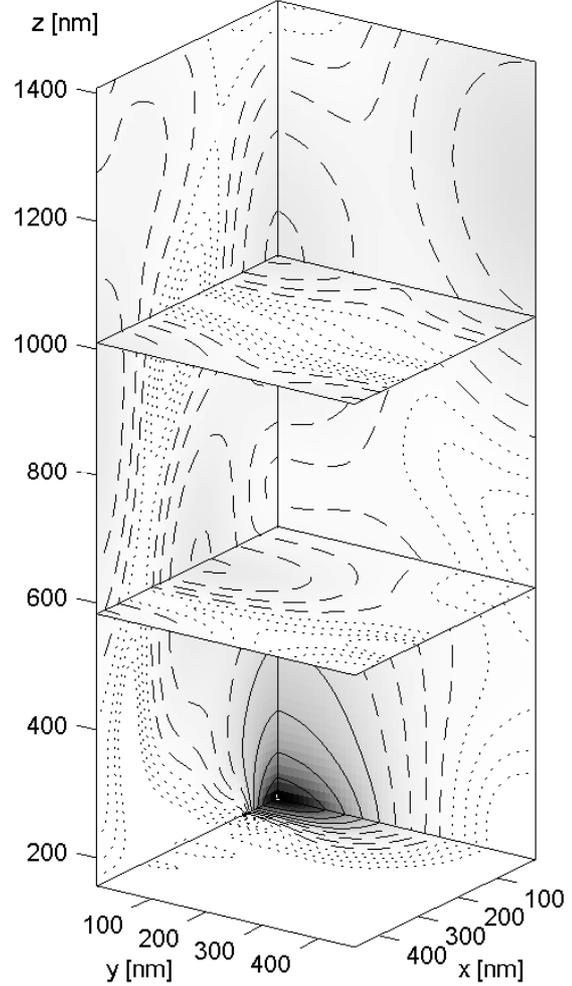


Figure 39: Intensity for $\lambda_e = 488\text{nm}$

³¹This is a standard excitation for many confocal microscopes by Carl Zeiss for example.

Discussion:

1. At λ_{Ar} , we are able to handle a fluorophore concentration C_f that is about an order of magnitude higher than with standard confocal FCS. At λ_{HeNe} , we gain about two orders of magnitude.
2. The confocal spot size has a lower bound imposed by diffraction. Therefore, it grows for increasing wavelengths. In contrast, the spot size behind our hole mask tends to **shrink** for longer wavelengths. While diffraction is a problem in confocal microscopy, we use it to keep the light fields very small. The more the light diffracts in region *IV*, the smaller is r_z whereas r_x and r_y grow slightly.

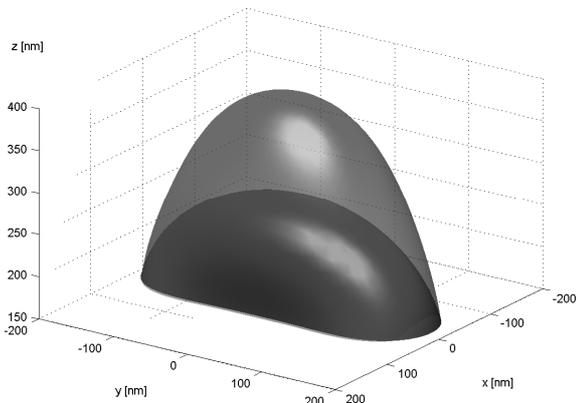


Figure 40: $I/I_{max} = e^{-2}$ intensity surface at λ_{Ar} and λ_{HeNe}

3. As figure 40 shows, the smaller spot at the HeNe laser wavelength λ_{HeNe} is contained to more than 98% in the larger spot at the Argon laser wavelength λ_{Ar} . Hence, our setup provides an elegant method for fluorescence cross-correlation spectroscopy.

9 Conclusions

By means of rigorous wave coupling, we elaborated an explicit numerical model for the electromagnetic field through an array of subwavelength apertures. The model characteristics were outlined and some example calculations were given in view of the target application. Due to delays in the sample fabrication, we are not yet able to present experimental results in this paper. Hence, the experimental verification remains in the pipeline and will be launched when the first sample is available.

On the theoretical part, we will generalise our numerical approach to handle any configuration in regions *I* and *III*. Especially, we will accommodate the model for circular holes and step into the calculation of rough interfaces $I \rightarrow II$ and $II \rightarrow IV$.

References

- [1] A. Perentes, *Nano-optic Structures produced by Focused Electron Beam induced Deposition*, Diploma Work at the Swiss Federal Institute of Technology Lausanne, Monitors: MER Dr. P. Hoffmann and Dr. I. Utke, Lausanne (2003)
- [2] K. Bacia, I. V. Majoul, P. Schwille, *Probing the endocytic pathway in live cells using dual-color fluorescence cross-correlation analysis*, *Biophys. J.* **83**#2 (2002), 1184..1193
- [3] A. Barbara et al., *Optical transmission through subwavelength metallic gratings*, *Phys. Rev. B* **66**#16 (2002), 1403
- [4] M. Born, E. Wolf, "Electromagnetic Theory of Propagation, Interference and Diffraction of Light" *Principles of Optics*, 7th edition, Cambridge University Press, Cambridge (2002)
- [5] F. J. García-Bidal et al., *Transmission and focusing of light in one-dimensional periodically nanostructured metals*, *Phys. Rev. B* **66**#15 (2002), 5412
- [6] H. J. Lezec et al., *Beaming Light from a Subwavelength Aperture*, *Science* **297** (2002), 820
- [7] T. Thio et al., *Giant optical transmission of sub-wavelength apertures: physics and applications*, *Nanotech.* **13** (2002), 429
- [8] S. Kawata (ed.), *Near-Field Optics and Surface Plasmon Polaritons*, Topics in Applied Physics **81**, Springer-Verlag, Berlin (2001)
- [9] A. Krishnan, *Evanesciently coupled resonance in surface plasmon enhanced transmission*, *Opt. Comm.* **200** (2001), 1
- [10] H. J. Lezec et al., *Enhanced light transmission through a single subwavelength aperture*, *Opt. Lett.* **26**#24 (2001), 1972
- [11] L. Martín-Moreno et al., *Theory of Extraordinary Optical Transmission through Subwavelength Hole Arrays*, *Phys. Rev. Lett.* **86**#6 (2001), 1114
- [12] R. Rigler, *Fluorescence correlation spectroscopy: theory and applications*, Springer series in chemical physics **65**, Springer-Verlag, Berlin (2001)
- [13] S. Wilhelm, B. Gröbler, M. Gluch, H. Heinz, "Principles" *Confocal Laser Scanning Microscopy*, Carl Zeiss R&D, Jena
- [14] J. D. Jackson, *Classical Electrodynamics*, 3rd edition, John Wiley & Sons Inc., New York (1998), Chapter 8, 352
- [15] W. Demtröder, *Experimentalphysik 2, Elektrizität und Optik*, Springer-Verlag, Berlin (1995), 3
- [16] DMK/DPK *Formel und Tafeln, Mathematik - Physik*, OrellFüssli Verlag, Zürich (1992)
- [17] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer Tracts in Modern Physics **111**, Springer-Verlag, Berlin (1988)
- [18] A. Wirgin, A. A. Maradudin, *Resonant response of a bare metallic grating to s-polarized light*, *Prog. Surf. Sci.* **22**#1 (1987), 1
- [19] L. D. Landau, E. M. Lifshitz, L. P. Pitaevskiĭ, *Electrodynamics of Continuous Media*, 2nd edition, Pergamon Press, Oxford (1984), 300

- [20] R. A. Depine et al., *Surface impedance boundary condition for metallic diffraction gratings in the optical infrared range*, *Optica Acta* **30**#3 (1983), 313
- [21] M. J. Adams, "1. Optical waveguide theory" *An introduction to optical waveguides*, John Wiley & Sons Inc., New York (1981), "6. Guides of rectangular cross-section"
- [22] R. A. Depine et al., *Diffraction grating efficiencies Conformal mapping method for a good real conductor*, *Optica Acta* **29**#11 (1982), 1459
- [23] P. B. Johnson et al., *Optical Constants of the Nobel Metals*, *Phys. Rev. B* **6**#12 (1972), 4370
- [24] E. Kretschmann, *Die Bestimmung optischer Konstanten von Metallen durch Anregung von Oberflächenplasmaschwingungen*, *Z. Physik* **241**, Springer-Verlag, (1971), 313

List of Figures

1	Model outline	4
2	The Gauss law	5
3	The Stokes law	6
4	Dispersion curves	12
5	Verification of the Drude model	15
6	Complex refraction index	15
7	Surface plasmon at a gold film	17
8	Schema of a surface plasmon	18
9	Kretschmann-Raether configuration	19
10	Otto configuration	19
11	Surface plasmon dispersion	20
12	Momentum match on a grating	21
13	Zero order transmittance	24
14	Reflected and transmitted intensity	24
15	Fresnel coefficients at a surface	28
16	Fresnel coefficients at a layer	28
17	Surface wave	30
18	Diffraction spectrum	31
19	Diffracted intensity	31
20	Propagation through a metal film	32
21	Field of a surface plasmon	34
22	TE slit modes	38
23	Coupling coefficients	38
24	Modal expansion error	40
25	Rayleigh expansion	43
26	Ewald sphere	43
27	Intensity through slits	46
28	Investigated geometry	47
29	Projection coefficients	50
30	Intensity through holes	51
31	Intensity through the slit mask.	52
32	Magnetic field component H_y through the slit mask.	52
33	3D display controls.	53
34	Phase of the magnetic field component H_y through the slit mask.	53
35	Intensity through the slit mask under symmetric illumination.	53

36	Intensity through the hole mask.	54
37	Proposed setup for parallel FCS.	55
38	Intensity for $\lambda_e = 633\text{nm}$	56
39	Intensity for $\lambda_e = 488\text{nm}$	57
40	$I/I_{max} = e^{-2}$ intensity surface at λ_{Ar} and λ_{HeNe}	58
41	Kretschmann setup	61
42	Fresnel coefficient of reflection	62
43	0.5° sampling in $[41.5^\circ, 80.0^\circ]$	62
44	1.0° sampling in $[41.5^\circ, 79.5^\circ]$	62
45	Random reflection error $\sigma_{ R_\theta ^2} = 0.05$	62
46	1.0° sampling in $[42.0^\circ, 80.0^\circ]$	63
47	Random angle error $\sigma_{\theta_i} = 0.05^\circ$	63
48	Angle offset $\Delta\theta_i = 0.1^\circ$	63
49	Angle scale $\Delta\theta_i = 5\%(\theta_i - 45^\circ)$	63
50	Reflection offset $\Delta R_\theta ^2 = -0.1$	63
51	Reflectance scale $\Delta R_\theta ^2 = -10\% R_\theta ^2$	63
52	Surface plasmon at a gold film	64
53	Surface plasmon at a silver film	64
54	Surface plasmon at a gold film	65
55	Surface plasmon at a gold film	65

List of Tables

1	Parameters of the Drude model	15
---	---	----

A Determination of the relative dielectric constant of metals

E. Kretschmann described in 1970 in his thesis [24] a new method to measure the optical constants of metals. His method relies on the anomalous low reflection at thin metal layers in a Kretschmann-Raether configuration³². Due to an excitation of surface plasmons, the incident electromagnetic energy is dissipated in the metal film instead of being reflected.

Conditions:

1. At a metal-dielectric interface, surface plasmons exist only if the dielectric constant ϵ_2 of the metal and ϵ_3 of the dielectric fulfil the conditions

$$\Re\epsilon_2 < -\Re\epsilon_3 \quad \text{and} \quad \Im\epsilon_2 < |\Re\epsilon_3| \quad (144)$$

assuring that the surface wave is effectively a plasmon and that the absorption is low enough for resonance.

2. The thickness h_2 of the metal film should be such that the reflectance drops to less than 50% at the resonance (see chapter 2.6). In the visible range for example, gold and silver films should have about 40nm to 60nm.

In general, a thickness between one and two skin-depths should be appropriate for a first measure. If necessary, a second measure should be made on a sample that better matches the resonance condition. This makes the determination more robust against measurement errors.

Setup and measure:

A plane TM wave excites surface waves on a thin metal film ϵ_2 . The metal film is evaporated on the flat of a BK7 glass cylinder ϵ_1 . As figure 41 shows, the wave is incident through the semi-cylinder on the glass-metal interface. The incidence angle θ_i is greater than the critical angle θ_c for total internal reflection at the opposite metal-air interface. Hence, the incident wave is either reflected or absorbed, but it cannot transit into the air ϵ_3 .

For correct measures, we need a well-collimated beam I_i incident on the glass-metal interface. Here, we work with a collimated HeNe laser beam ($\lambda_0 = 633\text{nm}$). We correct for the cylindrical air-glass interface by means of a divergent cylindrical lens. The reflected light is projected onto a photo diode to measure its intensity I_r .

A goniometer assures a 1:1 correspondence between θ_i and θ_o . It is used to turn the HeNe laser and the photo diode around the semi-cylinder while measuring $\theta_i \in [41.5^\circ, 80^\circ]$ and $|R_\theta|^2 = I_r/I_i$.

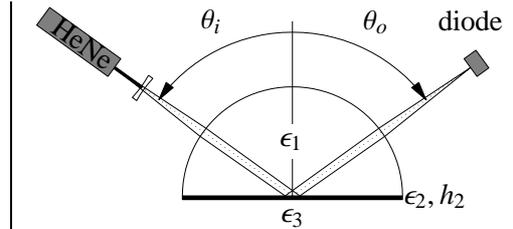


Figure 41: Kretschmann setup

Sensitivity and accuracy:

We have already seen that the excitation of surface plasmons is very sensitive on θ_i and h_2 . Hence, the distinction of two different metals mainly depends on the accuracy of the method. From the measured reflection $|R_\theta|^2$, we get the parameters ϵ_2 and h_2 by fitting $|R_\theta|^2$ to the Fresnel coefficient $|R_{13(\epsilon_2, h_2, \theta_i)}|^2$. We use a least square algorithm to find the best fit. Therefore

$$\text{error} = \sum_m \left(\frac{|R_{\theta(\theta_{im})}|^2}{|R_{\theta(\theta_{if})}|^2} - \frac{|R_{13(\epsilon_2, h_2, \theta_{im})}|^2}{|R_{13(\epsilon_2, h_2, \theta_{if})}|^2} \right)^2 \rightarrow 0 \quad (145)$$

³²Review figures 9 and 21.

where

$$|R_f|^2 = |R_{\theta(\theta_{if})}|^2 > 0.8 \max |R_{\theta(\theta_i)}|^2 \quad \text{and} \quad \left| \frac{d|R_{\theta}|^2}{d\theta_i} \right| = \text{minimal at } \theta_{if}$$

As proposed by E. Kretschmann, we fit the ratios $|R_{\theta}|^2/|R_f|^2$ instead of the coefficients $|R_{\theta}|^2$. This eliminates any linear scaling³³ on $|R_{\theta}|^2$ and we can directly replace it by I_r .

Next, we discuss the influence of random noise and some systematic errors on the fitted parameters ϵ_2 and h_2 at the example of a gold film. Figure 42 shows the parameters and the Fresnel coefficient $|R_{13}|^2$.

- Figure 43 outlines the parameter extraction if 78 values are used for the least square fit. The reference angle $\theta_f \approx 60^\circ$ and the extracted ϵ_2 and h_2 differ by about 0.6% from the exact values. Figure 44 shows the results for a fit on 39 values. ϵ_2 differs by about 0.7% and h_2 by 0.9%.
- Figure 45 outlines the influence of 5% random noise on the reflection values. Random noise is due to detector noise, intensity fluctuations of the HeNe laser or roughness at the interfaces. The standard deviations are about 3% for ϵ_2 and 5% for h_2 ³⁴.

If $|R_{\theta}|^2$ is measured with less than 1% of relative error and the incidence angle θ_i at $\pm 0.03^\circ$, $\Re\epsilon_2$ is easily determined to about 1% whereas $\Im\epsilon_2$ and h_2 are given to about 3%.

³³A very likely systematic error. For example, every air-glass interface reflects about 4% of the incident light.

³⁴Refer to appendix A for other random or systematic errors.

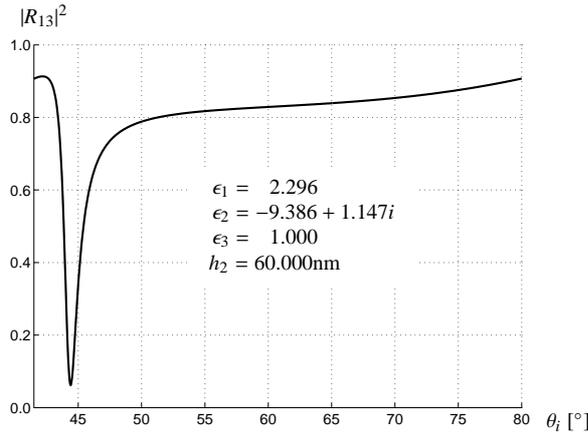


Figure 42: Fresnel coefficient of reflection

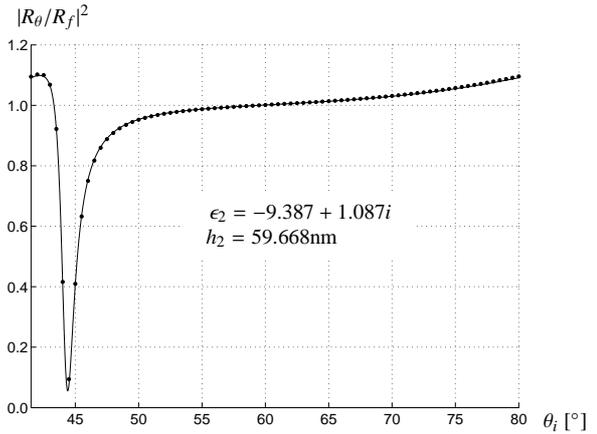


Figure 43: 0.5° sampling in $[41.5^\circ, 80.0^\circ]$

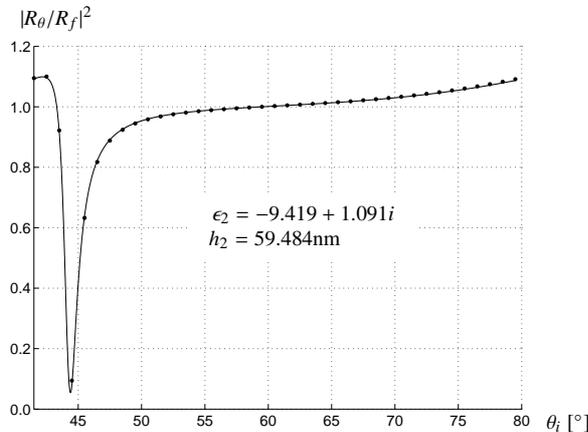


Figure 44: 1.0° sampling in $[41.5^\circ, 79.5^\circ]$

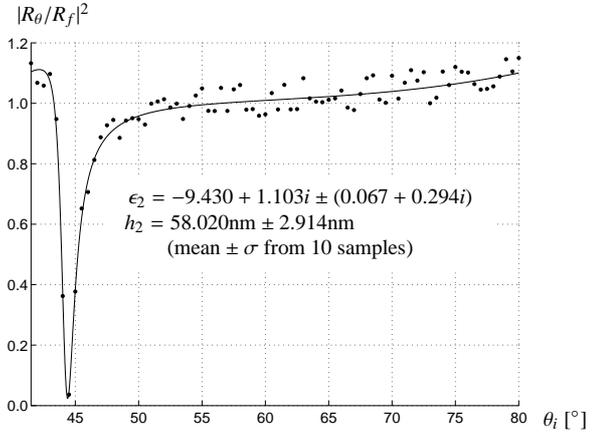


Figure 45: Random reflection error $\sigma_{|R_{\theta}|^2} = 0.05$

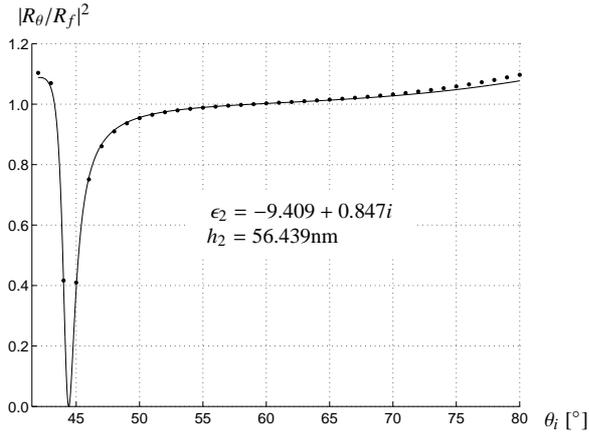


Figure 46: 1.0° sampling in [42.0°, 80.0°]

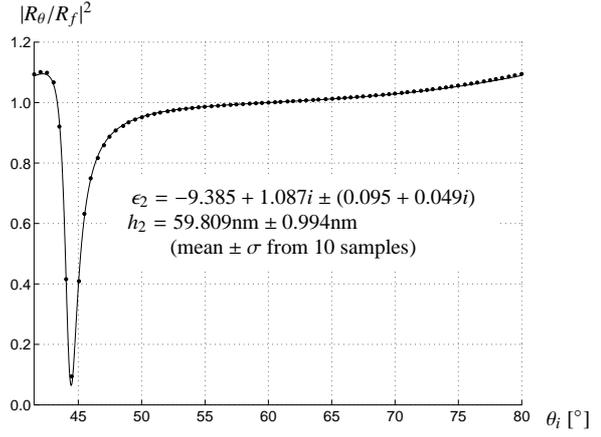


Figure 47: Random angle error $\sigma_{\theta_i} = 0.05^\circ$

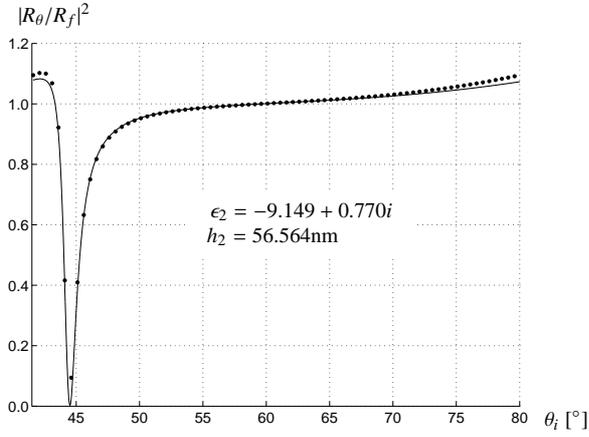


Figure 48: Angle offset $\Delta\theta_i = 0.1^\circ$

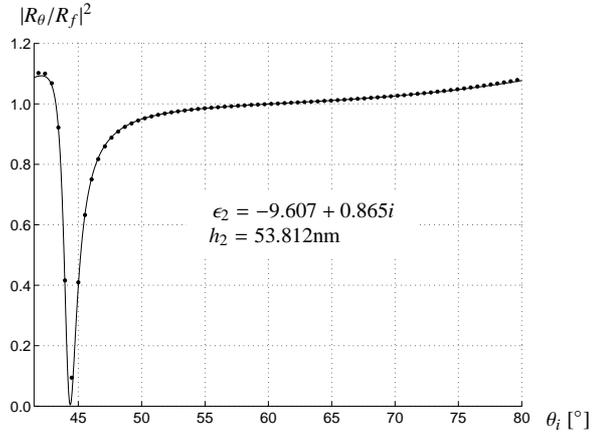


Figure 49: Angle scale $\Delta\theta_i = 5\%(\theta_i - 45^\circ)$

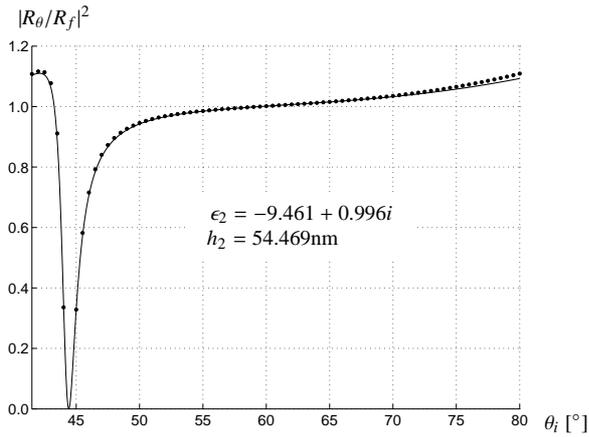


Figure 50: Reflection offset $\Delta|R_\theta|^2 = -0.1$

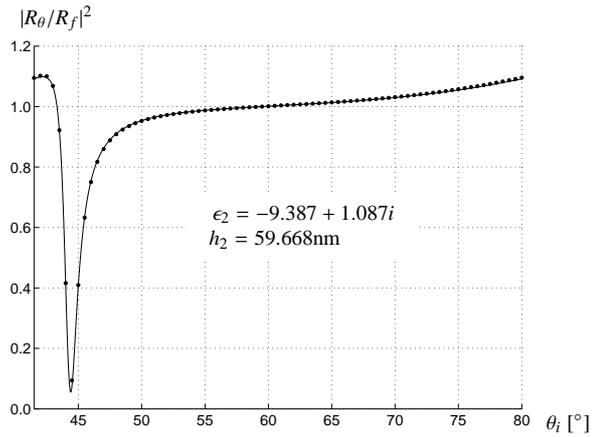


Figure 51: Reflectance scale $\Delta|R_\theta|^2 = -10\%|R_\theta|^2$

B Reflection and Transmission Spectrum of ATR devices

B.1 Kretschmann-Raether configuration

Gold

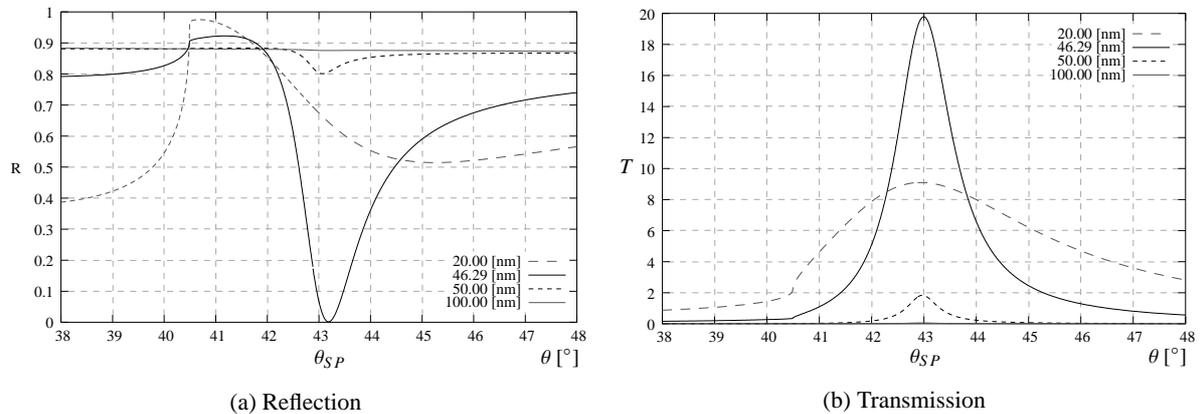


Figure 52: Reflected and transmitted intensity at a thin gold film in function of the thickness h . Incident light at $\lambda_0 = 617\text{nm}$, $e_1 = 1.54^2$ (glass), $e_2 = -10.662 + 1.374i$ (gold), $e_3 = 1$ (air). Plasmon excitation for $h_{min} = 46.29\text{nm}$ at $\theta_{SP} = 43.22^\circ$.

Silver

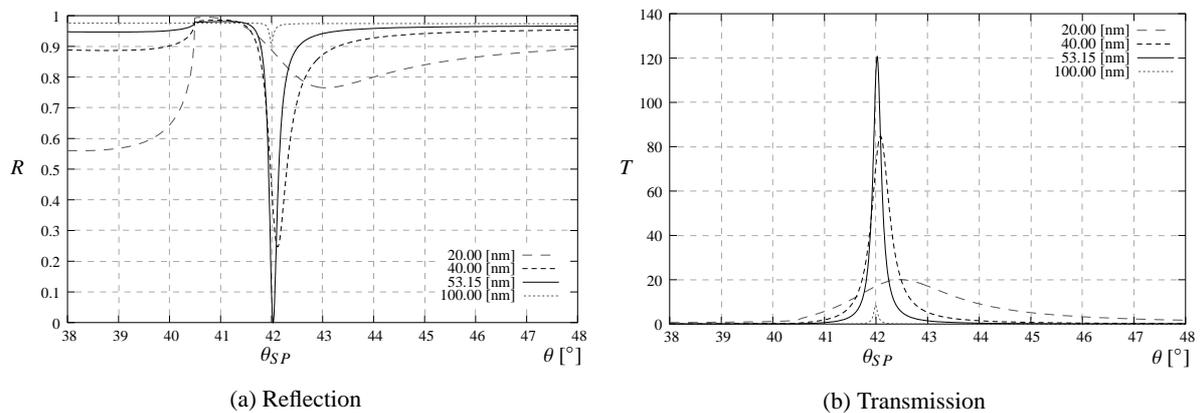


Figure 53: Reflected and transmitted intensity at a thin silver film in function of the thickness h . Incident light at $\lambda_0 = 617\text{nm}$, $e_1 = 1.54^2$ (glass), $e_2 = -17.236 + 0.498i$ (silver), $e_3 = 1$ (air). Plasmon excitation for $h_{min} = 53.15\text{nm}$ at $\theta_{SP} = 42.03^\circ$.

B.2 Otto configuration

Gold

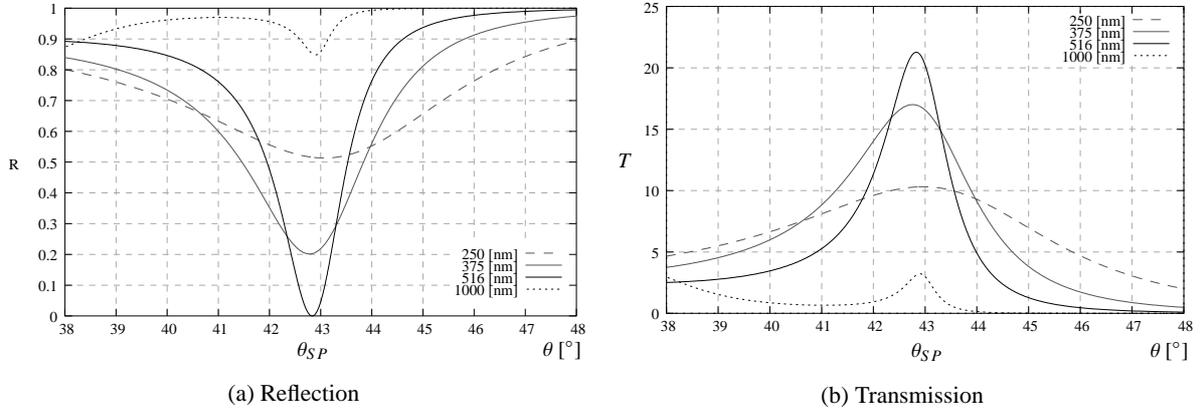


Figure 54: Reflected and transmitted intensity at a thin gold film in function of the thickness h . Incident light at $\lambda_0 = 617\text{nm}$, $e_1 = 1.54^2$ (glass), $e_2 = -10.662 + 1.374i$ (gold), $e_3 = 1$ (air). Plasmon excitation for $h_{min} \cong 516\text{nm}$ at $\theta_{SP} \cong 42.8^\circ$

Silver

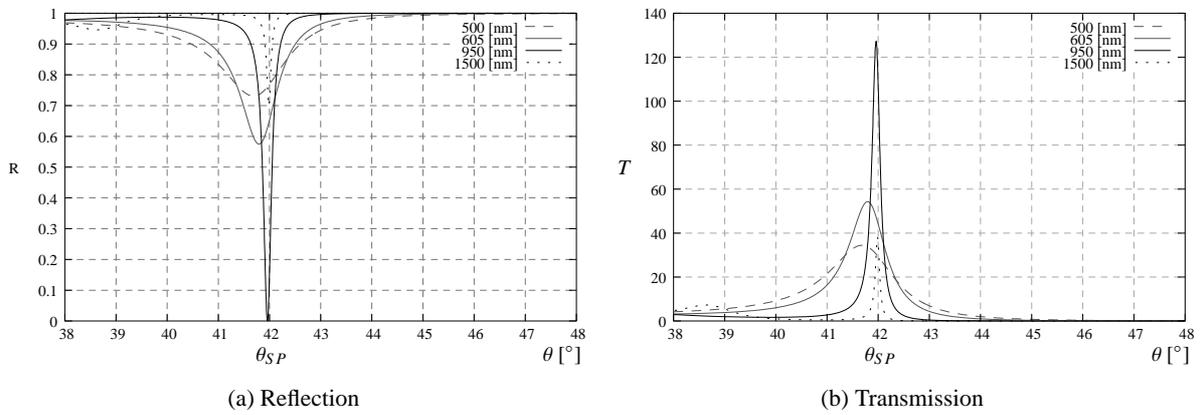


Figure 55: Reflected and transmitted intensity at a thin silver film in function of the thickness h . Incident light at $\lambda_0 = 617\text{nm}$, $e_1 = 1.54^2$ (glass), $e_2 = -17.236 + 0.498i$ (silver), $e_3 = 1$ (air). Plasmon excitation for $h_{min} \cong 950\text{nm}$ at $\theta_{SP} \cong 41.9^\circ$.

C MatLab scripts

A *structure* keeps the model parameters and the evaluation results together. Thus, a numerical evaluation is done in three to five steps.

1. Define all input parameters in a new *structure*.
2. Pass the *structure* to the desired *transmission* function.
3. Pass the result to the corresponding *field* function and define the drawing region and sampling. If the previous step was skipped, the *field* function automatically executes it with default parameters.
4. Inspect the final result with the *display3D* function.
5. If desired, post-treat the result to figure out additional information.

The input parameters are defined in a *structure* containing the following fields:

- .i Incidence angle θ_i [°], where $0^\circ = z$ -axis and $90^\circ = x$ -axis
- .l Wavelength λ_0 in free space [m]
- .h Thickness h of the metal film [m]
- .wx/y Width w_x and w_y of the apertures [m]
- .px/y Period p_x and p_y of the aperture array [m]
- .e1..4 Relative dielectric constants $\epsilon_{1..4}$ in region I to IV

C.1 2D computation of metallic slits

Our current implementation evaluates an array of metallic slits at TM and/or TE polarisation. Only the MatLab scripts for the TM polarisation are printed.

2D/slitFieldTM.m

Electromagnetic field through an array of slits in a metal layer. See figure 1(b) for an outline of the model geometry.

The incident plane wave has an intensity of $1\text{W}/\text{m}^2$.

- in:**
- r *Structure*
 - w Image width in pixels {256}
 - h Image height in pixels {256}
 - N Number of periods drawn {1}
 - P Compute up to $|k_x| = P \cdot k_0$ {all}
 - O Chooses the {'Rayleigh'} or 'Fourier' plane wave expansion method
- out:**
- r *Structure* completed with
 - .I Intensity [W/m^2]
 - .P Used limit for k_x
 - .S Drawing dimensions [m]
 - .W Image dimensions [pixel]
 - .x/y/z Coordinates of the computed points [m]
 - .Ex/z Electric field [V/m]
 - .Hy Magnetic field [A/m]

```

function r=slitFieldTM(r, w, h, N, P, 0)
if ~nargin | isempty(r) | isstruct(r)
    help slitFieldTM
    return;
end;
if nargin < 2 | isempty(w)
    w=256;
end;
if nargin < 3 | isempty(h)
    h=256;
end;
if nargin < 4 | isempty(N)
    N=1;
end;
if nargin < 5
    P=[];
end;
if nargin < 6
    O='';
end;
switch lower(O)
case 'fourier'
    O=1;
otherwise
    O=0;
end;
resetStatus;
if ~isfield(r,'A3')
    r=slitTransmissionTM(r);
end;
p=N*r.px/(w-1);
x=(1:w)-w/2)*p;
z=(1:h)-0.3*h)*p+r.h/2;
ki=i*r.k0;
%
%Field allocations
%
r.I=zeros(h,w);
r.P=abs(P);
r.S=[1 0 h/w]*N*r.px;
r.W=[w 1 h];
r.x=x;
r.y=0;
r.z=z;
r.Ex=-r.I;
r.Ez=r.I;
r.Hy=-r.I;
%
%Regions and slit positions
%
% Metal layer at z(a)...z(b)
% Holes at      x(c)
% Local x      d
% Hole phasors e
%
a=find(z >= 0);
a=a(1);
b=find(z <= r.h);
b=b(length(b));
c=[];
d=[];
e=[];
for j=-ceil(N):ceil(N)
    t=find(abs(x+j*r.px) < r.wx/2);
    c=[c t];
    d=[d x(t)+r.wx/2+j*r.px];
    e=[e repmat(exp(-j*ki*r.g0*r.px),size(t))];
end;
%
%Region 1: Input
%
% Hy = exp(i*k0*(g0*x + b1*z)) + R2*exp(i*k0*(g0*x - b1*z))
% + sum(Fn*exp(i*k0*(gn*x - bn*z)))
% Ex = b*sqrt(u0/e0)/er*Hy
% Ez = -g*sqrt(u0/e0)/er*Hy
%
s=exp(ki*r.g0*x);
n1=sqrt(r.e1);
if 0
    [gn,Fn]=fourierCtm(1+r.R2,r.g0,(1+r.R3).*r.Cm,r.g3,r.k0,r.wx,r.px,real(n1)*r.P);
    u=exp(ki*gn*(x+r.wx/2));
else
    [gn,Fn]=rayleighCtm(1+r.R2,r.g0,(1+r.R3).*r.Cm,r.g3,r.k0,r.wx,r.px,real(n1)*r.P);
    u=exp(ki*gn*x);
end;
bn=zSpeed(r.e1,gn);
r.I(1:a-1,:)=1;
for j=1:a-1
    l=ki*z(j)*conj(r.b1);
    t=r.R2*exp(conj(l));
    l=exp(l);
    v=exp(-ki*z(j)*bn);
    r.Hy(j,:)=(1+t)*s + (Fn.*v)*u;
    r.Ex(j,:)=(r.b1*(1-t)*s - (Fn.*bn.*v)*u)/r.e1;
    r.Ez(j,:)=(-r.g0*(1+t)*s - (Fn.*gn.*v)*u)/r.e1;
    updateStatus(j/h);
end;

```

```

end;
%
%Region 2: Layer
%
% Hy = A2*exp(i*k0*(g0*x + b2*z)) + B2*exp(i*k0*(g0*x - b2*z))
%
%Region 3: Slits
%
% Hy = sum((A3*exp(i*k0*b3*z) + B3*exp(-i*k0*b3*z))*cos(k0*g3*x));
%
n2=sqrt(r.e2);
n3=sqrt(r.e3);
r.I(a:b,:)=n1/n2;
r.I(a:b,c)=n1/n3;
u=exp(ki*r.g3*d);
for j=a:b
    t=exp(ki*z(j)*r.b2);
    r.Hy(j,:)=(r.A2*t+r.B2/t)*s;
    r.Ex(j,:)=r.b2*(r.A2*t-r.B2/t)/r.e2*s;
    r.Ez(j,:)=r.g0*(r.A2*t+r.B2/t)/r.e2*s;
    t=exp(ki*z(j)*r.b3);
    v=(r.A3.*t+r.B3./t).*r.Cm;
    r.Hy(j,c)=v*real(u).*e;
    r.Ex(j,c)=(r.b3.*(r.A3.*t-r.B3./t).*r.Cm)*real(u).*e/r.e3;
    r.Ez(j,c)=(r.g3.*v)*imag(u).*e*(-i/r.e3);
    updateStatus(j/h);
end;
%
%Region 4: Output
%
% Hy = T2*exp(i*k0*(g0*x + b4*z))
% + sum(Fn*exp(i*k0*(gn*x + bn*(z-h))))
%
n4=sqrt(r.e4);
if 0
    [gn,Fn]=fourierCtm(r.T2,r.g0,r.T3.*r.Cm,r.g3,r.k0,r.wx,r.px,real(n4)*r.P);
    u=exp(ki*gn*(x+r.wx/2));
else
    [gn,Fn]=rayleighCtm(r.T2,r.g0,r.T3.*r.Cm,r.g3,r.k0,r.wx,r.px,real(n4)*r.P);
    u=exp(ki*gn*x);
end;
bn=zSpeed(r.e4,gn);
r.I(1+b:h,:)=n1/n4;
for j=1+b:h
    t=exp(ki*(z(j)-r.h)*r.b4);
    v=exp(ki*(z(j)-r.h)*bn);
    r.Hy(j,:)=r.T2*t*s + (Fn.*v)*u;
    r.Ex(j,:)=(r.b4*r.T2*t*s + (Fn.*bn.*v)*u)/r.e4;
    r.Ez(j,:)=(-r.g0*r.T2*t*s - (Fn.*gn.*v)*u)/r.e4;
    updateStatus(j/h);
end;
%
%Intensity & normalization for 1W/m^2 input
%
% I = 1/2*real(H*conj(H))*sqrt(u0/e0*er)
% s = Hi(I = 1W/m^2)
%
e0=8.85418782e-12; % [As/Vm]
u0=1.25663706e-6; % [Vs/Am]
t=[2 3 1];
s=sqrt(2*n1*sqrt(e0/u0));
r.I=permute(single(real(r.I.*r.Hy.*conj(r.Hy))),t);
r.Hy=permute(single(r.Hy*s),t);
s=sqrt(2*n1*sqrt(u0/e0));
r.Ex=permute(single(r.Ex*s),t);
r.Ez=permute(single(r.Ez*s),t);
eval('display3D(matchIntensity(r))','return');
delete(updateStatus(1));

```

2D/slitTransmissionTM.m

Fresnel coefficients and coupling coefficients for a plane TM wave incident onto an array of slits. See figure 1(b) for an outline of the model geometry.

- in:** r Structure
- P Compute up to $|k_x| = P \cdot k_0$ {10}
- out:** r Structure completed with
- .k0 Wavevector k_0 in free space [rad/m]
- .g0 $\gamma_0 = k_x/k_0$ of the incident wave
- .g3 $\gamma_m = k_x/k_0$ of the TM slit modes
- .b1..4 Propagation constants $\beta_{1..4} = k_z/k_0$
- .Cm Coupling coefficients C_m to the slit modes
- .R2/3 Reflection coefficients R at the $I \rightarrow II \rightarrow IV$ and $I \rightarrow III \rightarrow IV$ layers
- .A2/3 Coefficients A in region II and III of the forward propagating wave

.B2/3 Coefficients B in region II and III of the backward propagating wave

.T2/3 Transmission coefficients T through the $I \rightarrow II \rightarrow IV$ and $I \rightarrow III \rightarrow IV$ layers

```
function r=slitTransmissionTM(r, P)
if nargin < 2 | isempty(P)
    P=10;
else
    P=max(1,P);
end;
%
% Wavevector and propagation constants
%
s=r.l/r.wx/2;
r.k0=2*pi/r.l;
r.g0=real(sqrt(r.e1)*sin(r.i*pi/180));
r.g3=s*(0:ceil(P/s));
r.b1=zSpeed(r.e1,r.g0);
r.b2=zSpeed(r.e2,r.g0);
r.b3=zSpeed(r.e3,r.g3);
r.b4=zSpeed(r.e4,r.g0);
%
% Fresnel reflection and transmission through the metal layer
%
[R,T,A,B]=layerRTtm(r.e1,r.e2,r.e4,r.g0,r.k0,r.h);
r.R2=R;
r.T2=T;
r.A2=A;
r.B2=B;
%
% Coupling coefficients to the slit modes
%
r.Cm=slitCtm(r.g0,r.g3,r.k0,r.wx);
%
% Fresnel reflection and transmission through the slit
%
[R,T,A,B]=layerRTtm(r.e1,r.e3,r.e4,r.g3,r.k0,r.h);
r.R3=R;
r.T3=T;
r.A3=A;
r.B3=B;
```

2D/private/fourierCtm.m

Fourier expansion coefficients F_n of the difference field ΔH_y at the slit boundary.

in: C0 Coefficient in region I or IV at the metal layer
g0 $\gamma_0 = k_x/k_0$ of the incident wave
Cm Coupling coefficients C_m to the slit modes
gm $\gamma_m = k_x/k_0$ of the TM slit modes
k0 Wavevector k_0 in free space [rad/m]
wx Width w_x of the slits [m]
px Period p_x of the slit array [m]
P Compute up to $|k_x| = P \cdot k_0$ {all}

out: gn $\gamma_n = k_x/k_0$ of the diffracted waves
Fn Fourier coefficients F_n

```
function [gn,Fn]=fourierCtm(C0,g0,Cm,gm,k0,wx,px,P)
t=max(100,10*length(gm));
t=2*ceil(log2(px/wx*t));
s=floor(t*wx/2/px);
x=px/t*(-s:s);
f=Cm*cos(k0*gm*(x+wx/2))-C0*exp(i*k0*g0*x);
gn=[-t/2:t/2-1]*2*pi/px/k0;
Fn=fftshift(fft(f,t)/t);
if length(P)
    t=find(abs(gn) < P);
    gn=gn(t);
    Fn=Fn(t);
end;
```

2D/private/rayleighCtm.m

Rayleigh expansion coefficients F_n of the difference field ΔH_y at the slit boundary.

in: C0 Coefficient in region I or IV at the metal layer
g0 $\gamma_0 = k_x/k_0$ of the incident wave
Cm Coupling coefficients C_m to the slit modes
gm $\gamma_m = k_x/k_0$ of the TM slit modes

in: k_0 Wavevector k_0 in free space [rad/m]
 w_x Width w_x of the slits [m]
 p_x Period p_x of the slit array [m]
 P Compute up to $|k_x| = P \cdot k_0$ {all}
out: g_n $\gamma_n = k_x/k_0$ of the diffracted waves
 F_n Rayleigh coefficients F_n

```

function [gn,Fn]=rayleighCtm(C0,g0,Cm,gm,k0,wx,px,P)
%
%Field difference
%
% dF = perturbation
%
t=max(50,5*length(gm));
s=wx/t/2;
x=s*(-t:t);
t=2*t*px/wx;
dF=Cm*cos(k0*gm'*(x+wx/2))-C0*exp(i*k0*g0*x);
%
%Avoid aliasing (Shannon)
%
P=min([P pi/k0/s]);
%
%Rayleigh expansion
%
gn = g0 + n*px/l for all |gn| < P
%
s=2*pi/k0/px;
gn=g0+(-floor((P+g0)/s):floor((P-g0)/s))*s;
Fn=dF*exp(-i*k0*x'*gn)/t;

```

2D/private/slitCtm.m

Coupling coefficients C_m representing a plane TM wave as a series of cosine TM slit modes.

in: g_0 $\gamma_0 = k_x/k_0$ of the incident wave
 g_m $\gamma_m = k_x/k_0$ of the TM slit modes
 k_0 Wavevector k_0 in free space [rad/m]
 w_x Width w_x of the slits [m]
out: C_m Coupling coefficients C_m to the slit modes

```

%Reconstruction formula: f=Cm*cos(k0*gm'*(x+wx/2))
%
function Cm=slitCtm(g0,gm,k0,wx)
t=max(50,5*length(gm));
s=wx/2/t;
x=(-t:t)*s;
f=exp(i*k0*g0*x);
Cm=f*cos(k0*(x'+wx/2)*gm)/t;
Cm(1)=Cm(1)/2;
%
%Total relative intensity mismatch Ptot
%
d=f-Cm*cos(k0*gm'*(x+wx/2));
fprintf('Total relative intensity mismatch Ptot = %g%\n',100*(d*d')/(f*f'));

```

2D/private/zSpeed.m

Propagation constant β with positive imaginary part.

in: e Relative dielectric constant ϵ
 g $\gamma = k_x/k_0$
out: b $\beta = k_z/k_0$

```

function b=zSpeed(e,g)
b=sqrt(e-g.^2);
b=complex(real(b),abs(imag(b)));

```

C.2 3D computation of metallic holes

Currently, the implementation evaluates an array of metallic holes at TE polarisation only.

3D/holeFieldTE.m

Electromagnetic field through an array of slits in a metal layer. See figure 28 for an outline of the model geometry.

The incident plane wave has an intensity of 1W/m^2 .

in:

- r *Structure*
- s Drawing dimensions [m] $\{[p_x, p_y, h + 3\lambda_0]\}$
- w Image dimensions [pixel] $\{[100, 100, 100]\}$
- P Compute up to $|k_x| = P \cdot k_0$ {all}
- O Chooses the {'Rayleigh'} or 'Fourier' plane wave expansion method

out:

- r *Structure* completed with
- .I Intensity $[\text{W/m}^2]$
- .P Used limit for k_x
- .S Drawing dimensions [m]
- .W Image dimensions [pixel]
- .x/y/z Coordinates of the computed points [m]
- .Ex/y/z Electric field $[\text{V/m}]$
- .Hx/y/z Magnetic field $[\text{A/m}]$ (not yet implemented)

```
function r=holeFieldTE(r, s, w, P, O)
if nargin | isempty(r) | ~isstruct(r)
    help holeFieldTE
    return;
end;
if nargin < 2 | length(s) ~= 3
    s=[r.px r.py r.h+3*r.l];
end;
if nargin < 3 | length(w) ~= 3
    w=[100 100 100];
else
    w=ceil(w);
end;
if nargin < 4 | length(P) ~= 1
    P=[];
end;
if nargin < 5
    O='';
end;
switch lower(O)
case 'fourier'
    O=1;
otherwise
    O=0;
end;
resetStatus;
if ~isfield(r, 'A3')
    r=holeTransmissionTE(r,P);
end;
ki=i*r.k0;
t=s./w;
%Field allocations
%
x=((1:w(1))-(1+w(1))/2)*t(1);
y=((1:w(2))-(1+w(2))/2)*t(2);
z=((1:w(3))-w(3)/3)*t(3) + r.h/2;
r.I=zeros(w);
r.P=abs(P);
r.S=s;
r.W=w;
r.x=x;
r.y=y;
r.z=z;
r.Ex=r.I;
r.Ey=r.I;
r.Ez=r.I;
%
%Regions and hole positions
%
% Metal layer at z(a)...z(b)
% Holes at x(c) X y(d)
% Hole phasors e
% Local x f
%
a=find(z >= 0);
a=a(1);
b=find(z <= r.h);
b=b(length(b));
c=[];
d=[];
e=[];
f=[];
```

```

n=ceil(s(1)/r.px);
for j=-n:n
    t=find(abs(x-j*r.px) < r.wx/2);
    c=[c t];
    e=[e repmat(exp(j*ki*r.g0*r.px),size(t))];
    f=[f x(t)+r.wx/2-j*r.px];
end;
n=ceil(s(2)/r.py);
for j=-n:n
    t=find(abs(y-j*r.py) < r.wy/2);
    d=[d t];
end;
%
%Region 1: Input
%
% Ex = sum(bx*Fn*exp(i*k0*(gx*x + gy*y - bn*z)))
% Ey = exp(i*k0*(g0*x + b1*z)) + R2*exp(i*k0*(g0*x - b1*z))
%       + sum(by*Fn*exp(i*k0*(gx*x + gy*y - bn*z)))
% Ez = sum(bz*Fn*exp(i*k0*(gx*x + gy*y - bn*z)))
%
s=exp(ki*x'*r.g0);
n1=sqrt(r.e1);
if 0
    [gx,gy,Fn]=fourierCte(1+r.R2,r.g0,(1+r.R3).*r.Cm,r.g3,r.k0,r.wx,r.wy,r.px,r.py,real(n1)*r.P);
    u=exp(ki*(x+r.wx/2)*gx');
else
    [gx,gy,Fn]=rayleighCte(1+r.R2,r.g0,(1+r.R3).*r.Cm,r.g3,r.k0,r.wx,r.wy,r.px,r.py,real(n1)*r.P);
    u=exp(ki*x'*gx');
end;
[bx,by,bz]=projectEy(gx,gy,r.e1);
bn=zSpeed(r.e1,gx,gy);
Fx=Fn.*bx;
Fy=Fn.*by;
Fz=Fn.*bz;
r.I(:, :, 1:a-1)=1;
for j=1:a-1
    t=ki*z(j)*conj(r.b1);
    t=s*(exp(t)+r.R2*exp(conj(t)));
    v=exp(-ki*z(j)*bn);
    for i=1:length(y)
        p=exp(ki*y(i)*gy).*v;
        r.Ex(:, i, j)=u*(Fx.*p);
        r.Ey(:, i, j)=t + u*(Fy.*p);
        r.Ez(:, i, j)=u*(Fz.*p);
    end;
    updateStatus(j/w(3));
end;
%
%Region 2: Layer
%
% Ex = 0
% Ey = A2*exp(i*k0*(g0*x + b2*z)) + B2*exp(i*k0*(g0*x - b2*z))
% Ez = 0
%
%Region 3: Holes
%
% Ex = 0
% Ey = A3*exp(i*k0*(g3*x + b3*z)) + B3*exp(i*k0*(g3*x - b3*z))
% Ez = 0
%
n2=sqrt(r.e2);
n3=sqrt(r.e3);
r.I(:, :, a:b)-n2/n1;
r.I(c, d, a:b)-n3/n1;
u=exp(ki*r.g3'*f);
for j=a:b
    t=exp(ki*z(j)*r.b2);
    r.Ey(:, :, j)=repmat(s*(r.A2*t+r.B2/t),size(y));
    t=exp(ki*z(j)*r.b3);
    r.Ey(c, d, j)=repmat(((r.A3.*t+r.B3./t).*r.Cm)*imag(u).*e).^',size(d));
    updateStatus(j/w(3));
end;
%
%Region 4: Output
%
% Ex = sum(bx*Fn*exp(i*k0*(gx*x + gy*y + bn*(z-h))))
% Ey = T2*exp(i*k0*(g0*x + b4*z))
%       + sum(by*Fn*exp(i*k0*(gx*x + gy*y + bn*(z-h))))
% Ez = sum(bz*Fn*exp(i*k0*(gx*x + gy*y + bn*(z-h))))
%
n4=sqrt(r.e4);
if 0
    [gx,gy,Fn]=fourierCte(r.T2,r.g0,r.T3.*r.Cm,r.g3,r.k0,r.wx,r.wy,r.px,r.py,real(n4)*r.P);
    u=exp(ki*(x+r.wx/2)*gx');
else
    [gx,gy,Fn]=rayleighCte(r.T2,r.g0,r.T3.*r.Cm,r.g3,r.k0,r.wx,r.wy,r.px,r.py,real(n4)*r.P);
    u=exp(ki*x'*gx');
end;
[bx,by,bz]=projectEy(gx,gy,r.e4);
bn=zSpeed(r.e4,gx,gy);
Fx=Fn.*bx;
Fy=Fn.*by;
Fz=Fn.*bz;
r.I(:, :, 1+b:w(3))=n4/n1;
for j=1+b:w(3)
    t=r.T2*exp(ki*(z(j)-r.h)*r.b4)*s;

```

```

v=exp(ki*(z(j)-r.h)*bn);
for i=1:length(y)
    p=exp(ki*y(i)*gy).*v;
    r.Ex(:,i,j)=u*(Fx.*p);
    r.Ey(:,i,j)=t + u*(Fy.*p);
    r.Ez(:,i,j)=u*(Fz.*p);
end;
updateStatus(j/w(3));
end;
%
%Intensity & normalization for 1W/m^2 input
%
% I = 1/2*(E*E')*sqrt(e0/u0)*real(sqrt(er))
% s = Ei(I = 1W/m^2)
%
e0=8.85418782e-12; % [As/Vm]
u0=1.25663706e-6; % [Vs/Am]
s=sqrt(2/n1*sqrt(u0/e0));
r.I=single(real(r.I.*(r.Ex.*conj(r.Ex)+r.Ey.*conj(r.Ey)+r.Ez.*conj(r.Ez))));
r.Ex=single(r.Ex*s);
r.Ey=single(r.Ey*s);
r.Ez=single(r.Ez*s);
eval('display3D(matchIntensity(r))','return');
delete(updateStatus(1));

```

3D/holeTransmissionTE.m

Fresnel coefficients and coupling coefficients for a plane TE wave incident onto an array of holes. See figure 28 for an outline of the model geometry.

in: *r* *Structure*
 P Compute up to $|k_x| = P \cdot k_0$ {10}

out: *r* *Structure* completed with
 .k0 Wavevector k_0 in free space [rad/m]
 .g0 $\gamma_0 = k_x/k_0$ of the incident wave
 .g3 $\gamma_{mx} = k_x/k_0$ of the TE hole modes
 .b1..4 Propagation constants $\beta_{1..4} = k_z/k_0$
 .Cm Coupling coefficients C_m to the hole modes
 .R2/3 Reflection coefficients R at the $I \rightarrow II \rightarrow IV$ and $I \rightarrow III \rightarrow IV$ layers
 .A2/3 Coefficients A in region II and III of the forward propagating wave
 .B2/3 Coefficients B in region II and III of the backward propagating wave
 .T2/3 Transmission coefficients T through the $I \rightarrow II \rightarrow IV$ and $I \rightarrow III \rightarrow IV$ layers

```

function r=holeTransmissionTE(r, P)
if nargin < 2 | isempty(P)
    P=10;
else
    P=max(1,P);
end;
%
%Wavevector and propagation constants
%
s=r.l/r.wx/2;
r.k0=2*pi/r.l;
r.g0=sqrt(r.e1)*sin(r.i*pi/180);
r.g3=s*(1:ceil(P/s));
r.b1=zSpeed(r.e1,r.g0);
r.b2=zSpeed(r.e2,r.g0);
r.b3=zSpeed(r.e3,r.g3);
r.b4=zSpeed(r.e4,r.g0);
%
%Fresnel reflection and transmission through the metal layer
%
[R,T,A,B]=layerRTte(r.e1,r.e2,r.e4,r.g0,r.k0,r.h);
r.R2=R;
r.T2=T;
r.A2=A;
r.B2=B;
%
%Coupling coefficients to the TE hole modes
%
r.Cm=holeCte(r.g0,r.g3,r.k0,r.wx);
%
%Fresnel reflection and transmission through the hole
%
[R,T,A,B]=layerRTte(r.e1,r.e3,r.e4,r.g3,r.k0,r.h);
r.R3=R;
r.T3=T;
r.A3=A;
r.B3=B;

```

3D/private/fourierCte.m

Fourier expansion coefficients F_n of the difference field ΔE_y at the hole boundary.

in: C0 Coefficient in region I or IV at the metal layer
g0 $\gamma_0 = k_x/k_0$ of the incident wave
Cm Coupling coefficients C_m to the hole modes
gm $\gamma_m = k_x/k_0$ of the TE hole modes
k0 Wavevector k_0 in free space [rad/m]
wx/y Width w_x and w_y of the holes [m]
px/y Period p_x and p_y of the hole array [m]
P Compute up to $|k_x| = P \cdot k_0$ {all}

out: gx/y $\gamma_x = k_x/k_0$ and $\gamma_y = k_y/k_0$ of the diffracted waves
Fn Fourier coefficients F_n

```
function [gx,gy,Fn]=fourierCte(C0,g0,Cm,gm,k0,wx,wy,px,py,P)
t=max(100,10*length(gm));
t=2*ceil(log2(px/wx*t));
s=floor(t*wx/2/px);
x=px/t*(-s:s);
f=Cm*sin(k0*gm*(x+wx/2))-C0*exp(i*k0*g0*x);
n=(-t/2:t/2-1);
gx=n^2*pi/px/k0;
Fx=fftshift(fft(f,t)/t);
gy=n^2*pi/py/k0;
Fy=wy/py*sinc(wy/py*n);
if length(P)
    t=find(abs(gx) < P);
    gx=gx(t);
    Fx=Fx(t);
    t=find(abs(gy) < P);
    gy=gy(t);
    Fy=Fy(t);
end;
gx=gx';
Fn=Fx.*Fy;
t=size(gx);
gx=repmat(gx,size(gy));
gy=repmat(gy,t);
if length(P)
    t=find(gx.^2+gy.^2 < P^2);
    gx=gx(t);
    gy=gy(t);
    Fn=Fn(t);
else
    gx=gx(:);
    gy=gy(:);
    Fn=Fn(:);
end;
```

3D/private/holeCte.m

Coupling coefficients C_m representing a plane TE wave as a series of sine TE hole modes.

in: g0 $\gamma_0 = k_x/k_0$ of the incident wave
gm $\gamma_m = k_x/k_0$ of the TE hole modes
k0 Wavevector k_0 in free space [rad/m]
wx Width w_x of the holes [m]

out: Cm Coupling coefficients C_m to the hole modes

```
%Reconstruction formula: f=Cm*sin(k0*gm*(x+wx/2))
%
function Cm=holeCte(g0,gm,k0,wx)
t=max(50,5*length(gm));
s=wx/2/t;
x=(-t:t)*s;
f=exp(i*k0*g0*x);
Cm=f*sin(k0*(x'+wx/2)*gm)/t;
%
%Total relative intensity mismatch Ptot
%
d=f-Cm*sin(k0*gm*(x+wx/2));
fprintf('Total relative intensity mismatch Ptot = %g%%\n',100*(d*d')/(f*f'));
```

3D/private/projectEy.m

Projection coefficients of E_y to $\vec{E} \perp \vec{k}$.

in: gx/y $\gamma_x = k_x/k_0$ and $\gamma_y = k_y/k_0$
 e Relative dielectric constant ϵ

out: bx/y/z Projection coefficients $\vec{G} = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix}$

```
function [bx,by,bz]=projectEy(gx,gy,e)
%
%Surface plasmons have imaginary betas
%
bn=real(sqrt(real(sqrt(e))^2-gx.^2-gy.^2));
%
%Normalize the truncated k-vectors
%
t=sqrt(gx.^2+gy.^2+bn.^2);
gx=gx./t;
gy=gy./t;
bn=bn./t;
%
%Projection coefficients
%
by=bn.^2+gx.^2;
bx=-gx.*gy.*by;
bz=-bn.*gy.*by;
by=by.*by;
```

3D/private/rayleighCte.m

Rayleigh expansion coefficients F_n of the difference field ΔE_y at the hole boundary.

in: C0 Coefficient in region I or IV at the metal layer
 g0 $\gamma_0 = k_x/k_0$ of the incident wave
 Cm Coupling coefficients C_m to the hole modes
 gm $\gamma_m = k_x/k_0$ of the TE hole modes
 k0 Wavevector k_0 in free space [rad/m]
 wx/y Width w_x and w_y of the holes [m]
 px/y Period p_x and p_y of the hole array [m]
 P Compute up to $|k_x| = P \cdot k_0$ {all}

out: gx/y $\gamma_x = k_x/k_0$ and $\gamma_y = k_y/k_0$ of the diffracted waves
 Fn Rayleigh coefficients F_n

```
function [gx,gy,Fn]=rayleighCte(C0,g0,Cm,gm,k0,wx,wy,px,py,P)
%
%Along the x axis
%
t=max(50,5*length(gm));
s=wx./t/2;
x=s*(-t:t);
t=2*t*px/wx;
dF=Cm*sin(k0*gm.*(x+wx/2))-C0*exp(i*k0*g0*x);
%
%Avoid aliasing (Shannon)
%
P=min([P pi/k0/s]);
%
%Rayleigh expansion
%
% gx = g0 + n*px/l    for all |gx| < P
%
s=2*pi/k0/px;
gx=g0+(-floor((P+g0)/s):floor((P-g0)/s))*s;
Fx=dF*exp(-i*k0*x.*gx)/t;
%
%Along the y axis
%
s=2*pi/k0/py;
t=floor(P/s);
gy=(-t:t)*s;
Fy=wy/py*sinc(wy/py*(-t:t));
%
%Combine the results
%
gx=gx';
Fn=Fx.'*Fy;
t=size(gx);
gx=repmat(gx,size(gy));
gy=repmat(gy,t);
t=find(gx.^2+gy.^2 < P^2);
gx=gx(t);
gy=gy(t);
Fn=Fn(t);
```

3D/private/zSpeed.m

Propagation constant β with positive imaginary part.

in: e Relative dielectric constant ϵ
 gx/y $\gamma_x = k_x/k_0$ and $\gamma_y = k_y/k_0$

out: b $\beta = k_z/k_0$

```
function b=zSpeed(e,gx,gy)
if nargin < 3
    gy=0;
end;
b=sqrt(e-gx.^2-gy.^2);
b=complex(real(b),abs(imag(b)));
```

C.3 Fresnel coefficients at interfaces and layers

Reflection and transmission coefficients at an interface and at a material layer.

All/layerRTte.m

Fresnel reflection and transmission coefficients for a plane TE wave traveling through a material layer.

in: e1..3 Relative dielectric constants $\epsilon_{1,3}$ before, in and behind the layer.
 g $\gamma_{||} = k_x/k_0$ of the incident wave
 k0 Wavevector k_0 in free space [rad/m]
 h2 Thickness h_2 of the layer [m]

out: R13 Reflection coefficient R_{13} at the layer
 T13 Transmission coefficient T_{13} through the layer
 A2 Coefficients A_2 of the forward propagating wave in the layer
 B2 Coefficients B_2 of the backward propagating wave in the layer

```
function [R13,T13,A2,B2]=layerRTte(e1,e2,e3,g,k0,h2)
[R12,T12]=surfaceRTte(e1,e2,g);
[R23,T23]=surfaceRTte(e2,e3,g);
a=exp(i*k0*h2*sqrt(e2-g.^2));
b=1+R12.*R23.*a.^2;
A2=T12./b;
B2=R23.*A2.*a.^2;
R13=(R12+R23.*a.^2)./b;
T13=T23.*A2.*a;
```

All/layerRTtm.m

Fresnel reflection and transmission coefficients for a plane TM wave traveling through a material layer. See *layerRTte* for a list of parameters and results.

```
function [R13,T13,A2,B2]=layerRTtm(e1,e2,e3,g,k0,h2)
[R12,T12]=surfaceRTtm(e1,e2,g);
[R23,T23]=surfaceRTtm(e2,e3,g);
a=exp(i*k0*h2*sqrt(e2-g.^2));
b=1+R12.*R23.*a.^2;
A2=T12./b;
B2=R23.*A2.*a.^2;
R13=(R12+R23.*a.^2)./b;
T13=T23.*A2.*a;
```

All/private/surfaceRTte.m

Fresnel reflection and transmission coefficients for a plane TE wave at an interface.

in: e1/2 Relative dielectric constants ϵ_1 before and ϵ_2 after the interface.
 g $\gamma_{||} = k_x/k_0$ of the incident wave

out: R12 Reflection coefficient R_{12} at the interface
 T12 Transmission coefficient T_{12} through the interface

```
function [R12,T12]=surfaceRTte(e1,e2,g)
a=sqrt(e1-g.^2);
b=sqrt(e2-g.^2);
n=find(a-b ~= a+b);
R12=ones(size(a));
R12(n)=(a(n)-b(n))./(a(n)+b(n));
T12=R12+1;
```

All/private/surfaceRTtm.m

Fresnel reflection and transmission coefficients for a plane TM wave at an interface. See *surfaceRTte* for a list of parameters and results.

```
function [R12,T12]=surfaceRTtm(e1,e2,g)
a=e2*sqrt(e1-g.^2);
b=e1*sqrt(e2-g.^2);
n=find(a-b ~= a+b);
R12=ones(size(a));
R12(n)=(a(n)-b(n))./(a(n)+b(n));
T12=R12+1;
```

C.4 Progress information display

Status bar showing the evaluation progress.

All/createStatus.m

Creates and/or activates the status window.

out: f The status' window handle

```
function f=createStatus
f=findobj(get(0,'Children'),'flat','Tag','holeStatus');
if isempty(f) | ~ishandle(f)
r.t=cputime;
s=[256 64];
t=get(0,'ScreenSize');
f=figure('Colormap',[0 0 0;1 0 0;1 1 1],'CloseRequestFcn','', 'DoubleBuffer','on','MenuBar','none','Name','Field computation', ...
'IntegerHandle','off','NumberTitle','off','Resize','off','Position',[floor((t(3:4)-s)/2) s],'Tag','holeStatus','ToolBar','none');
a=axes('Parent',f,'Position',[0 0 1 1],'Visible','off','XLim',[0 256],'YLim',[0 64],'ZLim',[-1 1]);
%
%Horizontal bar
%
surface([4 252],[38 60],[-ones(2),3*ones(2)],[0 0 0;1 0 0;1 1 1],'DataMapping','direct','EdgeColor','none');
r.bar=surface([4 4],[38 60],zeros(2),2*ones(2),'DataMapping','direct','EdgeColor','none');
line([4 4 4 252;4 252 252 252],[38 38 60 38;60 38 60 60],ones(2,4),'Color',[0 0 0]);
%
%Description texts
%
o.FontWeight='bold';
o.Parent=a;
o.Units='pixels';
o.VerticalAlignment='middle';
text(136,49,1,'% ',o);
text(16,20,'Elapsed time: ',o);
text(200,20,'hours',o);
%
%Information texts
%
o.HorizontalAlignment='right';
r.ran=text(136,49,1,'0',o);
r.hms=text(198,20,'0:00:00',o);
set(f,'UserData',r);
else
figure(f);
end;
drawnow;
```

All/resetStatus.m

Resets the status bar to zero.

```
function resetStatus
f=createStatus;
r=get(f,'UserData');
r.t=cputime;
set(f,'CloseRequestFcn','', 'UserData',r);
updateStatus(0);
```

All/updateStatus.m

Updates the status bar.

in: p Relative progress $\in [0, 1]$

out: f The status' window handle

```

function f=updateStatus(p)
f=createStatus;
if p >= 1
    set(f,'CloseRequestFcn','delete(gcbo)');
end;
r=get(f,'UserData');
t=floor(cputime-r.t+0.5);
s=mod(t,60);
h=floor(t/60);
m=mod(h,60);
h=floor(h/60);
o='String';
p=max(0,min(1,p));
set(r.bar,'XData',[4 4+248*p]);
set(r.ran,o,sprintf('%u',floor(p*100+0.5)));
set(r.hms,o,sprintf('%u:%02u:%02u',[h;m;s]));
drawnow;

```

C.5 Graphical user interface

A graphical user interface to explore the evaluation results. Displays the intensity profile, the electromagnetic field and its phase at arbitrary xy , xz and yz cross-sections.

All/display3D.m

Starts exploring the electromagnetic field through an array of metallic apertures.

in: r *Structure*
 F Field to display first {'Ey' | 'Hy'}

```

function display3D(r, F)
if nargin < 2 | isempty(F) | ~isfield(r,F)
    F=[];
    s={'Ey','Hy','Ex','Hx','Ez','Hz'};
    for j=1:length(s)
        if isfield(r,s{j})
            F=s{j};
            break;
        end;
    end;
end;
if isempty(F) | ~isfield(r,F)
    disp(['There is no ' F ' field to display.']);
    return;
end;
r.F=F;
%
%Reduce memory consumption
%
s={'I','Ex','Ey','Ez','Hx','Hy','Hz'};
for n=1:length(s)
    if isfield(r,s{n})
        setfield(r,s{n},single(getfield(r,s{n})));
    end;
end;
%
%Display an xz, xy or yz section.
%
t=find(r.W > 1);
r.V=[t(1) t(length(t))];
%
%Display a centered section
%
t=r.W;
t(r.V)=1;
r.U=ceil(prod(t)/2);
%
%Precompute maximal intensity and field strength
%
r.M=double(max(max(max(r.I))));
if isinf(r.M)
    r.M=1;
end;
r.N=max(real(sqrt([r.e1 r.e2 r.e3 r.e4])));
%
%Creates the display and its controls window.
%
h=createDisplay3D(r);
s=get(0,'ScreenSize');
set(h,'Position',[s(3:4)-r.W(r.V))/2 r.W(r.V)];
updateControls3D(h);
updateDisplay3D(h);

```

All/eventControls3D.m

Callback function dispatcher for the 3D display controls.

in: n Callback function to execute

```
function eventControls3D(n)
if nargin
    feval(n,gcbo);
else
    [o,f]=gcbo;
    h=get(f,'UserData');
    r=get(h,'UserData');
    feval(get(o,'Tag'),f,h,o,r);
    updateDisplay3D(h);
end;

%
%Request to close the figure
%
function Close(f)
delete(findobj(get(0,'Children'),'flat','Tag','holeDisplay3D'));
delete(f);

%
%Change of the field component
%
function List(f,h,o,r)
s=get(o,'String');
r.F=s{get(o,'Value')};
switch get(get(h,'CurrentAxes'),'Tag')
case {'Field','Start'}
    set(h,'Name',['Field ' r.F]);
case 'Phase'
    set(h,'Name',['Phase ' r.F]);
end;
set(h,'UserData',r);

%
%Change of the view axes
%
function View(f,h,o,r)
s=get(o,'String');
switch s{get(o,'Value')}
case 'xy'
    r.V=[1 2];
case 'xz'
    r.V=[1 3];
case 'yz'
    r.V=[2 3];
end;
r.U=ceil(r.W(6-sum(r.V))/2);
set(h,'UserData',r);
updateSlider(f,r);
updateAxes(h,r);

%
%Change of the view position
%
function Slider(f,h,o,r)
s={'x','y','z'};
s=s{6-sum(r.V)};
p=getfield(r,s);
u=get(o,'Value');
t=abs(p-u);
r.U=find(t == min(t));
set(o,'Value',p(r.U));
set(h,'UserData',r);
updateText(f,s,p(r.U));

%
%Displays an electromagnetic field component
%
function Field(f,h,o,r)
set(h,'Name',['Field ' r.F]);
a=get(h,'CurrentAxes');
s='Field';
if strcmp(get(a,'Tag'),s)
    s='Start';
end;
set(a,'Tag',s);

%
%Displays the field intensity
%
function Intensity(f,h,o,r)
set(h,'Name','Intensity');
set(get(h,'CurrentAxes'),'Tag','Intensity');

%
%Displays the phase of a field component
%
function Phase(f,h,o,r)
set(h,'Name',['Phase ' r.F]);
set(get(h,'CurrentAxes'),'Tag','Phase');
```

All/eventDisplay3D.m

Callback function dispatcher for a 3D display figure.

in: n Callback function to execute
 h Display handle if manually called

```
function eventDisplay3D(n, h)
if nargin < 2
    h=gcbo;
end;
feval(n,h);

%
%Request to activate the figure.
%
function Activate(h)
updateControls3D(h);

%
%Request to close the figure.
%
function Close(h)
delete(h);
updateControls3D;

%
%Notification that the figure has been resized.
%
function Resize(h)
r=get(h,'UserData');
p=get(h,'Position');
p(3:4)=max(r.W(r.V),p(3:4));
t=r.S(r.V);
s=t./p(3:4);
if s(1) < s(2)
    p(3)=p(4)*t(1)/t(2);
else
    p(4)=p(3)*t(2)/t(1);
end;
set(h,'Position',p);
```

All/private/createControls3D.m

Creates and/or activates the 3D display controls window.

out: f The controls' window handle

```
function f=createControls3D
f=findobj(get(0,'Children'),'flat','Tag','holeControls3D');
if isempty(f) | ~ishandle(f)
    r='TooltipString';
    p='Position';
    s='String';
    y='Style';
    t='Tag';
    b=get(0,'ScreenSize');
    f=figure('CloseRequestFcn','eventControls3D(''Close''),'MenuBar','none','Name','Display controls','IntegerHandle','off', ...
            'NumberTitle','off','Resize','off',p,[50 b(4)-146 200 96],t,'holeControls3D','ToolBar','none');
    o.Parent=f;
    uicontrol(o,p,[72 64 120 24],s,' ',t,'Text',y,'text');
    o.Callback='eventControls3D';
    o.Enable='off';
    uicontrol(o,p,[8 64 56 24],s,{' '},t,'List',r,'Selects the electromagnetic field component',y,'popupmenu');
    uicontrol(o,p,[8 40 56 24],s,{'xy','xz','yz'},t,'View',r,'Selects the view axes',y,'popupmenu','Value',2);
    uicontrol(o,p,[72 40 120 24],t,'Slider',r,'Selects the section position',y,'slider');
    uicontrol(o,p,[8 8 56 24],s,'Field',t,'Field',r,'Displays the electromagnetic field');
    uicontrol(o,p,[72 8 56 24],s,'Intensity',t,'Intensity',r,'Displays the intensity');
    uicontrol(o,p,[136 8 56 24],s,'Phase',t,'Phase',r,'Displays the phase');
else
    figure(f);
end;
```

All/private/createDisplay3D.m

Creates a 3D display window.

in: r *Structure*

out: h The displays' window handle

```
function h=createDisplay3D(r)
o.NextPlot='replacechildren';
h=figure('CloseRequestFcn','eventDisplay3D(''Close''),'Colormap',[],'DoubleBuffer','on','HandleVisibility','callback', ...
        'IntegerHandle','off','MenuBar','none','Name','Intensity',o,'NumberTitle','off','ResizeFcn','eventDisplay3D(''Resize'')', ...
        'Tag','holeDisplay3D','ToolBar','none','UserData',r,'WindowButtonDownFcn','eventDisplay3D(''Activate'')');
axes('Parent',h,o,'DataAspectRatio',[1 1 1],'Position',[0 0 1 1],'Tag','Intensity','Visible','off');
eventDisplay3D('Resize',h);
updateAxes(h,r);
```

All/private/getControl.m

Fetches the handle of a specific user control.

in: f Handle of the figure window
t Tag of the control looked for

out: o Handle of the specified control

```
function o=getControl(f,t)
o=findobj(get(f,'Children'),'flat','Tag',t);
```

All/private/updateAxes.m

Updates the axes limits.

in: h Handle of the 3D display window
r *Structure*

```
function updateAxes(h,r)
s={'x','y','z'};
x=getfield(r,s{r.V(1)});
y=getfield(r,s{r.V(2)});
set(get(h,'CurrentAxes'),'XLim',[x(1) x(length(x))],'YLim',[y(1) y(length(y))]);
```

All/private/updateControls3D.m

Updates the 3D display controls to reflect the active figure.

in: h The window handle that is active

out: f The controls' window handle

```
function f=updateControls3D(h)
f=createControls3D;
s='holeDisplay3D';
u='UserData';
t='Tag';
if nargin & ishandle(h) & strcmp(get(h,t),s)
set(f,u,h);
else
h=get(f,u);
if isempty(h) | ~ishandle(h)
h=findobj(get(0,'Children'),'flat',t,s);
if isempty(h)
delete(f);
return;
end;
h=h(1);
set(f,u,h);
end;
end;
%
%Enable the controls
%
set(findobj(get(f,'Children'),'flat','Enable','off'),'Enable','on');
a=get(f,'CurrentAxes');
r=get(h,'UserData');
%
%Update the field components list
%
j=[];
s=fieldnames(r);
for n=1:length(s)
switch s{n}
case {'Ex','Ey','Ez','Hx','Hy','Hz'}
j=[j n];
end;
end;
s=s(j);
for n=1:length(s)
if strcmp(s{n},r.F)
j=n;
break;
end;
end;
set(getControl(f,'List'),'String',s,'Value',j);
%
%Update the axes view list
%
j=1;
switch sum(find(r.W > 1))
case 3
s={'xy'};
case 4
s={'xz'};
```

```

case 5
    s={'yz'};
case 6
    s={'xy','xz','yz'};
    j=sum(r.V)-2;
otherwise
    s=' ';
end;
set(getControl(f,'View'),'String',s,'Value',j);
%
%Update the slider and its text
%
updateSlider(f,r);

```

All/private/updateDisplay3D.m

Updates a 3D display window.

in: h Handle of the 3D display window

```

function updateDisplay3D(h)
a=get(h,'CurrentAxes');
feval(get(a,'Tag'),h,a,get(h,'UserData'));

%
%Extracts a section of the electromagnetic field
%
function [x,y,F]=getSection(r,f)
s={'x','y','z'};
x=getfield(r,s{r.V(1)});
y=getfield(r,s{r.V(2)});
F=zeros(r.W(r.V));
switch sum(r.V)
case 3
    F(:,.)=f(:,.,r.U);
case 4
    F(:,.)=f(:,r.U,:);
case 5
    F(:,.)=f(r.U,.,.);
end;
y=fliplr(y);
F=double(F.');

%
%Displays the electromagnetic field strength
%
function Field(h,a,r)
if ~isfield(r,'wt')
    r.wt=0;
    set(h,'UserData',r);
end;
t=zeros(256,3);
t(1:128,3)=(255:-2:1)'/255;
t(129:256,1)=(1:2:255)'/255;
set(h,'Colormap',t);
Frame(a,r);

%
%Animates the electromagnetic field
%
function Start(h,a,r)
s=cputime;
for l=1:72
    s=s+0.2;
    r.wt=r.wt+pi/18i;
    set(h,'UserData',r);
    Frame(a,r);
    drawnow;
    if cputime < s
        pause(s-cputime);
    end;
    if ~ishandle(h) | ~strcmp(get(a,'Tag'),'Start')
        return;
    end;
    r=get(h,'UserData');
end;
set(a,'Tag','Field');

%
%Displays the field component
%
function Frame(a,r)
[x,y,F]=getSection(r,getfield(r,r.F));
e0=8.85418782e-12; % [As/Vm]
u0=1.25663706e-6; % [Vs/Am]
if r.F(1) == 'E'
    t=127.5/sqrt(2*r.M*sqrt(u0/e0)/r.N);
else
    t=127.5/sqrt(2*r.M*sqrt(e0/u0));
end;
image(x,y,uint8(127.5+real(t*exp(r.wt)*F)),'Parent',a);

%
%Displays the electromagnetic field intensity

```

```

%
function Intensity(h,a,r)
[x,y,I]=getSection(r,r.I);
set(h,'Colormap',repmat((0:255)/255,1,3));
image(x,y,uint8(I*255/r.M),'Parent',a);

%
%Displays the electromagnetic field phase
%
function Phase(h,a,r)
[x,y,F]=getSection(r,getfield(r,r.F));
set(h,'Colormap',repmat((0:255)/255,1,3));
image(x,y,uint8((angle(F)+pi)*127.5/pi),'Parent',a);

```

All/private/updateSlider.m

Updates the slider and its corresponding text.

in: f Handle of the 3D display controls' window
r *Structure*

```

function updateSlider(f,r)
s={'x','y','z'};
s=s{6-sum(r.V)};
t=getfield(r,s);
set(getControl(f,'Slider'),'Min',t(1)-1e-15,'Max',t(length(t))+1e-15,'SliderStep',[1 10]/r.W(6-sum(r.V)),'Value',t(r.U));
updateText(f,s,t(r.U));

```

All/private/updateText.m

Updates the position text in the 3D display controls.

in: f Handle of the 3D display controls' window
s Coordinate axis string
p Position

```

function updateText(f,s,p)
set(getControl(f,'Text'),'String',[s ' = ' sprintf('%12.4g',p) 'm']);

```

C.6 Extraction of the relative dielectric constant of metals

In chapter A, we resumed the determination of the dielectric constant of metals. To extract the dielectric constant and the thickness of the metal film, we use them as parameters and fit the measured to the theoretical reflectance. The best fit is considered as the least square difference between the measured and the computed reflectance. Its parameter values yield the dielectric constant and the film thickness looked for.

Here, we use the Fresnel reflection coefficient at a thin metal layer as computed in subchapter 4.2.

All/globalReh.m

Fitted parameters from a Kretschmann setup. See *localReh* for details.

in: e1 Dielectric constant ϵ_1 of the sample holder
e3 Dielectric constant ϵ_3 behind the metal film
k0 Wavevector k_0 in free space [rad/m]
Rm Measured reflectance R_m in intensity
tm Incidence angles θ_m [°]
s Window title

out: e2 Dielectric constant ϵ_2 of the metal film
h2 Thickness h_2 of the metal film [m]

```

function [e2,h2]=globalReh(e1,e3, k0, Rm,tm, s)
m=[1+floor((0:4)/4*(length(Rm)-1)) find(Rm == min(Rm)) find(Rm == max(Rm))];
gm=sqrt(e1)*sin(tm(:)*pi/180);
Rm=Rm(:);
%
%Retrieve the angle and reflectance of the flattest curve part
%
g0=abs(derive(Rm,1,gm));
g0=find(g0 == min(g0));
g0=[max(1,g0-1):min(g0+1,length(Rm))];
%
%Compensate the damping due to the sample holder
%
% The sample holder is assumed to reduce the reflection
% efficiency by a constant factor (about 0.92 for a glass
% cylinder due to two glass-air interfaces for example).
%
Rm=Rm/mean(Rm(g0));
g0=g0(g0);
%
%Look for the global error minimum
%
h2=10e-9:10e-9:100e-9;
er=-1.5:-0.5:-20;
ei=0.1:0.1:5;
e2=complex(repmat(er',size(ei)),repmat(ei,size(er')));
e=zeros(length(h2),length(er),length(ei));
for a=1:length(h2)
    for b=1:length(e2(:))
        e(a,b)=localReh([e2(b);h2(a)],e1,e3,k0,g0,gm(m),Rm(m));
    end;
end;
[a,b]=find(e == min(e(:)));
%
%Refine the global minimum looking for the local minimum
%
[m,e]=fminsearch(@localReh,[e2(b);h2(a)],optimset('Display','off','TolX',1e-4),e1,e3,k0,g0,gm,Rm);
e2=m(1);
h2=m(2);
%
%Display results and diagnostics
%
tt=min(tm):0.01:max(tm);
gm=sqrt(e1)*sin(tt*pi/180);
Rt=layerRTtm(e1,e3,gm,k0,h2);
R0=layerRTtm(e1,e3,g0,k0,h2);
figure('Name','s','NumberTitle','off','ToolBar','none');
plot(tm,Rm,'ro',tt,real(Rt.*conj(Rt)/mean(R0.*conj(R0))),'b-');
fprintf('Found:\n\te2 = %g %gi\n\th2 = %gmm\n\nMean error is %g%%.\n',[real(e2) imag(e2) 1e9*h2 100*sqrt(e/length(Rm))]);
set(gca,'Box','off','FontSize',14,'Position',[0.08 0.08 0.8 0.8],'XLim',[min(tm) max(tm)]);
text(1.05,-0.04,'ti','FontSize',14,'HorizontalAlignment','left','VerticalAlignment','middle','Units','Normalized');
text(-0.02,1.06,'R','FontSize',14,'HorizontalAlignment','center','VerticalAlignment','baseline','Units','Normalized');
grid on;
%
%Special
%
h=get(gca,'Children');
set(gca,'XLim',[41.5 80],'YLim',[0 1.2],'YTick',[0:0.2:1.2],'YTickLabel',{'0.0';'0.2';'0.4';'0.6';'0.8';'1.0';'1.2'});
set(h(4),'MarkerEdgeColor','black','MarkerFaceColor','black','MarkerSize',3);
set(h,'Color','black');

```

All/localReh.m

Squaresum of the differences between the computed and measured reflectance in a Kretschmann setup. Collimated, monochromatic light is incident on a thin metal film at TM polarisation. The metal film is deposited on a sample holder - in general a glass semi-cylinder. The reflection efficiency $R_m = I_r/I_i$ is measured at various incidence angles.

in: x Fitting parameters $\begin{pmatrix} \epsilon_2 \\ h_2 \end{pmatrix}$ [m]

 e1 Dielectric constant ϵ_1 of the sample holder

 e3 Dielectric constant ϵ_3 behind the metal film

 k0 Wavevector k_0 in free space [rad/m]

 gf Incidence angles γ_f of the flat

 gm Incidence angles γ_m [$k_x/k_0 > 1$]

 Rm Measured reflectance R_m in intensity

out: e Square-summed error between R_m and the theoretical reflectance

```

function e=localReh(x,e1,e3,k0,gf,gm,Rm)
Rt=layerRTtm(e1,x(1),e3,gm,k0,x(2));
Rf=layerRTtm(e1,x(1),e3,gf,k0,x(2));
Rt=Rt.*conj(Rt)/mean(Rf.*conj(Rf));
e=sum((Rm-real(Rt)).^2);

```