

# Light Transmission through Sub-Wavelength Apertures



VRIJE UNIVERSITEIT

**Light Transmission through  
Sub-Wavelength Apertures**

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan  
de Vrije Universiteit Amsterdam,  
op gezag van de rector magnificus  
prof.dr. T. Sminia,  
in het openbaar te verdedigen  
ten overstaan van de promotiecommissie  
van de faculteit der Exacte Wetenschappen  
op dinsdag 22 november 2005 om 15.45 uur  
in het auditorium van de universiteit,  
De Boelelaan 1105

door

Hugo Frederik Schouten

geboren te Amsterdam

promotoren: prof.dr. D. Lenstra  
                  prof.dr.ir. H. Blok  
copromotor: dr. T.D. Visser

This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

# Contents

<b>Contents</b>	<b>5</b>
<b>1 Introduction</b>	<b>7</b>
1.1 Historical introduction . . . . .	7
1.2 Outline of this thesis . . . . .	8
1.3 The Maxwell equations in matter . . . . .	9
1.3.1 Constitutive relations . . . . .	10
1.3.2 Boundary conditions . . . . .	11
1.4 The steady-state Maxwell equations . . . . .	12
1.5 The Poynting vector . . . . .	15
1.6 Two-dimensional electromagnetic fields . . . . .	16
1.7 Guided modes . . . . .	18
1.7.1 Guided modes inside a slit in a metal plate . . . . .	18
1.7.2 Surface plasmons . . . . .	20
1.8 Phase singularities . . . . .	21
1.8.1 Singular and stationary points . . . . .	22
1.8.2 Topological charge and index . . . . .	25
1.8.3 Phase singularities in two-dimensional electromagnetic waves	29
<b>2 The Green's Tensor Formalism</b>	<b>33</b>
2.1 Introduction . . . . .	34
2.2 The scattering model . . . . .	35
2.3 The derivation of the Green's tensors . . . . .	39
2.3.1 The Green's tensors for a homogeneous medium . . . . .	39
2.3.2 The Green's tensor for a layered medium . . . . .	42
2.4 The collocation method for solving the domain integral equation . .	46
2.5 The conjugate gradient method for solving the linear system . . . .	51
2.6 Conclusions . . . . .	56
2.A Plane wave incident on a stratified medium . . . . .	56

---

<b>3</b>	<b>Light Transmission through a Single Sub-wavelength Slit</b>	<b>59</b>
3.1	Introduction . . . . .	60
3.2	The configuration . . . . .	61
3.3	Transmission through a slit in a thin metal plate . . . . .	63
3.4	Transmission through a slit in a thin semiconductor plate . . . . .	68
3.5	Transmission through a slit in a thick metal plate . . . . .	71
3.6	Conclusions . . . . .	73
<b>4</b>	<b>The Radiation Pattern of a Single Slit in a Metal Plate</b>	<b>77</b>
4.1	Introduction . . . . .	78
4.2	The definition of the radiation pattern . . . . .	78
4.3	The radiation pattern of a slit for normal incidence . . . . .	81
4.4	Oblique angle of incidence . . . . .	87
4.5	Conclusions . . . . .	88
<b>5</b>	<b>Plasmon-assisted Light Transmission through Two Slits</b>	<b>89</b>
5.1	Introduction . . . . .	90
5.2	Plasmon-assisted transmission through two slits . . . . .	91
5.3	A surface plasmon Fabry-Pérot effect . . . . .	97
5.4	Conclusions . . . . .	103
<b>6</b>	<b>Young's Interference Experiment with Partially Coherent Light</b>	<b>105</b>
6.1	Introduction . . . . .	106
6.2	Partially coherent fields . . . . .	106
6.3	Coherence properties of light in Young's interference experiment . . . . .	108
6.4	Planes of full coherence . . . . .	109
6.5	Phase singularities of the coherence functions . . . . .	112
6.6	Plasmon-induced coherence in Young's interference experiment . . . . .	116
6.7	Conclusions . . . . .	119
	<b>Bibliography</b>	<b>121</b>
	<b>Samenvatting</b>	<b>131</b>
	<b>List of Publications</b>	<b>133</b>
	<b>Dankwoord</b>	<b>135</b>

# Chapter 1

## Introduction

### 1.1 Historical introduction

The diffraction of light by an aperture in a screen is one of the classical subjects in Physical Optics. For example, in the nineteenth century it was the main application of both Fraunhofer and Fresnel diffraction. The diffraction by an aperture was usually treated by approximating the field in the aperture by the incident field and using the Huygens-Fresnel principle to calculate the diffracted field. This approach was put on a more rigorous mathematical basis by Kirchhoff [BORN AND WOLF, 1999, Chap. 8]. However, he still needed to approximate the field in the aperture by the incident field, an approximation that is only valid for apertures with dimensions much larger than the wavelength. For the experiments and applications of those days involving light this approximation was justified, because the wavelength of visible light ( $\sim 500$  nm) is much smaller than the aperture sizes that were typically used.<sup>1</sup>

However, this was not always the case for the similar problem of the transmission of *sound* through apertures. Rayleigh [1897] was the first to calculate the diffraction of sound by apertures with dimensions much smaller than the wavelength. In the early twentieth century, this kind of approach also became relevant for Optics, because of the discovery of radio waves and microwaves. This became especially relevant at the time of the World War II, because of the many applications of such waves that then emerged. There are several studies devoted to this problem, the most famous one by Bethe [1944]. For a review of this early work on transmission problems, see [BOUWKAMP, 1954]. All these studies have in common that they assume that the screen is *perfectly conducting*, an assumption that is quite reasonable for metals at radio or microwave frequencies. Perfect con-

---

<sup>1</sup>However, experiments involving sub-wavelength slits were already performed by Fizeau [1861].

duction means essentially that the electromagnetic waves cannot penetrate into the metal plate. Usually, the assumption of perfect conduction is accompanied by the assumption of an infinitely thin metal plate.

The first interest in the *light* transmission through sub-wavelength apertures arose in the eighties, due to the invention of the near-field optical microscope [POHL *et al.*, 1984; BETZIG AND TRAUTMAN, 1992]. In such a microscope a sub-wavelength-sized tip is scanned very close (i.e., at distances smaller than the wavelength) across a sample, to obtain sub-wavelength resolution. One of the disadvantages was that the light throughput of the tip was very low. To obtain a better insight into this problem, the similar configuration of an aperture in a metal plate was studied again (see e.g. [BETZIG *et al.*, 1986; LEVIATAN, 1986; ROBERTS, 1987]). In these studies usually the assumption of perfect conductivity was still applied, although at optical wavelengths it is questionable.

Quite recently, light transmission through sub-wavelength apertures has turned out to be a hot topic in Optics. This is due to the observation by Ebbesen *et al.* [1998] of extraordinarily large transmission through hole arrays in a metal plate. This effect was attributed to the occurrence of *surface plasmons*, which are surface waves on a metal-dielectric interface [RAETHER, 1988].<sup>2</sup> However, this explanation was questioned by some other authors (see e.g. [CAO AND LALANNE, 2002]), which resulted in an intense debate about this subject.

In this thesis, the light transmission through a single slit is studied. Contrary to most studies, we will take into account both the finite thickness and finite conductivity of the metal plate, by making use of a rigorous Green's tensor method. Another topic addressed in this thesis is the light transmission through two apertures, with the aim of clarifying the role of surface plasmons in the interaction between the two apertures.

## 1.2 Outline of this thesis

In the remainder of this Chapter we discuss the necessary background material to understand the rest of the thesis. It consists of a brief summary of the Maxwell equations, guided modes, surface plasmons and an introduction to Singular Optics.

The second Chapter describes the scattering model which is used in later Chapters to calculate the field near sub-wavelength slits. In this scattering model, the Maxwell equations are converted into an integral equation with a Green tensor as a kernel. This Green tensor is derived for a multi-layered background medium. Also the numerical procedure to solve the integral equation is described.

---

<sup>2</sup>See also [BARNES *et al.*, 2003; ZAYATS AND SMOLYANINOV, 2003] for reviews about the recent surge of interest in surface plasmons.

The third Chapter discusses the light transmission through a single sub-wavelength slit. The influence of several parameters such as the slit width, the plate thickness, the material properties of the plate and the polarization of the incident field are discussed. For the explanation of the results the concepts of guided modes inside the slit and phase singularities of the Poynting vector are used.

In the fourth Chapter the radiation pattern of a single narrow slit is investigated. The radiation pattern describes how the light is diffracted in different directions. The results can, as in the preceding Chapter, be explained in terms of waveguiding and phase singularities.

The light transmission through *two* sub-wavelength slits is the topic of the fifth Chapter. The results are explained by a heuristic model involving the local excitation of surface plasmons at the slits.

The sixth Chapter gives a description of the transmission of *partially coherent* light through two apertures, i.e., in contrast to the other Chapters the electromagnetic field is not taken to be coherent and monochromatic. In the first part of Chapter 6 the coherence properties of Young's interference experiment are described. In the second part of the Chapter the consequences of the presence of surface plasmons on the coherence properties is described.

### 1.3 The Maxwell equations in matter

The Maxwell equations in matter are given by<sup>3</sup>

$$-\nabla \times \mathbf{H}(\mathbf{r}, t) + \mathbf{J}^{(\text{ind})}(\mathbf{r}, t) + \partial_t \mathbf{D}(\mathbf{r}, t) = -\mathbf{J}^{(\text{ext})}(\mathbf{r}, t), \quad (1.1)$$

$$\nabla \times \mathbf{E}(\mathbf{r}, t) + \partial_t \mathbf{B}(\mathbf{r}, t) = 0, \quad (1.2)$$

where  $\nabla = (\partial_x, \partial_y, \partial_z)$  denotes differentiation with respect to the spatial Cartesian coordinates  $\mathbf{r} = (x, y, z)$ , and  $\partial_t$  denotes differentiation with respect to the time of observation  $t$ . Furthermore,

$\mathbf{E}$	=	the <i>electric field strength</i> (V/m),
$\mathbf{H}$	=	the <i>magnetic field strength</i> (A/m),
$\mathbf{D}$	=	$\varepsilon_0 \mathbf{E} + \mathbf{P}$ = the <i>electric flux density</i> (C/m <sup>2</sup> ),
$\mathbf{B}$	=	$\mu_0 (\mathbf{H} + \mathbf{M})$ = the <i>magnetic flux density</i> (T),
$\mathbf{J}^{(\text{ind})}$	=	the <i>induced volume density of electric conduction current</i> (A/m <sup>2</sup> ),
$\mathbf{P}$	=	the <i>electric polarization</i> (C/m <sup>2</sup> ),
$\mathbf{M}$	=	the <i>magnetization</i> (A/m),
$\mathbf{J}^{(\text{ext})}$	=	the <i>external volume density of electric current</i> (A/m <sup>2</sup> ),
$\varepsilon_0$	=	the <i>permittivity in vacuum</i> (F/m),
$\mu_0$	=	the <i>permeability in vacuum</i> (H/m).

---

<sup>3</sup>Only SI units will be used in this thesis.

$\mathbf{J}^{(\text{ind})}$ ,  $\mathbf{P}$  and  $\mathbf{M}$  describe the reaction of matter to the presence of electromagnetic fields.

$\mathbf{J}^{(\text{ext})}$  describes the current sources that, together with the electric charges, generate the fields.<sup>4</sup>

$\varepsilon_0$  and  $\mu_0$  are constants which are related by

$$\varepsilon_0 = \frac{1}{\mu_0 c_0^2}, \quad (1.3)$$

where  $c_0 = 2.99792458 \times 10^8$  m/s is the *speed of light in vacuum*. The value of  $\mu_0$  is  $\mu_0 = 4\pi \times 10^{-7}$  H/m. So one obtains  $\varepsilon_0 = 8.8541878 \times 10^{-12}$  F/m.

The Maxwell equations are supplemented by the following *compatibility relations*,

$$\nabla \cdot \mathbf{B}(\mathbf{r}, t) = 0, \quad (1.4)$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t) - \rho^{(\text{ind})}(\mathbf{r}, t) = \rho^{(\text{ext})}(\mathbf{r}, t), \quad (1.5)$$

where the *induced volume density of electric charge*  $\rho^{(\text{ind})}$  and the *external volume density of electric charge*  $\rho^{(\text{ext})}$  are introduced. They are related to the current densities  $\mathbf{J}^{(\text{ind})}$  and  $\mathbf{J}^{(\text{ext})}$  by the *continuity relations*

$$\nabla \cdot \mathbf{J}^{(\text{ind})}(\mathbf{r}, t) + \partial_t \rho^{(\text{ind})}(\mathbf{r}, t) = 0, \quad (1.6)$$

$$\nabla \cdot \mathbf{J}^{(\text{ext})}(\mathbf{r}, t) + \partial_t \rho^{(\text{ext})}(\mathbf{r}, t) = 0. \quad (1.7)$$

$\mathbf{J}^{(\text{ext})}$  and  $\rho^{(\text{ext})}$  are called the *sources*, and are considered to be field-independent.

### 1.3.1 Constitutive relations

The Maxwell equations (1.1) and (1.2) constitute an incomplete system of equations since the number of equations is less than the number of unknown quantities. Therefore supplementing equations, known as the *constitutive relations*, are needed, which describe the reaction of matter to the electric and magnetic fields. These relations express  $\mathbf{J}^{(\text{ind})}$ ,  $\mathbf{D}$  and  $\mathbf{B}$  in terms of  $\mathbf{E}$  and  $\mathbf{H}$ . We assume that the medium is *linear, time invariant, locally reacting, isotropic* and *causal* [DE HOOP, 1995, Chap. 19]. Furthermore, we assume that  $\mathbf{J}^{(\text{ind})}$  and  $\mathbf{D}$  are only dependent on  $\mathbf{E}$ , whereas  $\mathbf{B}$  depends only on  $\mathbf{H}$ . In that case the constitutive

---

<sup>4</sup>Some authors put on the right-hand side of (1.2) an *external volume density of magnetic current*  $\mathbf{K}^{(\text{ext})}$  for reasons of symmetry. See, e.g., [BLOK AND VAN DEN BERG, 1999].

relations are given by

$$\mathbf{J}^{(\text{ind})}(\mathbf{r}, t) = \int_0^\infty \kappa_c(\mathbf{r}, t') \mathbf{E}(\mathbf{r}, t - t') dt', \quad (1.8)$$

$$\mathbf{D}(\mathbf{r}, t) = \int_0^\infty \kappa_e(\mathbf{r}, t') \mathbf{E}(\mathbf{r}, t - t') dt', \quad (1.9)$$

$$\mathbf{B}(\mathbf{r}, t) = \int_0^\infty \kappa_m(\mathbf{r}, t') \mathbf{H}(\mathbf{r}, t - t') dt', \quad (1.10)$$

where  $\kappa_c$ ,  $\kappa_e$  and  $\kappa_m$  are the *conduction relaxation function*, the *dielectric relaxation function* and the *magnetic relaxation function*, respectively.

### 1.3.2 Boundary conditions

Consider an interface  $\partial D$  between two adjacent media,  $D^{(1)}$  and  $D^{(2)}$  with different constitutive functions (see Eqs. (1.8–1.10)). Assume that this interface has at every point an unique tangent plane and let  $\mathbf{n}_{1 \rightarrow 2}$  denote the unit normal vector of  $\partial D$ , pointing from  $D^{(1)}$  into  $D^{(2)}$ .

We now want to investigate the behavior of the electromagnetic field at the interface. First the electromagnetic field is split in a normal part and a part that is parallel to the interface. To derive the behavior of the normal parts of the field, Eqs. (1.4) and (1.5) are used. We take an infinitesimal “pillbox”  $V$ , positioned half in  $D^{(1)}$  and half in  $D^{(2)}$ , as is drawn in Fig. 1.1. Eqs. (1.4) and (1.5) are integrated over  $V$  and if Gauss’ theorem is applied to the result, one obtains

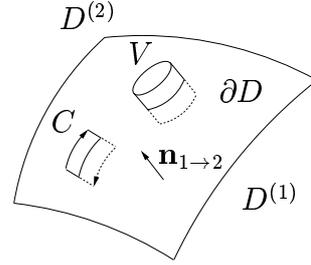


Figure 1.1: Impression of the interface between  $D^{(1)}$  and  $D^{(2)}$ .

$$\int_V \nabla \cdot \mathbf{B}(\mathbf{r}, t) d^3r = \int_{\partial V} \mathbf{n} \cdot \mathbf{B}(\mathbf{r}, t) d^2r = 0, \quad (1.11)$$

$$\int_V \nabla \cdot \mathbf{D}(\mathbf{r}, t) d^3r = \int_{\partial V} \mathbf{n} \cdot \mathbf{D}(\mathbf{r}, t) d^2r = \int_V \rho^{(\text{ind})}(\mathbf{r}, t) + \rho^{(\text{ext})}(\mathbf{r}, t) d^3r. \quad (1.12)$$

Here  $\mathbf{n}$  denotes the outward normal to  $V$  and  $\partial V$  denotes the boundary of  $V$ . In the limit of a very shallow pillbox the side surface does not contribute to the integrals in (1.11) and (1.12). If the top and bottom of  $V$  are tangential to the interface  $\partial D$ , then (1.11) and (1.12) become:

$$(\mathbf{B}_2 - \mathbf{B}_1) \cdot \mathbf{n}_{1 \rightarrow 2} = 0, \quad (1.13)$$

$$(\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n}_{1 \rightarrow 2} = \Sigma, \quad (1.14)$$

where  $\Sigma$  denotes the *surface charge density* at the interface and the subscripts 1 and 2 denote the field in  $D^{(1)}$  and  $D^{(2)}$ , respectively.

In order to analyze the tangential components we use a rectangular closed contour  $C$  that crosses the interface and has its plane perpendicular to it, such that its normal  $\mathbf{t}$  is tangential to the interface, see Fig. 1.1. The arms of the contour are chosen such that the two long arms are tangential to  $\partial D$  and the other two short arms are perpendicular to the interface. On integrating (1.1) and (1.2) over  $C$ , one obtains with the help of Stokes' theorem:

$$\oint_C \mathbf{H}(\mathbf{r}, t) \cdot d\mathbf{l} = \int_S [\mathbf{J}^{(\text{ind})}(\mathbf{r}, t) + \mathbf{J}^{\text{ext}}(\mathbf{r}, t) + \partial_t \mathbf{D}(\mathbf{r}, t)] \cdot \mathbf{t} \, d^2r, \quad (1.15)$$

$$\oint_C \mathbf{E}(\mathbf{r}, t) \cdot d\mathbf{l} = - \int_S \partial_t \mathbf{B}(\mathbf{r}, t) \cdot \mathbf{t} \, d^2r, \quad (1.16)$$

where  $S$  denotes the surface inside  $C$ .

If the short arms of  $C$  are of negligible size, then (1.15) and (1.16) become:

$$(\mathbf{H}_1 - \mathbf{H}_2) \times \mathbf{n}_{1 \rightarrow 2} = \mathbf{J}^{(\text{sur})}, \quad (1.17)$$

$$(\mathbf{E}_1 - \mathbf{E}_2) \times \mathbf{n}_{1 \rightarrow 2} = 0, \quad (1.18)$$

where  $\mathbf{J}^{(\text{sur})}$  is the *surface current density*.

## 1.4 The steady-state Maxwell equations

If  $f(\mathbf{r}, t)$  denotes an electromagnetic quantity that is causally related to the action of some source that is switched on at time  $t = 0$ , then its one-sided Laplace transform with respect to time is given by

$$\hat{f}(\mathbf{r}, s) = \int_0^\infty e^{-st} f(\mathbf{r}, t) \, dt, \quad (1.19)$$

where  $s$  is a complex number such that  $\text{Re}(s) > s_0$ . Here is  $s_0$  chosen such that for sufficiently large  $t$ ,  $|e^{-s_0 t} f(\mathbf{r}, t)| \leq M$ , where  $M$  is a positive constant. When  $\hat{f}(\mathbf{r}, t)$  is evaluated,  $f(\mathbf{r}, t)$  can be recovered by the Bromwich integral [ARFKEN AND WEBER, 1995, p. 908]:

$$f(\mathbf{r}, t) = \frac{1}{2\pi i} \int_{s_0 - i\infty}^{s_0 + i\infty} e^{st} \hat{f}(\mathbf{r}, s) \, ds, \quad (1.20)$$

where  $i$  is the imaginary unit.

If the Maxwell equations (1.1) and (1.2) are subjected to a one-sided Laplace transformation with respect to time, one obtains

$$-\nabla \times \hat{\mathbf{H}}(\mathbf{r}, s) + \hat{\mathbf{J}}^{(\text{ind})}(\mathbf{r}, s) + s\hat{\mathbf{D}}(\mathbf{r}, s) = -\hat{\mathbf{J}}^{(\text{ext})}(\mathbf{r}, s), \quad (1.21)$$

$$\nabla \times \hat{\mathbf{E}}(\mathbf{r}, s) + s\hat{\mathbf{B}}(\mathbf{r}, s) = 0, \quad (1.22)$$

where it was used that the Laplace transform of  $\partial_t f(t)$  is given by  $s\hat{f}(s) - f(0)$ . In this case  $f(0) = 0$ , because  $f$  is causally related to the source that was switched on at  $t = 0$ .

The constitutive relations (1.8–1.10) simplify significantly if one takes the Laplace transform:

$$\hat{\mathbf{J}}^{(\text{ind})}(\mathbf{r}, s) = \sigma(\mathbf{r}, s)\hat{\mathbf{E}}(\mathbf{r}, s), \quad (1.23)$$

$$\hat{\mathbf{D}}(\mathbf{r}, s) = \varepsilon(\mathbf{r}, s)\hat{\mathbf{E}}(\mathbf{r}, s), \quad (1.24)$$

$$\hat{\mathbf{B}}(\mathbf{r}, s) = \mu(\mathbf{r}, s)\hat{\mathbf{H}}(\mathbf{r}, s), \quad (1.25)$$

where  $\sigma$ ,  $\varepsilon$  and  $\mu$  are the *conductivity*, the *permittivity* and the *permeability* of the medium defined by

$$\sigma(\mathbf{r}, s) = \int_0^\infty \kappa_c(\mathbf{r}, t)e^{-st} dt, \quad (1.26)$$

$$\varepsilon(\mathbf{r}, s) = \int_0^\infty \kappa_e(\mathbf{r}, t)e^{-st} dt, \quad (1.27)$$

$$\mu(\mathbf{r}, s) = \int_0^\infty \kappa_m(\mathbf{r}, t)e^{-st} dt. \quad (1.28)$$

In the derivation of Eqs. (1.23–1.25) the convolution theorem for Laplace transforms was used [ARFKEN AND WEBER, 1995, Sec. 15.11].

The *relative permittivity*  $\varepsilon_r$  and the *relative permeability*  $\mu_r$  are defined by  $\varepsilon_r = \varepsilon/\varepsilon_0$  and  $\mu_r = \mu/\mu_0$ . If the constitutive relations hold and the medium is *homogeneous*, then  $\sigma$ ,  $\varepsilon$  and  $\mu$  do not depend on  $\mathbf{r}$ . When a medium has  $\sigma = 0$ , it is called *nonconducting*. If a medium has  $\mu_r = 1$ , it is called *nonmagnetic*.

If we assume that  $s_0 = 0$  and consider  $\hat{f}$  for imaginary values  $s = -i\omega$ , where  $\omega$  is the *angular frequency*, Eqs. (1.19) and (1.20) can be written as

$$\hat{f}(\mathbf{r}, -i\omega) = \int_0^\infty e^{i\omega t} f(\mathbf{r}, t) dt, \quad (1.29)$$

$$f(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^\infty e^{-i\omega t} \hat{f}(\mathbf{r}, -i\omega) d\omega. \quad (1.30)$$

In the case of *steady-state fields*, all electromagnetic field quantities are assumed to be sinusoidally varying in time with a common angular frequency  $\omega$ . Then each

real quantity  $f(\mathbf{r}, t)$  is related to the complex quantity  $\hat{f}(\mathbf{r}, -i\omega)$  via

$$f(\mathbf{r}, t) = \text{Re}[\hat{f}(\mathbf{r}, -i\omega)e^{-i\omega t}]. \quad (1.31)$$

The steady-state analysis may be considered as the limiting case of the Laplace transform where  $s \rightarrow -i\omega$  via  $\text{Re}(s) > 0$  [BLOK AND VAN DEN BERG, 1999, Sec. 2.7]. From Eqs. (1.21) and (1.22) it follows that the steady-state Maxwell equations are given by

$$-\nabla \times \hat{\mathbf{H}}(\mathbf{r}) + \sigma(\mathbf{r})\hat{\mathbf{E}}(\mathbf{r}) - i\omega\varepsilon(\mathbf{r})\hat{\mathbf{E}}(\mathbf{r}) = -\hat{\mathbf{J}}^{(\text{ext})}(\mathbf{r}), \quad (1.32)$$

$$\nabla \times \hat{\mathbf{E}}(\mathbf{r}) - i\omega\mu(\mathbf{r})\hat{\mathbf{H}}(\mathbf{r}) = 0, \quad (1.33)$$

where the  $\omega$  dependence of all quantities is omitted, because we now consider  $\omega$  only as a parameter.

To describe a lossy medium, we could use a non-zero  $\sigma$ . Instead of this we will use a complex-valued  $\varepsilon$  and  $\sigma = 0$ . Eq. (1.32) shows that the effect is the same. The relation between the two representations is  $\text{Im}(\varepsilon) = \sigma/\omega$ . The motivation behind this is as follows. Consider a homogeneous medium without sources. If we take the curl of Eq. (1.33) and use Eqs. (1.32), (1.5) and (1.9), the result is:

$$\nabla^2 \hat{\mathbf{E}}(\mathbf{r}) + (\omega^2\varepsilon\mu + i\omega\sigma\mu)\hat{\mathbf{E}}(\mathbf{r}) = 0. \quad (1.34)$$

We try plane wave solutions of the form  $\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}$ , where  $\mathbf{k}$  is the *wave vector*. When this Ansatz is inserted in Eq. (1.34), one obtains the relation  $k^2 = \omega^2\varepsilon\mu + i\omega\sigma\mu$  for the *wave number*  $k$ . Condition (1.5) gives that  $\mathbf{k} \cdot \mathbf{E}_0 = 0$ . If we define the *index of refraction* as  $n := k/k_0$ , where  $k_0 = \omega/c_0$ , then

$$n = \sqrt{\varepsilon_r\mu_r + i\frac{\sigma\mu_r}{\omega\varepsilon_0}}, \quad (1.35)$$

with the square root chosen such that  $\text{Im}(n) \geq 0$ . In optics, it is common to work with a complex-valued index of refraction rather than with a conductivity  $\sigma$ . Therefore, only complex-valued permittivities or refractive indices are used in this thesis, and the conductivity is set to zero.

The boundary conditions in the steady-state analysis are similar as before (see Eqs. (1.13)–(1.14) and Eqs. (1.17)–(1.18)):

$$(\hat{\mathbf{B}}_2 - \hat{\mathbf{B}}_1) \cdot \mathbf{n}_{1 \rightarrow 2} = 0, \quad (1.36)$$

$$(\hat{\mathbf{D}}_2 - \hat{\mathbf{D}}_1) \cdot \mathbf{n}_{1 \rightarrow 2} = 0, \quad (1.37)$$

and

$$(\hat{\mathbf{H}}_1 - \hat{\mathbf{H}}_2) \times \mathbf{n}_{1 \rightarrow 2} = 0, \quad (1.38)$$

$$(\hat{\mathbf{E}}_1 - \hat{\mathbf{E}}_2) \times \mathbf{n}_{1 \rightarrow 2} = 0. \quad (1.39)$$

Here the surface charge and surface current densities are assumed to be zero. The surface current density is zero because we are considering media with finite conductivity. The surface charge density is then found to be zero because of the continuity relations (1.6) and (1.7) for the case of steady-state fields.

It follows from Eqs.(1.36–1.37) that the normal parts of  $\hat{\mathbf{B}}$  and  $\hat{\mathbf{D}}$  are continuous across the interface. But if Eqs. (1.24) and (1.25) hold and  $D^{(1)}$  and  $D^{(2)}$  are homogeneous media with different  $\varepsilon$  and  $\mu$ , then the normal parts of  $\mathbf{E}$  and  $\mathbf{H}$  can be discontinuous across the interface. Similarly it follows that the tangential parts of  $\mathbf{E}$  and  $\mathbf{H}$  are continuous. For the same reason as for the normal parts of  $\mathbf{E}$  and  $\mathbf{H}$ , the tangential parts of  $\mathbf{D}$  and  $\mathbf{B}$  do not have to be continuous.

## 1.5 The Poynting vector

The work done per second by the electromagnetic field in a volume  $\mathcal{D}$  is given by

$$\frac{dW(t)}{dt} = \int_{\mathcal{D}} \mathbf{J}^{(\text{ind})}(\mathbf{r}, t) \cdot \mathbf{E}(\mathbf{r}, t) \, d^3r. \quad (1.40)$$

If Eq. (1.1) is substituted into Eq. (1.40), one obtain

$$\begin{aligned} \frac{dW(t)}{dt} = & - \int_{\mathcal{D}} \mathbf{J}^{(\text{ext})}(\mathbf{r}, t) \cdot \mathbf{E}(\mathbf{r}, t) \, d^3r \\ & - \int_{\mathcal{D}} \nabla \cdot [\mathbf{E}(\mathbf{r}, t) \times \mathbf{H}(\mathbf{r}, t)] \, d^3r \\ & - \int_{\mathcal{D}} [\mathbf{E}(\mathbf{r}, t) \cdot \partial_t \mathbf{D}(\mathbf{r}, t) + \mathbf{H}(\mathbf{r}, t) \cdot \partial_t \mathbf{B}(\mathbf{r}, t)] \, d^3r, \end{aligned} \quad (1.41)$$

where Eq. (1.2) and the vector identity  $\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H})$  were used.

The first term in Eq. (1.41) represents the electromagnetic power generated by the sources in the volume  $\mathcal{D}$ . The second term can, with Gauss' theorem, be written as

$$\int_{\mathcal{D}} \nabla \cdot [\mathbf{E}(\mathbf{r}, t) \times \mathbf{H}(\mathbf{r}, t)] \, d^3r = \int_{\partial\mathcal{D}} \mathbf{n} \cdot \mathbf{S}(\mathbf{r}, t) \, d^2r, \quad (1.42)$$

where  $\partial\mathcal{D}$  is the boundary of  $\mathcal{D}$ ,  $\mathbf{n}$  is the outward normal of  $\partial\mathcal{D}$  and  $\mathbf{S} = \mathbf{E} \times \mathbf{H}$  is the *Poynting vector*, representing the energy per square meter per second flowing out of the volume. The last term of Eq. (1.41) represents the change of electromagnetic energy in the volume  $\mathcal{D}$ .

Eq. (1.41) holds for an arbitrary volume  $\mathcal{D}$ , so it can also be written in a differential form, namely

$$\begin{aligned} -\mathbf{E}(\mathbf{r}, t) \cdot \mathbf{J}^{(\text{ext})}(\mathbf{r}, t) = & [\mathbf{E}(\mathbf{r}, t) \cdot \partial_t \mathbf{D}(\mathbf{r}, t) + \mathbf{H}(\mathbf{r}, t) \cdot \partial_t \mathbf{B}(\mathbf{r}, t)] \\ & + \nabla \cdot \mathbf{S}(\mathbf{r}, t) + \mathbf{J}^{(\text{ind})}(\mathbf{r}, t) \cdot \mathbf{E}(\mathbf{r}, t), \end{aligned} \quad (1.43)$$

where Eq. (1.8) was used.

In the steady-state case one works with the *time-averaged Poynting vector*  $\langle \mathbf{S} \rangle_T$ , rather than with the Poynting vector  $\mathbf{S}$ . We define  $\langle \mathbf{S} \rangle_T$  by

$$\langle \mathbf{S}(\mathbf{r}) \rangle_T = \frac{1}{T} \int_{t'}^{t'+T} \mathbf{S}(\mathbf{r}, t) dt, \quad (1.44)$$

where  $T = 2\pi/\omega$  is the period of the field. If the definition of the Poynting vector is substituted into Eq. (1.31), one obtain

$$\langle \mathbf{S}(\mathbf{r}) \rangle_T = \frac{1}{2} \text{Re}[\hat{\mathbf{E}}(\mathbf{r}) \times \hat{\mathbf{H}}^*(\mathbf{r})]. \quad (1.45)$$

To obtain the steady-state variant of Eq. (1.43), we take the dot product of the complex conjugate of Eq. (1.32) with  $\hat{\mathbf{E}}$  and add the dot product of Eq. (1.33) with  $\hat{\mathbf{H}}^*$  to obtain

$$\nabla \cdot \langle \mathbf{S}(\mathbf{r}) \rangle_T + \frac{1}{2} \omega \text{Im}[\varepsilon(\mathbf{r})] |\hat{E}(\mathbf{r})|^2 = -\frac{1}{2} \text{Re}[\hat{\mathbf{E}}(\mathbf{r}) \cdot \hat{\mathbf{J}}^{ext*}(\mathbf{r})], \quad (1.46)$$

where we used a complex-valued permittivity, which takes into account the conductivity as described in the previous section.

## 1.6 Two-dimensional electromagnetic fields

Consider a configuration which is assumed to be independent of one variable, say  $y$ . Then the solutions  $\{\hat{\mathbf{E}}(\mathbf{r}_{\parallel}), \hat{\mathbf{H}}(\mathbf{r}_{\parallel})\}$  of the Maxwell equations will also be independent of  $y$ . Here we have introduced the notation  $\mathbf{r}_{\parallel} \equiv (x, 0, z)$ . If this assumption is inserted into the Maxwell equations (1.32) and (1.33), one finds that they split into two independent sets:

$$\partial_z \hat{H}_y(\mathbf{r}_{\parallel}) - i\omega \varepsilon(\mathbf{r}_{\parallel}) \hat{E}_x(\mathbf{r}_{\parallel}) = -\hat{J}_x^{(ext)}(\mathbf{r}_{\parallel}), \quad (1.47)$$

$$-\partial_x \hat{H}_y(\mathbf{r}_{\parallel}) - i\omega \varepsilon(\mathbf{r}_{\parallel}) \hat{E}_z(\mathbf{r}_{\parallel}) = -\hat{J}_z^{(ext)}(\mathbf{r}_{\parallel}), \quad (1.48)$$

$$\partial_z \hat{E}_x(\mathbf{r}_{\parallel}) - \partial_x \hat{E}_z(\mathbf{r}_{\parallel}) - i\omega \mu(\mathbf{r}_{\parallel}) \hat{H}_y(\mathbf{r}_{\parallel}) = 0, \quad (1.49)$$

for  $\hat{E}_x$ ,  $\hat{H}_y$  and  $\hat{E}_z$ . For  $\hat{H}_x$ ,  $\hat{E}_y$  and  $\hat{H}_z$ , one obtains

$$-\partial_z \hat{H}_x(\mathbf{r}_{\parallel}) + \partial_x \hat{H}_z(\mathbf{r}_{\parallel}) - i\omega \varepsilon(\mathbf{r}_{\parallel}) \hat{E}_y(\mathbf{r}_{\parallel}) = -\hat{J}_y^{(ext)}(\mathbf{r}_{\parallel}), \quad (1.50)$$

$$-\partial_z \hat{E}_y(\mathbf{r}_{\parallel}) - i\omega \mu(\mathbf{r}_{\parallel}) \hat{H}_x(\mathbf{r}_{\parallel}) = 0, \quad (1.51)$$

$$\partial_x \hat{E}_y(\mathbf{r}_{\parallel}) - i\omega \mu(\mathbf{r}_{\parallel}) \hat{H}_z(\mathbf{r}_{\parallel}) = 0. \quad (1.52)$$

If  $\hat{E}_x = \hat{H}_y = \hat{E}_z = 0$  the field is called *E-polarized* and if  $\hat{H}_x = \hat{E}_y = \hat{H}_z = 0$  the field is called *H-polarized* [BORN AND WOLF, 1999, p. 638].

An E-polarized field is completely determined by  $\hat{E}_y$ . This can be seen by substituting from Eqs. (1.51) and (1.52) into Eq. (1.51), which yields

$$\nabla^2 \hat{E}_y(\mathbf{r}_{\parallel}) - \frac{\nabla \mu(\mathbf{r}_{\parallel})}{\mu(\mathbf{r}_{\parallel})} \cdot \nabla \hat{E}_y(\mathbf{r}_{\parallel}) + k_0^2 n^2(\mathbf{r}_{\parallel}) \hat{E}_y(\mathbf{r}_{\parallel}) = -i\omega \mu(\mathbf{r}_{\parallel}) \hat{J}_y^{(\text{ext})}(\mathbf{r}_{\parallel}), \quad (1.53)$$

where Eq. (1.35) was used. If  $\hat{E}_y$  is known, the other field components follow from Eqs. (1.51) and (1.52). The boundary conditions at an interface between two different media (Eqs. (1.38) and (1.39)) reduce to the requirement that at the interface

$$\hat{E}_{1y} = \hat{E}_{2y}, \quad (1.54)$$

$$\frac{1}{\mu_1} \partial_n \hat{E}_{1y} = \frac{1}{\mu_2} \partial_n \hat{E}_{2y}, \quad (1.55)$$

where the subscripts 1 and 2 denote the two different media and  $\partial_n \hat{E}_y / \mu = \mathbf{n} \cdot \nabla \hat{E}_y / \mu$ , where  $\mathbf{n}$  is the normal of the interface.

In the same way an H-polarized field is determined by  $\hat{H}_y$ :

$$\nabla^2 \hat{H}_y(\mathbf{r}_{\parallel}) - \frac{\nabla \varepsilon(\mathbf{r}_{\parallel})}{\varepsilon(\mathbf{r}_{\parallel})} \cdot \nabla \hat{H}_y(\mathbf{r}_{\parallel}) + k_0^2 n^2(\mathbf{r}_{\parallel}) \hat{H}_y(\mathbf{r}_{\parallel}) = \partial_x \hat{J}_z^{(\text{ext})}(\mathbf{r}_{\parallel}) - \partial_z \hat{J}_x^{(\text{ext})}(\mathbf{r}_{\parallel}). \quad (1.56)$$

The other field components now follow from Eqs. (1.48) and (1.49), when  $\hat{H}_y$  is known. The boundary conditions for an H-polarized field at a interface between two different media (Eqs. (1.38) and (1.39)) reduce to the requirement that at the interface

$$\hat{H}_{1y} = \hat{H}_{2y}, \quad (1.57)$$

$$\frac{1}{\varepsilon_1} \partial_n \hat{H}_{1y} = \frac{1}{\varepsilon_2} \partial_n \hat{H}_{2y}. \quad (1.58)$$

The reduction of the Maxwell Equations for the case of a two-dimensional configuration to two independent scalar equations (viz. (1.53) and (1.56)) is sometimes called the *scalar nature of two-dimensional electromagnetic fields* [BORN AND WOLF, 1999, Sec. 11.4].

## 1.7 Guided modes

Consider a configuration in which the constitutive parameters ( $\varepsilon$  and  $\mu$ )<sup>5</sup> are independent of  $z$ . A *guided mode* or *waveguide mode* is a field of the following form

$$\hat{\mathbf{E}}(\mathbf{r}) = \mathbf{e}(x, y)e^{ik_{\text{eff}}z}, \quad (1.59)$$

$$\hat{\mathbf{H}}(\mathbf{r}) = \mathbf{h}(x, y)e^{ik_{\text{eff}}z}, \quad (1.60)$$

with  $k_{\text{eff}}$  the *effective wave number* and  $\mathbf{e}$  and  $\mathbf{h}$  represent the profile of the guided mode.<sup>6</sup> Here  $\mathbf{e}$  and  $\mathbf{h}$  are functions such that  $|\mathbf{e}|, |\mathbf{h}| \rightarrow 0$  if  $|x|, |y| \rightarrow \infty$ . An exception are two-dimensional configurations, as discussed in the previous Section. In that case  $\mathbf{e}$  and  $\mathbf{h}$  do not depend on  $y$  at all.

### 1.7.1 Guided modes inside a slit in a metal plate

An example encountered in this thesis is the case of guided modes inside a slit of width  $2a$  in a metal plate. The configuration is non-magnetic (i.e.,  $\mu = \mu_0$ ) and has a permittivity specified by

$$\varepsilon(x) = \begin{cases} \varepsilon_m, & \text{if } |x| > a, \\ \varepsilon_0, & \text{if } |x| \leq a, \end{cases} \quad (1.61)$$

where  $\varepsilon_m$  is the complex-valued permittivity of the metal. The configuration is two-dimensional, and so the field splits into an E-polarized part and an H-polarized part, as described in the previous Section.

For an E-polarized mode, substitution of Eq. (1.59) into Eq. (1.53) yields

$$(\partial_x^2 + k_{0x}^2)\hat{E}_y(\mathbf{r}_{\parallel}) = 0, \quad \text{if } |x| < a, \quad (1.62)$$

$$(\partial_x^2 + k_{mx}^2)\hat{E}_y(\mathbf{r}_{\parallel}) = 0, \quad \text{if } |x| > a, \quad (1.63)$$

with

$$k_{0x} = \sqrt{k_0^2 - k_{\text{eff}}^2} = \sqrt{\omega^2 \varepsilon_0 \mu_0 - k_{\text{eff}}^2}, \quad (1.64)$$

$$k_{mx} = \sqrt{k_m^2 - k_{\text{eff}}^2} = \sqrt{\omega^2 \varepsilon_m \mu_0 - k_{\text{eff}}^2}, \quad (1.65)$$

where the square roots are taken such that  $\text{Im}(\sqrt{k_{0x}}) \geq 0$  and  $\text{Im}(\sqrt{k_{mx}}) \geq 0$ . Eqs. (1.62–1.63) have solutions which are symmetric with respect to  $x$  given by

$$\hat{E}_y(\mathbf{r}_{\parallel}) = \begin{cases} A \cos(k_{0x}x)e^{ik_{\text{eff}}z} & \text{if } |x| < a, \\ B e^{ik_{mx}|x|}e^{ik_{\text{eff}}z} & \text{if } |x| > a, \end{cases} \quad (1.66)$$

<sup>5</sup>Note that  $\varepsilon$  may be complex-valued to take into account losses in the medium.

<sup>6</sup>For an overview of guided modes, see [SNYDER AND LOVE, 1983].

and they have antisymmetric solutions given by

$$\hat{E}_y(\mathbf{r}_{\parallel}) = \begin{cases} C \sin(k_{0x}x)e^{ik_{\text{eff}}z} & \text{if } |x| < a, \\ \text{sign}(x)De^{ik_{mx}|x|}e^{ik_{\text{eff}}z} & \text{if } |x| > a. \end{cases} \quad (1.67)$$

At  $|x| = a$ ,  $\hat{E}_y$  and  $\partial_x \hat{E}_y$  have to be continuous (see Eqs. (1.54–1.55)). This implies that  $k_{\text{eff}}$  has to satisfy

$$-k_{0x} \tan(k_{0x}a) = ik_{mx}, \quad (1.68)$$

for symmetric modes, whereas for antisymmetric modes it has to satisfy

$$k_{0x} \cot(k_{0x}a) = ik_{mx}, \quad (1.69)$$

These two equations are only satisfied by certain discrete values of  $k_{\text{eff}}$ , which correspond with different guided modes. These values can be computed by numerically solving Eq. (1.68) or (1.69).

For an H-polarized mode, one obtains the equations

$$(\partial_x^2 + k_{0x}^2)\hat{H}_y(\mathbf{r}_{\parallel}) = 0, \quad \text{if } |x| < a, \quad (1.70)$$

$$(\partial_x^2 + k_{mx}^2)\hat{H}_y(\mathbf{r}_{\parallel}) = 0, \quad \text{if } |x| > a, \quad (1.71)$$

but now  $\hat{H}_y$  and  $\partial_x \hat{H}_y/\varepsilon$  have to be continuous at  $|x| = a$  (see Eqs. (1.57–1.58)). Eqs. (1.70–1.71) have solutions which are symmetric with respect to  $x$  given by

$$\hat{H}_y(\mathbf{r}_{\parallel}) = \begin{cases} A \cos(k_{0x}x)e^{ik_{\text{eff}}z} & \text{if } |x| < a, \\ B e^{ik_{mx}|x|}e^{ik_{\text{eff}}z} & \text{if } |x| > a, \end{cases} \quad (1.72)$$

and they have antisymmetric solutions given by

$$\hat{H}_y(\mathbf{r}_{\parallel}) = \begin{cases} C \sin(k_{0x}x)e^{ik_{\text{eff}}z} & \text{if } |x| < a, \\ \text{sign}(x)De^{ik_{mx}|x|}e^{ik_{\text{eff}}z} & \text{if } |x| > a. \end{cases} \quad (1.73)$$

The continuity of  $\hat{H}_y$  and  $\partial_x \hat{H}_y$  at  $|x| = a$  implies for symmetric modes that  $k_{\text{eff}}$  has to satisfy

$$-\varepsilon_m k_{0x} \tan(k_{0x}a) = \varepsilon_0 ik_{mx}, \quad (1.74)$$

whereas for antisymmetric modes it has to satisfy

$$\varepsilon_m k_{0x} \cot(k_{0x}a) = \varepsilon_0 ik_{mx}. \quad (1.75)$$

In the case of a slit in a perfectly conducting metal plate, one finds that modes have *cut-off frequencies* [JACKSON, 1999, Sec. 8.3]. This cut-off frequency is a critical frequency  $\omega_c$  below which the mode is *evanescent* ( $\text{Re}(k_{\text{eff}}) = 0$  and

$\text{Im}(k_{\text{eff}}) > 0$ ), whereas for larger frequencies the mode is *propagating* ( $\text{Re}(k_{\text{eff}}) > 0$  and  $\text{Im}(k_{\text{eff}}) = 0$ ).<sup>7</sup> In later Chapters, we will not change the frequency, but instead change the width of the slit. Therefore, we rather work with a *cut-off width*  $w_c$ . In the case that the metal has a finite conductivity, a mode always has a hybrid character, i.e., both  $\text{Re}(k_{\text{eff}}) > 0$  and  $\text{Im}(k_{\text{eff}}) > 0$ . However, the concept of a cut-off width is still meaningful because for  $w > w_c$  one has  $\text{Re}(k_{\text{eff}}) > 0$  and  $\text{Im}(k_{\text{eff}}) \approx 0$ , whereas for  $w < w_c$  one has  $\text{Re}(k_{\text{eff}}) \approx 0$  and  $\text{Im}(k_{\text{eff}}) > 0$ .<sup>8</sup>

### 1.7.2 Surface plasmons

An important special case of a guided mode is a *surface plasmon* (see [RAETHER, 1988]), which is a guided mode of a configuration consisting of a half-space ( $x > 0$ ) consisting of a metal with permittivity  $\varepsilon_m$ , such that  $\text{Re}(\varepsilon_m) < 0$ , and the other half-space ( $x < 0$ ) consisting of a dielectric with permittivity  $\varepsilon_d$ , such that  $\text{Re}(\varepsilon_d) > 0$ . The configuration is again assumed to be invariant in the  $y$ -direction, and therefore the E-polarized and H-polarized parts can be treated separately. It is found that an E-polarized surface plasmon is impossible due to the boundary conditions at  $x = 0$ , as we show later. Therefore a surface plasmon is always H-polarized. Its magnetic field is given by

$$\hat{H}_y = \begin{cases} Ae^{i(k_{\text{sp}}z + k_{mx}x)}, & \text{if } x > 0, \\ Ae^{i(k_{\text{sp}}z - k_{dx}x)}, & \text{if } x < 0, \end{cases} \quad (1.76)$$

where  $A$  is some arbitrary amplitude and we have used the notation  $k_{\text{sp}}$  instead of  $k_{\text{eff}}$ . Furthermore

$$k_{dx} = \sqrt{\omega^2 \varepsilon_d \mu_0 - k_{\text{sp}}^2}, \quad (1.77)$$

$$k_{mx} = \sqrt{\omega^2 \varepsilon_m \mu_0 - k_{\text{sp}}^2}, \quad (1.78)$$

where the square roots are chosen such that  $\text{Im}(k_{dx}) > 0$  and  $\text{Im}(k_{mx}) > 0$ . This implies that the field decays exponentially if one moves away from the interface. The boundary conditions at  $x = 0$ , i.e., the requirement that both  $\hat{H}_y$  and  $\partial_x \hat{H}_y / \varepsilon$  are continuous (see Eqs. (1.57–1.58)) yield the expression

$$k_{\text{sp}} = \omega \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d} \mu_0}. \quad (1.79)$$

<sup>7</sup>However, there is one H-polarized guided mode possible which is propagating for all frequencies and so does not have a cut-off frequency. This is a so-called *TEM-mode* [JACKSON, 1999, p. 358], and is only possible for the H-polarization case.

<sup>8</sup>This behavior can be observed in Fig. 3.3.

The field for a hypothetical E-polarized surface plasmon is given by

$$\hat{E}_y = \begin{cases} Ae^{i(k_{\text{sp}}z + k_{mx}x)}, & \text{if } x > 0, \\ Ae^{i(k_{\text{sp}}z - k_{dx}x)}, & \text{if } x < 0. \end{cases} \quad (1.80)$$

This field already satisfies the continuity requirement of  $\hat{E}_y$ . The requirement that  $\hat{E}_y/\mu$  is continuous at  $x = 0$  yields the condition

$$k_{mx}/\mu_0 = -k_{dx}/\mu_0. \quad (1.81)$$

The imaginary part of the right hand side of this equation is positive, whereas the imaginary part of the left-hand side of the same equation is negative. It follows that Eq. (1.81) can never be satisfied, and that an E-polarized surface plasmon cannot exist.<sup>9</sup>

## 1.8 Phase singularities

Consider a smooth vector field  $V(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , where  $V = (V_x, V_y)$ . We change to polar coordinates by writing  $V_x = \rho \cos(\phi)$  and  $V_y = \rho \sin(\phi)$ , where  $\rho = \rho(x, y)$  is the amplitude and  $\phi = \phi(x, y)$  denotes the phase. A point  $(x, y) \in \mathbb{R}^2$  is called a *phase singularity* if  $\phi$  is not continuous in  $(x, y)$ . In this section some properties of these phase singularities are discussed. Phase singularities can be observed in, e.g., the tides [BERRY, 1981] or the quantum mechanical wave function [HIRSCHFELDER *et al.*, 1974a; HIRSCHFELDER *et al.*, 1974b]. In Optics they are found, e.g., near the edge of a perfectly conducting half-plane [BRAUNBEK AND LAUKIEN, 1952] and near the focus of a convergent beam [BOIVIN *et al.*, 1967]. However, the systematic study of phase singularities did not start until the classic paper of Nye and Berry [1974]. These endeavors resulted in a new branch of optics called *Singular Optics* [NYE, 1999; SOSKIN AND VASNETSOV, 2001; ALLEN *et al.*, 2003].

The discontinuity of  $\phi$  in  $(x, y)$  together with the smoothness of  $V$ , implies that  $V(x, y) = 0$ . Therefore, we first give some examples of points where a smooth vector field  $V$  is zero. In the second subsection we discuss some important indices that can be assigned to phase singularities.

In the third subsection the phase singularities in two-dimensional electromagnetic fields are treated. These can be either phase singularities in the time-averaged Poynting vector, or phase singularities in a field component of the electric or magnetic field. In the latter case we have instead of a real vector field, a complex

---

<sup>9</sup>However, at an interface between two media with different signs of  $\text{Re}(\mu)$ , as is the case if one medium is a so-called “left-handed” material, an E-polarized surface plasmon is indeed possible [RUPPIN, 2000].

field from  $\mathbb{R}^2$  into  $\mathbb{C}$ . The phase  $\phi$  can then be defined by changing to the polar representation  $\rho e^{i\phi}$ . A real vector field  $V(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  and a complex field  $\Psi(x, y) : \mathbb{R}^2 \rightarrow \mathbb{C}$  can be identified by the relations  $\text{Re}(\Psi) = V_x$  and  $\text{Im}(\Psi) = V_y$ . So our discussion of phase singularities in real vector fields is also valid for complex fields.

### 1.8.1 Singular and stationary points

A *singular point* of a smooth vector field  $V(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is defined as a point  $(x, y)$  where  $V(x, y) = 0$ . As already mentioned, a necessary condition for a point  $(x, y)$  to be a phase singularity is that it is singular point, i.e., the amplitude is zero. In Fig. 1.2 a few examples of vector fields with a singular point are given. All vectors are scaled to one. The first vector field (a) in Fig. 1.2 is an example of a singular point which is not a phase singularity. The other five vector fields are the most commonly met singular points.

It becomes clearer whether or not a point is a phase singularity by looking at the equiphase-lines. Whenever the phase  $\phi(x, y)$  equals some constant  $c$ , with  $c \in [0, 2\pi]$ , and  $\nabla\phi(x, y) \neq 0$ , the set

$$W = \{(x', y') \in \mathbb{R}^2 : \phi(x', y') = c\}, \quad (1.82)$$

is locally at  $(x, y)$  a line.<sup>10</sup> We can therefore plot the equiphase-lines. Near a phase singularity the equiphase-lines will (in general) look like as in Fig. 1.3 (a) or (b).

A point  $(x, y)$  with (in contrast to the above case)  $\nabla\phi(x, y) = 0$  is called a *stationary point*. Around a stationary point  $(x, y)$ , with  $\phi(x, y) = c$ , the set  $W$  will in general not be a line at  $(x, y)$ . Note that the Hessian  $\mathcal{H}$  of a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  in a point  $(x, y)$  is given by

$$\mathcal{H}(x, y) = \begin{vmatrix} \partial_x^2 f(x, y) & \partial_x \partial_y f(x, y) \\ \partial_x \partial_y f(x, y) & \partial_y^2 f(x, y) \end{vmatrix}. \quad (1.83)$$

If we assume that the Hessian of  $\phi$  is not equal to zero at  $(x, y)$ , then Morse's lemma can be applied [MARSDEN AND HOFFMAN, 1993, p. 412]. This lemma states that at a stationary point where the Hessian is unequal to zero,  $\phi(x, y)$  locally equals, after a change of coordinates, one of the following functions

$$f_1(x', y') = c + x'^2 + y'^2, \quad (1.84)$$

$$f_2(x', y') = c - x'^2 - y'^2, \quad (1.85)$$

$$f_3(x', y') = c + x'^2 - y'^2. \quad (1.86)$$

<sup>10</sup>By the implicit function theorem, see p. 397 of [MARSDEN AND HOFFMAN, 1993].

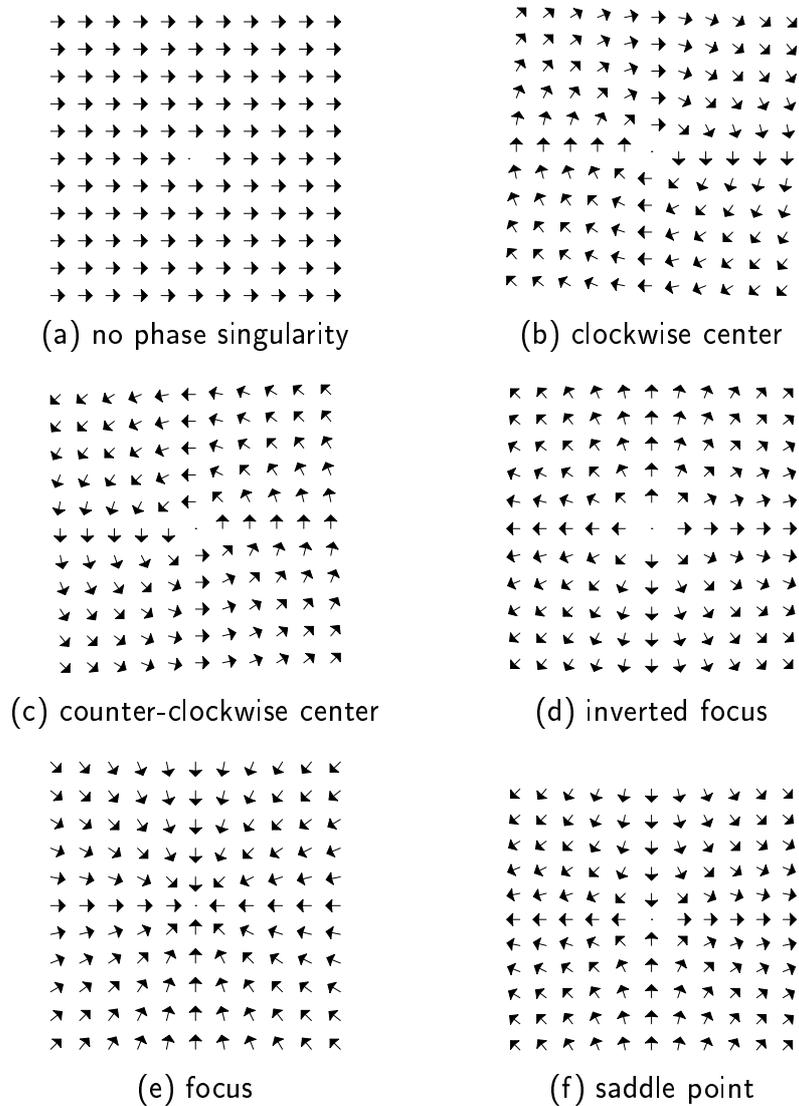


Figure 1.2: Vector fields with singular points. The centers and foci have topological charge 1 (see section 1.8.2), whereas the saddle point (f) has topological charge  $-1$ . The singular point which is not a phase singularity (a) has topological charge zero.

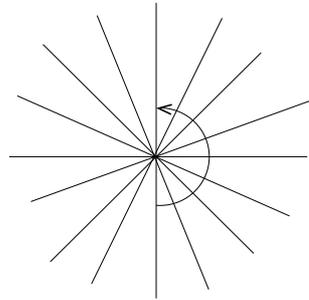
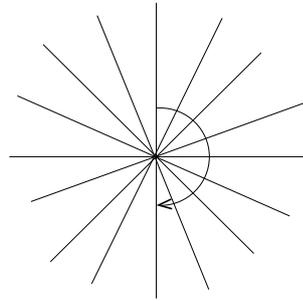
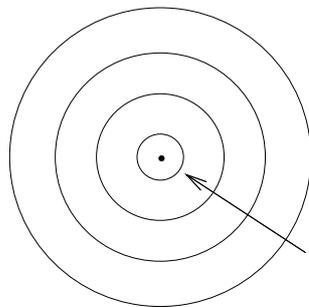
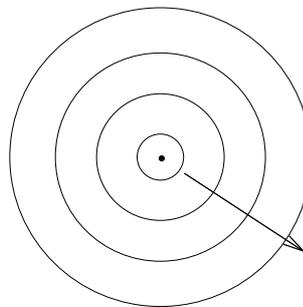
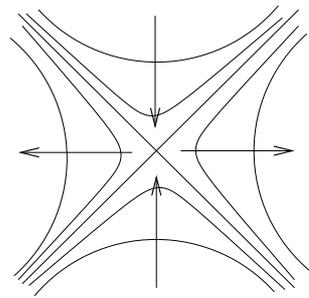
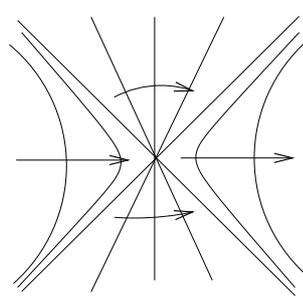
(a) phase singularity with  $s > 0$ (b) phase singularity with  $s < 0$ (c) maximum ( $s = 0$ )(d) minimum ( $s = 0$ )(e) phase saddle ( $s = 0$ )(f) phase singularity with  $s = 0$ 

Figure 1.3: Equiphase-lines around phase singularities and stationary points. The arrows point in the direction of increasing phase.  $s$  is the topological charge, defined in subsection 1.8.2

Hence we obtain that the point  $(x, y)$  is a *minimum* of the phase (corresponding to equation (1.84)) or a *maximum* (corresponding to (1.85)) and so the set  $W$  contains locally only the point  $(x, y)$ , as is depicted in Fig. 1.3 (c) and (d). The third possibility is that the point  $(x, y)$  is a saddle point for the phase, and then the set  $W$  is locally given by two crossing lines (corresponding to Eq. (1.86), see Fig. 1.3) (e). We call a saddle point of the phase a *phase saddle*.

## 1.8.2 Topological charge and index

It is seen from Fig. 1.3 that the equiphase-lines in some cases displays a vortex-like behavior around the phase singularity. With the concept of *topological charge* the change of phase when we go around the phase singularity is measured [NYE AND BERRY, 1974]. More formally, let  $C$  be a closed curve with winding number<sup>11</sup> 1 around a phase singularity  $(x, y)$  of a vector field  $V$  with phase  $\phi$ , such that  $(x, y)$  is the only phase singularity inside  $C$  and there are no phase singularities on  $C$ . Then the topological charge  $s$  of the phase singularity  $(x, y)$  is given by [NYE, 1999, p. 104]

$$s := \frac{1}{2\pi} \oint_C d\phi = \frac{1}{2\pi} \oint_C \nabla\phi \cdot d\mathbf{r}. \quad (1.87)$$

Because the phase is continuous on  $C$ ,  $s$  has an integer value, i.e.,

$$s = 0, \pm 1, \pm 2, \dots \quad (1.88)$$

The topological charge of a point  $(x, y)$  is independent of the choice of the curve  $C$ , as long as it fulfills the conditions mentioned above. This can be seen by realizing that with Stokes' theorem the integral (1.87) gives zero for a curve  $C$  which contains no phase singularities in its interior, and using a proof as in Cauchy's theorem. This also implies that only phase singularities have topological charge. Now consider a closed curve  $C$  with winding number 1, with no phase singularities on  $C$ . Assume that there are  $n$  phase singularities  $(x_1, z_1), \dots, (x_n, z_n)$  with topological charges  $s_1, \dots, s_n$  inside  $C$ , then the integral in Eq. (1.87) is given by

$$s_{tot} = \frac{1}{2\pi} \oint_C d\phi = \sum_{i=1}^n s_i, \quad (1.89)$$

i.e., the sum of the topological charges of the phase singularities lying inside  $C$  is obtained.

An important property of topological charge is that it is conserved under smooth changes of the vector field [NYE AND BERRY, 1974]. This is very important because in many problems the vector field (or complex field) smoothly

<sup>11</sup>The winding number is the number of times that the curve wraps around  $(x, y)$ , measured counter-clockwise [FULTON, 1995, Chap. 3].

depends on the relevant parameters. Then the only way that a phase singularity, with charge unequal to zero, can disappear, is for it to annihilate with other phase singularities such that the total topological charge is zero. Likewise, the only way that a phase singularity, with charge unequal to zero, can be created is together with other phase singularities such that the sum of their topological charges is zero. The most common birth (or annihilation) of phase singularities is when a phase singularity with charge 1 is created together with (or annihilated by) a phase singularity with charge  $-1$ . However, as we shall demonstrate, more complex processes are possible as well.

In Fig. 1.2 the topological charges of some singular points of vector fields are given. An example of a complex field with topological charge  $\pm s$  is  $\Psi(r, \phi) = re^{\pm is\phi}$ , so phase singularities with an arbitrary topological charge exist. However, phase singularities with charges unequal to  $\pm 1$  are seldom seen because in most problems they decay in phase singularities with charges equal to  $\pm 1$ , if some parameter is changed. In other words, phase singularities with charges unequal to  $\pm 1$  are very unstable. An important example of a phase singularity with charge 0 occurs at the creation of two singularities with charges  $\pm 1$ . The equiphase-lines of this kind of phase singularity are given in Fig. 1.3 (f).<sup>12</sup> The *monkey saddle* [HSIUNG, 1981, p. 266] is a phase singularity with topological charge  $-2$ . A monkey saddle is similar to a saddle point, but possesses three attracting and three repulsing directions, rather than two of each.

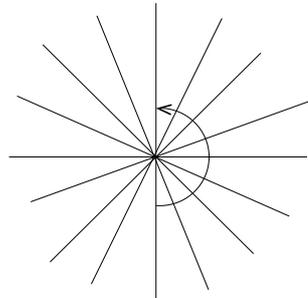
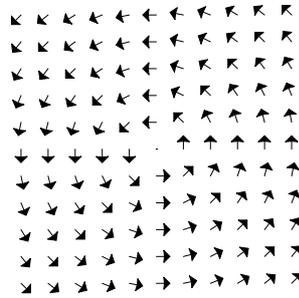
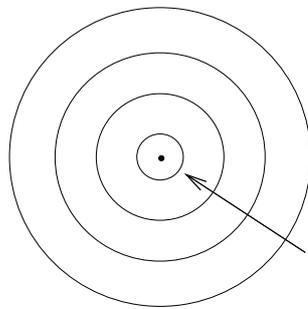
To the phase singularities and the stationary points we can assign a *topological index*<sup>13</sup>  $t$  [NYE *et al.*, 1988], which is defined as the topological charge of the phase singularities of the vector field  $\nabla\phi$ . Around a phase singularity with a positive or negative topological charge, the field  $\nabla\phi$  looks like a counter-clockwise or clockwise center, respectively, see Fig. 1.4 (a) and (b), so the topological index of both a positive and a negative vortex is  $+1$ . Note that this statement, remains true<sup>14</sup> even for phase singularities with charges unequal to  $\pm 1$ , as long as the equiphase-lines have the star-like structure as is shown in Fig. 1.4 (a). For a maximum or minimum for the phase,  $\nabla\phi$  looks like a focus (also called “sink”) or an inverted focus, respectively, and therefore both have a topological index  $t = 1$  (see Fig. 1.4 (c) and (d)). The field  $\nabla\phi$  around a phase saddle forms a saddle point, so the index  $t = -1$  (see Fig. 1.4 (e) and (f)). Because  $\nabla\phi$  is again a vector field, it is possible to define higher order indices [FREUND, 1995], but these are of less importance.

Because the topological index is the topological charge of the vector field  $\nabla\phi$ , it too is conserved under smooth variations of the vector field. The conservation of topological index poses an additional constraint on the creation of phase

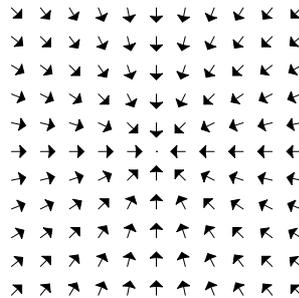
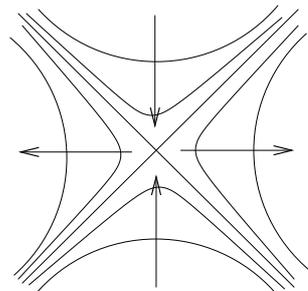
<sup>12</sup>This plot is after [NYE, 1998].

<sup>13</sup>Also called the Poincaré-Hopf index.

<sup>14</sup>However there are exotic phase singularities possible for which this is not true, see [FREUND, 2001].

(a) phase singularity with  $s > 0$ (b)  $\nabla\phi$  around phase singularity with  $s > 0$ 

(c) maximum of phase

(d)  $\nabla\phi$  around maximum

(e) phase saddle

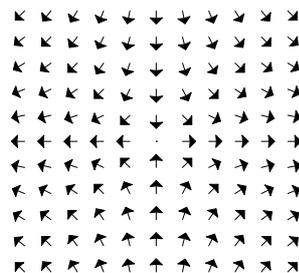
(f)  $\nabla\phi$  around phase saddle

Figure 1.4: Equiphase-lines (left-handed column) and the corresponding vector field  $\nabla\phi$  (right-handed column) around a phase singularity (a), a maximum of the phase (c) and a phase saddle (e). The arrows in (a), (c) and (e) indicate the direction of increasing phase  $\phi$ .

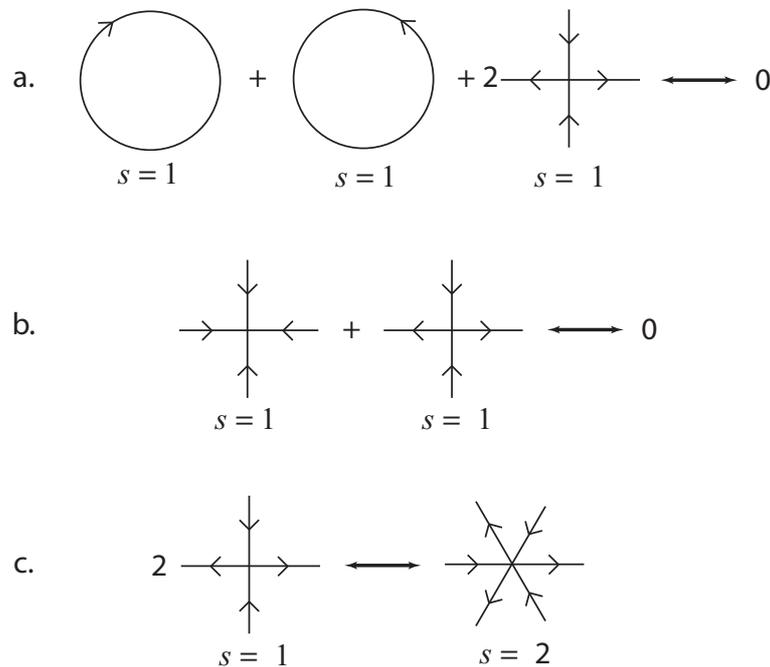


Figure 1.5: Illustrating some of the possible reactions between phase singularities of a vector field: (a) The annihilation (creation) of two vortices of opposite direction and two saddle points; (b) The annihilation (creation) of a saddle point and a sink; (c) The creation (decay) of a monkey saddle out of two saddle points.

singularities: e.g. the birth of a phase singularity with charge 1 (and index 1) and a phase singularity with charge  $-1$  (and index 1), has to be combined with the creation of two phase saddles (each with index  $-1$ ), because otherwise the conservation of topological index would be violated (see Fig. 1.5 (a)). Another possible reaction is the creation of a phase saddle ( $s = 0, t = -1$ ), together with a maximum or minimum of the phase ( $s = 0, t = 1$ ), as is depicted in Fig. 1.5 (b). These are the simplest reactions, there are of course more complicated ones possible. An example is the reaction of a phase singularity with charge 1 ( $s = 1, t = 1$ ) with a phase saddle ( $s = 0, t = -1$ ), which results in two phase singularities with charge 1 (each with  $s = 1, t = 1$ ), one phase singularities with charge  $-1$  ( $s = -1, t = 1$ ) and three phase saddles (each with  $s = 0, t = -1$ ) [BEIJERSBERGEN, 1996; BERRY, 1998; NYE, 1998]. This reaction has been experimentally observed in the focal region of a lens [KARMAN *et al.*, 1997; KARMAN *et al.*, 1998]. Finally, we mention the creation of a monkey saddle out of two saddle points, as is depicted in Fig. 1.5 (c).

### 1.8.3 Phase singularities in two-dimensional electromagnetic waves

In the previous two subsections the mathematics of phase singularities was introduced. Next we discuss phase singularities in two-dimensional electromagnetic waves. The coordinates in this subsection will be indicated as  $\mathbf{r}_{\parallel}$  instead of  $(x, y)$ , as used in the remaining chapters. It is assumed that the configuration is non-magnetic and homogeneous.

The time-averaged Poynting vector for an E-polarized field is given by

$$\langle \mathbf{S}(\mathbf{r}_{\parallel}) \rangle_T = \frac{1}{2} \operatorname{Re} \begin{pmatrix} \hat{E}_y(\mathbf{r}_{\parallel}) \hat{H}_z^*(\mathbf{r}_{\parallel}) \\ 0 \\ -\hat{E}_y(\mathbf{r}_{\parallel}) \hat{H}_x^*(\mathbf{r}_{\parallel}) \end{pmatrix}, \quad (1.90)$$

where Eq. (1.45) was used. For an H-polarized field, the time-averaged Poynting vector is given by

$$\langle \mathbf{S}(\mathbf{r}_{\parallel}) \rangle_T = \frac{1}{2} \operatorname{Re} \begin{pmatrix} -\hat{E}_z(\mathbf{r}_{\parallel}) \hat{H}_y^*(\mathbf{r}_{\parallel}) \\ 0 \\ \hat{E}_x(\mathbf{r}_{\parallel}) \hat{H}_y^*(\mathbf{r}_{\parallel}) \end{pmatrix}. \quad (1.91)$$

If Eqs. (1.51), (1.52), (1.47) and (1.48) are used, one obtains for an E-polarized field,

$$\langle \mathbf{S}(\mathbf{r}_{\parallel}) \rangle_T = -\frac{1}{2\omega\mu_0} \operatorname{Im}(\hat{E}_y(\mathbf{r}_{\parallel}) \nabla \hat{E}_y^*(\mathbf{r}_{\parallel})), \quad (1.92)$$

whereas for an H-polarized field one obtains

$$\langle \mathbf{S} \rangle_T = -\frac{1}{2\omega|\epsilon|^2} \operatorname{Im}(\epsilon \hat{H}_y(\mathbf{r}_{\parallel}) \nabla \hat{H}_y(\mathbf{r}_{\parallel})^*). \quad (1.93)$$

So we see that the energy flow is determined by the scalar field  $\hat{E}_y(\mathbf{r}_{\parallel})$  for an E-polarized field and by the scalar field  $\hat{H}_y(\mathbf{r}_{\parallel})$  for an H-polarized field.

By writing  $\hat{E}_y = |\hat{E}_y| e^{i\phi_E}$ , it is found for an E-polarized field that

$$\langle \mathbf{S}(\mathbf{r}_{\parallel}) \rangle_T = \frac{1}{2\omega\mu_0} |\hat{E}_y(\mathbf{r}_{\parallel})|^2 \nabla \phi_E(\mathbf{r}_{\parallel}), \quad (1.94)$$

and for an H-polarized field in a medium with a real-valued permittivity one obtains that

$$\langle \mathbf{S} \rangle_T = \frac{1}{2\omega\epsilon} |\hat{H}_y(\mathbf{r}_{\parallel})|^2 \nabla \phi_H(\mathbf{r}_{\parallel}), \quad (1.95)$$

where it was used that  $\hat{H}_y = |\hat{H}_y| e^{i\phi_H}$ . Note that the relation between  $\hat{H}_y$  and  $\langle \mathbf{S} \rangle_T$  is not this simple in the case that  $\epsilon$  is a complex number.

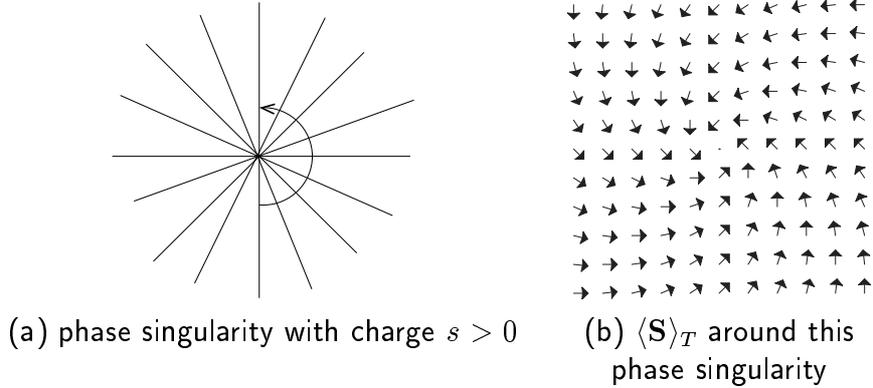


Figure 1.6: Equiphase-lines of  $\hat{H}_y$  (a) and  $\langle \mathbf{S} \rangle_T$  (b) around a phase singularity in a medium with a complex-valued permittivity with a positive real part.

Now consider a phase singularity in  $\langle \mathbf{S} \rangle_T$  for an E-polarized field. Eq. (1.94) shows that this will either be a phase singularity in  $\hat{E}_y$  or a point where  $\nabla\phi_E = 0$ . The same equation also shows that the topological *charge* of  $\langle \mathbf{S} \rangle_T$  equals the topological *index* of  $\hat{E}_y$ . This can be seen as follows: the direction of  $\langle \mathbf{S} \rangle_T$  is along  $\nabla\phi_E$  and the topological *index* of  $\hat{E}_y$  is defined as the topological *charge* of  $\nabla\phi_E$  (see the discussion in subsection 1.8.2). Therefore a phase singularity of the field  $\hat{E}_y$  with positive topological charge corresponds to a counter-clockwise center in  $\langle \mathbf{S} \rangle_T$ , whereas a phase singularity in  $\hat{E}_y$  with a negative topological charge corresponds to a clockwise center in  $\langle \mathbf{S} \rangle_T$ ,<sup>15</sup> see Fig. 1.4 (a) and (b). A maximum or a minimum of the phase of  $\hat{E}_y$  corresponds to a focus or an inverted focus in  $\langle \mathbf{S} \rangle_T$ , respectively, see Fig. 1.4 (c) and (d). A phase saddle in the phase of  $\hat{E}_y$ , finally, corresponds to a saddle point in  $\langle \mathbf{S} \rangle_T$ , see Fig. 1.4 (e) and (f).

As can be seen from equation (1.93) a phase singularity in  $\hat{H}_y$ , even when  $\varepsilon$  is complex, still corresponds to a phase singularity in  $\langle \mathbf{S} \rangle_T$ . Note that this will not be exactly a center, but will in general be more spiral-like, i.e., something between a center and a focus, see Fig. 1.6. This can be shown by a direct computation of equation (1.93), when for  $\hat{H}_y$  the field  $x \pm iz$  is taken, which describes a phase singularity with charge  $\pm 1$  at the origin.<sup>16</sup> If  $\text{Re}(\varepsilon) > 0$  the spirals are, like the

<sup>15</sup>See also [TOTZECK AND TIZIANI, 1997b]

<sup>16</sup>The reader might argue that this field component does not satisfy the Helmholtz equation. However, because it is only the linearized form (around the zero) of a solution of the Helmholtz equation, it does *not* have to satisfy the Helmholtz equation. The following example illustrates this: take  $f(x, z) = e^{ikz} + 2 \cos(kx)$ , which is simply the sum of three plane waves. It is easy to check that this scalar field is a solution of the Helmholtz equation. The singular points of  $f$  are solutions of the equations  $\sin(kz) = 0$  and  $\cos(kz) + 2 \cos(kx) = 0$ . Take e.g. the point  $(0, \arccos(-0.5)/k)$ . The linearized form of  $f$  around this point is  $ikz - \sqrt{3}kx$ , which shows that

centers, counter-clockwise if the topological charge of  $\hat{H}_y$  is 1, and clockwise if the topological charge of  $\hat{H}_y$  is  $-1$ . If  $\text{Re}(\varepsilon) < 0$ , the spirals are clockwise if the topological charge of  $\hat{H}_y$  is 1, and counter-clockwise if the topological charge of  $\hat{H}_y$  is  $-1$ . The spirals will look more like a center, the more  $\text{Im}(\varepsilon)$  is smaller compared to  $\text{Re}(\varepsilon)$ . If  $\text{Im}(\varepsilon)$  is large compared to  $\text{Re}(\varepsilon)$  the spiral will look more like a focus.

Eq. (1.86) shows that the field  $e^{i(x^2-z^2)}$  can be taken as the local approximation for a phase saddle in  $\hat{H}_y$  at origin.<sup>17</sup> By using Eq. (1.95) it is easy to show that the field of power flow exhibits then a saddle point. In the same way, one can locally approximate  $\hat{H}_y$  by  $e^{\pm i(x^2+z^2)}$  in the case that its phase has a maximum or a minimum at the origin. Eq.(1.95) shows then that this corresponds with a focus or an inverted focus, depending on the sign  $\text{Re}(\varepsilon)$ .

An important example that we will study in Chapter 3 is the following. Due to the conservation of topological charge of  $\hat{E}_y$  and  $\hat{H}_y$ , the birth of a phase singularity with charge 1, a phase singularity with charge  $-1$  and two phase saddles in  $\hat{E}_y$  or  $\hat{H}_y$ , corresponds to the birth of a clockwise and a counter-clockwise center (or spiral) with two saddle points in the time-averaged Poynting vector (see Fig.1.5 (a)).

A minimum in the phase of  $\hat{E}_y$  or  $\hat{H}_y$  corresponds to an inverted focus (or “source”) in  $\langle \mathbf{S} \rangle_T$ . Due to energy conservation this is not possible outside the region where there are electromagnetic sources (see equation (1.46)). So, wherever such sources are absent, neither  $\hat{E}_y$  or  $\hat{H}_y$  can have a minimum in the phase. Likewise, if there are no losses, i.e.,  $\text{Im}(\varepsilon) = 0$ , then there are also no maxima in the phase of  $\hat{E}_y$  or  $\hat{H}_y$ , because these correspond to foci in  $\langle \mathbf{S} \rangle_T$ .

In a medium with  $\text{Im}(\varepsilon) > 0$ , i.e., a medium with absorption, it is possible that a focus in  $\langle \mathbf{S} \rangle_T$  occurs, so also maxima in the phase of  $\hat{E}_y$  or  $\hat{H}_y$  are possible. In Chapter 3 we will present examples of this. In this case a reaction that can occur is the annihilation of a phase saddle with a maximum of the phase of  $\hat{E}_y$  or  $\hat{H}_y$ . For the time-averaged Poynting vector this corresponds to the annihilation of a focus with a saddle point (see Fig. 1.5 (b)). An example of this reaction will also be given in Chapter 3.

---

the singular point is indeed a phase singularity with charge  $-1$ . Note that the linearized form does not satisfy the Helmholtz equation.

<sup>17</sup>This is only valid if  $\nabla \hat{H}_y = 0$ , because in the case of a complex permittivity it is no longer true that  $\nabla \phi = 0$  implies that  $\nabla \hat{H}_y = 0$ .



# Chapter 2

## The Green's Tensor Formalism

This Chapter is based on the following publication:

- H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “An efficient numerical method for solving electromagnetic domain integral equations”, to be submitted.

### **Abstract**

We present an efficient numerical technique to obtain the (time-harmonic) electromagnetic field in configurations in which a two-dimensional scattering structure is embedded in a stratified, non-magnetic medium. This is accomplished by numerically solving the domain integral equation for the electric field inside the scatterer. The kernel of this integral equation is a Green's tensor with respect to the stratified embedding medium. By exploiting the symmetry properties of this tensor we are able to significantly improve the efficiency of this method.

## 2.1 Introduction

Green's tensor techniques are commonly used to compute the electromagnetic field for a wide variety of problems. Examples are the modeling of near-field optical microscopes [DEREUX *et al.*, 1991; GIRARD AND DEREUX, 1996], the analysis of channel and ridge waveguides in stratified media [KOLK *et al.*, 1990; BAKEN *et al.*, 1990], the scattering of light by non-spherical interstellar particles [PURCELL AND PENNYPACKER, 1973; DRAINE, 1988], the transmission of light through a sub-wavelength slit [SCHOUTEN *et al.*, 2003a; SCHOUTEN *et al.*, 2003b], the scattering of surface plasmons by rough surfaces [MARADUDIN AND MILLS, 1975; MILLS AND MARADUDIN, 1975], the modeling of optical lithography [MARTIN *et al.*, 1998; PAULUS *et al.*, 2001] and the analysis of semiconductor laser amplifiers [VISSER *et al.*, 1999]. Although there are different variants of the Green's tensor technique, they all have in common that only the scattering body has to be discretized, and no computation window is needed. Other advantages of this method are that complex-valued and anisotropic dielectric constants are allowed, and even gain media can be modelled with it. Quite often one deals with scatterers that are embedded in a stratified medium. The Green's tensor technique can then still be used, but now with a Green's tensor with respect to the stratified configuration. Such a tensor can be derived analytically in an angular spectrum representation [MILLS AND MARADUDIN, 1975; TSANG *et al.*, 1975; REED *et al.*, 1987; TOMAŠ, 1995; VISSER *et al.*, 1999; PAULUS *et al.*, 2000].

A commonly used variant of the Green's tensor technique converts the Maxwell equations into an integral equation over the scattering domain. This integral equation is sometimes called the *domain integral equation*. In this Chapter we present a numerical technique for solving the domain integral equation for the case of a stratified configuration. The integral equation is converted into a linear system of equations by using the collocation method. This system is then solved with a variant of the conjugate gradient method. In this way we are able to use the symmetry properties of the stratified embedding configuration to significantly reduce the requirements for data storage and computation time.

In Section 2.2 the scattering configuration is described, and the domain integral equations are derived. In Section 2.3 the derivation of the Green's tensor with respect to a layered configuration is presented. Section 2.4 deals with the collocation method. The conjugate gradient method for solving the resulting linear system is explained in Section 2.5. The exploitation of the symmetries of the embedding structure, together with the use of the fast Fourier transform (FFT), are shown to significantly improve the numerical performance. In Appendix 2.A the field due to a plane wave incident on a stratified medium is calculated.

## 2.2 The scattering model

The configuration at hand is a stratified “background” medium in which a scatterer, which occupies a bounded volume  $\mathcal{D}$ , is embedded (see Fig. 2.1). The structure is invariant in the  $y$ -direction and the materials that make up the configuration are assumed to be nonmagnetic. The *background structure* is defined as the embedding structure *without* the scatterer. It is stratified in the  $z$ -direction and invariant in the  $x$ - and  $y$ -directions, and is characterized by its permittivity  $\varepsilon_b(z)$  which is given by

$$\varepsilon_b(z) = \varepsilon_i \quad \text{if } z_{i-1} \leq z < z_i, \quad (2.1)$$

with  $i = 1, \dots, N$ , see Fig. 2.2. Here  $\varepsilon_i$  is the permittivity of the  $i$ th layer,  $N$  is the number of layers, and  $z = z_i$  indicates the position of the interface between layer  $i$  and layer  $i + 1$ . Therefore,  $z_{i-1} < z_i$  for all  $i$ ,  $z_0 = -\infty$  and  $z_N = \infty$ . The actual configuration consists of this background configuration with a scattering volume  $\mathcal{D}$  characterized by its permittivity  $\varepsilon(\mathbf{r}_{\parallel})$ , where  $\mathbf{r}_{\parallel} = (x, 0, z)$  (see Sec. 1.6. This scattering volume is assumed to be bounded in the  $x$  and  $z$ -directions), and invariant in the  $y$ -direction. Later, in Sections 2.4 and 2.5, we make use of the assumption that the scatterer consist of  $m$  homogeneous “blocks”, i.e.,

$$\mathcal{D} = \bigcup_{j=1}^m \mathcal{D}_j = \bigcup_{j=1}^m (x_j^-, x_j^+) \times (z_j^-, z_j^+), \quad (2.2)$$

where  $x_j^-$  and  $x_j^+$  are the lower and upper bound in the  $x$ -direction of block  $j$ ,  $z_j^-$  and  $z_j^+$  are the lower and upper bound in the  $z$ -direction of block  $j$ . Furthermore,  $\varepsilon(\mathbf{r}_{\parallel}) = \varepsilon^{(j)}$  if  $\mathbf{r}_{\parallel} \in \mathcal{D}_j$ . Also, we assume that each block  $\mathcal{D}_j$  lies in only one layer, i.e., for all  $j = 1, \dots, m$ , there is an integer  $i$ , which will depend on  $j$ , such that  $z_{i-1} \leq z_j^- < z_j^+ \leq z_i$ .

The configuration is illuminated by a monochromatic plane wave with time-dependence  $\exp(-i\omega t)$ , where  $\omega$  denotes the angular frequency. The total electric field  $\hat{\mathbf{E}}(\mathbf{r}_{\parallel})$  and the total magnetic field  $\hat{\mathbf{H}}(\mathbf{r}_{\parallel})$  are written as the sum of the incident field and a scattered field, i.e.,

$$\hat{\mathbf{E}}(\mathbf{r}_{\parallel}) = \hat{\mathbf{E}}^{(\text{inc})}(\mathbf{r}_{\parallel}) + \hat{\mathbf{E}}^{(\text{sca})}(\mathbf{r}_{\parallel}), \quad (2.3)$$

$$\hat{\mathbf{H}}(\mathbf{r}_{\parallel}) = \hat{\mathbf{H}}^{(\text{inc})}(\mathbf{r}_{\parallel}) + \hat{\mathbf{H}}^{(\text{sca})}(\mathbf{r}_{\parallel}). \quad (2.4)$$

The incident field is defined as the solution of the steady-state Maxwell equations for the background configuration, i.e.,

$$-\nabla \times \hat{\mathbf{H}}^{(\text{inc})}(\mathbf{r}_{\parallel}) - i\omega\varepsilon_b(z)\hat{\mathbf{E}}^{(\text{inc})}(\mathbf{r}_{\parallel}) = 0, \quad (2.5)$$

$$\nabla \times \hat{\mathbf{E}}^{(\text{inc})}(\mathbf{r}_{\parallel}) - i\omega\mu_0\hat{\mathbf{H}}^{(\text{inc})}(\mathbf{r}_{\parallel}) = 0. \quad (2.6)$$

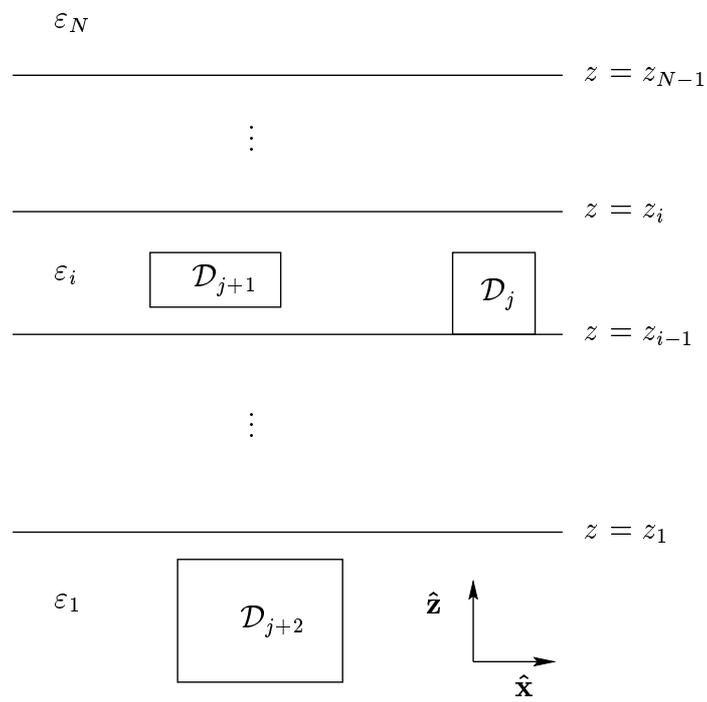


Figure 2.1: The actual configuration: a scattering volume  $\mathcal{D}$ , consisting of  $m$  blocks  $\mathcal{D}_1, \dots, \mathcal{D}_m$ , embedded in a stratified medium. Only three blocks are drawn.

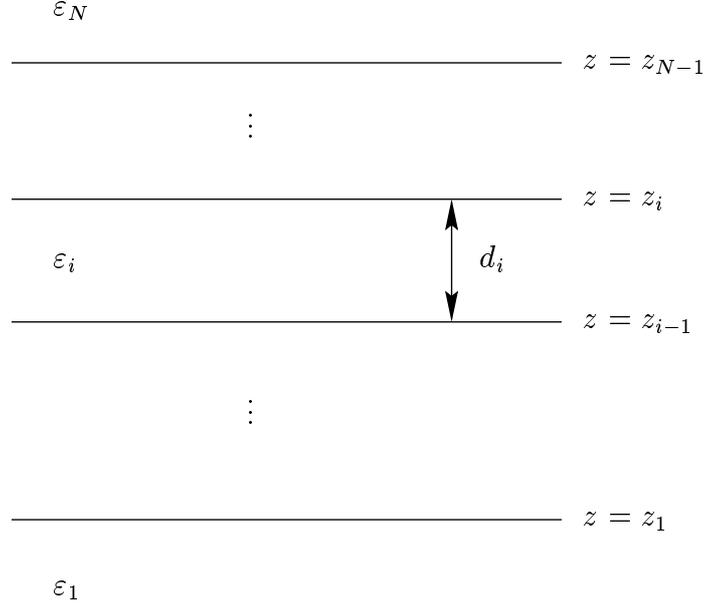


Figure 2.2: The background configuration: a stratified medium.

The incident field is produced by sources far away from the structure, which implies that these sources do not have to be taken into account explicitly. It can be calculated analytically by a recursive procedure, as is explained in Appendix 2.A.

Maxwell's equations for the total field (with respect to the actual configuration) can be written as

$$-\nabla \times \hat{\mathbf{H}}(\mathbf{r}_{\parallel}) - i\omega\varepsilon_b(z)\hat{\mathbf{E}}(\mathbf{r}_{\parallel}) = -\hat{\mathbf{J}}^{(\text{con})}(\mathbf{r}_{\parallel}), \quad (2.7)$$

$$\nabla \times \hat{\mathbf{E}}(\mathbf{r}_{\parallel}) - i\omega\mu_0\hat{\mathbf{H}}(\mathbf{r}_{\parallel}) = 0, \quad (2.8)$$

where the contrast (or polarization) current density  $\hat{\mathbf{J}}^{(\text{con})}(\mathbf{r}_{\parallel})$  is defined by

$$\hat{\mathbf{J}}^{(\text{con})}(\mathbf{r}_{\parallel}) = -i\omega\Delta\varepsilon(\mathbf{r}_{\parallel})\hat{\mathbf{E}}(\mathbf{r}_{\parallel}), \quad (2.9)$$

with  $\Delta\varepsilon(\mathbf{r}_{\parallel}) = \varepsilon(\mathbf{r}_{\parallel}) - \varepsilon_b(z)$  for points  $\mathbf{r}_{\parallel} \in \mathcal{D}$  and  $\Delta\varepsilon(\mathbf{r}_{\parallel}) = 0$  otherwise. Subtracting Eqs. (2.5) and (2.6) from Eqs. (2.7) and (2.8) yields

$$-\nabla \times \hat{\mathbf{H}}^{(\text{sca})}(\mathbf{r}_{\parallel}) - i\omega\varepsilon_b(z)\hat{\mathbf{E}}^{(\text{sca})}(\mathbf{r}_{\parallel}) = -\hat{\mathbf{J}}^{(\text{con})}(\mathbf{r}_{\parallel}), \quad (2.10)$$

$$\nabla \times \hat{\mathbf{E}}^{(\text{sca})}(\mathbf{r}_{\parallel}) - i\omega\mu_0\hat{\mathbf{H}}^{(\text{sca})}(\mathbf{r}_{\parallel}) = 0. \quad (2.11)$$

The *electric Green's tensor*  $\mathbf{G}^E$  and the *magnetic Green's tensor*  $\mathbf{G}^H$  are defined by the expressions

$$-\nabla \times \mathbf{G}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) - i\omega\varepsilon_b(z)\mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = -\mathbf{l}\delta(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel}), \quad (2.12)$$

$$\nabla \times \mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) - i\omega\mu_0\mathbf{G}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = 0. \quad (2.13)$$

Here  $\mathbf{l}$  is the  $3 \times 3$  unit tensor and  $\delta(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel})$  is the two-dimensional Dirac delta function. Note that  $G_{ij}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel})$  [ $G_{ij}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel})$ ] is the  $i$ -th component of the electric (magnetic) field in the background configuration at  $\mathbf{r}_{\parallel}$  due to a point current source located at  $\mathbf{r}'_{\parallel}$  and pointing in the  $j$ -direction, with  $i, j = x, y, z$ .

Using the definition of the electric and magnetic Green's tensors, the solution of Eqs. (2.10) and (2.11) can be written as

$$\hat{\mathbf{E}}^{(\text{sca})}(\mathbf{r}_{\parallel}) = \int_{\mathcal{D}} \mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) \cdot \hat{\mathbf{J}}^{(\text{con})}(\mathbf{r}'_{\parallel}) d^2r'_{\parallel}, \quad (2.14)$$

$$\hat{\mathbf{H}}^{(\text{sca})}(\mathbf{r}_{\parallel}) = \int_{\mathcal{D}} \mathbf{G}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) \cdot \hat{\mathbf{J}}^{(\text{con})}(\mathbf{r}'_{\parallel}) d^2r'_{\parallel}. \quad (2.15)$$

For points  $\mathbf{r}_{\parallel} \notin \mathcal{D}$ , these equations can be verified by inserting them into Eqs. (2.10) and (2.11), and using Eqs. (2.12) and (2.13). However, for  $\mathbf{r}_{\parallel} \in \mathcal{D}$ , Eq. (2.14), contrary to Eq. (2.15), is not valid in the classical function sense [YAGHJIAN, 1980; CHEW, 1989]. The reason is that the integral, which is singular at  $\mathbf{r}'_{\parallel} = \mathbf{r}_{\parallel}$ , is not convergent. However, if one excludes from the integration domain a “principal volume” around the point  $\mathbf{r}'_{\parallel} = \mathbf{r}_{\parallel}$ , the resulting integral is found to converge. To correct for the exclusion of the principal volume, one has to add a source tensor to the right-hand side of Eq. (2.14), which depends on the geometry of the exclusion volume [YAGHJIAN, 1980]. However, Eq. (2.14) is valid in a distributional function sense, which is clearly observed if one derives the electric Green's tensor in an angular spectrum representation (as is done in the next Section). In that case no source tensor term needs to be added, because it is already included in the spectral representation of the electric Green's tensor [CHEW, 1989].<sup>1</sup>

If Eqs. (2.14) and (2.15) are substituted in Eqs. (2.3–2.4), one obtains

$$\hat{\mathbf{E}}(\mathbf{r}_{\parallel}) = \hat{\mathbf{E}}^{(\text{inc})}(\mathbf{r}_{\parallel}) - i\omega \int_{\mathcal{D}} \Delta\varepsilon(\mathbf{r}'_{\parallel})\mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) \cdot \hat{\mathbf{E}}(\mathbf{r}'_{\parallel}) d^2r'_{\parallel}, \quad (2.16)$$

$$\hat{\mathbf{H}}(\mathbf{r}_{\parallel}) = \hat{\mathbf{H}}^{(\text{inc})}(\mathbf{r}_{\parallel}) - i\omega \int_{\mathcal{D}} \Delta\varepsilon(\mathbf{r}'_{\parallel})\mathbf{G}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) \cdot \hat{\mathbf{E}}(\mathbf{r}'_{\parallel}) d^2r'_{\parallel}, \quad (2.17)$$

where Eq. (2.9) was used. These equations are sometimes called the *domain integral equations*. For points  $\mathbf{r}_{\parallel} \in \mathcal{D}$ , Eq. (2.16) is a Fredholm equation of the second

<sup>1</sup>For a more detailed discussion about the singular behavior of the Green's tensor, see Chapter 7 of [CHEW, 1995] or Chapter 3 of [VAN BLADEL, 2000].

kind for the electric field. If the solution of this integral equation is found, one can use Eqs. (2.16) and (2.17) to obtain both the electric and the magnetic field at an arbitrary point in space.

## 2.3 The derivation of the Green's tensors

In this Section the Green's tensors for a stratified medium are derived. The electric Green's tensor is subject to the equation

$$-\nabla \times [\nabla \times \mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel})] + \omega^2 \mu_0 \varepsilon_b(z) \mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = -i\omega \mu_0 \mathbf{l} \delta(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel}), \quad (2.18)$$

which can be derived by substituting from Eq. (2.13) into Eq. (2.12). The magnetic Green's tensor can be obtained from the electric Green's tensor, by using Eq. (2.13). In subsection 2.3.1 the Green's tensors with respect to a homogeneous background are derived.<sup>2</sup> In subsection 2.3.2 these tensors are used to obtain the Green's tensors pertaining to a stratified medium.

### 2.3.1 The Green's tensors for a homogeneous medium

Consider a homogeneous background with permittivity  $\varepsilon_b(z) = \varepsilon$  and permeability  $\mu_0$ . The Fourier transform of the electric Green's tensor with respect to  $\mathbf{r}_{\parallel}$  is defined as

$$\tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) = \int_{-\infty}^{\infty} \mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) e^{i\mathbf{k}_{\parallel} \cdot \mathbf{r}_{\parallel}} d^2 r_{\parallel}, \quad (2.19)$$

where  $\mathbf{k}_{\parallel} = (k_x, 0, k_z)$  and with the inverse transform given by

$$\mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = \left( \frac{1}{2\pi} \right)^2 \int_{-\infty}^{\infty} \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) e^{-i\mathbf{k}_{\parallel} \cdot \mathbf{r}_{\parallel}} d^2 k_{\parallel}. \quad (2.20)$$

On taking the Fourier transform, Eq. (2.18) reduces to

$$\mathbf{k}_{\parallel} \times [\mathbf{k}_{\parallel} \times \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel})] + \omega^2 \mu_0 \varepsilon \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) = -i\omega \mu_0 \mathbf{l} e^{i\mathbf{k}_{\parallel} \cdot \mathbf{r}'_{\parallel}}. \quad (2.21)$$

Using the identity  $\mathbf{k}_{\parallel} \times [\mathbf{k}_{\parallel} \times \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel})] = \mathbf{k}_{\parallel} \mathbf{k}_{\parallel} \cdot \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) - k_{\parallel}^2 \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel})$ , this equation can be converted into

$$[(k_{\parallel}^2 - k^2) \mathbf{l} - \mathbf{k}_{\parallel} \mathbf{k}_{\parallel}] \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) = i\omega \mu_0 \mathbf{l} e^{i\mathbf{k}_{\parallel} \cdot \mathbf{r}'_{\parallel}}, \quad (2.22)$$

with  $k^2 = \omega^2 \varepsilon \mu_0$  and  $k_{\parallel} = |\mathbf{k}_{\parallel}|$ . Note that a tensor  $\mathbf{C} = \mathbf{a}\mathbf{b}$  has components given by  $C_{ij} = a_i b_j$ . To solve this matrix equation, we make use of the following

<sup>2</sup>This derivation is based on [CHEW, 1989].

identity, which can be verified by direct substitution: if  $\mathbf{A}$  is a  $n \times n$  matrix such that  $\mathbf{A}^2 = \beta\mathbf{A}$ , then

$$(\alpha\mathbf{I} - \mathbf{A})^{-1} = \frac{1}{\alpha}\left(\mathbf{I} + \frac{1}{\alpha - \beta}\mathbf{A}\right), \quad (2.23)$$

with  $\alpha$  such that  $\beta \neq \alpha \neq 0$ . Now  $(\mathbf{k}_{\parallel}\mathbf{k}_{\parallel})^2 = k_{\parallel}^2\mathbf{k}_{\parallel}\mathbf{k}_{\parallel}$ , so if Eq. (2.23) is applied to Eq. (2.22) with  $\mathbf{A} = \mathbf{k}_{\parallel}\mathbf{k}_{\parallel}$ ,  $\alpha = k^2 - k_{\parallel}^2$  and  $\beta = k_{\parallel}^2$ , one obtains

$$\tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) = i\omega\mu_0 \frac{|k^2 - \mathbf{k}_{\parallel}\mathbf{k}_{\parallel}|}{k^2(k_{\parallel}^2 - k^2)} e^{i\mathbf{k}_{\parallel}\cdot\mathbf{r}'_{\parallel}}. \quad (2.24)$$

If Eq. (2.24) is substituted into Eq. (2.20), one finds

$$\mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = \frac{i\omega\mu_0}{(2\pi)^2} \int_{-\infty}^{\infty} \frac{|k^2 - \mathbf{k}_{\parallel}\mathbf{k}_{\parallel}|}{k^2(k_{\parallel}^2 - k^2)} e^{i\mathbf{k}_{\parallel}\cdot(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel})} d^2k_{\parallel}, \quad (2.25)$$

where the transformation  $\mathbf{k}_{\parallel} \rightarrow -\mathbf{k}_{\parallel}$  has been made.

The magnetic Green's tensor can easily be derived by taking the Fourier transform of Eq. (2.13), which leads to

$$i\mathbf{k}_{\parallel} \times \tilde{\mathbf{G}}^E(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) - i\omega\mu_0 \tilde{\mathbf{G}}^H(\mathbf{k}_{\parallel}, \mathbf{r}'_{\parallel}) = 0. \quad (2.26)$$

It follows that

$$\mathbf{G}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = \frac{i}{(2\pi)^2} \int_{-\infty}^{\infty} \frac{\mathbf{k}_{\parallel} \times \mathbf{l}}{k_{\parallel}^2 - k^2} e^{i\mathbf{k}_{\parallel}\cdot(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel})} d^2k_{\parallel}. \quad (2.27)$$

In the following section we will need these Green's tensors in an *angular spectrum representation* [MANDEL AND WOLF, 1995]. Therefore the integration over  $k_z$  in Eq. (2.25) is explicitly performed. First note that one component of the integrand does *not* tend to zero in the limit  $k_z \rightarrow \infty$ , i.e.,

$$\lim_{k_z \rightarrow \infty} \frac{|k^2 - \mathbf{k}_{\parallel}\mathbf{k}_{\parallel}|}{k^2(k_{\parallel}^2 - k^2)} = -\frac{\hat{\mathbf{z}}\hat{\mathbf{z}}}{k^2}, \quad (2.28)$$

with  $\hat{\mathbf{z}}$  the unit vector in the  $z$ -direction. Therefore the integral in Eq. (2.25) is split into two parts, viz.

$$\begin{aligned} \mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) &= \frac{i\omega\mu_0}{(2\pi)^2} \int_{-\infty}^{\infty} \left[ \frac{|k^2 - \mathbf{k}_{\parallel}\mathbf{k}_{\parallel}|}{k^2(k_{\parallel}^2 - k^2)} + \frac{\hat{\mathbf{z}}\hat{\mathbf{z}}}{k^2} \right] e^{i\mathbf{k}_{\parallel}\cdot(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel})} d^2k_{\parallel} \\ &+ \frac{\hat{\mathbf{z}}\hat{\mathbf{z}}}{i\omega\varepsilon(2\pi)^2} \int_{-\infty}^{\infty} e^{i\mathbf{k}_{\parallel}\cdot(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel})} d^2k_{\parallel}. \end{aligned} \quad (2.29)$$

The second integral in the equation above is just the Fourier representation of the two-dimensional Dirac delta function. The first integral can be evaluated by use of Jordan's lemma and Cauchy's theorem. Therefore one obtains the residue contributions of the poles  $k_z = \pm\sqrt{k^2 - k_x^2}$  in the complex  $k_z$  plane, where the plus sign applies for  $z - z' > 0$ , and the minus sign for  $z - z' < 0$ . Finally, this yields

$$\mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = \frac{-\omega\mu_0}{4\pi} \int_{-\infty}^{\infty} \left( \frac{|k^2 - \mathbf{k}^S \mathbf{k}^S}{k^2 k_z} \right) e^{ik_x(x-x') + ik_z|z-z'|} dk_x + \frac{\hat{\mathbf{z}}\hat{\mathbf{z}}}{i\omega\varepsilon} \delta(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel}), \quad (2.30)$$

where  $k_z = \sqrt{k^2 - k_x^2}$ , chosen such that  $\text{Im}(k_z) \geq 0$  and  $\mathbf{k}^S = (k_x, 0, Sk_z)$ , with  $S = \text{sign}(z - z')$ . If we introduce the notation  $\hat{\mathbf{e}}^S = \mathbf{k}^S \times \hat{\mathbf{y}}/k$ , the identity tensor  $\mathbf{l}$  can be written as

$$\mathbf{l} = \mathbf{k}^S \mathbf{k}^S / k^2 + \hat{\mathbf{y}}\hat{\mathbf{y}} + \hat{\mathbf{e}}^S \hat{\mathbf{e}}^S, \quad (2.31)$$

where it was used that a tensor  $\hat{a}\hat{a}$  is a projection operator if  $\hat{a}$  is a unit vector and that  $\mathbf{k}^S$ ,  $\hat{\mathbf{y}}$  and  $\hat{\mathbf{e}}^S$  are three mutually orthogonal vectors. By substitution of the identity (2.31), Eq. (2.30) can be rewritten as

$$\mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = \frac{-\omega\mu_0}{4\pi} \int_{-\infty}^{\infty} \frac{1}{k_z} (\hat{\mathbf{y}}\hat{\mathbf{y}} + \hat{\mathbf{e}}^S \hat{\mathbf{e}}^S) e^{i\mathbf{k}^S \cdot (\mathbf{x}_{\parallel} - \mathbf{x}'_{\parallel})} dk_x + \mathbf{G}^{(\text{sin})}(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}), \quad (2.32)$$

with  $\hat{\mathbf{e}}^S = \mathbf{k}^S \times \hat{\mathbf{y}}/k$  and the singular part of the tensor is given by

$$\mathbf{G}^{(\text{sin})}(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = \frac{\hat{\mathbf{z}}\hat{\mathbf{z}}}{i\omega\varepsilon} \delta(\mathbf{r}_{\parallel} - \mathbf{r}'_{\parallel}). \quad (2.33)$$

The first term within the brackets of Eq. (2.32) represents the E-polarized part of the Green's tensors, whereas the second term within the brackets represents the H-polarized part.

The angular spectrum representation for the magnetic Green's tensor follows similarly as

$$\mathbf{G}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = \frac{-k}{4\pi} \int_{-\infty}^{\infty} \frac{1}{k_z} (\hat{\mathbf{e}}^S \hat{\mathbf{y}} - \hat{\mathbf{y}} \hat{\mathbf{e}}^S) e^{i\mathbf{k}^S \cdot (\mathbf{x}_{\parallel} - \mathbf{x}'_{\parallel})} dk_x. \quad (2.34)$$

The same splitting into an E-polarized and an H-polarized part can be observed as in the electric Green's tensor given by Eq. (2.32). Note that in contrast to the electric Green's tensor, the magnetic Green's tensor does not have a singular part. This is related to the different kind of singular behavior of these tensors at  $\mathbf{r} = \mathbf{r}'$  (see the discussion below Eq. (2.15)).

### 2.3.2 The Green's tensor for a layered medium

In the preceding subsection the Green's tensors for a homogeneous background were derived in the form of an angular spectrum of plane waves. In the case of a layered medium, the Green's tensors consist of this source term plus a term that describes the reflections and transmissions of the field at the interfaces between the layers. This extra term can be written as an angular spectrum representation too, i.e., we write the Green's tensor as a sum of plane waves with different, unknown, coefficients. Therefore, if the source point  $\mathbf{r}'_{\parallel}$  is located in layer  $s$  and the observation point  $\mathbf{r}_{\parallel}$  is located in layer  $i$ , then the electric Green's tensor can be written as

$$\begin{aligned} \mathbf{G}^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = & \frac{-\omega\mu_0}{4\pi} \int_{-\infty}^{\infty} \frac{1}{k_{sz}} \left[ (\hat{\mathbf{y}}\mathbf{A}_i^E + \hat{\mathbf{e}}_i^+\mathbf{A}_i^H) e^{ik_{iz}(z-z_{i-1})} e^{ik_x x} + \right. \\ & (\hat{\mathbf{y}}\mathbf{B}_i^E + \hat{\mathbf{e}}_i^-\mathbf{B}_i^H) e^{ik_{iz}(z_i-z)} e^{ik_x x} + \\ & \left. \delta_{is}(\hat{\mathbf{y}}\hat{\mathbf{y}} + \hat{\mathbf{e}}_s^S\hat{\mathbf{e}}_s^S) e^{i\mathbf{k}^S \cdot (\mathbf{x}_{\parallel} - \mathbf{x}'_{\parallel})} \right] dk_x \\ & + \mathbf{G}^{(\text{sin})}(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}). \end{aligned} \quad (2.35)$$

Here we have introduced coefficient vectors  $\mathbf{A}_i^E$ ,  $\mathbf{A}_i^H$ ,  $\mathbf{B}_i^E$  and  $\mathbf{B}_i^H$  for  $i = 1, \dots, N$ , which represent the amplitudes of the upgoing or downgoing, E-polarized or H-polarized, plane waves in the different layers. These coefficients are yet undetermined. The magnetic Green's tensor is then given by

$$\begin{aligned} \mathbf{G}^H(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) = & \frac{-k_i}{4\pi} \int_{-\infty}^{\infty} \frac{1}{k_{sz}} \left[ (\hat{\mathbf{e}}_i^+\mathbf{A}_i^E - \hat{\mathbf{y}}\mathbf{A}_i^H) e^{ik_{iz}(z-z_{i-1})} e^{ik_x x} + \right. \\ & (\hat{\mathbf{e}}_i^-\mathbf{B}_i^E - \hat{\mathbf{y}}\mathbf{B}_i^H) e^{ik_{iz}(z_i-z)} e^{ik_x x} + \\ & \left. \delta_{is}(\hat{\mathbf{e}}_i^S\hat{\mathbf{y}} - \hat{\mathbf{y}}\hat{\mathbf{e}}_i^S) e^{i\mathbf{k}^S \cdot (\mathbf{x}_{\parallel} - \mathbf{x}'_{\parallel})} \right] dk_x. \end{aligned} \quad (2.36)$$

Due to the mode decomposition, the Green's tensors already satisfy the Maxwell equations. The only thing that has to be done is that the coefficient vectors  $\mathbf{A}_i^E$ ,  $\mathbf{A}_i^H$ ,  $\mathbf{B}_i^E$  and  $\mathbf{B}_i^H$  have to be chosen in such a way that the boundary conditions at  $z_i$  for  $i = 1, \dots, N-1$  are satisfied. Also note that  $\mathbf{A}_1^E$ ,  $\mathbf{A}_1^H$ ,  $\mathbf{B}_N^E$  and  $\mathbf{B}_N^H$  are all zero, because the field cannot generate waves coming in from infinity.

To obtain the coefficient vectors, we note that the plane waves with different values of  $k_x$  generated by the source, can be treated independently. We now introduce two complementary configurations. First, consider the configuration consisting of the layers 1 until  $s$ , and assume that the source layer is now filling the half-space  $z > z_{s-1}$ , i.e., it is a medium consisting of  $s$  layers. If a plane wave is incident from above, all the coefficients of the plane waves within this medium can be determined with the aid of the method described in the Appendix 2.A. These

fields are denoted by  $\{\hat{\mathbf{E}}_-^E(\mathbf{r}_\parallel), \hat{\mathbf{H}}_-^E(\mathbf{r}_\parallel)\}$  and  $\{\hat{\mathbf{E}}_-^H(\mathbf{r}_\parallel), \hat{\mathbf{H}}_-^H(\mathbf{r}_\parallel)\}$ , where the first is for the E-polarized case, and the second for the H-polarized case. The electric fields are given by

$$\hat{\mathbf{E}}_-^E(\mathbf{r}_\parallel) = \begin{cases} (\hat{\mathbf{y}}e^{ik_{sz}(z_{s-1}-z)} + \alpha_s^E \hat{\mathbf{y}}e^{ik_{sz}(z-z_{s-1})})e^{ik_x x}, & z > z_{s-1}, \\ (\alpha_i^E \hat{\mathbf{y}}e^{ik_{iz}(z-z_{i-1})} + \beta_i^E \hat{\mathbf{y}}e^{ik_{iz}(z_i-z)})e^{ik_x x}, & z_{i-1} < z < z_i, \\ \beta_1^E \hat{\mathbf{y}}e^{ik_{1z}(z_1-z)}e^{ik_x x}, & z_1 > z, \end{cases} \quad (2.37)$$

and

$$\hat{\mathbf{E}}_-^H(\mathbf{r}_\parallel) = \begin{cases} (\hat{\mathbf{e}}_s^- e^{ik_{sz}(z_{s-1}-z)} + \alpha_s^H \hat{\mathbf{e}}_s^+ e^{ik_{sz}(z-z_{s-1})})e^{ik_x x}, & z > z_{s-1}, \\ (\alpha_i^H \hat{\mathbf{e}}_i^+ e^{ik_{iz}(z-z_{i-1})} + \beta_i^H \hat{\mathbf{e}}_i^- e^{ik_{iz}(z_i-z)})e^{ik_x x}, & z_{i-1} < z < z_i, \\ \beta_1^H \hat{\mathbf{e}}_1^- e^{ik_{1z}(z_1-z)}e^{ik_x x}, & z_1 > z, \end{cases} \quad (2.38)$$

with  $i = 3, \dots, s-1$ . Similarly, we introduce a configuration consisting of the layers  $s$  until  $N$ , but now with the source layer filling the half-space  $z < z_s$ . So, this configuration consists of  $N-s+1$  layers. Again, the fields  $\{\hat{\mathbf{E}}_+^E(\mathbf{r}_\parallel), \hat{\mathbf{H}}_+^E(\mathbf{r}_\parallel)\}$  and  $\{\hat{\mathbf{E}}_+^H(\mathbf{r}_\parallel), \hat{\mathbf{H}}_+^H(\mathbf{r}_\parallel)\}$  consisting of E-polarized or H-polarized plane waves incident from below on this medium can be computed by the method given in the Appendix 2.A, and are given by

$$\hat{\mathbf{E}}_+^E(\mathbf{r}_\parallel) = \begin{cases} (\hat{\mathbf{y}}e^{ik_{sz}(z-z_s)} + \eta_s^E \hat{\mathbf{y}}e^{ik_{sz}(z_s-z)})e^{ik_x x}, & z < z_s, \\ (\zeta_i^E \hat{\mathbf{y}}e^{ik_{iz}(z-z_{i-1})} + \eta_i^E \hat{\mathbf{y}}e^{ik_{iz}(z_i-z)})e^{ik_x x}, & z_{i-1} < z < z_i, \\ \zeta_N^E \hat{\mathbf{y}}e^{ik_{Nz}(z-z_{N-1})}, & z_{N-1} < z, \end{cases} \quad (2.39)$$

and

$$\hat{\mathbf{E}}_+^H(\mathbf{r}_\parallel) = \begin{cases} (\hat{\mathbf{e}}_s^+ e^{ik_{sz}(z-z_s)} + \eta_s^H \hat{\mathbf{e}}_s^- e^{ik_{sz}(z_s-z)})e^{ik_x x}, & z < z_s, \\ (\zeta_i^H \hat{\mathbf{e}}_i^+ e^{ik_{iz}(z-z_{i-1})} + \eta_i^H \hat{\mathbf{e}}_i^- e^{ik_{iz}(z_i-z)})e^{ik_x x}, & z_{i-1} < z < z_i, \\ \zeta_N^H \hat{\mathbf{e}}_N^+ e^{ik_{Nz}(z-z_{N-1})}, & z_{N-1} < z, \end{cases} \quad (2.40)$$

with  $i = s+1, \dots, N-1$ .

The fields introduced above do satisfy all the boundary conditions at the interfaces  $z = z_i$ . Therefore, the Green's tensors can also be written as

$$\mathbf{G}^E(\mathbf{r}, \mathbf{r}_\parallel) = \frac{-\omega\mu_0}{4\pi} \begin{cases} \int_{-\infty}^{\infty} \frac{1}{k_{sz}} \left[ \hat{\mathbf{E}}_-^E(\mathbf{r}_\parallel) \mathbf{C}_-^E(\mathbf{r}'_\parallel) + \hat{\mathbf{E}}_-^H(\mathbf{r}_\parallel) \mathbf{C}_-^H(\mathbf{r}'_\parallel) \right] dk_x, & z < z', \\ \int_{-\infty}^{\infty} \frac{1}{k_{sz}} \left[ \hat{\mathbf{E}}_+^E(\mathbf{r}_\parallel) \mathbf{C}_+^E(\mathbf{r}'_\parallel) + \hat{\mathbf{E}}_+^H(\mathbf{r}_\parallel) \mathbf{C}_+^H(\mathbf{r}'_\parallel) \right] dk_x, & z > z', \end{cases} \quad (2.41)$$

and

$$\mathbf{G}^H(\mathbf{r}, \mathbf{r}_{\parallel}) = \frac{-k_i}{4\pi} \begin{cases} \int_{-\infty}^{\infty} \frac{1}{k_{sz}} \left[ \hat{\mathbf{H}}_-^E(\mathbf{r}_{\parallel}) \mathbf{C}_-^E(\mathbf{r}'_{\parallel}) + \hat{\mathbf{H}}_-^H(\mathbf{r}_{\parallel}) \mathbf{C}_-^H(\mathbf{r}'_{\parallel}) \right] dk_x, & z < z', \\ \int_{-\infty}^{\infty} \frac{1}{k_{sz}} \left[ \hat{\mathbf{H}}_+^E(\mathbf{r}_{\parallel}) \mathbf{C}_+^E(\mathbf{r}'_{\parallel}) + \hat{\mathbf{H}}_+^H(\mathbf{r}_{\parallel}) \mathbf{C}_+^H(\mathbf{r}'_{\parallel}) \right] dk_x, & z > z', \end{cases} \quad (2.42)$$

where  $\mathbf{C}_{\pm}^E(\mathbf{r}_{\parallel})$  and  $\mathbf{C}_{\pm}^H(\mathbf{r}_{\parallel})$  are vectors yet to be determined. These Green's tensors satisfy the Maxwell's equations for  $z \neq z'$  and satisfy all the boundary conditions at  $z = z_i$ . Furthermore, it is consistent with Eq. (2.35) in all layers, except for layer  $s$  which requires a separate treatment, if the coefficients are chosen as

$$\mathbf{A}_i^E(\mathbf{r}'_{\parallel}) = \begin{cases} \alpha_i^E \mathbf{C}_-^E(\mathbf{r}'_{\parallel}), & i < s, \\ \zeta_i^E \mathbf{C}_+^E(\mathbf{r}'_{\parallel}), & i > s, \end{cases} \quad (2.43)$$

and

$$\mathbf{B}_i^E(\mathbf{r}'_{\parallel}) = \begin{cases} \beta_i^E \mathbf{C}_-^E(\mathbf{r}'_{\parallel}), & i < s, \\ \eta_i^E \mathbf{C}_+^E(\mathbf{r}'_{\parallel}), & i > s. \end{cases} \quad (2.44)$$

To have consistency between Eq.(2.41) and Eq. (2.35) in the source layer, one obtains for the E-polarized components the conditions

$$\mathbf{A}_s^E(\mathbf{r}'_{\parallel}) e^{ik_{sz}(z-z_{s-1})} + \hat{\mathbf{y}} e^{ik_{sz}(z-z')} e^{-ik_x x'} = \mathbf{C}_+^E(\mathbf{r}'_{\parallel}) e^{ik_{sz}(z-z_s)}, \quad (2.45)$$

$$\mathbf{B}_s^E(\mathbf{r}'_{\parallel}) e^{ik_{sz}(z_s-z)} = \mathbf{C}_+^E(\mathbf{r}'_{\parallel}) \alpha_s^E e^{ik_{sz}(z_s-z)}, \quad (2.46)$$

$$\mathbf{A}_s^E(\mathbf{r}'_{\parallel}) e^{ik_{sz}(z-z_{s-1})} = \mathbf{C}_-^E(\mathbf{r}'_{\parallel}) \eta_s^E e^{ik_{sz}(z-z_{s-1})}, \quad (2.47)$$

$$\mathbf{B}_s^E(\mathbf{r}'_{\parallel}) e^{ik_{sz}(z_s-z)} + \hat{\mathbf{y}} e^{ik_{sz}(z'-z)} e^{-ik_x x'} = \mathbf{C}_-^E(\mathbf{r}'_{\parallel}) e^{ik_{sz}(z_{s-1}-z)}. \quad (2.48)$$

These equations have the solution

$$\mathbf{C}_+^E(\mathbf{r}'_{\parallel}) = \hat{\mathbf{y}} e^{-ik_x x'} \frac{e^{ik_{sz}(z_s-z')} + \eta_s^E \gamma_s e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}, \quad (2.49)$$

$$\mathbf{C}_-^E(\mathbf{r}'_{\parallel}) = \hat{\mathbf{y}} e^{-ik_x x'} \frac{\alpha_s^E \gamma_s e^{ik_{sz}(z_s-z')} + e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}, \quad (2.50)$$

with  $\gamma_s = \exp(2ik_{sz}d_s)$  and

$$\mathbf{A}_s^E(\mathbf{r}'_{\parallel}) = \eta_s^E \hat{\mathbf{y}} e^{-ik_x x'} \frac{\alpha_s^E \gamma_s e^{ik_{sz}(z_s-z')} + e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}, \quad (2.51)$$

$$\mathbf{B}_s^E(\mathbf{r}'_{\parallel}) = \alpha_s^E \hat{\mathbf{y}} e^{-ik_x x'} \frac{e^{ik_{sz}(z_s-z')} + \eta_s^E \gamma_s e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}. \quad (2.52)$$

For the H-polarized components, the following conditions are obtained

$$\mathbf{A}_s^H(\mathbf{r}'_{\parallel})e^{ik_{sz}(z-z_{s-1})} + \hat{\mathbf{e}}_s^+ e^{ik_{sz}(z-z')}e^{-ik_x x'} = \mathbf{C}_+^H(\mathbf{r}'_{\parallel})e^{ik_{sz}(z-z_s)}, \quad (2.53)$$

$$\mathbf{B}_s^H(\mathbf{r}'_{\parallel})e^{ik_{sz}(z_s-z)} = \mathbf{C}_+^H(\mathbf{r}'_{\parallel})\alpha_s^H e^{ik_{sz}(z_s-z)}, \quad (2.54)$$

$$\mathbf{A}_s^H(\mathbf{r}'_{\parallel})e^{ik_{sz}(z-z_{s-1})} = \mathbf{C}_-^H(\mathbf{r}'_{\parallel})\eta_s^H e^{ik_{sz}(z-z_{s-1})}, \quad (2.55)$$

$$\mathbf{B}_s^H(\mathbf{r}'_{\parallel})e^{ik_{sz}(z_s-z)} + \hat{\mathbf{e}}_s^- e^{ik_{sz}(z'-z)}e^{-ik_x x'} = \mathbf{C}_-^H(\mathbf{r}'_{\parallel})e^{ik_{sz}(z_{s-1}-z)}, \quad (2.56)$$

with as the solution

$$\mathbf{C}_+^H(\mathbf{r}'_{\parallel}) = e^{-ik_x x'} \frac{\hat{\mathbf{e}}_s^+ e^{ik_{sz}(z_s-z')} + \eta_s^E \gamma_s \hat{\mathbf{e}}_s^- e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}, \quad (2.57)$$

$$\mathbf{C}_-^H(\mathbf{r}'_{\parallel}) = e^{-ik_x x'} \frac{\alpha_s^E \gamma_s \hat{\mathbf{e}}_s^+ e^{ik_{sz}(z_s-z')} + \hat{\mathbf{e}}_s^- e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}, \quad (2.58)$$

and

$$\mathbf{A}_s^H(\mathbf{r}'_{\parallel}) = \eta_s^H e^{-ik_x x'} \frac{\alpha_s^E \gamma_s \hat{\mathbf{e}}_s^+ e^{ik_{sz}(z_s-z')} + \hat{\mathbf{e}}_s^- e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}, \quad (2.59)$$

$$\mathbf{B}_s^H(\mathbf{r}'_{\parallel}) = \alpha_s^H e^{-ik_x x'} \frac{\hat{\mathbf{e}}_s^+ e^{ik_{sz}(z_s-z')} + \eta_s^E \gamma_s \hat{\mathbf{e}}_s^- e^{ik_{sz}(z'-z_{s-1})}}{1 - \eta_s^E \alpha_s^E \gamma_s^2}. \quad (2.60)$$

This completely determines the Green's tensor.

In the following Sections it is convenient to have the electric Green's tensor rewritten in the following form, which expresses the spatial dependence more explicitly,

$$\begin{aligned} \mathbf{G}^E(\mathbf{r}, \mathbf{r}_{\parallel}) = & \int_{-\infty}^{\infty} \left[ \mathbf{u}_1(k_x) e^{i(k_{iz}z - k_{sz}z')} + \mathbf{u}_2(k_x) e^{i(k_{iz}z + k_{sz}z')} \right. \\ & + \mathbf{d}_1(k_x) e^{-i(k_{iz}z - k_{sz}z')} + \mathbf{d}_2(k_x) e^{-i(k_{iz}z + k_{sz}z')} \\ & \left. + \delta_{is} \mathbf{s}(k_x) e^{ik_{sz}|z-z'|} \right] e^{ik_x(x-x')} dk_x \\ & + \delta_{is} \mathbf{G}^{(\text{sin})}(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}), \end{aligned} \quad (2.61)$$

where the tensors  $\mathbf{u}_1$ ,  $\mathbf{u}_2$ ,  $\mathbf{d}_1$ ,  $\mathbf{d}_2$  and  $\mathbf{s}$  are defined by

$$\begin{aligned} \mathbf{u}_1 = & \begin{cases} -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\zeta_i^E e^{ik_{sz}z_s}/D_E + \hat{\mathbf{e}}_i^+ \hat{\mathbf{e}}_s^+ \zeta_i^H e^{ik_{sz}z_s}/D_H), & i > s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\alpha_i^E \alpha_s^E \gamma_s e^{ik_{sz}z_s}/D_E + \hat{\mathbf{e}}_i^+ \hat{\mathbf{e}}_s^+ \alpha_i^H \alpha_s^H \gamma_s e^{ik_{sz}z_s}/D_H), & i < s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\eta_s^E \alpha_s^E \gamma_s e^{ik_{sz}z_s}/D_E + \hat{\mathbf{e}}_s^+ \hat{\mathbf{e}}_s^+ \eta_s^H \alpha_s^H \gamma_s e^{ik_{sz}z_s}/D_H), & i = s, \end{cases} \quad (2.62) \\ \mathbf{u}_2 = & \begin{cases} -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\zeta_i^E \eta_s^E \gamma_s e^{-ik_{sz}z_{s-1}}/D_E + \hat{\mathbf{e}}_i^+ \hat{\mathbf{e}}_s^- \zeta_i^H \eta_s^H \gamma_s e^{-ik_{sz}z_{s-1}}/D_H), & i > s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\alpha_i^E e^{-ik_{sz}z_{s-1}}/D_E + \hat{\mathbf{e}}_i^+ \hat{\mathbf{e}}_s^- \alpha_i^H e^{-ik_{sz}z_{s-1}}/D_H), & i < s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\eta_s^E e^{-ik_{sz}z_{s-1}}/D_E + \hat{\mathbf{e}}_s^+ \hat{\mathbf{e}}_s^- \eta_s^H e^{-ik_{sz}z_{s-1}}/D_H), & i = s, \end{cases} \quad (2.63) \end{aligned}$$

$$\mathbf{d}_1 = \begin{cases} -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\eta_i^E \eta_s^E \gamma_s e^{-ik_{sz}z_{s-1}}/D_E + \hat{\mathbf{e}}_i^- \hat{\mathbf{e}}_s^- \eta_i^H \eta_s^H \gamma_s e^{-ik_{sz}z_{s-1}}/D_H), & i > s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\beta_i^E e^{-ik_{sz}z_{s-1}}/D_E + \hat{\mathbf{e}}_i^- \hat{\mathbf{e}}_s^- \beta_i^H e^{-ik_{sz}z_{s-1}}/D_H), & i < s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\alpha_s^E \eta_s^E \gamma_s e^{-ik_{sz}z_{s-1}}/D_E + \hat{\mathbf{e}}_s^- \hat{\mathbf{e}}_s^- \alpha_s^H \eta_s^H \gamma_s e^{-ik_{sz}z_{s-1}}/D_H), & i = s, \end{cases} \quad (2.64)$$

$$\mathbf{d}_2 = \begin{cases} -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\eta_i^E e^{ik_{sz}z_s}/D_E + \hat{\mathbf{e}}_i^- \hat{\mathbf{e}}_s^+ \eta_i^H e^{ik_{sz}z_s}/D_H), & i > s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\beta_i^E \alpha_s^E \gamma_s e^{ik_{sz}z_s}/D_E + \hat{\mathbf{e}}_i^- \hat{\mathbf{e}}_s^+ \beta_i^H \alpha_s^H \gamma_s e^{ik_{sz}z_s}/D_H), & i < s, \\ -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}}\alpha_s^E e^{ik_{sz}z_s}/D_E + \hat{\mathbf{e}}_s^- \hat{\mathbf{e}}_s^+ \alpha_s^H e^{ik_{sz}z_s}/D_H), & i = s, \end{cases} \quad (2.65)$$

and

$$\mathbf{s} = -\frac{\omega\mu_0}{4\pi k_{sz}} (\hat{\mathbf{y}}\hat{\mathbf{y}} + \hat{\mathbf{e}}_s^S \hat{\mathbf{e}}_s^S). \quad (2.66)$$

## 2.4 The collocation method for solving the domain integral equation

In the collocation method<sup>3</sup> the solution  $\hat{\mathbf{E}}(\mathbf{r}_{\parallel})$  of the integral equation (2.16), with  $\mathbf{r}_{\parallel} \in \mathcal{D}$ , is approximated by a function of the form

$$\check{\mathbf{E}}(\mathbf{r}_{\parallel}) = \sum_{i=1}^p \mathbf{a}_i u_i(\mathbf{r}_{\parallel}), \quad (2.67)$$

where  $u_1, \dots, u_p$  are linear independent functions and  $\mathbf{a}_1, \dots, \mathbf{a}_p$  are complex coefficients vectors. To obtain conditions for the coefficients, a residual  $\mathbf{R}$  is defined as follows

$$\begin{aligned} \mathbf{R}(\mathbf{r}_{\parallel}) := & \check{\mathbf{E}}(\mathbf{r}_{\parallel}) - \hat{\mathbf{E}}^{(\text{inc})}(\mathbf{r}_{\parallel}) \\ & + i\omega \int_{\mathcal{D}} \Delta\varepsilon(\mathbf{r}'_{\parallel}) G^E(\mathbf{r}_{\parallel}, \mathbf{r}'_{\parallel}) \check{\mathbf{E}}(\mathbf{r}'_{\parallel}) d^2r'_{\parallel}. \end{aligned} \quad (2.68)$$

This residual will be unequal to zero, unless  $\check{\mathbf{E}} \equiv \hat{\mathbf{E}}$ , i.e., the approximation equals exactly the actual solution (see Eq. (2.16)). In the collocation method the residual  $\mathbf{R}$  is required to be zero at  $p$  points,  $\mathbf{r}_{\parallel}^{(1)}, \dots, \mathbf{r}_{\parallel}^{(p)}$ , which have to be chosen in such a way that the linear system  $\{u_i(\mathbf{r}_{\parallel}^{(j)})\}$  has full rank. From this one obtains the following  $3p$  linear equations for the  $3p$  coefficients  $a_{1x}, \dots, a_{pz}$ ,

$$\sum_{i=1}^p \left[ u_i(\mathbf{r}_{\parallel}^{(j)}) \mathbf{l} + i\omega \mathbf{C}_{ij} \right] \cdot \mathbf{a}_i = \hat{\mathbf{E}}^{(\text{inc})}(\mathbf{r}_{\parallel}^{(j)}), \quad (2.69)$$

<sup>3</sup>See p. 54 of [ATKINSON, 1976]. This method is sometimes presented as a special form of the Method of Moments, see p. 9 of [HARRINGTON, 1999].

with  $j = 1, \dots, p$  and

$$\mathbf{C}_{ij} = \int_{\mathcal{D}} \Delta \varepsilon(\mathbf{r}'_{\parallel}) \mathbf{G}^E(\mathbf{r}'_{\parallel}, \mathbf{r}'_{\parallel}) u_i(\mathbf{r}'_{\parallel}) d^2 r'_{\parallel}. \quad (2.70)$$

The simplest choice for the basis  $u_1, \dots, u_p$  are piecewise-constant functions.<sup>4</sup> That is, for every block  $\mathcal{D}_i$  (with  $i = 1, \dots, m$ ), we choose a suitable partition,  $\mathcal{D}_i^1, \dots, \mathcal{D}_i^{p_i}$ , with  $p_i$  positive numbers such that

$$p = \sum_{i=1}^m p_i. \quad (2.71)$$

It is convenient to introduce a double index for denoting the basis function: we define for every  $i = 1, \dots, p$ , two unique integers  $i_1$  and  $i_2$  such that

$$i = i_2 + \sum_{j=1}^{i_1-1} p_j, \quad (2.72)$$

with  $i_1 = 1, \dots, m$  numbering the blocks, and  $i_2 = 1, \dots, p_{i_1}$  numbering the particular basis function within that block. The partition is chosen in such a way that  $\mathcal{D}_i$  is divided into  $n_{ix} \times n_{iz}$  (with  $p_i = n_{ix} n_{iz}$ ) rectangular pieces with sizes  $\Delta_{ix} = (x_i^+ - x_i^-)/n_{ix}$  and  $\Delta_{iz} = (z_i^+ - z_i^-)/n_{iz}$ . The numbering of the elements  $\mathcal{D}_i$  is showed in Fig. 2.3 Now we define

$$u_i(\mathbf{r}_{\parallel}) = \begin{cases} 1 & \text{if } \mathbf{r}_{\parallel} \in \mathcal{D}_{i_1}^{i_2}, \\ 0 & \text{otherwise.} \end{cases} \quad (2.73)$$

In this case the integral in Eq. (2.70) is given by

$$\mathbf{C}_{ij} = \int_{\mathcal{D}_{i_1}^{i_2}} (\varepsilon^{(i_1)} - \varepsilon_{i_1}) \mathbf{G}^E(\mathbf{r}_{\parallel}^{(j)}, \mathbf{r}_{\parallel}) d^2 r'_{\parallel}, \quad (2.74)$$

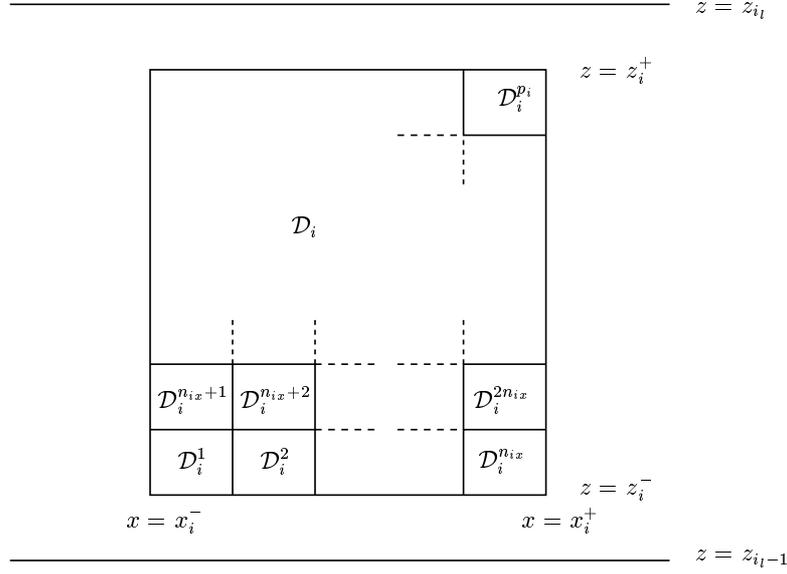
where  $\mathbf{r}_{\parallel}^{(j)}$  is chosen at the baricenter of  $\mathcal{D}_{i_1}^{j_2}$  and  $i_l$  denotes the layer in which  $\mathcal{D}_{i_1}$  is embedded.

Now the matrix elements  $\mathbf{C}_{ij}$  have to be computed. Therefore Eq. (2.61) is substituted into Eq. (2.74) and, with the help of Fubini's theorem, the order of integration is interchanged to get

$$\mathbf{C}_{ij} = \int_{-\infty}^{\infty} (\varepsilon^{(i_1)} - \varepsilon_{i_1}) \mathbf{T}_{ij}(k_x) e^{ik_x(x^{(j)} - x^{(i)})} dk_x + \delta_{ij} \frac{\hat{\mathbf{z}}\hat{\mathbf{z}}}{i\omega\varepsilon_{i_1}}, \quad (2.75)$$

---

<sup>4</sup>This method has been applied in the context of the Green's tensor method in, e.g., [VISSER *et al.*, 1999].

Figure 2.3: Illustration of the discretization of the block  $\mathcal{D}_i$ .

where

$$\begin{aligned} \Gamma_{ij}(k_x) = & \frac{4}{k_x k_{i_1 z}} \sin(k_x \Delta_{i_1 x} / 2) \sin(k_{i_1 z} \Delta_{i_1 z} / 2) \\ & \left[ \mathbf{u}_1(k_x) e^{i(k_{j_1 z} z^{(j)} - k_{i_1 z} z^{(i)})} + \mathbf{u}_2(k_x) e^{i(k_{j_1 z} z^{(j)} + k_{i_1 z} z^{(i)})} \right. \\ & \left. + \mathbf{d}_1(k_x) e^{-i(k_{j_1 z} z^{(j)} - k_{i_1 z} z^{(i)})} + \mathbf{d}_2(k_x) e^{-i(k_{j_1 z} z^{(j)} + k_{i_1 z} z^{(i)})} \right], \\ & + \delta_{ij} \mathbf{l}_{ij}(k_x), \end{aligned} \quad (2.76)$$

with (see Eq. 2.66)

$$\mathbf{l}_{ij} = -\frac{1}{4\pi\omega\varepsilon_{i_1}} \begin{pmatrix} I_s k_{i_1 z} & 0 & -I_a k_x \\ 0 & I_s k_{i_1}^2 / k_{i_1 z} & 0 \\ -I_a k_x & 0 & I_s k_x^2 / k_{i_1 z} \end{pmatrix}, \quad (2.77)$$

where

$$I_s = \begin{cases} \frac{2}{k_{i_1 z}} \sin(k_{i_1 z} \Delta_{i_1 z} / 2) e^{i k_{i_1 z} |z^{(j)} - z^{(i)}|}, & \text{if } |z^{(j)} - z^{(i)}| > \Delta_{i_1 z} / 2, \\ \frac{2}{i k_{i_1 z}} \{ \cos[k_{i_1 z} (z^{(j)} - z^{(i)})] e^{i k_{i_1 z} \Delta_{i_1 z} / 2} - 1 \}, & \text{if } |z^{(j)} - z^{(i)}| < \Delta_{i_1 z} / 2, \end{cases} \quad (2.78)$$

and

$$I_a = \begin{cases} \frac{2}{k_{i_1 z}} \sin(k_{i_1 z} \Delta_{i_1 z}/2) e^{ik_{i_1 z}|z^{(j)}-z^{(i)}|}, & \text{if } (z^{(j)} - z^{(i)}) > \Delta_{i_1 z}/2, \\ -\frac{2}{k_{i_1 z}} \sin(k_{i_1 z} \Delta_{i_1 z}/2) e^{ik_{i_1 z}|z^{(j)}-z^{(i)}|}, & \text{if } (z^{(j)} - z^{(i)}) < -\Delta_{i_1 z}/2, \\ \frac{2}{k_{i_1 z}} \sin[k_{i_1 z}(z^{(j)} - z^{(i)})] e^{ik_{i_1 z} \Delta_{i_1 z}/2}, & \text{if } |z^{(j)} - z^{(i)}| < \Delta_{i_1 z}/2. \end{cases} \quad (2.79)$$

The coefficients  $C_{ij}$  can now be determined by numerically performing the Fourier transformation of Eq. (2.75) by use of the Fast Fourier Transform. However, one encounters here two problems: first some of the coefficients  $T_{ij}(k_x)$  tend to converge rather slowly to zero if  $k_x \rightarrow \infty$ , this is due to the constant term occurring in the terms involving  $I_s$ . Furthermore the integrand  $T_{ij}(k_x)$  can contain singularities for some values of  $k_x$ . In the case that the media are lossy, the singularities only occur for complex-values, but this can still give rise to numerically poor behavior.

The first problem, i.e., the slow convergence of  $T_{ij}(k_x)$ , is solved by subtracting this slowly converging part, and integrate it analytically. Therefore Eq. (2.75) is rewritten as

$$C_{ij} = \int_{-\infty}^{\infty} (\varepsilon^{(i_1)} - \varepsilon_{i_1}) [T_{ij}(k_x) - \delta_{ij} J_{ij}(k_x)] e^{ik_x(x^{(j)}-x^{(i)})} dk_x + \delta_{ij} A_{ij}, \quad (2.80)$$

with

$$A_{ij} = \int_{-\infty}^{\infty} (\varepsilon^{(i_1)} - \varepsilon_{i_1}) J_{ij}(k_x) e^{ik_x(x^{(j)}-x^{(i)})} dk_x + \delta_{ij} \frac{\hat{\mathbf{z}}\hat{\mathbf{z}}}{i\omega\varepsilon_{i_1}}, \quad (2.81)$$

and

$$J_{ij} = -\frac{i \sin(k_x \Delta_{i_1 x})}{2\pi\omega\varepsilon_{i_1} k_x} \begin{pmatrix} 1 & 0 & 0 \\ 0 & k_{i_1}^2/k_{i_1 z}^2 & 0 \\ 0 & 0 & k_x^2/k_{i_1 z}^2 \end{pmatrix}, \quad (2.82)$$

if  $|z^{(j)} - z^{(i)}| < \Delta_{i_1 z}/2$  and  $J_{ij} = 0$  otherwise. The integrals in Eq. (2.81) can be evaluated analytically with the aid of Cauchy's theorem. This yields

$$A_{ijxx} = -\delta_{ij} \frac{\varepsilon^{(i_1)} - \varepsilon_{i_1}}{\omega\varepsilon_{i_1}}, \quad (2.83)$$

and

$$A_{ijyy} = \begin{cases} \frac{1}{\omega\varepsilon_{i_1}} e^{ik_{i_1}|x^{(j)}-x^{(i)}|} \sin(k_{i_1} \Delta_{i_1 x}/2), & \text{if } |x^{(j)} - x^{(i)}| > \Delta_{i_1 x}/2, \\ \frac{i}{\omega\varepsilon_{i_1}} \{ \cos[k_{i_1}(x^{(j)} - x^{(i)})] e^{ik_{i_1} \Delta_{i_1 x}/2} - 1 \}, & \text{if } |x^{(j)} - x^{(i)}| < \Delta_{i_1 x}/2, \end{cases} \quad (2.84)$$

and  $A_{ijzz} = A_{ijyy}$ .

The second problem with the numerical evaluation of Eq. (2.75), has to with the occurrence of singularities in the integrand. These singularities are actually

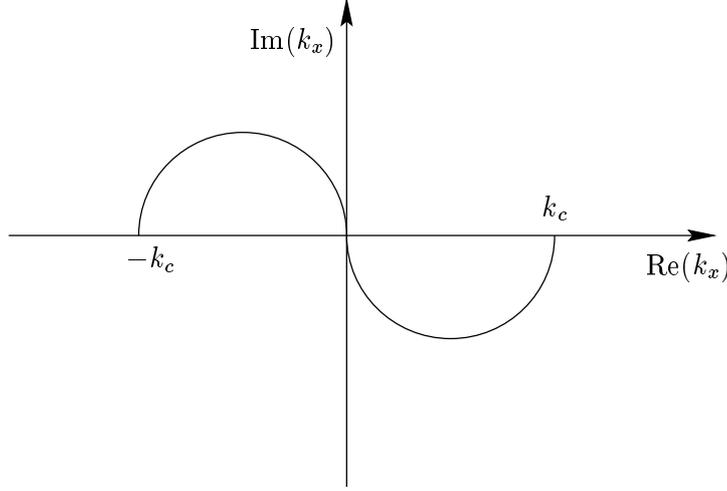


Figure 2.4: Illustration of the change of integration path in the complex  $k_x$  plane.

singularities in the electric Green's tensor and are related with the occurrence of guided modes in the stratified medium (see Chap. 2 of [CHEW, 1995]), i.e., there is a singularity in the integrand of Eq. (2.75) at  $k_x^0$  if there is a guided mode possible with effective wavenumber  $k_x^0$  that travels in the  $x$ -direction with respect to the background configuration. In the case that the background configuration supports surface plasmons, as will be often the case in the later chapters, this singularity lies close to the real  $k_x$ -axis. To avoid this singular behavior a contour deformation in the complex  $k_x$  plane is made to a path consisting of four parts (see Fig. 2.4): two “end parts” of the form  $(-\infty, -k_c)$  and  $(k_c, \infty)$  and two elliptic parts [PAULUS *et al.*, 2000] to avoid the singularities. The elliptic parts are calculated numerically with Gauss-Kronrod quadrature, whereas the end parts are calculated with a Fast Fourier Transform. The position of  $k_c$  depends on the particular configuration. In the configurations of Chapters 3, 4 and 5, i.e., a three layered background with the sub- and superstrate consisting of air, the guided modes lying close to the  $\text{Re}(k_x)$  axis are the surface plasmons (see Sec. 1.7.2). Therefore, for the H-polarization the choice  $k_c = 2k$  is made to avoid these guided modes. For the the E-polarization case, there are no surface plasmons, and the contour deformation can be omitted.

In the case of an E-polarized field, the only non-zero component of  $\hat{\mathbf{E}}$  is  $\hat{E}_y$ . Therefore only the following  $p \times p$  system of equations

$$\sum_{i=1}^p \left[ u_i(\mathbf{r}_{\parallel}^{(j)}) + i\omega C_{ijyy} \right] \cdot a_{iy} = \hat{E}_y^{(\text{inc})}(\mathbf{r}_{\parallel}^{(j)}), \quad (2.85)$$

has to be solved for the coefficients  $a_{1y}, \dots, a_{py}$ . To obtain the coefficients  $C_{ijyy}$

explicitly, the Fourier transform of  $T_{ijkl}$  in Eq. (2.75) have to be performed. In general, the number of Fourier transform that have to be performed equals  $n_z^2$ , where  $n_z$  is the number of possible values of  $z^{(i)}$  given by

$$n_z = \sum_{i=1}^m n_{iz}. \quad (2.86)$$

However, this number can be significantly reduced by realizing that  $T_{ij}$  only depends on  $z^{(i)}$  and  $z^{(j)}$  through the combinations  $k_{j1z}z^{(j)} - k_{i1z}z^{(i)}$  and  $k_{j1z}z^{(j)} + k_{i1z}z^{(i)}$ , see Eq. (2.76). Therefore, if  $k_{i1z} = k_{j1z}$ , i.e., if  $\varepsilon_{i1} = \varepsilon_{j1}$ , the number of different Fourier transforms that have to be performed can be reduced to  $4n_z$ . The Fourier transforms are computed numerically by using the fast Fourier transform.

For an H-polarized field, a  $2p \times 2p$  system of equations is obtained:

$$\sum_{i=1}^p \sum_{q=x,z} \left[ u_i(\mathbf{r}_{\parallel}^{(j)}) \delta_{qr} + i\omega C_{ijqr} \right] \cdot a_{iy} = \hat{E}_r^{(\text{inc})}(\mathbf{r}_{\parallel}^{(j)}), \quad (2.87)$$

for  $j = 1, \dots, p$  and  $r = x, z$ . The same argument about the number of Fourier transforms as used in the case of E-polarized fields, applies to H-polarized fields.

When the matrix equations (2.85) or (2.85) are constructed, they can be solved with Gaussian elimination or, for some special cases, with the method discussed in Sec. 2.5. To obtain the field outside  $\mathcal{D}$ , the calculated field in  $\mathcal{D}$  is substituted into equations (2.16) or (2.17) to obtain the electric or magnetic field, respectively. The number of discretization points needed for the collocation method to converge is in each direction ( $x$  and  $z$ ) approximately 10 to 20 points per wavelength<sup>5</sup> for the E-polarization case and at least 20 points per wavelength for the H-polarization case. This difference in convergence is also found by other authors [TOTZECK AND TIZIANI, 1997a].

## 2.5 The conjugate gradient method for solving the linear system

The collocation method, discussed in Sec. 2.4, converts the domain integral equation into a linear system of equations. This system could of course be solved with Gaussian elimination, but because the data storage grows quadratically and the time to required grows cubically as function of  $p$ , the size of the matrix, this is not possible for very large systems of equations. Therefore we will employ an variant of the *conjugate gradient* method (see Sec. 10.2 of [GOLUB AND VAN LOAN, 1990]),

<sup>5</sup>This is taken as the maximum of all the wavelength in the materials involved.

which can be used in a special case, which includes the configurations treated in the following Chapters. This method can be used when the scatterer  $\mathcal{D}$  consists of equally sized blocks with the same permittivity, all embedded in layers with the same background permittivity. Furthermore, the sizes of all the discretization elements  $\Delta_{i_1x}$  or  $\Delta_{i_2z}$  should all be same. For an optimal use of this method, it is necessary to use the symmetries of the matrix. Therefore we will first discuss them.

Let  $A = [a_{ij}]_{i,j=1,\dots,p}$  be a complex matrix resulting from the collocation method for an E-polarized field, see Eq. 2.85. Now  $A$  can be written as a block matrix build from  $m^2$  blocks:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{bmatrix}, \quad (2.88)$$

where  $A_{ij}$  consists of  $p_i^2$  elements, because  $p_i = p_j$ . Moreover,  $A_{ij} = A_{ji}$ . Each element  $A_{i_1j_1}$  can be written as the sum of two parts, one which is a function of  $z^{(i)} - z^{(j)}$ , whereas the second is a function of  $z^{(i)} + z^{(j)}$ . This implies that  $A_{i_1j_1}$  can be written as

$$A_{i_1j_1} = T_{i_1j_1} + H_{i_1j_1}, \quad (2.89)$$

where

$$T_{i_1j_1} = \begin{bmatrix} T_{i_1j_1}^0 & T_{i_1j_1}^1 & \dots & T_{i_1j_1}^{n_{iz}-1} \\ T_{i_1j_1}^{-1} & T_{i_1j_1}^0 & \dots & T_{i_1j_1}^{n_{iz}-2} \\ \vdots & \vdots & \ddots & \vdots \\ T_{i_1j_1}^{1-n_{iz}} & T_{i_1j_1}^{2-n_{iz}} & \dots & T_{i_1j_1}^0 \end{bmatrix}, \quad (2.90)$$

and

$$H_{i_1j_1} = \begin{bmatrix} H_{i_1j_1}^{1-n_{iz}} & H_{i_1j_1}^{2-n_{iz}} & \dots & H_{i_1j_1}^0 \\ H_{i_1j_1}^{2-n_{iz}} & H_{i_1j_1}^{3-n_{iz}} & \dots & H_{i_1j_1}^1 \\ \vdots & \vdots & \ddots & \vdots \\ H_{i_1j_1}^0 & H_{i_1j_1}^1 & \dots & H_{i_1j_1}^{n_{iz}-1} \end{bmatrix}, \quad (2.91)$$

i.e.,  $T$  is a *Toeplitz matrix* (the matrix-elements are constant along the usual northwest-southeast diagonals) and  $H$  is a *Hankel matrix* (the matrix-elements are constant along the northeast-southwest diagonals), both with as elements  $n_{ix} \times n_{ix}$  matrices. Due to the  $x$ -dependence of the electric Green's tensor and the rectangular mesh that is used (see Eq. 2.75), it follows that each element  $T_{i_1j_1}^q$  and  $H_{i_1j_1}^q$ , with  $q = 1 - n_{iz}, \dots, n_{iz} - 1$  is a Toeplitz matrix.

For an H-polarized field, the matrix corresponding with Eq. (2.87) consists of four parts:

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \quad (2.92)$$

where each of the  $p \times p$  sub-matrices has the same structure as the matrix  $A$  in Eq. (2.88) for the E-polarization case.

The conjugate gradient method is an iterative method with converges in, at most,  $p$  steps for a  $p \times p$  matrix. In each step a matrix-vector product has to be computed, besides some inner products. However, the convergence is only guaranteed for Hermitian matrices, i.e.,  $x^\dagger Ax = (Ax)^\dagger x$ . Now the matrix  $A$  obtained from the collocation method is a symmetric complex matrix, so it is certainly not self-adjoint and the conjugate gradient method does not have to converge. To still use a conjugate method like method, the following trick is employed: the inner product  $x^\dagger y$  is replaced by the product  $x^T y$ . It can be shown that in that case the conjugate gradient does converge, see p. 57 of [BARRETT *et al.*, 1994]. This can be proven by closely following the proof of the convergence of the conjugate gradient method.

As told, in each iteration the matrix  $A$  is multiplied with some vector. Now by using the block-Toeplitz structure of the matrix  $A$  the time required for doing this can be brought back from  $O(p^2)$  to  $O(p \log(p))$ , whereas the data storage of the matrix is reduced to  $O(p)$ , instead of  $O(p^2)$ . For this, the concept of a *circulant matrix* is needed: a  $n \times n$  matrix  $C$  is circulant if it has the following form

$$C = \begin{bmatrix} c_0 & c_1 & \dots & c_{n-1} \\ c_{n-1} & c_0 & \dots & c_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & \dots & c_0 \end{bmatrix}, \quad (2.93)$$

that is,  $C$  is a Toeplitz matrix where each row is identical to the previous row, but shifted one to the right and wrapped around. An useful property of circulant matrices is that they can be diagonalized with a Fourier transform (see [DAVIS, 1979]). that is  $C = F^\dagger \Lambda F$ , where  $F_{st} = e^{2\pi jst/n}$  and  $\Lambda$  is a diagonal matrix. This diagonal matrix can be obtained by taking the Fourier transform of the first column of  $C$ .

The matrix-vector product  $Tx$ , where  $T$  is a  $n \times n$  Toeplitz matrix can be performed with  $O(n \log(n))$  operations by embedding  $T$  in a  $2n \times 2n$  circulant matrix  $C$ , i.e.,

$$C = \begin{bmatrix} T & * \\ * & T \end{bmatrix}, \quad (2.94)$$

where the  $*$  matrices are chosen such that  $C$  is circulant. By employing a fast Fourier transform to  $C$ , the diagonal matrix  $\Lambda$  is obtained. If we now want to multiply a vector  $x$  with  $T$ , the vector  $x$  has to be extended with zeros in a vector with  $2n$  elements, in which the elements of  $x$  are the first  $n$  elements. Subsequently, a fast Fourier transform is applied to this extended vector, and the result is multiplied by the diagonal matrix  $\Lambda$ . If now the (inverse) fast Fourier

transform is taken, the resulting vector contains as its first  $n$  elements the vector  $Tx$  (see [CHAN AND NG, 1996]). The same trick can also be used for a Hankel matrix by noting that  $H$  is a Toeplitz matrix times a permutation represented by the matrix

$$J = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}. \quad (2.95)$$

By employing this methods for a matrix-vector multiplication to the matrices (2.88) or (2.92), for the E-polarization case or H-polarization case, respectively, the cost of the matrix-vector product reduces to  $O(p_i \log(p_i))$  for each sub-matrix, while the data storage reduces to  $8p_i$  or  $24p_i$  elements, for the E-polarization or H-polarization, respectively.

The convergence rate of the conjugate gradient method depends on the spectral properties of the matrix  $A$ ,<sup>6</sup> therefore if  $A$  is in some way “approximated” by a matrix  $M$ , the equation

$$M^{-1}Ax = M^{-1}b \quad (2.96)$$

has the same solution  $x$  but the spectral properties of  $M^{-1}A$  may be better, which hopefully leads to a faster convergence, i.e., less iterations have to be performed to achieve an acceptable solution. The matrix  $M$  is called the *preconditioner*. Instead of actually performing the multiplication of  $M^{-1}A$ , the conjugate gradient method can be adapted such that only the matrix equation  $Ms = t$  have to be solved in each iteration. The conjugate gradient method applied to this system is called the *preconditioned conjugate gradient method* (see Sec. 10.3 of[GOLUB AND VAN LOAN, 1990]). We want to construct  $M$  such that it can be inverted in  $O(n_{ix}^2 \log(n_{ix}))$  operations. Because a  $n \times n$  circulant matrix can be inverted with  $O(n \log(n))$  operations, we would like to use block-circulant-circulant-block matrix as preconditioner, i.e., a  $n \times n$  circulant matrix, with as elements  $m \times m$  circulant matrices (see [CHAN AND NG, 1996]). This can be done by approximating the  $n_{ix} \times n_{ix}$  Toeplitz blocks

$$G = \begin{bmatrix} g_0 & g_1 & \dots & g_{n_{ix}-1} \\ g_{-1} & g_0 & \dots & g_{n_{ix}-2} \\ \vdots & \vdots & \ddots & \vdots \\ g_{1-n_{ix}} & g_{2-n_{ix}} & \dots & g_0 \end{bmatrix}, \quad (2.97)$$

---

<sup>6</sup>The spectrum of a matrix  $A$  consists of the complex numbers  $\lambda$  such that  $\lambda - A$  is not invertible.

which can be either  $T_{i_1 j_1}^q$  or  $H_{i_1 j_1}^q$ , by the circulant matrix with elements given by [CHAN, 1988]

$$c_j = \begin{cases} \frac{(m-j)g_j + jg_{j-m}}{m}, & 0 < j < m, \\ g_0, & j = 0 \end{cases} \quad (2.98)$$

So, a circulant matrix  $M_{i_1 j_1}^q$  is obtained. Now let

$$\Delta = \begin{bmatrix} FM_{11}^q F^\dagger & FM_{12}^q F^\dagger & \dots & FM_{1n_{iz}}^q F^\dagger \\ FM_{21}^q F^\dagger & FM_{22}^q F^\dagger & \dots & FM_{2n_{iz}}^q F^\dagger \\ \vdots & \vdots & \ddots & \vdots \\ FM_{n_{iz}1}^q F^\dagger & FM_{n_{iz}2}^q F^\dagger & \dots & FM_{n_{iz}n_{iz}}^q F^\dagger \end{bmatrix}, \quad (2.99)$$

where  $F$  is the  $n_{ix} \times n_{ix}$  Fourier transform, then each block  $FM_{ij}^q F^\dagger$  is diagonal so  $\Delta$  can be permuted such that the result is a block diagonal matrix:

$$P^\dagger \Delta^q P = \begin{bmatrix} \tilde{M}_1^q & 0 & \dots & 0 \\ 0 & \tilde{M}_2^q & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tilde{M}_{n_{ix}}^q \end{bmatrix}, \quad (2.100)$$

where each block  $\tilde{M}_i^q$  is the sum of a Toeplitz part and a Hankel part. If  $\tilde{M}_i^q$  is given by

$$M_i^q = \begin{bmatrix} m_{11}^q & m_{12}^q & \dots & m_{1n_{iz}}^q \\ m_{21}^q & m_{22}^q & \dots & m_{2n_{iz}}^q \\ \vdots & \vdots & \ddots & \vdots \\ m_{n_{iz}1}^q & m_{n_{iz}2}^q & \dots & m_{n_{iz}n_{iz}}^q \end{bmatrix}, \quad (2.101)$$

then a circulant matrix is constructed with elements given by the following formula<sup>7</sup>

$$d_l^q = \frac{1}{n_{iz}} \sum_{j-k=l(\text{mod } n_{iz})} m_{jk}^q, \quad l = 0, \dots, n_{iz} - 1. \quad (2.102)$$

Again Fourier transforms are applied to the resulting circulant matrices, to obtain a diagonal matrix.

The initial cost of constructing this circulant matrix is  $O(mn \log(mn))$ , while in each iteration step the cost of solving  $Ms = t$ , with  $s$  and  $t$  vectors with  $mn$  elements, is  $O(mn \log(mn))$ . To store the conditioner  $M$ , one only needs  $mn$  elements.

<sup>7</sup>See [CHAN AND NG, 1996].

## 2.6 Conclusions

In this Chapter, a scattering model was described to calculate the field in a number of rectangular scatterers embedded in a layered background medium. Such configurations include for example the single and double slit structures discussed in the following Chapters. The Green's tensor pertaining to such a stratified background medium was derived. A numerical method to solve the domain integral equations, the Collocation method, was described. Furthermore, it was shown that the symmetries of the Green's tensor allowed for an improved solution of the resulting linear system, by using the Conjugate Gradient method together with the Fast Fourier transform.

### 2.A Plane wave incident on a stratified medium

Consider a medium stratified in the  $z$ -direction, as discussed in Sec. 2.2. In this Appendix, the field induced by a plane wave incident from below is calculated. Due to the two-dimensional character of the configuration, the E-polarized case and the H-polarized case can be treated separately, see Sec. 1.6. For an incident E-polarized plane wave with wave vector  $(k_x, 0, k_{1z})$  and unit amplitude at  $\mathbf{r} = (0, 0, z_1)$ , the electric field can be written as

$$\hat{\mathbf{E}}(\mathbf{r}_{\parallel}) = \begin{cases} (\hat{\mathbf{y}}e^{ik_{1z}(z-z_1)} + d_1^E \hat{\mathbf{y}}e^{-ik_{1z}(z-z_1)})e^{ik_x x}, & z < z_1, \\ (u_i^E \hat{\mathbf{y}}e^{ik_{iz}(z-z_{i-1})} + d_i^E \hat{\mathbf{y}}e^{-ik_{iz}(z-z_i)})e^{ik_x x}, & z_{i-1} < z < z_i, \\ u_N^E \hat{\mathbf{y}}e^{ik_{Nz}(z-z_{N-1})}e^{ik_x x}, & z_{N-1} < z, \end{cases} \quad (2.103)$$

with  $2 \leq i \leq N-1$  and where we have introduced the amplitudes  $u_i^E$  and  $d_i^E$ . The magnetic field is then given by

$$\hat{\mathbf{H}}(\mathbf{r}_{\parallel}) = \frac{1}{\omega\mu_0} \begin{cases} (\hat{\mathbf{e}}_1^+ e^{ik_{1z}(z-z_1)} + d_1^E \hat{\mathbf{e}}_1^- e^{-ik_{1z}(z-z_1)})e^{ik_x x}, & z < z_1, \\ (u_i^E \hat{\mathbf{e}}_i^+ e^{ik_{iz}(z-z_{i-1})} + d_i^E \hat{\mathbf{e}}_i^- e^{-ik_{iz}(z-z_i)})e^{ik_x x}, & z_{i-1} < z < z_i, \\ u_N^E \hat{\mathbf{e}}_N^+ e^{ik_{Nz}(z-z_{N-1})}e^{ik_x x}, & z_{N-1} < z, \end{cases} \quad (2.104)$$

with  $\mathbf{k}_{\pm}^{(i)} = (k_x, 0, \pm k_{iz})$  and  $\hat{\mathbf{e}}_{\pm}^i = \mathbf{k}_{\pm}^{(i)} \times \hat{\mathbf{y}}/k_i$ .

The field given by Eqs. (2.103) and (2.104) does satisfy the Maxwell's equations (2.5) and (2.6). However, the amplitudes  $u_i^E$  and  $d_i^E$  must be chosen in such a way that the field satisfies the boundary conditions at the interfaces  $z = z_i$ , which in this case reduce to the continuity of  $\hat{E}_y$  and  $\hat{H}_x$ , see Eqs. (1.38) and (1.39). Applying this to Eqs. (2.103) and (2.104), one obtain for  $i = 2, \dots, N-1$ ,

$$u_i^E \gamma_i + d_i^E = u_{i+1}^E + d_{i+1}^E \gamma_{i+1}, \quad (2.105)$$

$$k_{iz} u_i^E \gamma_i - k_{iz} d_i^E = k_{(i+1)z} u_{i+1}^E - k_{(i+1)z} d_{i+1}^E \gamma_{i+1}, \quad (2.106)$$

with  $\gamma_i \equiv \exp(ik_{iz}d_i)$ . For the special cases of  $z = z_1$  and  $z = z_{N-1}$ , one obtains

$$1 + d_1^E = u_2^E + d_2^E \gamma_2, \quad (2.107)$$

$$k_{1z} - k_{1z}d_1^E = k_{2z}u_2^E - k_{2z}d_2^E \gamma_2, \quad (2.108)$$

and

$$u_{N-1}^E \gamma_{N-1} + d_{N-1}^E = u_N^E, \quad (2.109)$$

$$k_{(N-1)z} u_{N-1}^E \gamma_{N-1} - k_{(N-1)z} d_{N-1}^E = k_{Nz} u_N^E. \quad (2.110)$$

To calculate the amplitudes, a recursive procedure is used (see [VISSER *et al.*, 1999]). We start by defining generalized reflection and transmission coefficients  $r_i^E$  and  $t_i^E$  through the expressions

$$u_{i+1}^E = t_i^E u_i^E \gamma_i, \quad (2.111)$$

$$d_i^E = r_i^E u_i^E \gamma_i, \quad (2.112)$$

with  $i = 1, \dots, N-1$ . Furthermore, we define  $u_1^E = 1$  and  $\gamma_1 = 1$  for consistency in the substrate. Inserting these equations into the boundary conditions (2.105) and (2.106) yields

$$(1 + r_i^E) = t_i^E (1 + r_{i+1}^E \gamma_{i+1}), \quad (2.113)$$

$$k_z^{(i)} (1 - r_i^E) = k_z^{(i+1)} t_i^E (1 - r_{i+1}^E \gamma_{i+1}). \quad (2.114)$$

These two equations enable us to obtain a recursive relation for  $r_i^E$ :

$$r_i^E = \frac{k_{iz}(1 + r_{i+1}^E \gamma_{i+1}) - k_{(i+1)z}(1 - r_{i+1}^E \gamma_{i+1})}{k_{iz}(1 + r_{i+1}^E \gamma_{i+1}) + k_{(i+1)z}(1 - r_{i+1}^E \gamma_{i+1})}. \quad (2.115)$$

This recursive relation is initialized by defining  $r_N^E = 0$ . If Eqs. (2.113) and (2.114) are used, the transmission coefficients can be expressed in the terms of the reflection coefficients:

$$t_i^E = \frac{k_{(i+1)z} + k_{iz}}{2k_{(i+1)z}} + \frac{k_{(i+1)z} - k_{iz}}{2k_{(i+1)z}} r_i^E. \quad (2.116)$$

For an incident H-polarized plane wave with wave vector  $(k_x, 0, k_{1z})$  and unit amplitude at  $\mathbf{r} = (0, 0, z_1)$ , the procedure is similar. In this case the electric field is given by

$$\hat{\mathbf{E}}(\mathbf{r}_{\parallel}) = \begin{cases} (\hat{\mathbf{e}}_1^+ e^{ik_{1z}(z-z_1)} + d_1^H \hat{\mathbf{e}}_1^- e^{-ik_{1z}(z-z_1)}) e^{ik_x x}, & z < z_1, \\ (u_i^H \hat{\mathbf{e}}_i^+ e^{ik_{iz}(z-z_{i-1})} + d_i^H \hat{\mathbf{e}}_i^- e^{-ik_{iz}(z-z_i)}) e^{ik_x x}, & z_{i-1} < z < z_i, \\ u_N^H \hat{\mathbf{e}}_N^+ e^{ik_{Nz}(z-z_{N-1})} e^{ik_x x}, & z_{N-1} < z, \end{cases} \quad (2.117)$$

where the amplitudes  $u_i^H$  of the upward propagating plane waves and  $d_i^H$  of the downward propagating are introduced. The magnetic field is given by

$$\hat{\mathbf{H}}(\mathbf{r}_{\parallel}) = -\frac{1}{\omega\mu_0} \begin{cases} k^{(1)}(\hat{\mathbf{y}}e^{ik_{1z}(z-z_1)} + d_1^H\hat{\mathbf{y}}e^{-ik_{1z}(z-z_1)})e^{ik_x x}, & z < z_1, \\ k^{(i)}(u_i^H\hat{\mathbf{y}}e^{ik_{iz}(z-z_{i-1})} + d_i^H\hat{\mathbf{y}}e^{-ik_{iz}(z-z_i)})e^{ik_x x}, & z_{i-1} < z < z_i, \\ k^{(N)}u_N^H\hat{\mathbf{y}}e^{ik_{Nz}(z-z_{N-1})}e^{ik_x x}, & z_{N-1} < z. \end{cases} \quad (2.118)$$

For H-polarized fields, one needs the continuity of  $\hat{E}_x$  and  $\hat{H}_y$  at the boundaries  $z = z_i$ , see Eqs. (1.38) and (1.39). If Eqs. (2.117) and (2.118) are used, one obtains for  $i = 2, \dots, N-1$ ,

$$k^{(i)}(u_i^H\gamma_i + d_i^H) = k^{(i+1)}(u_{i+1}^H + d_{i+1}^H\gamma_{i+1}), \quad (2.119)$$

$$(k_{iz}u_i^E e^{ik_{iz}d_i} - k_{iz}d_i^E)/k^{(i)} = (k_{(i+1)z}u_{i+1}^E - k_{(i+1)z}d_{i+1}^E e^{ik_{(i+1)z}d_{i+1}})/k^{(i+1)}. \quad (2.120)$$

For the special cases of  $z = z_1$  and  $z = z_{N-1}$ , one obtains

$$k^{(1)}(1 + d_1^H) = k^{(2)}(u_2^E + d_2^E\gamma_2), \quad (2.121)$$

$$(k_{1z} - k_{1z}d_1^E)/k^{(1)} = (k_z^{(2)}u_2^E - k_z^{(2)}d_2^E\gamma_2)/k^{(2)}, \quad (2.122)$$

and

$$k^{(N-1)}(u_{N-1}^E\gamma_{N-1} + d_{N-1}^E) = k^{(N)}u_N^E, \quad (2.123)$$

$$(k_{(N-1)z}u_{N-1}^E\gamma_{N-1} - k_{(N-1)z}d_{N-1}^E)/k^{(N-1)} = k_{Nz}u_N^E/k^{(N)}. \quad (2.124)$$

As in the case of the E-polarization we define generalized reflection and transmission coefficients by

$$u_{i+1}^H = t_i^H u_i^H \gamma_i, \quad (2.125)$$

$$d_i^H = r_i^H u_i^H \gamma_i, \quad (2.126)$$

with  $i = 1, \dots, N-1$ . If an approach strictly similar to the case of an E-polarized plane wave is followed, one obtains the following recursive relation for the reflection coefficients:

$$r_i^H = \frac{k_{iz}\varepsilon_{i+1}(1 + r_{i+1}^H\gamma_{i+1}) - k_{(i+1)z}\varepsilon_i(1 - r_{i+1}^H\gamma_{i+1})}{k_{iz}\varepsilon_{i+1}(1 + r_{i+1}^H\gamma_{i+1}) + k_{(i+1)z}\varepsilon_i(1 - r_{i+1}^H\gamma_{i+1})}. \quad (2.127)$$

This recursive relation is initialized by defining  $r_N^H = 0$ . The transmission coefficients are expressed in the terms of the reflection coefficients by the relation

$$t_i^H = \frac{k_{(i+1)z}\varepsilon_i + k_{iz}\varepsilon_{i+1}}{2k_{(i+1)z}\sqrt{\varepsilon_i\varepsilon_{i+1}}} + \frac{k_{(i+1)z}\varepsilon_i - k_{iz}\varepsilon_{i+1}}{2k_{(i+1)z}\sqrt{\varepsilon_i\varepsilon_{i+1}}}r_i^H. \quad (2.128)$$

# Chapter 3

## Light Transmission through a Single Sub-wavelength Slit

This Chapter is based on the following publications:

- H.F. Schouten, T.D. Visser, D. Lenstra and H. Blok, “Light transmission through a sub-wavelength slit: waveguiding and optical vortices”, *Physical Review E*, vol. 67, 036608 (2003).
- H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “Creation and annihilation of phase singularities near a sub-wavelength slit”, *Optics Express*, vol. 11, pp. 371–380 (2003).
- H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “Diffraction of light at slits in plates of different materials”, *Journal of Optics A: Pure and Applied Optics*, vol. 6, pp. S277–S280 (2004).
- H.F. Schouten, T.D. Visser and D. Lenstra, “Optical vortices near sub-wavelength structures”, *Journal of Optics B: Quantum and Semiclassical Optics*, vol. 6, pp. S404–S409 (2004).

### Abstract

We analyze the diffraction of light incident on a sub-wavelength slit in a thin plate. The anomalous (i.e., more than 100%) light transmission through the slit is connected to both waveguiding and phase singularities of the field of power flow near the slit. It is demonstrated that the enhanced transmission is accompanied by the annihilation of these phase singularities. Furthermore, it is found that plates with different material properties, such as conductivity and thickness, show a fundamentally different behaviour of the field near the slit.

## 3.1 Introduction

The analysis of light transmission through a slit with a sub-wavelength width in a thin plate is a subject with a venerable history, dating back to Lord Rayleigh [RAYLEIGH, 1897; BOUWKAMP, 1954; BORN AND WOLF, 1999]. Because of its importance for near-field optics and semiconductor technology, it continues to attract attention. Ebbesen *et al.* demonstrated experimentally [EBBESEN *et al.*, 1998; GHAEMI *et al.*, 1998] that certain arrays of cylindrical cavities in metal plates allow much more light transmission than predicted by the standard theory of aperture diffraction [BETHE, 1944]. These authors suggested that this enhanced transmission was generated by the coupling of the light to surface plasmons on the metal/air interface. More recently, it was demonstrated [THIO *et al.*, 2001] that enhanced transmission can be achieved even with the use of a single aperture. Broadly speaking there are two mechanisms proposed to explain the extraordinary light transmission: the coupling of light with surface plasmons [EBBESEN *et al.*, 1998; GHAEMI *et al.*, 1998; SCHRÖTER AND HEITMANN, 1998; PORTO *et al.*, 1999; MARTÍN-MORENO *et al.*, 2001], and Fabry-Pérot-like resonances inside the apertures [ASTILEAN *et al.*, 2000; TAKAKURA, 2001; YANG AND SAMBLES, 2002]. A good understanding of the causes and requirements for enhanced transmission could lead to, among other things, novel near-field optical measurement devices as well as optical storage devices with a density not restricted by the diffraction limit.

Most theoretical studies that deal with the diffraction of light by a narrow slit in a metal plate assume as simplifying conditions that the plate is vanishingly thin and perfectly conducting (see, for example, [STAMNES, 1995; NYE, 2002] and the references therein). In Chapter 2, we have developed a rigorous Green's tensor formalism that allows us to analyze the diffraction of light at slits in plates of finite thickness and finite conductivity. In Section 3.3 the transmission of light through a single sub-wavelength slit in a metal plate is analyzed. The slit is illuminated by a normally incident plane wave, which can be either E-polarized or H-polarized. E-polarization differs from H-polarization in that no surface plasmons are excited (see Sec. 1.7.2). Furthermore, for E-polarization the first waveguide mode in a perfect conductor has a cut-off width of  $w_{\text{cut-off}} = \lambda/2$ , with  $\lambda$  the wavelength. However, due to their finite conductivity, efficient energy transport may be possible at smaller slit widths in realistic metal plates. A rigorous computation of the field demonstrates that near these cut-off widths, there is an enhanced transmission through the slit. Transmission efficiencies as high as 300% are found for special configurations. We emphasize that these remarkable enhancement effects occur even though there is no coupling to surface plasmons. To elucidate why the field couples so effectively with the propagating waveguide modes, we have analyzed the field of power flow (i.e., the time-averaged Poynting vector) near the slit. It is found

that this field exhibits *optical vortices* and other kinds of *phase singularities* (see Sec. 1.8) which are arranged in an array-like pattern. We find that the location and the annihilation or creation of these phase singularities are intimately connected with the phenomenon of enhanced transmission.

In Section 3.4 we examine the diffraction of light and its singular behavior near a slit in a semiconductor plate. It is found that there is a significant difference in the behavior of the field near metal plates and near those made out of semiconductor material. In the latter case, frustrated transmission may occur—that is, the transmission through a narrow slit in such a plate can be *less* than the transmission through a semiconductor plate *without* a slit. It is shown by numerical simulation that frustrated transmission is accompanied by a handedness of optical vortices that is opposite to that which occurs in the case of metal plates for which enhanced transmission takes place.

Furthermore, in Section 3.5 we study the dependence of the light transmission on the thickness of the plate. The results found are qualitatively explained by considering Fabry-Pérot resonances of waveguide modes.

## 3.2 The configuration

The configuration under consideration is illustrated in Fig. 3.1. An monochromatic electromagnetic wave is normally incident upon a metal plate of thickness  $d$  and permittivity  $\varepsilon_{\text{plate}}$  from the negative  $z$ -direction. A single slit of width  $w$ , infinitely long in the  $y$ -direction, is present in the plate. Because the system is invariant with respect to  $y$ -translations, we may treat the problem as two-dimensional, with relevant coordinates  $x$  and  $z$ . For this reason, the field splits into two mutually independent parts, namely, the E-polarized part and the H-polarized part (see Sec. 1.6).

To calculate the field, the Green's tensor method as discussed in Chapter 2 is employed. The background medium consists in this case of three non-magnetic layers with permittivities given by

$$\varepsilon_b(z) = \begin{cases} \varepsilon_0, & z < 0, \\ \varepsilon_{\text{plate}} & 0 < z < d, \\ \varepsilon_0, & d < z. \end{cases} \quad (3.1)$$

The scatterer  $\mathcal{D}$  consist of the slit with permittivity  $\varepsilon_0$  and has dimensions  $[-a, a] \times [0, d]$ , where  $a = w/2$ . The methods described in Chapter 2 were applied to obtain the field.

In order to study the transmission properties of a slit, we need to define its transmission coefficient. The energy flow is given by the time-averaged Poynting

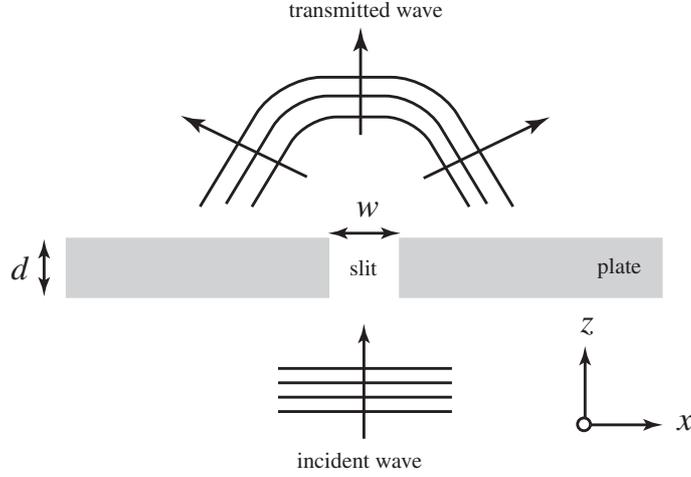


Figure 3.1: Illustrating the notation relating to transmission through a slit.

vector (see Sec. 1.5). We would like to define the transmission by integrating the flux of the time-averaged Poynting vector along the plane  $z = d$ , which represents the total energy transmitted into the upper half-space. The transmission would be given by

$$\int_{-\infty}^{\infty} \langle S_z(x, d) \rangle_T dx. \quad (3.2)$$

The problem with this definition is that it not only gives the transmission of the slit, but also takes also into account the transmission of light through the metal plate which (for a very thin plate) would be also present in the absence of the slit. As a consequence, this definition yields an infinite transmission. This is due to the fact that

$$\langle S_z(x, d) \rangle_T \rightarrow \langle S_z^{(\text{inc})}(x, d) \rangle_T \neq 0 \quad (\text{as } x \rightarrow \pm\infty), \quad (3.3)$$

i.e., in the  $z = d$  plane, far away from the slit, it converges to the constant value of the Poynting vector  $S_z^{(\text{inc})}$  in the absence of the slit. In order to remove this singularity and only take into account the effect of the slit on the transmission, the flux of the time-averaged Poynting vector of the incident field along the “dark side” of the metal plate is subtracted from expression (3.2), i.e., the following function

$$f(x) := \begin{cases} \langle S_z(x, d) \rangle_T & |x| < a, \\ \langle S_z(x, d) \rangle_T - \langle S_z^{(\text{inc})}(x, d) \rangle_T & |x| \geq a, \end{cases} \quad (3.4)$$

is subtracted from the integrand in the expression (3.2). Note that we do not subtract the incident field at  $|x| < a$ . The transmission is then given by

$$\int_{-\infty}^{\infty} f(x) dx. \quad (3.5)$$

The reason for not subtracting the incident field when  $|x| < a$  in expression (3.4), is that otherwise for very wide slits ( $a \gg \lambda$ ), the normalized transmission coefficient (the normalization will be given in the next lines) would be less than one. Apart from this, we like to make the transmission independent of the amplitude of the incident field. Therefore we divide expression (3.5) by the following integral

$$\int_{-a}^a \langle S_z^{(0)}(x, 0) \rangle_T dx, \quad (3.6)$$

where the entity  $\langle \mathbf{S}^{(0)} \rangle_T$  represents the time-averaged Poynting vector of the field that is emitted by the laser. Note that this entity does not equal the time-averaged Poynting vector of the incident field because the incident field is partly reflected by the metal plate. So finally, the transmission coefficient of the slit is given by the following equation

$$T := \frac{\int_{\text{slit}} \langle S_z(x, d) \rangle_T dx + \int_{\text{plate}} \langle S_z(x, d) \rangle_T - \langle S_z^{(\text{inc})}(x, d) \rangle_T dx}{\int_{\text{slit}} \langle S_z^{(0)}(x, 0) \rangle_T dx}. \quad (3.7)$$

The normalization is such that for a very wide slit, the transmission equals one. However, note that it is possible for the transmission to be larger than one. This does not constitute a violation of energy conservation, because the normalization (3.6) contains the flux of the emitted field, and not that of the actual field. A comparable phenomenon occurs in the scattering of light by small particles, where such a particle can under certain circumstances absorb more light than is incident on it [BOHREN, 1983; BOHREN AND HUFFMAN, 1983].

### 3.3 Transmission through a slit in a thin metal plate

In Fig. 3.2 (lower curve) the transmission coefficient is shown as a function of the width of a slit in a thin silver plate. The upper curve is discussed later. We observe a damped resonance behavior as a function of the width  $w$ , with maxima at  $w \approx 0.5\lambda, 1.5\lambda, \dots$ , where the transmission is enhanced, i.e., the transmission coefficient is greater than one, with its largest value  $T = 1.33$  at  $w = 0.5\lambda$ .

If the dispersion relation is computed for the first guided mode (see Sec. 1.7.1) of a silver waveguide, it is found that, due to the finite conductivity, the cut-off width is less than that for a perfectly conducting waveguide, viz.  $w_{\text{cut-off}} = 0.4\lambda$  (see Fig. 3.3). The position of the maxima in the transmission as a function of the

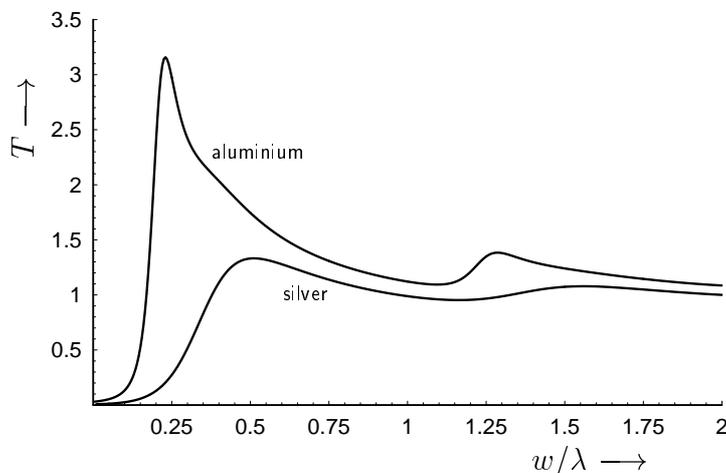


Figure 3.2: The transmission coefficient  $T$  of a narrow slit in a thin plate as a function of the slit width  $w$ , expressed in wavelengths. The lower curve is for a slit in a 100 nm thick silver plate and  $\lambda = 500$  nm. The upper curve is for a slit in a 100 nm thick aluminium plate and  $\lambda = 91.8$  nm. At these wavelengths the refractive indices are taken as  $n_{\text{silver}} = 0.05 + i2.87$  and  $n_{\text{aluminium}} = 0.041 + i0.517$ , respectively. The slit is illuminated by a E-polarized, normally incident plane wave.

slit width is found to be close to the cut-off width; see Fig. 3.2. This also holds for the higher order modes.

If the conductivity of the metal decreases, the cut-off width shifts to lower values. This is illustrated in Fig. 3.3 (left-hand curves), where the effective index is plotted for aluminium at  $\lambda = 91.8$  nm. It is seen that the cut-off width is shifted to  $w \approx 0.15\lambda$ . Because the normalization factor of the transmission coefficient (see Eq. 3.7) will be smaller for smaller values of the width, it is to be expected that the anomalous transmission will be greater for materials with a smaller cut-off width. In Fig. 3.2 the transmission coefficient for a slit in aluminium is shown (upper curve). Both the expected shift of the maximum to a lower value is observed, as well as the enhancement in the transmission, with a maximum of  $T \approx 3.2$  at  $w \approx 0.25\lambda$ .

To understand this anomalous transmission, we have analyzed the field of power flow near the slit. A typical example of these calculations is shown in Fig. 3.4 (for the thin silver plate discussed before), where the field is seen to exhibit phase singularities, i.e., points where the amplitude of the time-averaged Poynting vector is zero and as a consequence its direction, or equivalently its phase, is undetermined. It is seen that the anomalous transmission (namely  $T = 1.11$ ) coincides with the presence of two optical vortices (a and b) *within* the plate, and a “funnel-

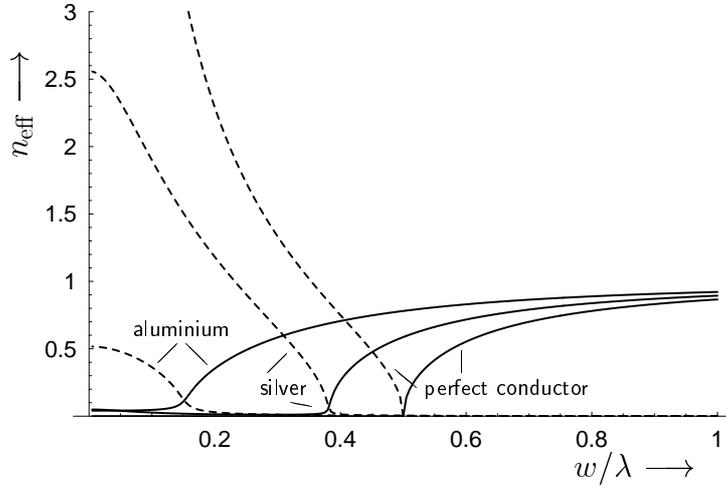


Figure 3.3: Effective indices  $n_{\text{eff}}$  of the first E-polarized waveguide mode inside a narrow slit as a function of the slit width  $w$ , expressed in wavelengths. The full lines denote the real part of the effective index, the dashed lines denote the imaginary part. The two curves for aluminium are for  $\lambda = 91.8$  nm, the two curves for silver are for  $\lambda = 500$  nm, whereas the last two curves are for a perfectly conducting material. The refractive indices were taken to be  $n_{\text{silver}} = 0.05 + i2.87$  and  $n_{\text{aluminium}} = 0.041 + i0.517$ .

like” power flow into the slit. This funneling effect corresponds to a transmission coefficient of more than one. In addition, four other phase singularities are visible just below the slit (c,d,e and f; two saddle points and two vortices). In Fig. 3.5 the location of the phase singularities is shown on a larger scale. It is seen that they are arranged in an array-like pattern. It is to be noted that only part of the phase singularities are shown—the pattern is continuous in a periodic way to the left and right, and also downwards. It was found that at least 5000 phase singularities are present for this particular example.

Changing the slit width in a continuous manner causes the phase singularities to move through space. Near the cut-off width (at  $w \approx 0.45\lambda$ ) the array of phase singularities along the symmetry-axis annihilate, each annihilation consisting of two vortices (one left-handed and one right-handed) and two saddle points. In Fig. 3.6 the resulting arrangement for  $w = 0.5\lambda$  is shown. Because the annihilation of phase singularities leads to a smoother field of power flow, an increased transmission is observed. Near the other cut-off widths at  $w \approx 1.4\lambda, 2.4\lambda, \dots$  additional annihilations occur. In such processes the total topological charge is always conserved, see Sec. 1.8.2.

If the field of power flow is analyzed for a slit in an aluminium plate (at  $\lambda = 91.8$

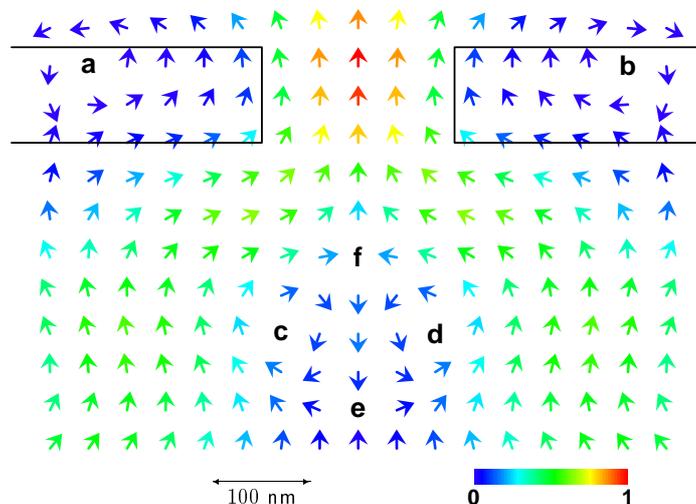


Figure 3.4: Behavior of the time-averaged Poynting vector near a 200 nm wide slit in a 100 nm thick silver plate. The incident light (coming from below) has a wavelength  $\lambda = 500$  nm and is E-polarized. The left-handed (a and d) and right-handed (b and c) optical vortices (b and c) each have a topological charge of +1, whereas the topological charge of the saddle points (e and f) is  $-1$ . The transmission coefficient  $T = 1.11$ . The color coding indicates the modulus of the (normalized) Poynting vector (see legend).

nm) the same pattern and behavior of the phase singularities as function of the slit width is found as described above. Below the cut-off width, the same pattern of phase singularities is found as in Figs. 3.4 and 3.5. At a width slightly larger than the cut-off width (i.e., at  $w = 0.2\lambda$ ) an array of phase singularities along the symmetry axis annihilates. Again, at the cut-off widths of the higher modes similar annihilations occur. The described behavior of the pattern of phase singularities was also found for other materials with  $\text{Re}(n) \ll \text{Im}(n)$  (i.e., good conductors, at a frequency below a plasma resonance frequency).

In Fig. 3.7 the transmission coefficient as a function of the slit width is shown for an H-polarized incident plane wave. Contrary to the E-polarized case, we observe that even for widths much smaller than the wavelength there is significant light transmission through the slit. However, note that for very small slit widths ( $w < 0.1\lambda$ ), the transmission coefficient does drop to almost zero. For higher values of the width, the transmission coefficient rises as a function the width, with a periodic pattern of maxima at  $w \approx \lambda, 2\lambda, \dots$  (the latter maxima are not visible in the plot). Note that in contrast tot the E-polarized case, the transmission remains

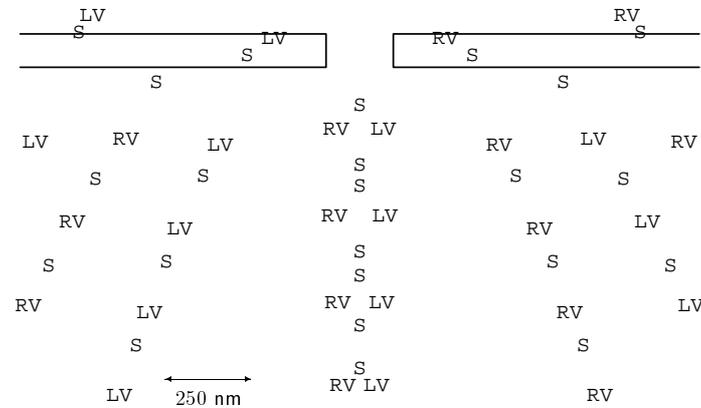


Figure 3.5: Location of phase singularities in the field of power flow for the same configuration as in Fig. 3.4, i.e., for a slit width of  $w = 0.4\lambda = 200$  nm. The left- and right-handed optical vortices are denoted by lv and rv, respectively; s denotes a saddle point. Notice the larger scale as compared to Fig. 3.4.

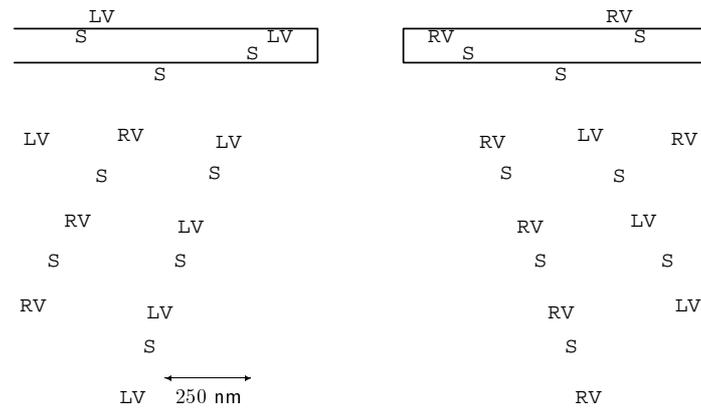


Figure 3.6: Location of phase singularities in the field of power flow for the same configuration as in Fig. 3.5, but now for a slit width  $w = 0.5\lambda = 250$  nm. Notice that the central array of phase singularities which was visible below the slit in Fig. 3.5 has now been annihilated.

less than one after the first peak.

The effective index as a function of the slit width is given in Fig. 3.8 for the first two symmetric guided modes. The first mode, as in the case of a slit in a perfectly conducting material, does not have a cut-off width. However, due to the penetration of the field into the silver plate, the real part of the effective index

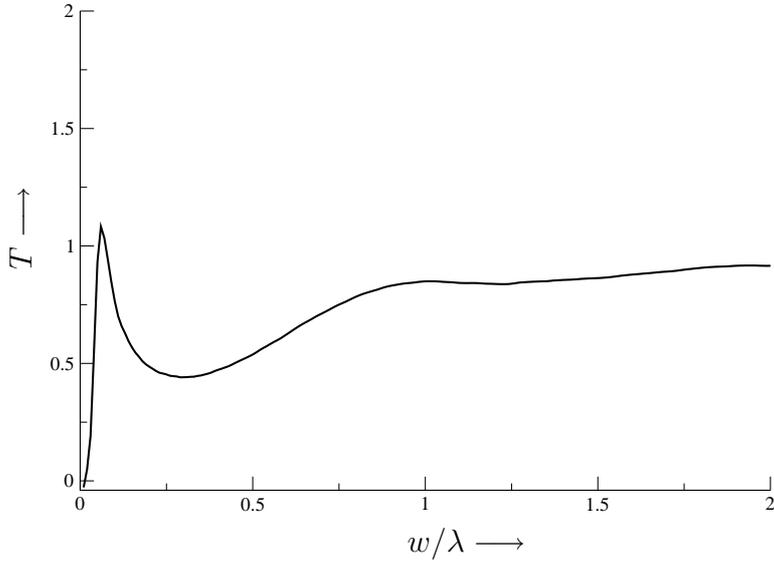


Figure 3.7: The transmission coefficient  $T$  of a narrow slit in a thin plate as a function of the slit width  $w$ , expressed in wavelengths. The plate is 100 nm thick and consists of silver ( $n_{\text{silver}} = 0.05 + i2.87$ ). The incident field is a normally incident, H-polarized plane wave with a wavelength of  $\lambda = 500$  nm.

rises to very high values if the slit width is decreased. This implies that the drop of the transmission for decreasing slit widths at  $w \approx 0.1\lambda$  is due to a less efficient coupling to this guided mode. This drop of the transmission has recently been observed at microwave frequencies [SUCKLING *et al.*, 2004]. The position of the cut-off width of the second symmetric mode seems to correspond to the position of the maximum of the transmission, just as in the E-polarized case. This however does not explain why the transmission remains smaller than one for larger slit widths. A possible explanation of this effect could be that the surface plasmons excited at the edges of the slits, which propagate at the air-silver interface away from the slit in the  $\pm x$ -directions, carry away some of the energy. This would indeed explain why the transmission for the H-polarized case (where there are surface plasmons possible) is smaller than for the E-polarized case (where there are no surface plasmons possible).

### 3.4 Transmission through a slit in a thin semiconductor plate

In Fig. 3.9 the transmission coefficient is shown as a function of the slit width for a thin silver plate and for a thin silicon plate (the middle curve will be discussed in

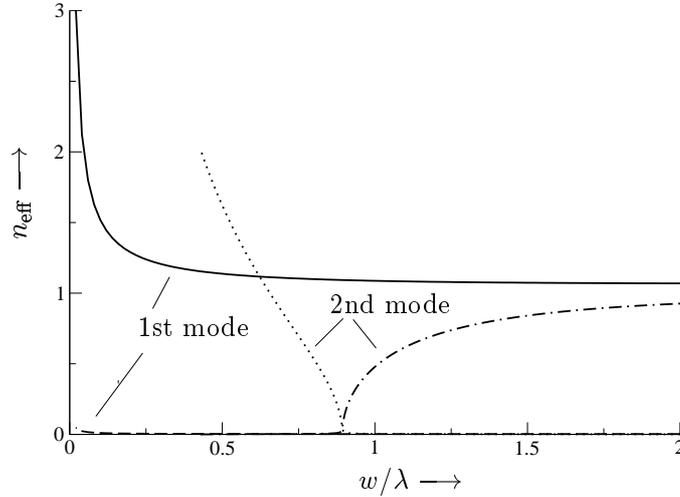


Figure 3.8: Effective indices  $n_{\text{eff}}$  of the first two symmetric H-polarized waveguide modes inside a narrow slit in a silver plate as a function of the slit width  $w$ , expressed in wavelengths. The full lines denote the real part of the effective index of the first mode, the dashed line denotes its imaginary part. The dashed-dotted line denotes the real part of the effective index of the second mode, whereas the dotted line denotes the imaginary part of its effective index. The wavelength is  $\lambda = 500$  nm and the refractive index is taken as  $n_{\text{silver}} = 0.05 + i2.87$ .

the next section). For the silver plate we observe a damped oscillating behavior of the transmission coefficient. This is explained in detail in the preceding Section. The behavior of the transmission coefficient of a slit in a silicon plate is seen from Fig. 3.9 to be quite different: for small slit widths the transmission is frustrated (i.e.,  $T < 0$ ). From Eq. (3.7) it follows that in this case the power flow through a plate *with* a slit is less than that through a plate *without* a slit. Also, in contrast to what we found for a slit in a silver plate, the transmission coefficient now never exceeds unity.

To better understand the behavior of the transmission coefficient, we have plotted in Fig. 3.10 the effective index of the first guided mode for both silicon and silver waveguides as a function of the slit width. For the first guided mode in silver (discussed in the preceding Section), it is found that there is a cut-off width at  $w \approx 0.4\lambda$ , i.e. for a width smaller than this critical width the mode is evanescent, whereas for a width larger than the cut-off width the mode is propagating. The silicon waveguide has a completely different behavior: for small slit widths ( $w < 0.15\lambda$ ) the first guided mode is not present (i.e., only radiation modes [SNYDER AND LOVE, 1983] exist), whereas for a somewhat larger slit width (i.e.,  $w \approx 0.15\lambda$ –

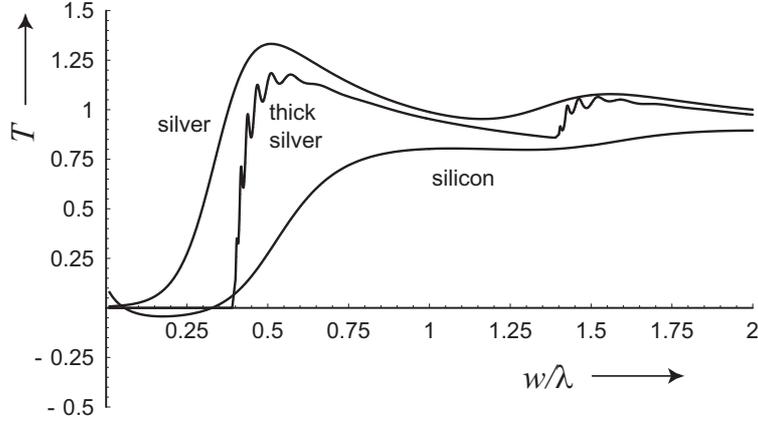


Figure 3.9: The transmission coefficient  $T$  of a narrow slit as a function of the slit width  $w$ , expressed in wavelengths. The lower curve is for a slit in a 100 nm thick silicon plate, the upper curve is for a slit in a 100 nm thick silver plate, whereas the middle curve is for a slit in a  $2.5\mu\text{m}$  thick silver plate. The wavelength is  $\lambda = 500$  nm, and the refractive indices are taken as  $n_{\text{silver}} = 0.05 + i2.87$  and  $n_{\text{silicon}} = 4.3 + i0.74$ , respectively. The slit is illuminated by a E-polarized, normally incident plane wave.

$0.4\lambda$ ) the first guided mode is more damped than a plane wave through the bulk material. The extraordinarily high damping of this guided mode is caused by its extension into the lossy cladding. It should be noted that at these widths the transmission coefficient is smaller than that of a silicon plate without a slit (see Fig. 3.9). For slit widths  $w > 0.4\lambda$ , the guided mode has a propagating character, which corresponds with a positive transmission coefficient.

In Fig. 3.11, the time-averaged Poynting vector is plotted around a 100 nm wide slit in a silicon plate. In the case of a silver plate (Fig. 3.12), two vortices are visible, which correspond with a funnel-like power flow into the slit (see the preceding Section). In contrast, the power flow near a slit in a silicon plate exhibits two vortices and two saddle points located inside the slit, coinciding with a power flow into the silicon plate rather than into the slit. It should be noted that this coincides with a different handedness of the vortices: vortex  $\mathbf{b}$  in Fig. 3.12 is left-handed, whereas vortex  $\mathbf{b}$  in Fig. 3.11 is right-handed. In Fig. 3.13 a detail of Fig. 3.11 is shown. There it can be observed that in this region a sink and a saddle point are present. Due to conservation of energy, a sink is only possible inside a lossy material (see Sec. 1.8.3). If the width of the slit is decreased, the sink and the saddle point annihilate each other, a process in which the topological charge is conserved (see Sec. 1.8.2).

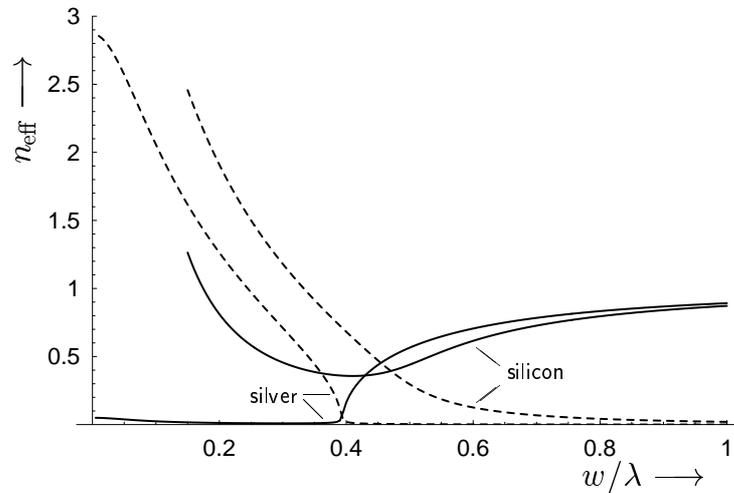


Figure 3.10: Effective indices  $n_{\text{eff}}$  of the first E-polarized waveguide mode inside a narrow slit as a function of the slit width  $w$ , expressed in wavelengths. The full lines denote the real part of the effective index, the dashed lines denote the imaginary part.

In Fig. 3.14 another detail of Fig. 3.11 is presented. If the width of the slit is slightly increased, the saddle points ( $a$  and  $d$ ) move together to form a *monkey saddle singularity* [HSIUNG, 1981, p. 266]. This is shown in Fig. 3.15. A monkey saddle is similar to a saddle point, but possesses three attracting and three repulsing directions, instead of two attracting and two repulsing directions. The monkey saddle is unstable, as it exists for one value of the slit width only; for a larger width it decays into two saddle points (see Fig. 3.16). With the aid of symmetry considerations, one can show that the singularity in Fig. 3.15 is indeed a monkey saddle, and not two closely spaced saddle points.

### 3.5 Transmission through a slit in a thick metal plate

Next we discuss the influence of the plate thickness on the light transmission process. In Fig. 3.9 the transmission coefficient is plotted as a function of the slit width for a thick (i.e.,  $5\lambda$ ) silver plate. Several differences with the thin silver plate can be observed: below the cut-off width at  $w = 0.4\lambda$  there is hardly any transmission through the thick plate. At the cut-off width a steep rise of the transmission as a function of the width is seen. Furthermore we note that there are some fast oscillations of the transmission near the cut-off widths.

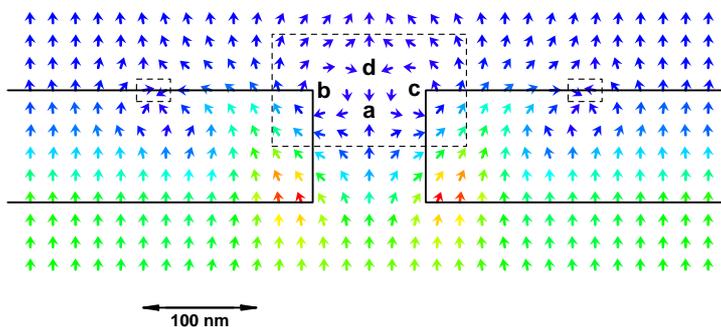


Figure 3.11: Behavior of the time-averaged Poynting vector near a 100 nm wide slit in a 100 nm thick silicon plate. The incident light (coming from below) has a wavelength  $\lambda = 500$  nm. Two vortices (b and c) and two saddle points (a and d) are present in the middle region. In the dashed regions on the right and the left a saddle point and a sink are present (see Fig. 3.13). The dashed box in the middle denotes the region depicted in Fig. 3.14.

The negligible transmission below  $w \approx 0.4\lambda$  can be explained by noting that all guided modes are evanescent in that region, i.e., the transmission will decrease exponentially as a function of the thickness of the plate. The explanation for the fast oscillations near the cut-off widths is more subtle: the guided mode travelling in one direction can be reflected into a guided mode travelling into the opposite direction at the end of the slit. Therefore a Fabry-Pérot type resonance exists, i.e., the transmission as a function of the thickness has an oscillating behavior, with a period determined by the real part of the effective index. Because the effective index as a function of the width changes rapidly near the cut-off width (see Fig. 3.10), the interference of the guided modes will change repeatedly from constructive to destructive and back. This results in the fast oscillations observed in Fig. 3.9.

In Fig. 3.17 the transmission coefficient of a thick silver plate is plotted as a function of the slit width, for the case of an H-polarized incident plane wave. Similar Fabry-Pérot-like resonances as for the E-polarized case can be observed around the cut-off widths at  $w \approx 0.9\lambda$  (see Fig. 3.8) and  $w \approx 1.9\lambda$ . Furthermore the same behaviour occurs at the decline of the real part of the effective index of the first mode at  $w \approx 0.1\lambda$  (see Fig. 3.8). However, note that the amplitude of the oscillations is much larger than for the E-polarized case, which seems to imply that the H-polarized modes have a larger reflection coefficient at the end of the slit than the E-polarized modes.

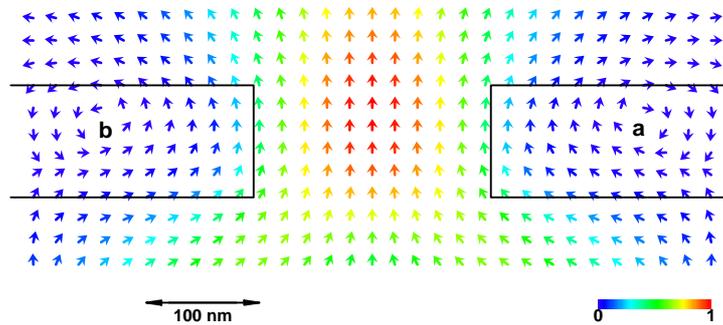


Figure 3.12: Behavior of the time-averaged Poynting vector near a 200 nm wide slit in a 100 nm thick silver plate. The incident light (coming from below) has a wavelength  $\lambda = 500$  nm. The color coding indicates the modulus of the (normalized) Poynting vector (see legend). Two vortices (a and b) can be observed.

### 3.6 Conclusions

In conclusion, we have shown that anomalous light transmission through a narrow slit in a thin metal plate is connected to both waveguiding and phase singularities of the field of power flow. The onset of guided modes yields the maxima of the transmission curve, while a qualitative understanding of the light transmission is obtained by charting the different phase singularities in the field of power flow. More explicitly, it was demonstrated that the enhanced transmission is accompanied by the annihilation of phase singularities. In particular, it was found for certain configurations that transmission efficiencies as high as 3 are possible. We notice that in this particular configuration no surface plasmons are excited. Our study demonstrates that anomalous light transmission can even occur in their absence.

Furthermore, we have shown that the process of light transmission through a narrow slit in a plate strongly depends on the material properties of the plate. In particular, a slit in a silver plate can give rise to enhanced transmission, whereas a slit in a comparable silicon plate exhibits frustrated transmission. This was found to coincide with a change in handedness of certain optical vortices. Our results suggest that the material properties of the plate and its thickness are as important as the width of the slit in the application of extraordinary light transmission in any practical optical system.

Because of the ability of sub-wavelength slits to “focus” a large amount of the power flow onto a small area, our findings are relevant for the design of novel sub-wavelength light sources. Further study of the sub-wavelength features of optical fields may lead to increased resolution in microscopy, or an ability to “write” information on an optical disk with a higher density than is presently possible.

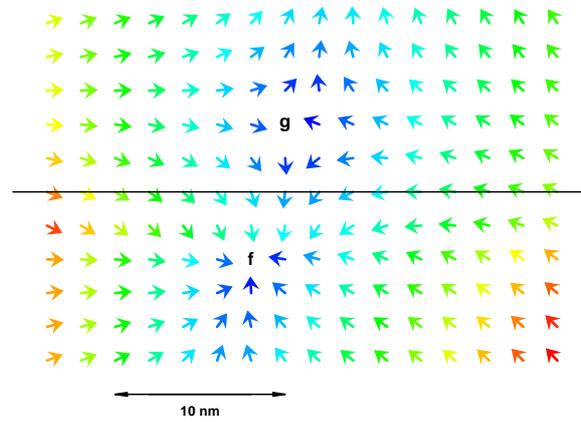


Figure 3.13: Detail (from the right-hand side) of Fig. 3.11. A sink (f) and a saddle point (g) are present.

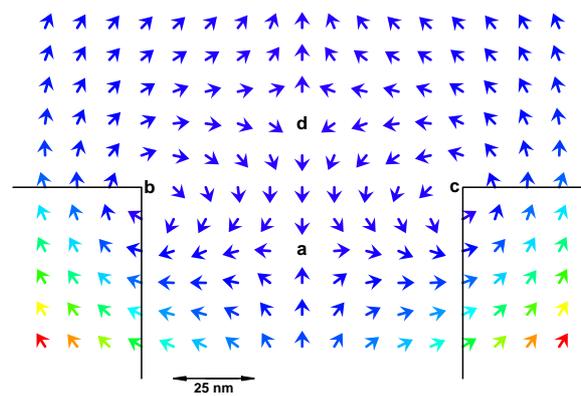


Figure 3.14: Detail of the time-averaged Poynting vector field near a 100 nm wide slit in a 100 nm thick silicon plate. The depicted region is indicated in Fig. 3.11.

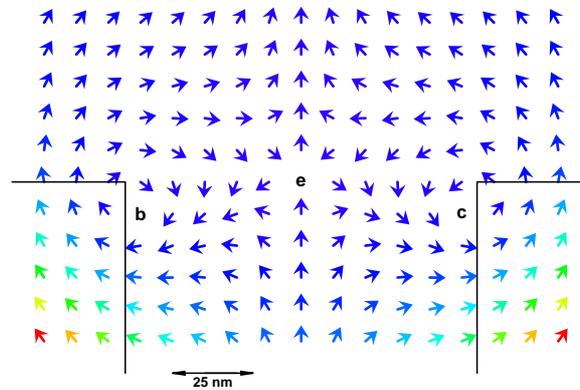


Figure 3.15: The time-averaged Poynting vector near a 109.5 nm wide slit in a 100 nm thick silicon plate. A monkey saddle (e) with topological charge  $-2$  can be seen.

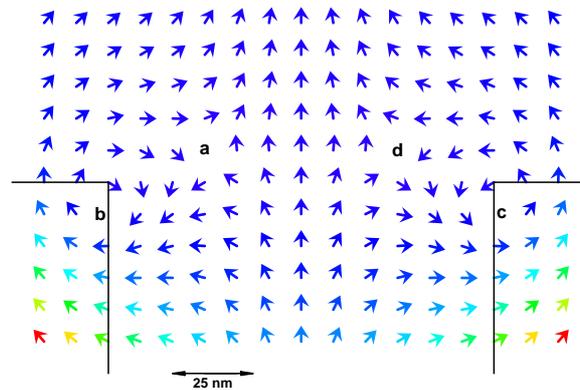


Figure 3.16: The time-averaged Poynting vector near a 120 nm wide slit in a 100 nm thick silicon plate. The monkey saddle (visible in Fig. 3.15) has decayed into two saddle points (a and d).

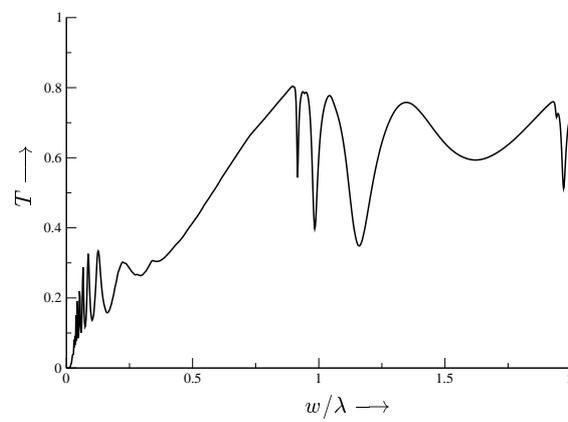


Figure 3.17: The transmission coefficient  $T$  of a narrow slit in a  $2 \mu\text{m}$  thick silver plate as a function of the slit width  $w$ , expressed in wavelengths. The wavelength is  $\lambda = 500 \text{ nm}$ , and the refractive index is taken as  $n_{\text{silver}} = 0.05 + i2.87$ . The slit is illuminated by a H-polarized, normally incident plane wave.

# Chapter 4

## The Radiation Pattern of a Single Slit in a Metal Plate

This Chapter is based on the following publication:

- H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “Connection between phase singularities and the radiation pattern of a slit in a metal plate”, *Physical Review Letters*, vol. 93, 173901 (2004).

### **Abstract**

We report a new fundamental relation between the minima of the far-zone radiation pattern of a narrow slit in a metal plate and the location of phase singularities in the intermediate field. If a system parameter such as the wavelength is changed, a previously unappreciated singular optics phenomenon occurs, namely the transition of a near-zone phase singularity into a singularity of the radiation pattern. Our results have significance for the design of novel nano-scale light sources and antennas.

## 4.1 Introduction

Surprising effects in the radiation pattern of sub-wavelength light sources have recently been observed in several studies, for example the “beaming effect” due to surface plasmons on grating structures around sub-wavelength apertures [LEZEC *et al.*, 2002; MARTÍN-MORENO *et al.*, 2003]. In this Chapter we report the prediction of a new phenomenon, namely the relationship between phase singularities in the near-field and intermediate field of the aperture, and the shape of the far-zone radiation pattern. Furthermore, a new singular optics process is described: the transition of a near-zone singularity into a singularity of the radiation pattern.

In previous articles [SCHOUTEN *et al.*, 2003a; SCHOUTEN *et al.*, 2003b; SCHOUTEN *et al.*, 2004] (see Chapter 3), a connection was made between the transmission coefficient of a sub-wavelength slit and phase singularities of the time-averaged Poynting vector field on the illuminated side of the metal plate. We now show that the phase singularities of the electric or magnetic field on the “dark” side of the metal plate are intimately connected with the radiation properties of the slit. More specifically, we find phase singularities in the intermediate field whose position is directly connected with the location of minima in the far-field radiation pattern. Changing a system parameter, such as the thickness of the metal plate or the wavelength of the incident field, can cause these phase singularities to move further away from the metal plate, corresponding to a minimum in the radiation pattern becoming more pronounced. Surprisingly, on further changing the parameter, these phase singularities can move all the way to the far-field and eventually disappear at infinity. This disappearance of a phase singularity is directly related to the occurrence of an exact zero of the radiation pattern, i.e., a phase singularity at infinity.

## 4.2 The definition of the radiation pattern

The configuration under consideration is the same as the one discussed in the preceding Chapter, and it is illustrated in Fig. 3.1: a plane monochromatic electromagnetic wave is incident upon a metal plate of thickness  $d$  and permittivity  $\epsilon_{\text{plate}}$ . A single slit of width  $w$ , infinitely long in the  $y$ -direction, is present in the plate. Because the system is invariant with respect to  $y$ -translations, we may treat the problem as two-dimensional, with relevant coordinates  $x$  and  $z$ . A rigorous scattering approach, which takes into account the finite conductivity and the finite thickness of the plate, is used to calculate the field in the vicinity of an infinitely long slit in a metal plate. This method, commonly referred to as the Green’s tensor method, is described in detail in Chapter 2.

The intensity radiation pattern is defined by the expression

$$I(\theta) \equiv \frac{\lim_{\rho \rightarrow \infty} \langle \mathbf{S}^{(\text{sca})}(\rho, \theta) \rangle \cdot \boldsymbol{\rho}}{\int_{\text{slit}} \langle S_z^{(0)} \rangle dx}, \quad (4.1)$$

with  $\boldsymbol{\rho} = (x, 0, z)$  and  $\cos \theta = z/\rho$ . Furthermore,  $\langle \mathbf{S}^{(\text{sca})} \rangle$  is the time-averaged Poynting vector associated with the scattered field (i.e., the field minus the incident field) and  $\langle S_z^{(0)} \rangle$  is the component of the time-averaged Poynting vector associated with the illuminating field (i.e., the field emitted by the laser source) that is perpendicular to the plate. For convenience we take the amplitude of the illuminating electric field equal to unity.

The radiation pattern may be calculated using the angular spectrum representation of the field. Let  $U$  be a component of the scattered electric or magnetic field. Because  $U$  satisfies the Helmholtz equation it can, for points in the half space  $z > d$  (where  $z = d$  is taken to be the exit plane of the slit) be written in the form of an angular spectrum [MANDEL AND WOLF, 1995, Sec. 3.2], i.e.,

$$U(x, z) = \int_{-\infty}^{\infty} \tilde{U}(k_x, z = d) e^{ik_x x} e^{ik_z(z-d)} dk_x, \quad (4.2)$$

where  $\tilde{U}$  is the (inverse) Fourier transform of  $U$  with respect to  $x$ , i.e.,

$$\tilde{U}(k_x, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} U(x, z) e^{-ik_x x} dx, \quad (4.3)$$

and  $k_z = \sqrt{k^2 - k_x^2}$ , where  $k$  is the wavenumber. For points far away from the slit, the method of stationary phase (see Sec. 3.3 of [MANDEL AND WOLF, 1995] or Chap. 8 of [STAMNES, 1986]) may be applied to Eq. (4.2) to obtain the asymptotic expression

$$U(x, z) \sim \sqrt{2\pi k \cos \theta} \tilde{U}(k \sin \theta, d) \frac{e^{ik\rho}}{\sqrt{\rho}} e^{-i\pi/4} \quad (k\rho \rightarrow \infty). \quad (4.4)$$

To calculate the intensity radiation pattern, we take for  $U$  in Eq. (4.4) the components of the scattered electric and magnetic field, and apply the result to Eq. (4.1). In this manner one obtains for the two polarizations the formulae

$$I_E(\theta) = \frac{2\pi k}{w} \cos^2 \theta |\tilde{E}_y(k \sin \theta, d)|^2. \quad (4.5)$$

$$I_H(\theta) = \frac{2\pi k}{w} \frac{\mu_0}{\varepsilon_0} \cos^2 \theta |\tilde{H}_y(k \sin \theta, d)|^2, \quad (4.6)$$

with  $\mu_0$  the vacuum permeability,  $\varepsilon_0$  the vacuum permittivity and  $w$  the slit width.

We now demonstrate the consistency of our definition with the definition of the transmission coefficient  $T$  prescribed in Eq. (3.7) of Chapter 3, i.e., it is to be shown that

$$T = \int_{-\pi/2}^{\pi/2} I_H(\theta) + I_E(\theta) d\theta. \quad (4.7)$$

For a metal plate with a thickness significantly larger than the skin depth, the transmission coefficient given by Eq. (3.7) can be approximated by

$$T = \frac{\int_{-\infty}^{\infty} \langle S_z^{(\text{sca})} \rangle_{z=d} dx}{\int_{\text{slit}} \langle S_z^{(0)} \rangle_{z=0} dx}, \quad (4.8)$$

where  $\langle \mathbf{S}^{(\text{sca})} \rangle$  is the time-averaged Poynting vector of the scattered field. If Parseval's theorem is applied to Eq. (4.8), one obtains

$$T = \frac{2\pi}{c\varepsilon_0 w} \int_{-\infty}^{\infty} \text{Re} \left[ \tilde{E}_x^{(\text{sca})}(k_x, d) \tilde{H}_y^{(\text{sca})*}(k_x, d) - \tilde{E}_y^{(\text{sca})}(k_x, d) \tilde{H}_x^{(\text{sca})*}(k_x, d) \right] dk_x, \quad (4.9)$$

where Eqs. (1.45) and (4.2) were used. Furthermore, it was used that

$$\int_{\text{slit}} \langle S_z^{(0)} \rangle dx = \frac{1}{2} c\varepsilon_0 w. \quad (4.10)$$

From the Maxwell equations (1.47) and (1.51), together with Eq. (4.2), it follows that

$$\tilde{E}_x^{(\text{sca})}(k_x, d) = \frac{k_z}{\omega\varepsilon_0} \tilde{H}_y^{(\text{sca})}(k_x, d), \quad (4.11)$$

$$\tilde{H}_x^{(\text{sca})}(k_x, d) = -\frac{k_z}{\omega\mu_0} \tilde{E}_y^{(\text{sca})}(k_x, d). \quad (4.12)$$

Inserting these equations into (4.9) gives

$$T = \frac{2\pi}{c\varepsilon_0 w} \int_{-k}^k \frac{k_z}{\omega} \left[ \frac{1}{\varepsilon_0} |\tilde{H}_y^{(\text{sca})}(k_x, d)|^2 + \frac{1}{\mu_0} |\tilde{E}_y^{(\text{sca})}(k_x, d)|^2 \right] dk_x, \quad (4.13)$$

where the fact is used that  $k_z = \sqrt{k^2 - k_x^2}$  is real for  $|k_x| < k$  and purely imaginary for  $|k_x| > k$ . On making the change of variables  $k_x = k \sin \theta$  one obtains

$$T = \frac{2\pi k}{w} \int_{-\pi/2}^{\pi/2} \cos^2 \theta \left( \frac{\mu_0}{\varepsilon_0} |\tilde{H}_y^{(\text{sca})}(-k\theta, d)|^2 + |\tilde{E}_y^{(\text{sca})}(-k\theta, d)|^2 \right) d\theta, \quad (4.14)$$

which corresponds with Eq. (4.7). This proves the consistency of the definition of the intensity radiation pattern with our earlier definition of the transmission coefficient.

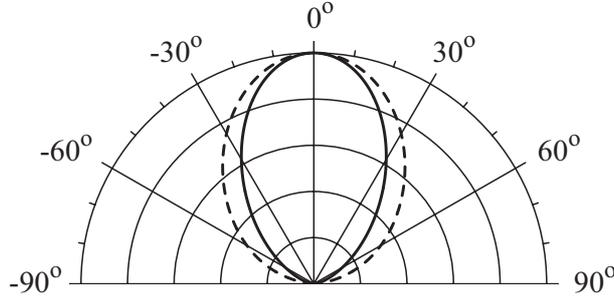


Figure 4.1: Polar plot of the (normalized) intensity radiation pattern of a 250 nm wide slit in a 100 nm thick silver plate with refractive index  $n = 0.05 + i2.87$ . The incident field has a wavelength  $\lambda = 500$  nm and is E-polarized (solid line) or H-polarized (dashed line).

### 4.3 The radiation pattern of a slit for normal incidence

In Fig. 4.1 the radiation pattern of a sub-wavelength slit is plotted both for E-polarized (i.e., with the electric field is parallel to the slit) and H-polarized fields. The pattern is found to be rather uniform, i.e., the light is diffracted into every direction, with no minima being present in the radiation pattern. This is typical for sub-wavelength slits, for both polarization cases — for such slits there is only one (propagating) guided mode possible.

The radiation pattern is drastically changed if the slit width is increased, because then more modes in the slit become propagating. In the configuration of Fig. 4.1, the second symmetric mode is propagating for a slit width  $w > 1.4\lambda$  for E-polarized fields, or  $w > 0.9\lambda$  for H-polarized fields [SCHOUTEN *et al.*, 2003a]. (No anti-symmetric modes are excited because the illuminating field is symmetric.) This implies that for such slit widths the intensity radiation pattern as a function of the plate thickness  $d$  is approximately given by (see Eqs. (4.5) and (4.6))

$$I(\theta) \sim \cos^2 \theta \left| \tilde{U}^{(1)}(k \sin \theta) + \tilde{U}^{(2)}(k \sin \theta) e^{i(k_2 - k_1)d} \right|^2, \quad (4.15)$$

with  $\tilde{U}^{(i)}$  the Fourier transform with respect to  $x$  of a field component ( $\hat{E}_y$  for E-polarization, and  $\hat{H}_y$  for H-polarization) at the exit plane of the slit of mode  $i$ , and  $k_i$  the effective wave number of mode  $i$  ( $i = 1, 2$ ). The exponential of the second term in Eq. (4.15) is periodic in  $k$  or  $d$  leading to alternate constructive and destructive interference of the two terms, which results in the radiation pattern

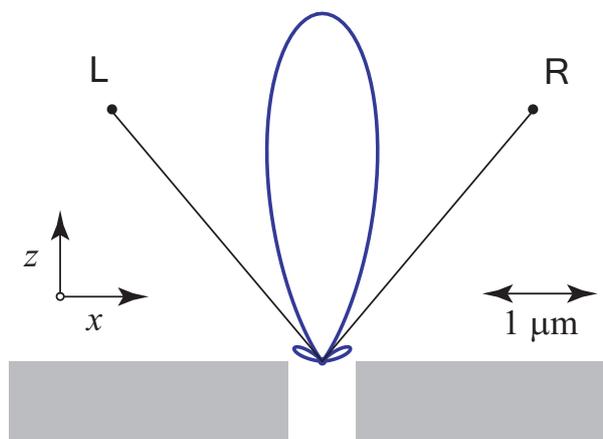


Figure 4.2: Polar plot of the intensity radiation pattern of a 600 nm wide slit in a 750 nm thick silver plate (shaded region) with refractive index  $n = 0.05 + i2.87$ . Also the positions of the phase singularities of  $\hat{H}_y$  in the region behind the slit are shown. L denotes a left-handed phase singularity, and R denotes a right-handed phase singularity. The incident light (coming from below) has a wavelength  $\lambda = 500$  nm and is H-polarized.

periodically changing from having minima to having no minima. In the case that there is a minimum, the radiation pattern is, of course, more directional.

If the field ( $\hat{E}_y$  for E-polarization and  $\hat{H}_y$  for H-polarization) behind the slit is plotted for cases such that the radiation pattern has minima, one typically finds a phase singularity *in the direction of the radiation minimum*, see Figs. 4.2 and 4.3.<sup>1</sup> In both figures two phase singularities are present, one with charge +1 (left-handed) and one with charge  $-1$  (right-handed). In Fig. 4.2 (H-polarization) phase singularities in  $\hat{H}_y$  are shown, whereas in Fig. 4.3 (E-polarization) phase singularities in  $\hat{E}_y$  are shown. For both cases, each of these phase singularities gives rise to a vortex (circulation) in the field of power flow. The connection between these two kinds of phase singularities is explained in detail in [SCHOUTEN *et al.*, 2003b]. The phase singularities in Figs. 4.2 and 4.3 are found by numerically calculating the topological charge (see Eq. (1.87)). In this way, we can determine if there actually is a phase singularity present, or just a minimum in the intensity.

In Figs. 4.3 and 4.4 the phase singularities near the slits as well as the radiation pattern are plotted for two different values of the wavelength of the incident field. It is seen that for  $\lambda = 510$  nm (Fig. 4.3) the phase singularities are closer to the plate and the minima are less pronounced than for  $\lambda = 500$  nm (Fig. 4.4).

<sup>1</sup>For a detailed description of phase singularities see Sec. 1.8.

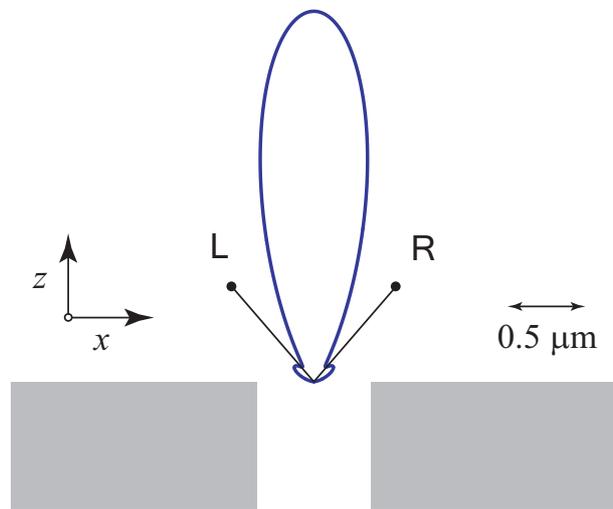


Figure 4.3: The intensity radiation pattern of a 750 nm wide slit in a 860 nm thick silver plate ( $n = 0.05 + i2.87$ ). Also the positions of the phase singularities of  $\hat{E}_y$  behind the slit are shown. The incident light has a wavelength  $\lambda = 510$  nm and is E-polarized.

This behavior was found to be typical for this kind of configuration (see, e.g., Fig. 4.2 for another case with pronounced minima and phase singularities behind the slit). If the wavelength is decreased even more, a curious phenomenon is observed: the phase singularities move away towards the far-zone, and at a certain value of the wavelength it is no longer possible to track them numerically. The minima in the intensity radiation pattern correspondingly become deeper with decreasing wavelength until at a certain critical wavelength, the minima evidently become true zeros. For wavelengths smaller than this critical wavelength, the value of the intensity at the minima rises as a function of decreasing wavelength. To investigate this behavior more closely, we quantify the phase behavior at infinity by first introducing for  $\hat{E}_y$  (or for  $\hat{H}_y$  for H-polarized fields) the *reduced field*  $\hat{E}_y^{\text{red}}$  defined by the expression

$$\hat{E}_y(\theta, \rho) = \frac{e^{ik\rho}}{\sqrt{\rho}} \hat{E}_y^{\text{red}}(\theta, \rho). \quad (4.16)$$

Next we take the limit for the phase  $\phi^{\text{red}}(\rho, \theta)$  of  $\hat{E}_y^{\text{red}}$ ,

$$\phi^{\text{inf}}(\theta) \equiv \lim_{\rho \rightarrow \infty} \phi^{\text{red}}(\rho, \theta) = \text{Arg} \left[ \tilde{E}_y(k \sin \theta) \right] - \pi/4, \quad (4.17)$$

where we used Eq. (4.4). Near the wavelength where the intensity of the radiation pattern for one particular angle of observation is almost zero, the phase at infinity shows the behavior plotted in Fig. 4.5. It is seen that for a wavelength  $\lambda = 490$  nm there is a rapid *increase* of the phase by  $\pi$  near the angle  $\theta \approx 36^\circ$  where a minimum of the radiation pattern occurs, whereas for  $\lambda = 489$  nm the phase rapidly *decreases* by an amount of  $\pi$  near the same angle. A similar behavior was found for several configurations for the H-polarization case.

A physical interpretation of this surprising behavior is provided by noting that in principle two cases can occur: either the phase singularities remain present at a finite distance from the slit (apparently far from it); or the phase singularities *disappear* at infinity. It is instructive to introduce the function  $(\theta, \rho) \rightarrow (\theta, \arctan \rho)$ , which maps the upper half-space behind the slit into the half-disk  $\{(\theta', \rho') : -\pi/2 \leq \theta' \leq \pi/2, 0 \leq \rho' = \arctan \rho \leq \pi/2\}$ . The phase of the reduced field  $\hat{E}_y^{\text{red}}$  is well defined on this space, even on the boundary  $\rho' = \pi/2$ , where it is given by the limit value  $\phi^{\text{inf}}$  of Eq. (4.17).

The disappearance of a phase singularity at infinity can be observed in the behavior of the phase at infinity  $\phi^{\text{inf}}$ . This effect is illustrated in Fig. 4.6: if the phase singularity is present at a large, but finite distance from the slit, the situation (a) on the left-hand side of the figure applies: near the angle where the phase singularity is present, the phase at infinity increases rapidly, but continuously, by  $\pi$  when the angle of observation  $\theta$  is increased (it increases by  $\pi$  if the topological

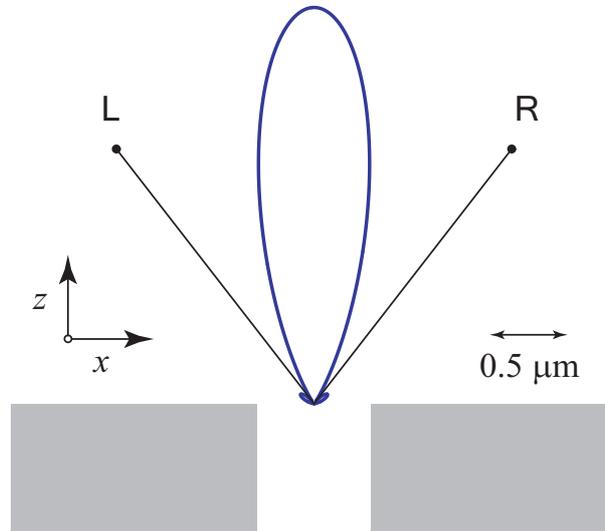


Figure 4.4: The intensity radiation pattern of a 750 nm wide slit in a 860 nm thick silver plate ( $n = 0.05 + i2.87$ ). Also the position of the phase singularities of  $\hat{E}_y$  behind the slit are shown. The incident light has a wavelength of 500 nm and is E-polarized.

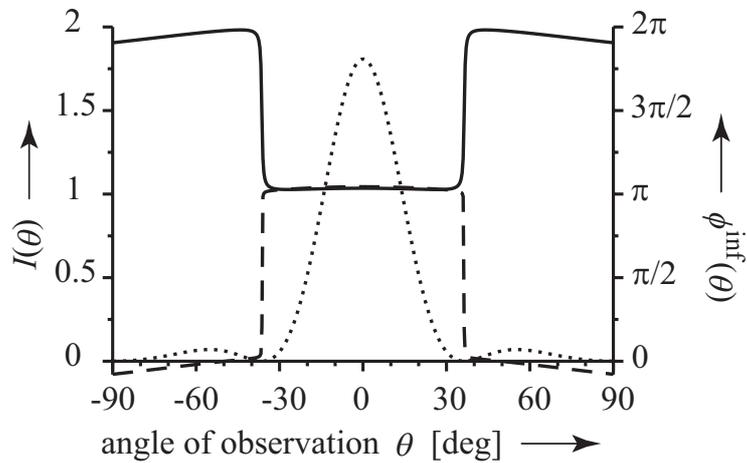


Figure 4.5: The intensity radiation pattern (dotted curve) and the far-field phase behavior of a 750 nm wide slit in a 860 nm thick silver plate ( $n = 0.05 + i2.87$ ). The incident light is E-polarized and has a wavelength of  $\lambda = 490 \text{ nm}$  (solid line) or  $\lambda = 489 \text{ nm}$  (dashed line).

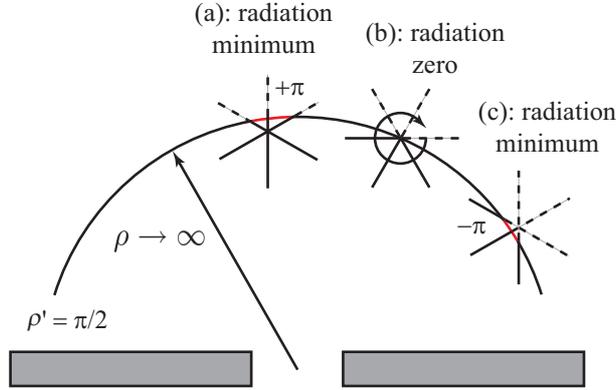


Figure 4.6: Illustration of the disappearance of a phase singularity at infinity when the wavelength of the field is gradually decreased. A sketch of the equiphase lines of  $\phi^{\text{red}}(\rho', \theta)$  is drawn, which on the semi-circle  $\rho' = \pi/2$  takes on the value  $\phi^{\text{inf}}(\theta)$ . Three cases are depicted: in (a) the phase singularity is still present, corresponding to a radiation minimum; in (b) the phase singularity is exactly at infinity, corresponding to a radiation zero; in (c) the phase singularity is no longer present, corresponding to a radiation minimum. In this example the singularity is taken to have a topological charge  $s = -1$ .

charge is  $-1$ , whereas it decreases by  $\pi$  if the topological charge is  $+1$ ). If the wavelength is decreased, the phase singularity can be exactly at infinity, i.e., at the boundary  $\rho' = \pi/2$ , as is shown in the middle (case (b)) of Fig. 4.6. In this case there is an exact zero in the radiation pattern, together with a  $\pi$  phase jump at this point. If the wavelength is decreased even further, then the right-hand side of Fig. 4.6 applies (case (c)): the phase singularity does not exist anymore. However, a “residual effect” can still be observed in the phase behavior at infinity: the phase rapidly changes by  $\pi$ . Note that if initially there was a  $\pi$  *increase* when the angle  $\theta$  is increased, now there is a  $\pi$  *decrease*.

It is to be further noted that this phase behavior at infinity is only possible if some phase singularity “crosses” the boundary  $\rho' = \pi/2$ , i.e.,  $\rho = \infty$ . However, it is impossible to distinguish between a phase singularity with charge  $s = -1$  *disappearing* at infinity, or a phase singularity with charge  $s = +1$  *appearing* at infinity. Because in our results (see Fig. 4.5) there was a phase singularity moving to the far-zone, this suggests that this phase singularity has disappeared at infinity, as discussed above. Note that this implies that the topological charge is *not* (locally) conserved. However, because of the mirror symmetry of the system, the positive and negative charges at infinity disappear together — that is, the total topological charge of the field is conserved, even though the singularities disappear

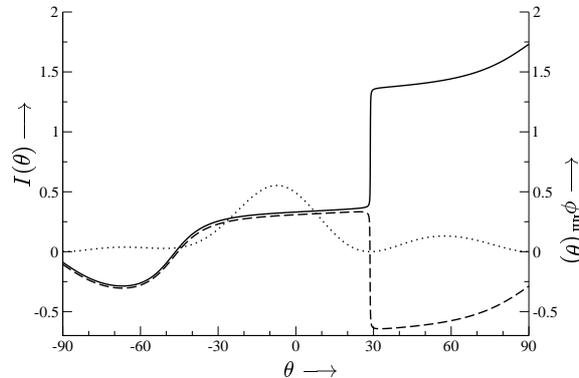


Figure 4.7: The intensity radiation pattern (dotted curve) and the far-field phase behavior of a 600 nm wide slit in a 600 nm (solid line) or 595 nm (dashed line) thick silver plate ( $n = 0.05 + i2.87$ ). The incident light is H-polarized and has a wavelength of  $\lambda = 500$  nm and an angle of incidence of  $10^\circ$ .

at widely separated spatial locations. These results suggest a new singular optics phenomenon, in which a singularity in the near- or intermediate-zone (a singularity in two-dimensional  $(x, z)$ -space) converts into a singularity in the radiation pattern (one-dimensional  $\theta$ -space). It is to be noted that other researchers [SOSKIN *et al.*, 1997] have studied the motion of phase singularities as system parameters are changed, in what is usually referred to as the theory of combined beams. However, that work, unlike our study, does not discuss the effects of phase singularities on the radiation pattern of the field.

## 4.4 Oblique angle of incidence

In the previous Section it was found that due to the symmetry in the  $x$ -direction, phase singularities were always disappearing in pairs. In this Section the behavior of the radiation pattern of a slit in a metal plate illuminated by an oblique incident plane wave is studied. The mirror-symmetry in the  $x$ -direction is then broken, and one can determine if the disappearance of a *single* phase singularity is possible.

In Fig. 4.7 the intensity radiation pattern is plotted for two, closely spaced, values of the thickness of the metal plate for an H-polarized plane wave incident under an angle of  $10^\circ$  with the normal of the plate. It can be seen that near  $\theta = 29^\circ$  there is indeed only one phase singularity appearing or disappearing.

## 4.5 Conclusions

We have shown that there is a connection between minima in the far-zone radiation pattern and phase singularities in the intermediate field. On changing a system parameter, such as the wavelength of the incident field, these singularities can move to infinity and become singularities of the radiation pattern. It was also shown that a single phase singularity could disappear at infinity, thereby breaking global topological charge conservation. A possible way to experimentally verify the predicted relation between phase singularities and the radiation pattern is to place a thin nano-wire at a phase singularity and align it parallel to the slit. The presence of the wire should not affect the observed radiation pattern in the far zone [LANDSTORFER *et al.*, 1972]. These results provide a new insight into the behavior of radiation patterns, and are therefore important for the design of nano-scale light sources and antennas [LANDSTORFER *et al.*, 1972; DE LA FUENTE, 1975].

# Chapter 5

## Plasmon-assisted Light Transmission through Two Slits

This Chapter is based on the following publication:

- H.F. Schouten, N. Kuzmin, G. Dubois, T.D. Visser, G. Gbur, P.F. Alkemade, H. Blok, G.W. 't Hooft, D. Lenstra and E.R. Eliel, “Plasmon-assisted two slit transmission: Young’s experiment revisited”, *Physical Review Letters*, vol. 94, 053901 (2005).

### **Abstract**

We present a theoretical study of the optical transmission of a thin metal screen perforated by two parallel sub-wavelength slits, separated by many optical wavelengths. The total intensity of the far-field double-slit pattern is shown to be reduced or enhanced as a function of the wavelength of the incident light beam. This modulation is attributed to an interference phenomenon at each of the slits, instead of at the detector. The interference arises as a consequence of the excitation of surface plasmons propagating from one slit to the other. Experimental data which confirm our predictions are also presented.

## 5.1 Introduction

Recently, there has been a surge of interest in the phenomenon of light transmission through sub-wavelength apertures in metal plates. This followed the observation by Ebbesen *et al.* [1998] that the transmission through a two-dimensional hole array can be much larger than predicted by conventional diffraction theory [BETHE, 1944]. This discovery has rekindled the interest in a similar but simpler problem, *viz.*, the transmission of a one-dimensional array of sub-wavelength slits in a metal film, *i.e.*, of a metal grating [EBBESEN *et al.*, 1998; SCHRÖTER AND HEITMANN, 1998; SOBNACK *et al.*, 1998; PORTO *et al.*, 1999; WENT *et al.*, 2000; ASTILEAN *et al.*, 2000; TAKAKURA, 2001; TREACY, 2002; CAO AND LALANNE, 2002; LALANNE *et al.*, 2003; BARBARA *et al.*, 2003; SUCKLING *et al.*, 2004; GARCÍA-VIDAL *et al.*, 2003; LEZEC AND THIO, 2004]. In many cases the enhanced transmission of hole or slit arrays has been explained in terms of the excitation of (coupled) surface plasmons on the metal film [SCHRÖTER AND HEITMANN, 1998; SOBNACK *et al.*, 1998; PORTO *et al.*, 1999; WENT *et al.*, 2000], an explanation that has recently been challenged [LEZEC AND THIO, 2004]. It has been shown that, for slit arrays, Fabry-Pérot-type waveguide resonances can also give rise to a considerably enhanced transmission [PORTO *et al.*, 1999; ASTILEAN *et al.*, 2000; TREACY, 2002; CAO AND LALANNE, 2002; BARBARA *et al.*, 2003].

In the present Chapter we study an even more fundamental system than the metallic grating, namely, a thin metal layer perforated by just two parallel sub-wavelength slits. In contrast to the systems that have recently attracted so much attention, our slits are separated by many optical wavelengths. Thus we study the light transmission of a setup that lies at the heart of wave physics, namely, that of Thomas Young [YOUNG, 1802; YOUNG, 1845]. We do, however, not focus on the well-known interference pattern named after him, but on the angle-integrated power transmission coefficient of the perforated screen, *i.e.*, the transmission integrated over many interference orders. We show that this transmission coefficient is strongly modulated as a function of the wavelength of the incident light for the case that that light is H-polarized, *i.e.*, with the electric field aligned perpendicular to the slits. In contrast, there is no such modulation when the incident light is E-polarized, or when the “wrong” metal is chosen. All our observations can be explained in terms of a model involving the coherent transport of electromagnetic energy between the slits by surface plasmons.

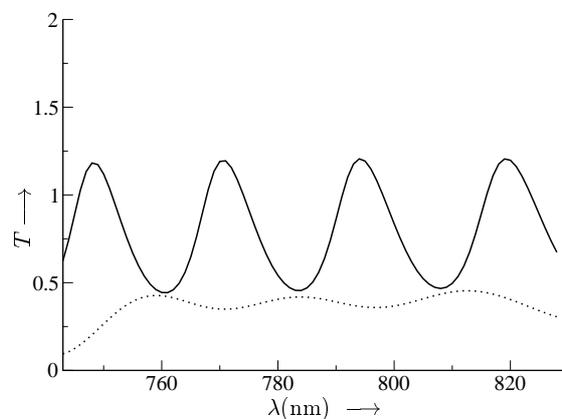


Figure 5.1: The transmission coefficient of a double slit in a 200 nm thick gold film as a function of the wavelength of the incident light. The slits are 200 nm wide and separated by 25  $\mu\text{m}$ . The full line displays the results for H polarization, while the dotted line (magnified 10 times) shows the results for the case of E polarization.

## 5.2 Plasmon-assisted transmission through two slits

We calculate the transmission of the double-slit system using a rigorous scattering model based on a Green's function approach, described in Chapter 2. The illuminating field is taken to be monochromatic and propagating perpendicular to the plate. The polarization of the electric field is taken as either perpendicular to the slits (H-polarization) or parallel to the slits (E-polarization). The H-polarization, in contrast to the E-polarization, permits the excitation of surface plasmons on the metal-air interfaces (see Sec. 1.7.2). The transmission coefficient is normalized on the geometrical optical transmission through the two slits (see Eq. (3.7)).

In Fig. 5.1 the transmission-coefficient of the two-slit configuration is shown as a function of the wavelength of the incident radiation. The wavelength dependence of the dielectric constant of the gold film is fully taken into account [JOHNSON AND CHRISTY, 1972]. When the incident field is E-polarized, the transmission of the double slit is small and weakly modulated as a function of wavelength. In contrast, for an H-polarized incident field, the transmission shows a strong modulation as a function of wavelength. In Fig. 5.2 a similar effect can be observed for different separation distances between the two slits.

The strong polarization anisotropy and the dependence on the screen material (as will be discussed later in the Section) both suggest that surface plasmons propagating along the gold-air interface lie at the heart of the phenomena. Expla-

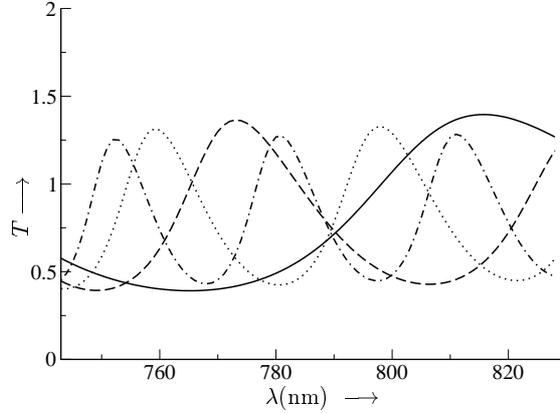


Figure 5.2: The transmission coefficient of a double slit in a 200 nm thick gold film as a function of the wavelength of the H-polarized incident light. The slits are 200 nm wide. The full line displays the results for 5  $\mu\text{m}$  separation between the slits, the dashed line for 10  $\mu\text{m}$ , the dotted line for 15  $\mu\text{m}$  and the dashed-dotted line for 20  $\mu\text{m}$ .

nations in terms of waveguide modes within the slit [PORTO *et al.*, 1999; ASTILEAN *et al.*, 2000; TREACY, 2002; CAO AND LALANNE, 2002; BARBARA *et al.*, 2003] or diffractive evanescent waves [LEZEC AND THIO, 2004] are excluded by the observed *dependence* of the spectral modulation period and the *independence* of the modulation depth on the slit separation.

The propagation constant  $k_{sp}$  of a surface plasmon is given by (see Eq. (1.79))

$$k_{sp} = k_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}}, \quad (5.1)$$

where  $\varepsilon_m$  and  $\varepsilon_d$  are the complex (relative) dielectric constants of the metal and dielectric, respectively, and  $k_0 = 2\pi/\lambda$  the free-space wavenumber. The surface-plasmon wavelength is related to the real part of  $k_{sp}$  by  $\lambda_{sp} = 2\pi/\text{Re}(k_{sp}) = \lambda_0/n_{sp}$ , while its (amplitude) decay length is given by  $1/\text{Im}(k_{sp})$ . For the gold-air interface at  $\lambda_0 = 800$  nm,  $n_{sp} = 1.02$  and  $1/\text{Im}(k_{sp}) = 80$   $\mu\text{m}$ , considerably larger than the separation of any pair of slits used in the simulation. Consequently, surface plasmons propagating along such an interface can easily cover the distance between the slits.

The function of the slits is threefold. First, the slits transmit part of the incident radiation, together giving rise to a conventional Young's-type interference pattern. Second, each slit scatters part of the incident radiation into a plasmonic channel, bridging the momentum gap between the surface plasmon and free-space radiation.

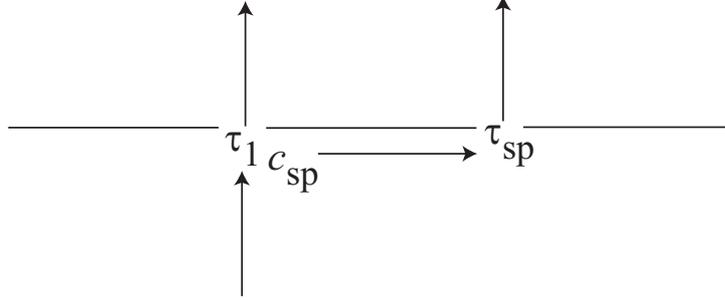


Figure 5.3: Schematic illustration of the heuristic model. For simplicity's sake, the plasmon generation and subsequent processes are only shown at one slit.

Third, each slit provides a mechanism for the back-conversion of a surface plasmon into free-space radiation. In Fig. 5.3 an illustration of these effects is given.

When the incident light is H-polarized the surface plasmon that is excited at one of the slits propagates towards its partner slit. There it is partially back-converted into light (see Fig. 5.3). The plasmon-mediated amplitude at this slit interferes with the amplitude of the light that is directly transmitted by that slit. Because of the sub-wavelength nature of our slits these two contributions are of comparable magnitude.

The field amplitude at the second slit's unilluminated side can now be written as

$$E_{\text{slit}}^{(2)} = E_1(\tau_1 + c_{\text{sp}}e^{ik_{\text{sp}}b}\tau_{\text{sp}}) = E_0\tau_1(1 + Ae^{ik_{\text{sp}}b}), \quad (5.2)$$

where  $b$  is the slit separation,  $c_{\text{sp}}$  is the coupling coefficient between the incident light and the surface plasmon,  $\tau_{\text{sp}}$  is the coupling coefficient between the surface plasmon and the transmitted light,  $A = c_{\text{sp}}\tau_{\text{sp}}/\tau_1$  is the (complex-valued) relative strength of the plasmon contribution,  $\tau_1$  is the coupling between the incident field and the transmitted field at the slit in the case that the other slit is not present and  $E_1$  is a constant needed for the conversion between the dimensionless transmission entities and the electric field. The normalization is such that the single slit transmission coefficient  $T_s = |\tau_1|^2$ . Note that a phase factor associated with the propagating field to plasmon scattering and back-conversion processes is taken into account by taking  $c_{\text{sp}}$  and  $\tau_{\text{sp}}$  to be complex-valued. The field amplitude  $E_{\text{slit}}^{(2)}$  behind the second slit is thus enhanced or suppressed, depending on the argument of the complex phase factor in Eq. (5.2). Because the laser beam is normally incident on the configuration and symmetrically illuminates the two slits, the field amplitude behind the first slit is given by  $E_{\text{slit}}^{(1)} = E_{\text{slit}}^{(2)}$ .

In the present setup the far-field two-slit pattern arises as a consequence of the interference of *four* paths, two of which are partially plasmonic, while the other

two are photonic all the way. Although the number of interfering channels is four in the present configuration, the far-field pattern that arises behind the plate is simply that of Young's experiment, i.e., a pattern of two interfering sources. The novel aspect is that the strength of each of these sources is enhanced or reduced due to the interference of a photonic and a plasmonic channel. The transmission coefficient<sup>1</sup> is equal to the power radiated by each slit separately, i.e.,

$$T = |\tau_1|^2 [1 + |A|^2 + 2|A| \cos(k_{sp}b + \arg(A))], \quad (5.3)$$

where we used Eq. (5.2) and the fact that the power transmitted by each slit is proportional with the field strength at the slit. Note that this equation indeed shows the oscillating dependence observed in Fig. 5.1.

Using the theoretical Green tensor model outlined above we have also calculated the intensity distribution, i.e., the value of  $|E|^2$ , on both sides of a freestanding perforated gold film (see Fig. 5.4). For calculational convenience we have taken values of the slit separation that are considerably smaller than those of the previous simulations, viz.  $5\lambda_{sp}/2$ , where the transmission is maximum, and  $4\lambda_{sp}/2$ , where the transmission is minimum. In the first case (maximum transmission) one can distinguish at the dark side of the metal film a well-developed standing-wave pattern along the interface, having six antinodes, two of which coincide with the slits themselves. In contrast, when the transmission is minimum the antinodes of the standing-wave pattern do not coincide with the slits; at these locations one rather finds a node of the standing-wave pattern. In both cases the intensity is seen to rapidly decay away from the air-metal interface.

The prediction of plasmon-assisted transmission was experimentally tested by Eliel *et al.* (Leiden University), see [SCHOUTEN *et al.*, 2005]. The sample consists of a 200 nm thick gold film, evaporated on top of a 0.5 mm thick fused quartz substrate with a 10 nm thick titanium adhesion layer between the gold and the glass. In the sample a two-slit pattern is written using a focused ion beam, each slit being 50  $\mu\text{m}$  long and 0.2  $\mu\text{m}$  wide. The slits are separated by a distance, as measured with a scanning electron microscope, of 4.9, 9.9, 14.8, 19.8, or 24.5  $\mu\text{m}$ , respectively. Such a two-slit pattern, with the metallized side facing the laser, is illuminated at normal incidence with the well-collimated output beam (2mm diameter) of a narrow-band cw Ti:sapphire laser, tunable between 740 and 830 nm. Detection takes place in transmission, integrating the double-slit pattern (shown at the top of Fig. 5.5) over a large number of interference orders. The zeroth order peak is considerably stronger than the other orders because of non-negligible leakage through the bulk metal, and is therefore blocked by an opaque

---

<sup>1</sup>Eq. (3.7) is used for the definition of the transmission coefficient. In this equation the integrals over the "slit" are replaced with integrals over the "slits", i.e., the normalization is such that it equals one in the geometric-optical limit.

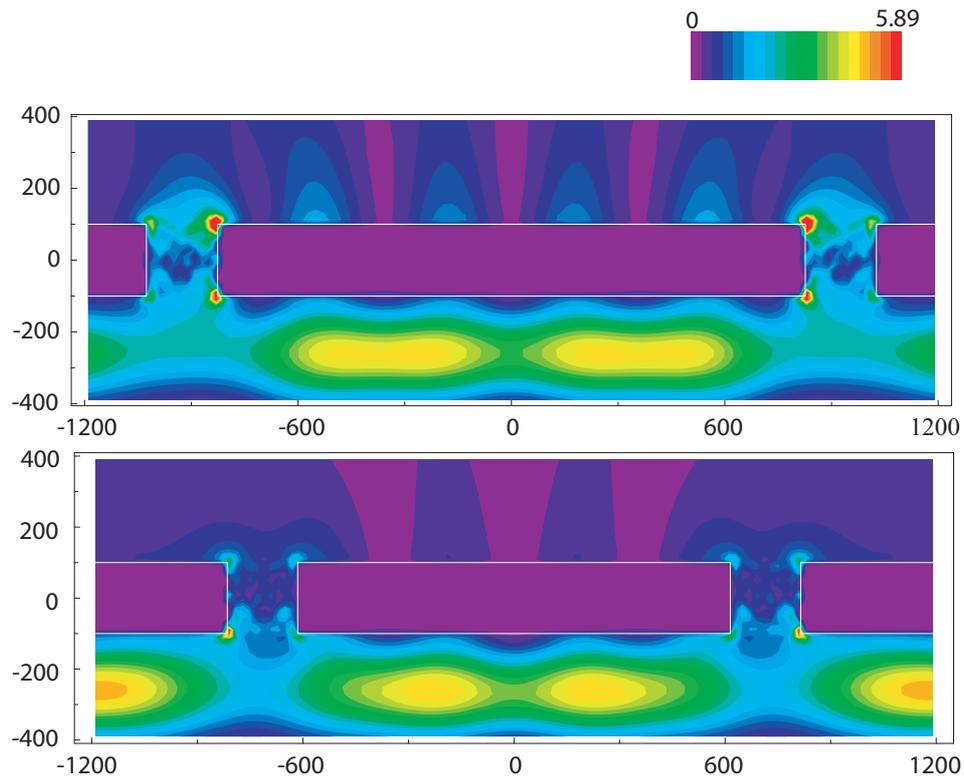


Figure 5.4: Intensity distribution in the immediate vicinity of the double-slit system for H-polarized incident radiation when the transmission is maximum (top frame, slit separation equal to  $5\lambda_{sp}/2$ ), and minimum (bottom frame, slit separation equal to  $4\lambda_{sp}/2$ ). The field is incident from below. All lengths are in nm.

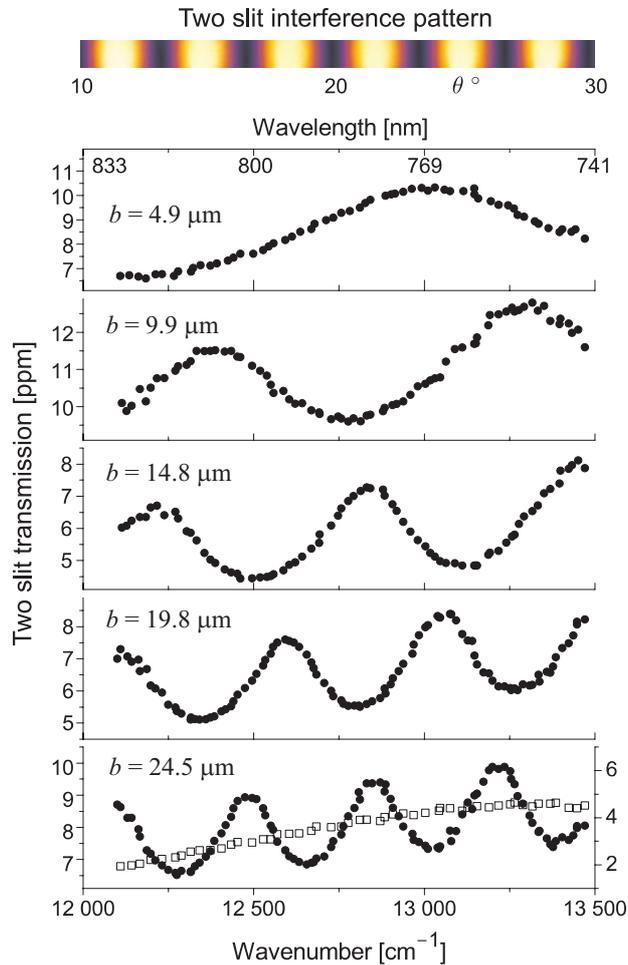


Figure 5.5: Experimental results obtained by Eliel *et al.* (Leiden University). The frame at the top shows the Young's-type interference pattern behind the screen, as recorded with a charge coupled device camera. The other frames display experimental transmission spectra for an H-polarized input beam (polarization perpendicular to the long axis of the 200 nm wide slits), recorded by integrating over the interference pattern. The value of the slit separation  $b$  is indicated in each of the frames. In the frame at the bottom ( $b = 24.5 \mu\text{m}$ ) the results for E-polarized incident light (open squares) are included; the scale at the right-hand side applies to this choice of polarization.

screen. The polarization of the incident light is either parallel (E-polarization) or perpendicular (H-polarization) to the long axis of the slits.

The experimental results for the case of the H-polarization are shown in Fig. 5.5.

The transmission is seen to be approximately sinusoidally modulated as a function of the wavenumber, the modulation period being inversely proportional to the slit separation. The visibility  $\mathcal{V}$  of the fringes is of order 0.2, roughly independent of the slit separation. Note that the fringes are superposed on an offset that gradually increases as a function of the wavenumber.

In contrast, for a E-polarized incident beam the detected signal shows no modulation whatsoever (see bottom frame of Fig. 5.5). Equally, no modulation is observed when the experiment is performed using a 200 nm thick titanium layer instead of gold, independent of the polarization of the incident radiation. On such a plate the surface plasmons have a decay length of  $\approx 7 \mu\text{m}$  on the air-titanium interface [JOHNSON AND CHRISTY, 1974], which is considerably smaller than the slit separation of 25  $\mu\text{m}$ .

Overall, the agreement between the experiment and the results of the Green's function model is seen to be good, the theoretical data having a somewhat larger visibility than the experimental ones (see Figs. 5.1 and 5.2). This difference can be attributed to the different embedding of the gold film in the experiment and in the calculation. While in the experiment the gold film is asymmetrically encapsulated, in the calculation the materials at either side of the film are identical, greatly enhancing the plasmonic effects.

### 5.3 A surface plasmon Fabry-Pérot effect

In the previous section, all results could be explained by a model which only contains one free parameter, viz. the relative coupling strength  $A$ . In the case when one considers slits with a smaller width, somewhat different effects occur. In Fig. 5.6 the transmission coefficient of the two-slit configuration is shown as a function of the distance between the two slits. For the E-polarization case, there is no significant modulation of the transmission beyond a slit separation distance larger than half a wavelength. In contrast, for the H-polarization case, a slightly damped oscillatory behavior can be observed. At the maxima the transmission is strongly enhanced with respect to the single-slit transmission ( $T = 1.1$ ), whereas at the minima the transmission is frustrated with respect to the single-slit transmission. The oscillation period is found to be approximately  $0.93\lambda \approx \lambda_{\text{sp}}$ . Note that the dependence is *not* sine-like, which implies that the results can not be directly explained by the model in the previous Section.

As an additional check that there are surface plasmons involved, we have calculated the field in a cross-section through the metal plate, far away from the slits, see Fig. 5.7. The exponential fall-off of  $E_z$  that is characteristic for surface plasmons is clearly seen on both sides of the metal plate. It was found that within the numerical accuracy ( $\sim 1\%$ ) the field behaves as predicted by Eq. (1.76). Because

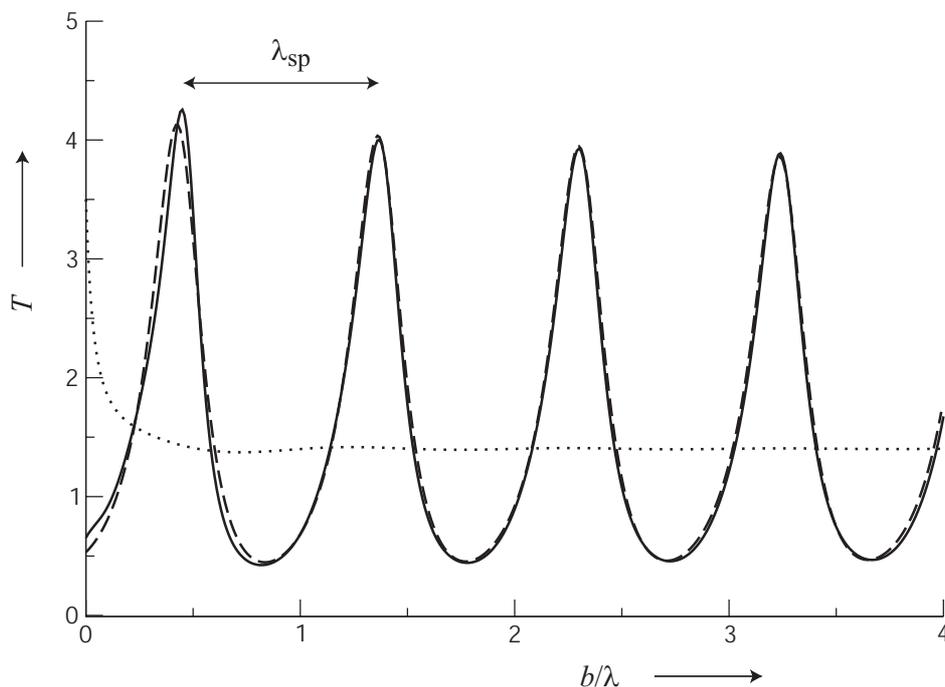


Figure 5.6: The transmission coefficient  $T$  of a two-slit configuration in a 100 nm thick silver plate as a function of the distance  $b$ , expressed in wavelengths, between the slits. The widths of the slits  $w = 30$  nm, the wavelength  $\lambda = 500$  nm and the refractive index is taken as  $n_{\text{silver}} = 0.05 + i2.87$ . The full line denotes the H-polarization case, the dotted line (magnified 100 times) denotes the E-polarization case, and the dashed line the denotes the fit of the simple model to the results for the H-polarization ( $c_{\text{sp}}\tau_{\text{sp}} = 0.59 \exp(i0.85\pi)$  and  $r_{\text{sp}} = 0.38 \exp(i0.94\pi)$ ).

surface plasmons cannot be excited by a plane wave on a flat surface, they must be excited by diffraction at one of the slits in the metal plate. In a single slit configuration, these surface plasmons do not contribute to the transmission; they travel away from the slit and eventually are absorbed. However in the two-slit configuration, as was shown in the previous Section, the surface plasmons excited by one slit, and traveling in the direction of the other slit, can contribute to the transmission.

Therefore we propose the following simple model to explain our numerical data (see Fig. 5.8): at each slit the incident field can directly couple to the transmitted field with a coupling factor  $\tau_1$ , which is related to the single slit transmission coefficient by  $T_s = |\tau_1|^2$ . However, at each slit the incident field can also couple, with a coefficient  $c_{\text{sp}}$ , to a surface plasmon, which travels in the direction of the

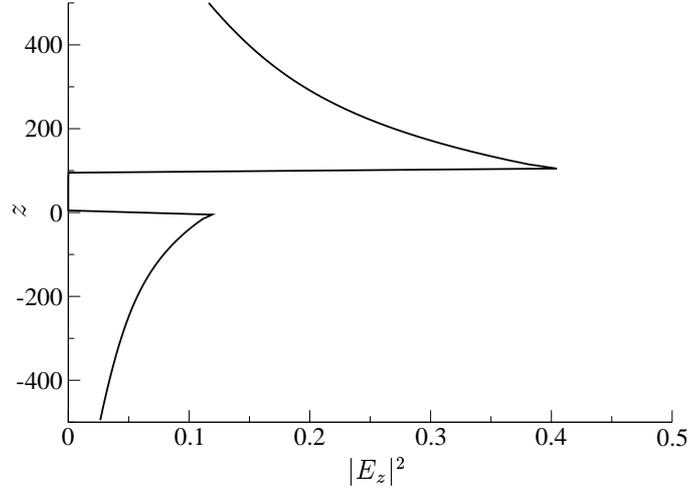


Figure 5.7: The transverse field component  $|E_z|^2$  in a cross-section taken through the plate at  $x = 1\mu\text{m}$ . The slits are separated by  $450\text{ nm}$ . The other parameters are as in Fig. 5.6.  $z$  is measured in nanometers.

other slit. If such a surface plasmon reaches the other slit, it can contribute to the transmitted field, with a coefficient  $\tau_{\text{sp}}$ , or it can be reflected with a coefficient  $r_{\text{sp}}$  as a surface plasmon travelling in the opposite direction. This reflected surface plasmon can, on reaching the other slit, again be reflected or transmitted. The same process, of course, also occurs at the other slit. In this way, one obtains for the transmission coefficient

$$T = \left| \tau_1 + \sum_{n=0}^{\infty} c_{\text{sp}} r_{\text{sp}}^n e^{i(n+1)k_{\text{sp}}b} \tau_{\text{sp}} \right|^2 = \left| \tau_1 + \frac{c_{\text{sp}} \tau_{\text{sp}} e^{ik_{\text{sp}}b}}{1 - r_{\text{sp}} e^{ik_{\text{sp}}b}} \right|^2, \quad (5.4)$$

with  $b$  the distance between the slits and  $n$  denotes the number of times that the surface plasmon is reflected before it couples to the transmitted field.

This model has two free parameters, viz., the product  $c_{\text{sp}} \tau_{\text{sp}}$  and  $r_{\text{sp}}$ . If we adjust these parameters to fit the model to the data, we obtain the curve displayed in Fig. 5.6. It is seen that the model is in excellent agreement with the rigorous calculation. Therefore, we conclude that the proposed heuristic model gives a good description of the transmission process. This model enables us to interpret our numerical results in the following manner: the transmission is due to two processes, one is the single-slit transmission, i.e., the transmission as if no other slit is present, the other is the contribution of the surface plasmons. These two contributions do not just add up, but rather interfere. This is the cause of the calculated enhanced

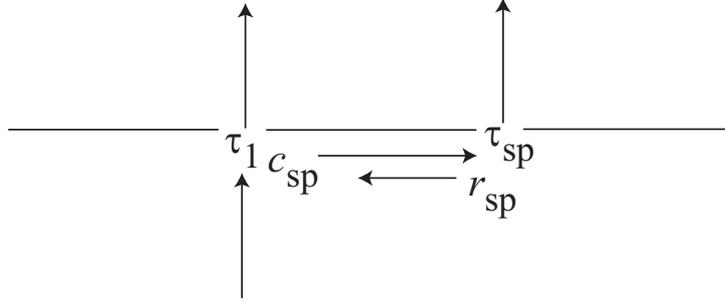


Figure 5.8: Schematic illustration of the heuristic model. For simplicity's sake, the plasmon generation and subsequent processes are only shown at one slit.

transmission (constructive interference) and frustrated transmission (destructive interference). We conclude from the value of  $r_{sp}$  that multiple plasmon scattering plays a significant role in the transmission process.

In all the results discussed in this Chapter, the light was normally incident on the two-slit system, which causes the transmission through the two slits to be identical. In the case of an oblique angle of incidence this symmetry is broken. Even for very small angles of incidence this can give notable results, as can be seen in Fig. 5.9, where the transmission as a function of the distance between the two slits is plotted for both a normally incident H-polarized plane wave, and an H-polarized plane wave propagating in a direction making an angle of  $2^\circ$  with the symmetry axis (the  $z$ -axis). Two effects can be observed: a kind of “beat” phenomenon which causes the resonance decrease in amplitude until  $b \approx 7\lambda$ , where it starts to increase again. However, at these distance the resonance is out of phase with the resonance visible at normal incidence. The second phenomenon is a doubling of the resonance frequency around  $b \approx 7\lambda$ , i.e., the distance where the amplitude of the resonance is minimum.

The observed phenomena can be easily explained with the heuristic model given above. First note that due to the very small angle of incidence, all the coupling constants (i.e.,  $c_{sp}$ ,  $\tau_{sp}$ , etc.) remain approximately the same compared with normal incidence. The main difference with normal incidence is that the light hitting the second slit, picks up an additional phase  $\phi$  (see Fig. 5.10) given by

$$\phi = 2\pi \sin \theta \frac{b+w}{\lambda}, \quad (5.5)$$

where  $w$  is the slit width and  $\theta$  is the angle of incidence. Therefore the transmission

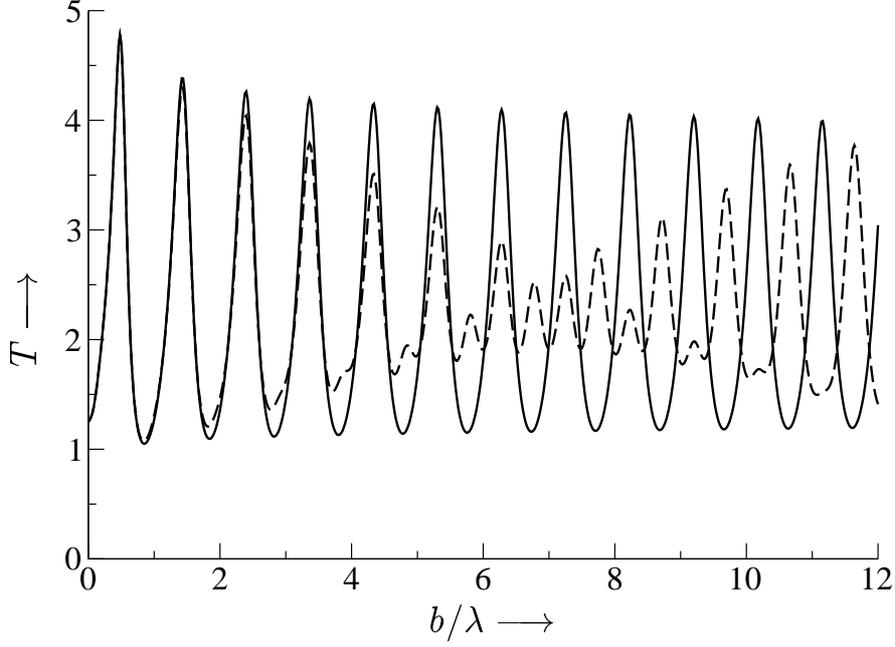


Figure 5.9: Transmission coefficient as a function of the distance between the two slits. The slits are 48 nm wide and located in a 200 nm thick gold plate. The incident plane wave is H-polarized and has a wavelength of 800 nm. The full line is for a normal incident plane wave, whereas the dashed line is incident under an angle of  $2^\circ$ .

through the first slit is given by

$$\begin{aligned}
 T_1 &= \left| \tau_1 + c_{\text{sp}} e^{2ik_{\text{sp}}b} r_{\text{sp}} \tau_{\text{sp}} \sum_{n=0}^{\infty} (r_{\text{sp}}^2 e^{2ik_{\text{sp}}b})^n + e^{i\phi} c_{\text{sp}} e^{ik_{\text{sp}}b} \tau_{\text{sp}} \sum_{n=0}^{\infty} (r_{\text{sp}}^2 e^{2ik_{\text{sp}}b})^n \right|^2 \\
 &= \left| \tau_1 + \frac{c_{\text{sp}} \tau_{\text{sp}} r_{\text{sp}} e^{2ik_{\text{sp}}b}}{1 - r_{\text{sp}}^2 e^{2ik_{\text{sp}}b}} + e^{i\phi} \frac{c_{\text{sp}} \tau_{\text{sp}} e^{ik_{\text{sp}}b}}{1 - r_{\text{sp}}^2 e^{2ik_{\text{sp}}b}} \right|^2, \quad (5.6)
 \end{aligned}$$

where the second term represents the surface plasmons that are both excited and de-excited at the first slit. The third term represents the plasmons that are excited at the second slit and de-excited at the first slit. The transmission through the

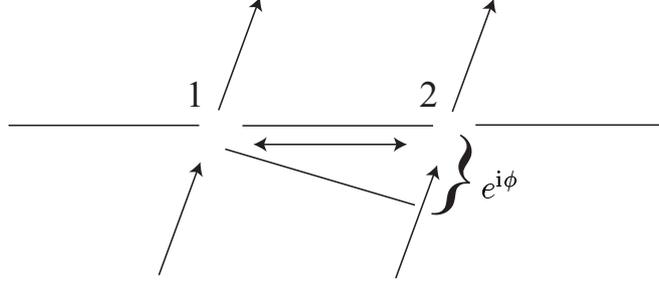


Figure 5.10: Schematic illustration of the extra phase factor due to the oblique angle of incidence in the heuristic model.

second slit is given by

$$\begin{aligned}
 T_2 &= \left| e^{i\phi} \tau_1 + e^{i\phi} c_{\text{sp}} e^{2ik_{\text{sp}}b} r_{\text{sp}} \tau_{\text{sp}} \sum_{n=0}^{\infty} (r_{\text{sp}}^2 e^{2ik_{\text{sp}}b})^n + c_{\text{sp}} e^{ik_{\text{sp}}b} \tau_{\text{sp}} \sum_{n=0}^{\infty} (r_{\text{sp}}^2 e^{2ik_{\text{sp}}b})^n \right|^2 \\
 &= \left| e^{i\phi} \tau_1 + e^{i\phi} \frac{c_{\text{sp}} \tau_{\text{sp}} r_{\text{sp}} e^{2ik_{\text{sp}}b}}{1 - r_{\text{sp}}^2 e^{2ik_{\text{sp}}b}} + \frac{c_{\text{sp}} \tau_{\text{sp}} e^{ik_{\text{sp}}b}}{1 - r_{\text{sp}}^2 e^{2ik_{\text{sp}}b}} \right|^2,
 \end{aligned} \tag{5.7}$$

where the second term represents the surface plasmons that are both excited and de-excited at the second slit. The third term represents the plasmons that are both excited at the first slit and de-excited at the second slit. Therefore the total transmission is given by

$$T = (T_1 + T_2)/2 = \left| \tau_1 + \frac{c_{\text{sp}} \tau_{\text{sp}} r_{\text{sp}} e^{2ik_{\text{sp}}b}}{1 - r_{\text{sp}}^2 e^{2ik_{\text{sp}}b}} + \cos \phi \frac{c_{\text{sp}} \tau_{\text{sp}} e^{ik_{\text{sp}}b}}{1 - r_{\text{sp}}^2 e^{2ik_{\text{sp}}b}} \right|^2, \tag{5.8}$$

as can be found after some manipulations. If this formula is compared with Eq. (5.4) for the case of normal incidence, one can see that the plasmon term in Eq. (5.4) is split into two terms, one of which depend on the angle of incidence  $\theta$ . Therefore in the case that  $\cos \phi \approx 1$  one finds that there are no differences between oblique and normal incidence. However, if  $\cos \phi \approx 0$ , the third term in Eq. (5.8) can be ignored, and one obtains a modulation of the transmission as a function of the distance between the slits with *half* of the surface plasmon wavelength as its period. For the case that  $\cos \phi \approx -1$ , one obtains a modulation of the transmission which is out of phase with the modulation observed for normal incidence. In general both the second and the third term of Eq. (5.8) are significant, and modulation of  $T$  with two different frequencies will be observed. Note that

both the “beat” phenomenon and the frequency doubling visible in the dashed curve of Fig. 5.9 are in agreement with this explanation.

## 5.4 Conclusions

In this Chapter we have shown that Young’s double-slit experiment, often seen as proof of the wave nature of light, can provide powerful evidence for the role of propagating surface plasmons in the transmission of perforated metal screens. The transport of electromagnetic energy by the surface plasmons over distances of many optical wavelengths gives rise to an interference phenomenon in the slits that enhances or reduces the intensity of the far-field pattern. Furthermore, we have predicted that for some cases, the multiple reflections of the surface plasmons by the slit are significant, leading to a “surface plasmon Fabry-Pérot effect”. Especially for an oblique angle of incidence, this leads to notable effects.



# Chapter 6

## Young's Interference Experiment with Partially Coherent Light

This Chapter is based on the following publications:

- H.F. Schouten, T.D. Visser and E. Wolf, “New effects in Young’s interference experiment with partially coherent light”, *Optics Letters*, vol. 28, pp. 1182–1184 (2003).
- H.F. Schouten, G. Gbur, T.D. Visser and E. Wolf, “Phase singularities of the coherence functions in Young’s interference pattern”, *Optics Letters*, vol. 28, pp. 968–970 (2003).
- N. Kuzmin, H.F. Schouten, G. Gbur, G.W. ’t Hooft, E.R. Eliel and T.D. Visser, “Surface-plasmon-induced Coherence”, to be submitted.

### **Abstract**

We analyze the coherence properties of a partially coherent field emerging from two pinholes in an opaque screen, i.e., we study Young’s interference experiment for the case that the illuminating wavefield is partially coherent. We show that at certain pairs of points in the region of superposition the light is fully coherent, regardless of the state of coherence of the light at the pinholes. Furthermore, we show that the spectral degree of coherence possesses phase singularities on certain surfaces in the region of superposition. We predict that it is possible to obtain a stationary interference pattern if one aperture is illuminated by a laser operating at frequency  $\omega_1$ , while the other aperture is being illuminated by a second laser operating at frequency  $\omega_2$ . Surface plasmons play a key role in this effect.

## 6.1 Introduction

Two hundred years after Thomas Young [1802; 1845] discussed the interference of light which passes through two pinholes, such experiments are still sources of new insights. Recent research involving Young's experiment with partially coherent light has predicted that if two pinholes are illuminated with broad-band light and, consequently, interference fringes are absent, strong spectral changes generally occur in the region of superposition [JAMES AND WOLF, 1991]. Such spectral changes may in turn be used to determine the spectral degree of coherence of the light at the two pinholes [JAMES AND WOLF, 1998]. These predictions have been verified experimentally [KANDPAL *et al.*, 1992a; KANDPAL *et al.*, 1992b; SANTARSIERO AND GORI, 1992; KUMAR AND RAO, 2001; BASANO *et al.*, 2002]. Somewhat analogous experiments with matter waves have been carried out using neutrons beams [RAUCH, 1993; JACOBSON *et al.*, 1994], see also [AGARWAL, 1995].

In a recent investigation of Young's interference experiment with partially coherent light [PONOMARENKO AND WOLF, 1999], expressions were derived for the *cross-spectral density* and the *spectral density* of the field in the region of superposition. In the present Chapter we derive somewhat more general expressions for such a situation in Sec. 6.3, after first giving a brief introduction to partially coherent fields in Sec. 6.2. In Sec. 6.4 it is shown that these relations imply remarkable properties of the spectral degree of coherence. For example, at any pair of points in certain planes of observation the light is found to be always completely coherent, irrespective of the state of coherence of the light at the two pinholes; in particular, the light could originate from independent lasers, each illuminating only one of the two pinholes. In [PONOMARENKO AND WOLF, 1999], pairs of points were found at which the spectral degree of coherence has zero value, implying the existence of phase singularities of the coherence function. In Sec. 6.5, we present a detailed analysis of the spectral degree of coherence and show that it possesses surfaces, defined by pairs of points, on which the phase is singular. The behavior of the field in the vicinity of these phase singularities is investigated. In Sec. 6.6, the effect of the coupling of light with surface plasmons at the apertures (see the preceding Chapter) on the coherence properties of the field is discussed. Note that this last subject is somewhat different from the rest of the Chapter in the sense that it takes into account the vectorial nature of the field, instead of assuming that the field is scalar.

## 6.2 Partially coherent fields

In the theory of partially coherent light, the fields are no longer considered to be deterministic, but may undergo random fluctuations. In this Section, we only

briefly summarize the main concepts used in the remainder of this Chapter, for more information about this subject, see Chapter 10 of [BORN AND WOLF, 1999] or [MANDEL AND WOLF, 1995].

We consider a fluctuating field  $U(\mathbf{r}, t)$ , which is assumed to be statistically stationary, at least in the wide sense [MANDEL AND WOLF, 1995, Sec. 2.2]. For simplicity, the vector-character of the field is not taken into account. Spatial and temporal correlations between pairs of points  $P_1(\mathbf{r}_1)$  and  $P_2(\mathbf{r}_2)$  may be characterized by use of the *mutual coherence function* [MANDEL AND WOLF, 1995, Sec. 4.3.1]

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle U^*(\mathbf{r}_1, t)U(\mathbf{r}_2, t + \tau) \rangle, \quad (6.1)$$

where the angular brackets denote time or ensemble averaging.

It is often advantageous to work instead with the Fourier transform of the mutual coherence function, the *cross-spectral density* [MANDEL AND WOLF, 1995, Sec. 4.3.2], defined as

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = \frac{1}{2\pi} \int \Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) e^{i\omega\tau} d\tau. \quad (6.2)$$

The cross-spectral density characterizes the spatial correlations of the field at two points at a single frequency  $\omega$ .

It can be shown that if the field is represented by an ensemble of space-frequency realizations  $\{U(\mathbf{r}, \omega)\}$ , then the cross-spectral density is given by [MANDEL AND WOLF, 1995, Sec. 4.7]

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = \langle U^*(\mathbf{r}_1, \omega)U(\mathbf{r}_2, \omega) \rangle, \quad (6.3)$$

where the angular brackets denote the ensemble average over all space-frequency realizations  $\{U(\mathbf{r}, \omega)\}$ .

The *spectral degree of coherence*  $\mu$  at frequency  $\omega$  is defined as

$$\mu(\mathbf{r}_1, \mathbf{r}_2, \omega) = \frac{W(\mathbf{r}_1, \mathbf{r}_2, \omega)}{\sqrt{S(\mathbf{r}_1, \omega)}\sqrt{S(\mathbf{r}_2, \omega)}}, \quad (6.4)$$

where the *spectral density*  $S$  at position  $\mathbf{r}$  at frequency  $\omega$  is defined by

$$S(\mathbf{r}, \omega) = W(\mathbf{r}, \mathbf{r}, \omega). \quad (6.5)$$

The modulus of the spectral degree of coherence satisfies the following inequalities [MANDEL AND WOLF, 1995, Sec. 4.3.2]

$$0 \leq |\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)| \leq 1, \quad (6.6)$$

where  $\mu = 0$  corresponds with complete incoherence and  $|\mu| = 1$  corresponds with complete coherence.

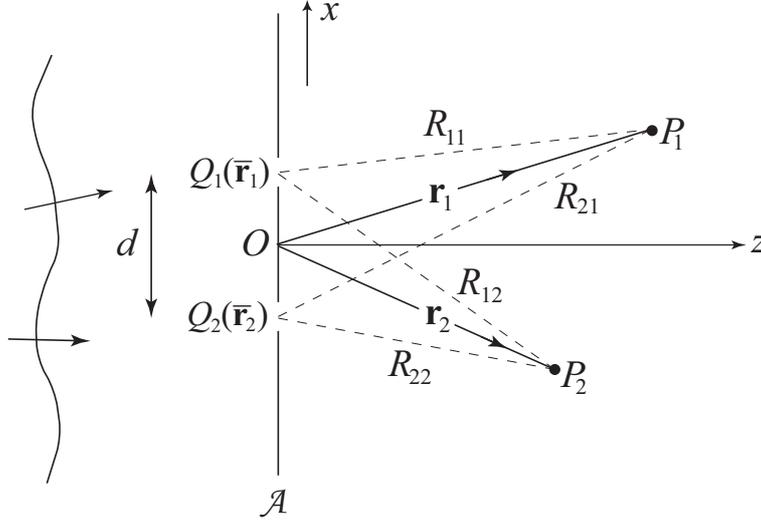


Figure 6.1: Illustrating the notation relating to interference patterns formed with partially coherent light in Young's experiment.

### 6.3 Coherence properties of light in Young's interference experiment

Consider a partially coherent field propagating into the half-space  $z > 0$ . The cross-spectral density  $W^{(0)}(\mathbf{r}'_1, \mathbf{r}'_2, \omega)$  of the field at frequency  $\omega$  at any two points  $\mathbf{r}'_1$  and  $\mathbf{r}'_2$  in the plane  $z = 0$  may be expressed in terms of the spectral density  $S^{(0)}(\mathbf{r}'_1, \omega)$ ,  $S^{(0)}(\mathbf{r}'_2, \omega)$  at the two points and the spectral degree of coherence  $\mu^{(0)}(\mathbf{r}'_1, \mathbf{r}'_2, \omega)$  of the light at these points in the form (See Eq. (6.4))

$$W^{(0)}(\mathbf{r}'_1, \mathbf{r}'_2, \omega) = \sqrt{S^{(0)}(\mathbf{r}'_1, \omega)S^{(0)}(\mathbf{r}'_2, \omega)}\mu^{(0)}(\mathbf{r}'_1, \mathbf{r}'_2, \omega). \quad (6.7)$$

The cross-spectral density of the field at any pair of points  $P_1(\mathbf{r}_1)$  and  $P_2(\mathbf{r}_2)$  in the half-space  $z > 0$ , is then given by the expression (Eqs. (4.4–15) and (4.4–16) of [MANDEL AND WOLF, 1995])

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = \left(\frac{1}{2\pi}\right)^2 \iint_{z=0} W^{(0)}(\mathbf{r}'_1, \mathbf{r}'_2, \omega) \left(ik + \frac{1}{R_1}\right) \left(-ik + \frac{1}{R_2}\right) e^{ik(R_2-R_1)} \times \frac{\cos \theta_1 \cos \theta_2}{R_1 R_2} d^2 r'_1 d^2 r'_2, \quad (6.8)$$

where  $k = \omega/c$  is the wavenumber associated with frequency  $\omega$ ,  $c$  is the speed of light *in vacuo*,  $R_i = |\mathbf{r}_i - \mathbf{r}'_i|$  and  $\theta_i$  is the angle between the vector  $\mathbf{r}_i - \mathbf{r}'_i$  and the positive  $z$ -direction.

Suppose now, that the plane  $z = 0$  is covered by an opaque screen  $\mathcal{A}$  with two small pinholes with area  $\delta A$  at points  $Q_1(\bar{\mathbf{r}}_1)$  and  $Q_2(\bar{\mathbf{r}}_2)$  (see Fig. 6.1). For this case the formula (6.8) reduces to (cf. the derivation as given in Ch. 4.3.2 of [MANDEL AND WOLF, 1995], but now with inclination factors)

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = \left(\frac{\delta A}{2\pi}\right)^2 \left\{ S_1(\omega)K_{11}^*K_{12} + S_2(\omega)K_{21}^*K_{22} \right. \\ \left. + \sqrt{S_1(\omega)S_2(\omega)}[\mu_{12}(\omega)K_{11}^*K_{22} \right. \\ \left. + \mu_{12}^*(\omega)K_{12}K_{21}^*] \right\}. \quad (6.9)$$

where  $S_i(\omega)$  is the spectral density at pinhole  $Q_i$  and  $\mu_{12}(\omega) = \mu^{(0)}(\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, \omega)$  is the spectral degree of coherence of the field at the two pinholes. Furthermore, we used the fact that  $\mu_{12}(\omega) = \mu_{21}^*(\omega)$ , and the factors  $K_{ij}$  are given by the expression

$$K_{ij} = \left(-ik + \frac{1}{R_{ij}}\right) \frac{e^{ikR_{ij}}}{R_{ij}} \cos \theta_{ij}, \quad (i, j = 1, 2) \quad (6.10)$$

where  $R_{ij}$  is the distance from the pinhole at  $Q_i(\bar{\mathbf{r}}_i)$  to the field point  $P_j(\mathbf{r}_j)$  and  $\theta_{ij}$  is the angle between the line  $Q_iP_j$  and the positive  $z$ -direction.

The spectral density at frequency  $\omega$  of the light at a point  $P_0(\mathbf{r}_0)$  in the region of superposition is given by the expression

$$S(\mathbf{r}_0, \omega) \equiv W(\mathbf{r}_0, \mathbf{r}_0, \omega) \\ = \left(\frac{\delta A}{2\pi}\right)^2 \left\{ S_1(\omega)|K_{10}|^2 + S_2(\omega)|K_{20}|^2 \right. \\ \left. + 2\sqrt{S_1(\omega)S_2(\omega)}\text{Re}[\mu_{12}(\omega)K_{10}^*K_{20}] \right\} \quad (6.11)$$

where  $R_{10}$  and  $R_{20}$  denote the distances from the pinholes  $Q_1$  and  $Q_2$ , respectively, to the point  $P_0(\mathbf{r}_0)$ .

## 6.4 Planes of full coherence

Let us choose the coordinate system with the origin  $O$  between the two pinholes, and with the plane  $z = 0$  coinciding with the plane containing the two pinholes. Let the pinholes be located symmetrically along the  $x$ -axis at distance  $d$  from each other, i.e., at points with position vectors

$$\bar{\mathbf{r}}_1 = (d/2, 0, 0), \quad \bar{\mathbf{r}}_2 = (-d/2, 0, 0), \quad (6.12)$$

referred to the origin  $O$ . For any pair of points  $P_1(\mathbf{r}_1), P_2(\mathbf{r}_2)$  in the plane  $x = 0$ , which we will refer to as the *bisecting plane*  $\Pi$ , i.e., for points

$$\mathbf{r}_1 = (0, y_1, z_1), \quad \mathbf{r}_2 = (0, y_2, z_2), \quad (6.13)$$

we have (see Fig. 6.1)

$$K_{11} = K_{21}, \quad K_{12} = K_{22}. \quad (6.14)$$

On substituting from Eqs. (6.14) into Eq. (6.9) we obtain for the cross-spectral density the expression

$$\begin{aligned} W(\mathbf{r}_1, \mathbf{r}_2, \omega) = & \left( \frac{\delta A}{2\pi} \right)^2 K_{11}^* K_{12} \left\{ S_1(\omega) + S_2(\omega) \right. \\ & \left. + 2\sqrt{S_1(\omega)S_2(\omega)} \operatorname{Re}[\mu_{12}(\omega)] \right\}. \end{aligned} \quad (6.15)$$

Next we substitute from Eq. (6.15) into Eq. (6.5), and find that the spectral density is given by

$$\begin{aligned} S(\mathbf{r}_i, \omega) = & \left( \frac{\delta A}{2\pi} \right)^2 |K_{1i}|^2 \left\{ S_1(\omega) + S_2(\omega) \right. \\ & \left. + 2\sqrt{S_1(\omega)S_2(\omega)} \operatorname{Re}[\mu_{12}(\omega)] \right\}, \quad (i = 1, 2). \end{aligned} \quad (6.16)$$

It immediately follows on using definition (6.4) for the spectral degree of coherence that

$$\begin{aligned} \mu(\mathbf{r}_1, \mathbf{r}_2, \omega) &= \frac{K_{11}^* K_{12}}{|K_{11}| |K_{12}|}, \\ &= e^{ik(R_{12}-R_{11})} \frac{(ik + 1/R_{11})(-ik + 1/R_{12})}{|ik + 1/R_{11}| | -ik + 1/R_{12}|}, \\ &= e^{ik(R_{12}-R_{11})} e^{i(\phi_1-\phi_2)}, \end{aligned} \quad (6.17)$$

where

$$\cos \phi_i = 1/R_{1i} D_i, \quad \sin \phi_i = k/D_i, \quad (6.18)$$

and

$$D_i = \sqrt{k^2 + 1/R_{1i}^2}. \quad (6.19)$$

Thus, we arrive at the conclusion that for any pair of points  $P_1(\mathbf{r}_1), P_2(\mathbf{r}_2)$  which lie in the bisecting plane  $\Pi$ , i.e., the plane bisecting the line joining the two pinholes and perpendicular to that line, the spectral degree of coherence of the field is unimodular, i.e.,

$$|\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)| = 1, \quad (6.20)$$

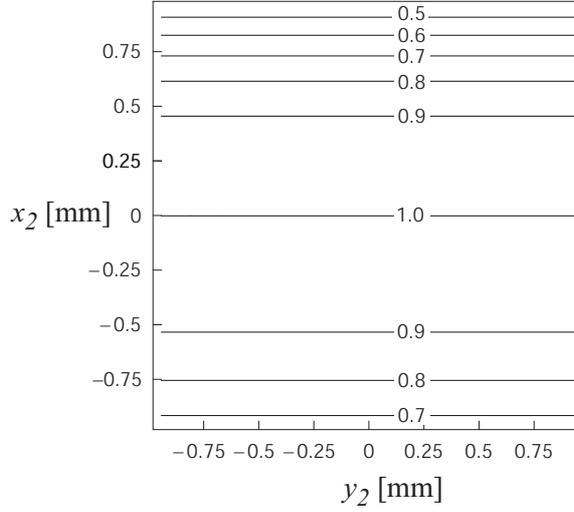


Figure 6.2: Contour lines of  $|\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)|$  with  $\mathbf{r}_1$  being kept fixed at  $(0, 0, 1.5 \text{ m})$  and  $\mathbf{r}_2$  varying in the plane  $z_2 = 1.5 \text{ m}$ . In this example  $d = 1 \text{ mm}$ ,  $\omega = 10^{15} \text{ s}^{-1}$  and  $\mu_{12}(\omega) = 0.2 + 0.2i$ .

implying that the light at these points is mutually spatially fully coherent, irrespective of the state of coherence of the field at the two pinholes.

This result is illustrated in Fig. 6.2 in which contours of  $|\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)|$  are shown. We note that the contours are not symmetric about the plane  $x = 0$ , even though the geometry is. This asymmetry is due to the fact that in this particular example  $\mu_{12}(\omega)$  is complex-valued and  $\mu_{12}(\omega) = \mu_{12}^*(\omega)$ .

Next consider a pair of points  $\mathbf{r}_1 = (x, y, z)$ ,  $\mathbf{r}_2 = (x, -y, z)$ . In this case (see Fig. 6.1) we find that

$$K_{11} = K_{12}, \quad K_{21} = K_{22}. \quad (6.21)$$

On substituting from Eqs. (6.21) into Eq. (6.9) we obtain for the cross-spectral density the expression

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = \left(\frac{\delta A}{2\pi}\right)^2 \left\{ S_1(\omega) |K_{11}|^2 + S_2(\omega) |K_{21}|^2 + 2\sqrt{S_1(\omega)S_2(\omega)} \text{Re}[K_{11}^* K_{22} \mu_{12}(\omega)] \right\}. \quad (6.22)$$

If we substitute from Eqs. (6.21) into Eq. (6.11) we see that the spectral densities

are given by the expression

$$\begin{aligned} S(\mathbf{r}_1, \omega) &= S(\mathbf{r}_2, \omega) \\ &= \left(\frac{\delta A}{2\pi}\right)^2 \left\{ S_1(\omega) |K_{11}|^2 + S_2(\omega) |K_{22}|^2 \right. \\ &\quad \left. + 2\sqrt{S_1(\omega)S_2(\omega)} \operatorname{Re}[K_{11}^* K_{22} \mu_{12}(\omega)] \right\}, \end{aligned} \quad (6.23)$$

On substituting from Eqs. (6.22)–(6.23) into the definition (6.4) for the spectral degree of coherence we find that

$$\mu(\mathbf{r}_1, \mathbf{r}_2, \omega) = 1. \quad (6.24)$$

The formula (6.24) shows that for any pair of points  $P_1(\mathbf{r}_1), P_2(\mathbf{r}_2)$  which are mirror images of each other in the plane containing the two pinholes and which is perpendicular to the screen, the spectral degree of coherence of the field is unity, irrespective of the state of coherence of the field at the two pinholes; i.e., the light is fully coherent and co-phasal at such a pair of points.

It is to be noted that the light incident on each of the two pinholes may originate in two different sources. In particular each pinhole might be illuminated by a different laser. Our results imply that even in such a case the light which two such independent lasers generate in the bisecting plane  $\Pi$  will be spatially completely coherent at every frequency contained in the spectra of both the lasers.

The analysis of the planes of full coherence was performed here under the approximation that the fields are scalar. After this study appeared, this analysis was extended by taking into account the vectorial character of the field [AGARWAL *et al.*, 2005]. In this article it was shown that the conclusions of this Section remain essentially unchanged, i.e., the same planes of full coherence were obtained, regardless of the state of coherence and the state of polarization of the field at the two pinholes.

## 6.5 Phase singularities of the coherence functions

Suppose again that two pinholes located at  $Q_1$  and  $Q_2$  in a plane opaque screen  $\mathcal{A}$  (see Fig. 6.1) are illuminated by partially coherent light. Under practical circumstances one has  $S_1(\omega) = S_2(\omega)$ . Furthermore, we assume that the angles of incidence and diffraction are small, so that Eq. (6.9) reduces to

$$\begin{aligned} W(\mathbf{r}_1, \mathbf{r}_2, \omega) &= \left(\frac{\delta A}{2\pi}\right)^2 S(\omega) [K_{11}^* K_{12} + K_{21}^* K_{22} \\ &\quad + K_{11}^* K_{22} \mu_{12}(\omega) + K_{12} K_{21}^* \mu_{12}^*(\omega)]. \end{aligned} \quad (6.25)$$

Here  $S(\omega)$  is the spectral density at  $Q_1$  and  $Q_2$  and  $K_{ij} \approx -ik \exp(ikR_{ij})/R_{ij}$  (see Eq. (6.10)). In most cases of practical interest the separation between the two pinholes  $d \ll R_{ij}$ , and then  $R_{1i} \approx R_{2i}$  and the spectral density takes on the simple form

$$S(\mathbf{r}_i, \omega) = \left(\frac{\delta A}{2\pi}\right)^2 \frac{2S(\omega)}{R_{1i}^2} \{1 + |\mu_{12}(\omega)| \cos[\beta + k(R_{2i} - R_{1i})]\}, \quad (6.26)$$

where  $\beta$  is the phase of the spectral degree of coherence, and  $i = 1, 2$ .

We now consider pairs of points in the far zone for which the phase of the spectral degree of coherence  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$  becomes singular; this happens at pairs of points for which  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega) = 0$ . It is to be noted from Eq. (6.4) that  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$  exhibits an additional singular behavior when  $S(\mathbf{r}_i, \omega) = 0$ . We exclude such cases from our consideration because the approximate form of the spectrum given by Eq. (6.26) suggests that it will not have zero value provided the light is not spatially fully coherent at the pinholes, i.e. provided that  $|\mu_{12}(\omega)| < 1$ .<sup>1</sup> We assume that the use of a far-zone approximation for the factors  $K_{ij}$  does not significantly alter the behavior of the singular points, an assumption that will later be supported by numerical calculations. In the far zone, the factors  $K_{ij}$  take on the approximate form

$$K_{ij} \approx -ik \frac{\exp[ik(R_j - \hat{\mathbf{r}}_j \cdot \mathbf{d}_i)]}{R_j}. \quad (6.27)$$

In this equation,  $R_j$  is the distance from the origin to the observation point  $P_j$ ,  $\hat{\mathbf{r}}_j$  is the unit vector pointing in the direction  $OP_j$ , and  $\mathbf{d}_i = \pm(d/2)\hat{\mathbf{x}}$ , where the positive or negative sign is taken accordingly as  $i = 1$  or  $2$ , respectively, and  $\hat{\mathbf{x}}$  is the unit vector in the positive  $x$ -direction (see Fig. 6.1). On substituting from Eq. (6.27) into Eq. (6.25), it readily follows that the cross-spectral density may be expressed in the form

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = 2 \left(\frac{k\delta A}{2\pi}\right)^2 S(\omega) \frac{\exp[ik(R_2 - R_1)]}{R_1 R_2} \left\{ \cos \left[ \frac{kd}{2} (\cos \theta_1 - \cos \theta_2) \right] + |\mu_{12}(\omega)| \cos \left[ \frac{kd}{2} (\cos \theta_1 + \cos \theta_2) + \beta \right] \right\}, \quad (6.28)$$

where  $\theta_i$  is the angle between  $\hat{\mathbf{r}}_i$  and the positive  $x$ -direction. It is readily seen that Eq. (6.28) implies the existence of phase singularities, i.e. the existence of pairs of points at which the cross-spectral density, and consequently the spectral degree of

<sup>1</sup>In [GBUR AND VISSER, 2003] it was indeed proven that the zeros of the spectral density are not generic, and will therefore only occur under some special circumstances, such as in [GORI *et al.*, 1998; PONOMARENKO, 2001].

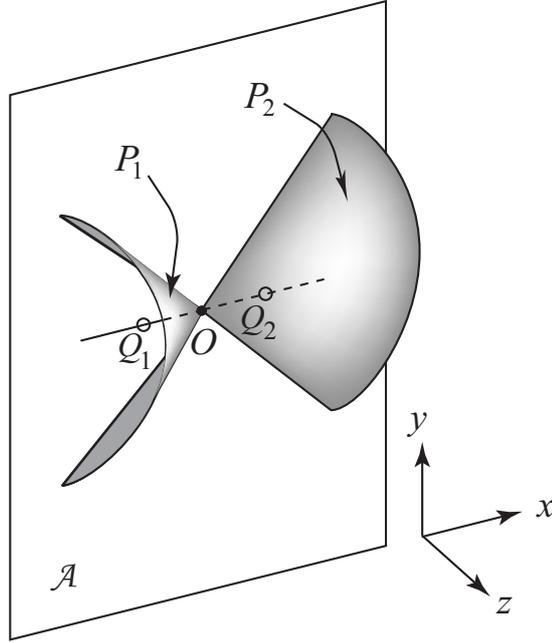


Figure 6.3: Schematic illustration of surfaces on which points of observation  $P_1$  and  $P_2$  in the far zone are located for which  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega) = 0$ , i.e., at which the phase of  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$  is singular.  $P_1$  and  $P_2$  lie on opposite cones.

coherence, has zero value. In particular,  $W(\mathbf{r}_1, \mathbf{r}_2, \omega) = 0$  at points for which the expression in the curly brackets vanishes. This expression is independent of the distances  $R_1$  and  $R_2$  and, in fact, depends only on the directions of observation. It follows that a given zero of the cross-spectral density requires that the observation points  $P_1$  and  $P_2$  both lie on conical surfaces  $\cos \theta_i = \text{constant}$ . A sketch of such surfaces is given in Fig. 6.3.

The behavior of the phase of  $\mu$  in the immediate vicinity of such surfaces can be readily found by noting that the expression in the curly brackets of Eq. (6.28) is real-valued, so that the only possible phase change of this factor on changing the angles  $\theta_1, \theta_2$  is a change in sign. This corresponds to a jump in phase of  $\pm\pi$ , and these are the only possible singular behaviors across the “singular surfaces”.

We have studied the spectral degree of coherence in the region of superposition numerically by using Eqs. (6.25) and (6.4). Let  $\mathbf{r}_1 = (x_1, y_1, z_1)$  and  $\mathbf{r}_2 = (x_2, y_2, z_2)$  specify the position of the observation points  $P_1(\mathbf{r}_1)$  and  $P_2(\mathbf{r}_2)$ , respectively. It is to be noted that the cross-spectral density is computed using Eq. (6.25), not the approximate form Eq. (6.28). By varying  $x_2$  and  $y_2$  while keeping  $z_2$  and  $\mathbf{r}_1$  fixed, the behavior of the phase,  $\phi_\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$ , of the spectral degree of coherence was studied in a plane parallel to the screen containing the apertures. An example is

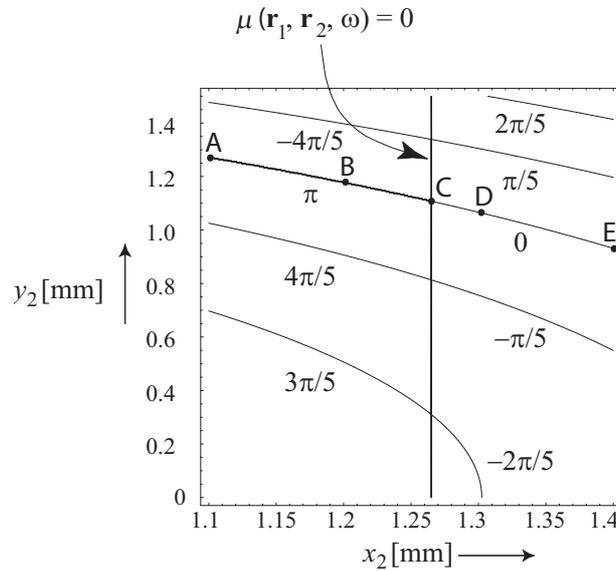


Figure 6.4: Contours of equal phase of the spectral degree of coherence  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$  near a singularity of its phase in a plane parallel to the screen. In this example  $k = 0.333 \times 10^7 \text{ m}^{-1}$ ,  $d = 0.1 \text{ cm}$ ,  $\mu_{12}(\omega) = 0.8 + 0.3i$ ,  $\mathbf{r}_1 = (0, 0, 1.5) \text{ m}$ , and  $z_2 = 1.5 \text{ m}$ .

shown in Fig. 6.4. The vertical line indicates the location of a phase singularity, i.e., a set of points  $P_2$  (with  $P_1$  fixed) for which  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega) = 0$  and hence the phase of the spectral degree of coherence is singular. It can be seen from the figure that the phase has a discontinuity of  $\pi$  across the singularity.

A detailed example of the behavior of the spectral degree of coherence is given in Fig. 6.5. The point  $P_1$  and the coordinates  $y_2$  and  $z_2$  are kept fixed while  $x_2$  is varied. The real and the imaginary part of  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$  are seen to change sign at the phase singularity, in accordance with a  $\pi$  phase jump.

We notice that these phase singularities of the spectral degree of coherence can easily be observed. This requires interfering the light from the vicinity of the pair of points  $P_1$  and  $P_2$ . This can be done, for instance, by bringing together the light from these points by means of another Young's interference experiment and observing the behavior of the interference fringes produced at frequency  $\omega$  by this additional experiment as the point  $P_1$  is kept fixed and the point  $P_2$  is moved across the phase singularity. Fig. 6.6 shows the fringe pattern that would be observed in this second experiment for a selection of points  $P_1, P_2$ . The point  $P_1$  was chosen as in Fig. 6.5, and the point  $P_2$  was taken along a line of constant phase at several points in the vicinity of the phase singularity. The choices of  $P_2$  are illustrated in

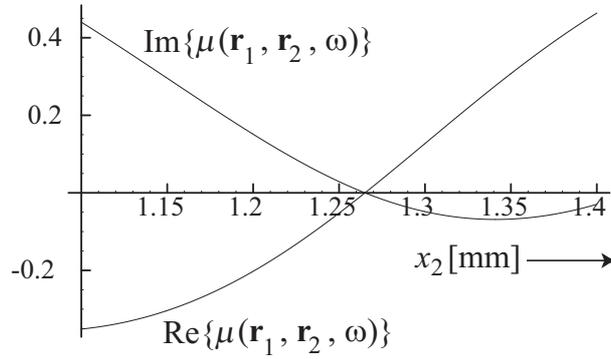


Figure 6.5: The real and the imaginary part of the spectral degree of coherence  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$ , with  $P_1$ ,  $y_2$  and  $z_2$  kept fixed whilst  $x_2$  is being varied. In this example  $y_2 = 0.9$  mm, all other parameters have the same value as in Fig. 6.4.

Fig. 6.4. It can be seen in Fig. 6.6 that the  $\pi$  phase change results in the minima of the secondary fringe pattern becoming maxima, and vice versa, in accordance with the spectral interference law [MANDEL AND WOLF, 1995, Sec. 4.3.2].

After this study was performed, it was shown by Gbur and Visser [2003] that in general the cross-spectral density possesses phase singularities in the form of lines in three-dimensional space (“coherence vortices”) and not surfaces as was found in this study. The phase singularities in the form of a surface are due to the special symmetries of Young’s interference experiment. The phenomenon of phase singularities in the context of partially coherent fields has recently also been experimentally investigated [PALACIOS *et al.*, 2004].

## 6.6 Plasmon-induced coherence in Young’s interference experiment

In optical coherence theory the correlation of the field at position  $\mathbf{r}_1$  at frequency  $\omega$  and the field at  $\mathbf{r}_2$  at the same frequency is described by the so-called *spectral degree of coherence*,  $\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)$ . The modulus of this correlation function is bound by zero and unity. If  $|\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)| = 0$ , the fields at  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are completely uncorrelated at frequency  $\omega$ . If  $|\mu(\mathbf{r}_1, \mathbf{r}_2, \omega)| = 1$  the fields are fully coherent. In all other cases the fields are said to be partially coherent. For narrowband light *the visibility of the interference fringes that are formed in Young’s experiment is proportional to the modulus of the spectral degree of coherence* [MANDEL AND WOLF, 1995]. Therefore, one would expect that if each aperture is illuminated by a separate laser, each tuned to a different frequency, the resulting fringes would have visibility

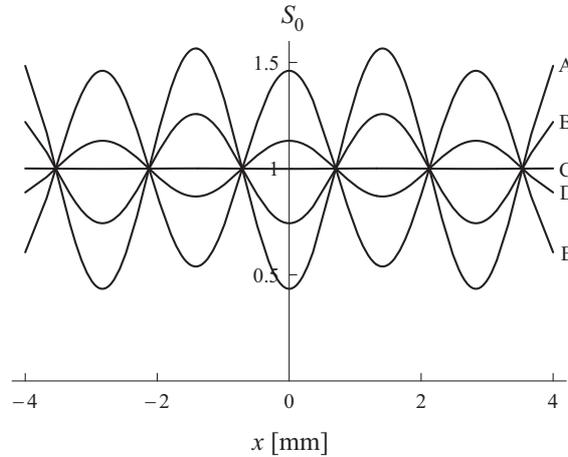


Figure 6.6: Illustrating the spectral interference pattern formed along the  $x$ -direction by combining the light from the pinholes  $P_1$  and  $P_2$  in a second Young's interference experiment. The observation plane was taken to be at  $z = 1.5$  m, and the spacing of the pinholes was taken to be  $d = 0.1$  cm. The positions of the points  $P_2$  are illustrated in Fig. 6.4.  $S_0$  is a spectral density normalized by the value of the spectral density on the curve C. All other parameters are the same as in Fig. 6.4.

zero. However, we predict here that under such circumstances fringes with a good visibility may be produced.

The explanation of this unexpected effect is the generation of surface plasmons at the two slits [RAETHER, 1988]. These electromagnetic surface modes are excited at each slit only when the incident light is H-polarized. They can travel from one slit to the other and then be converted back into a propagating light field [SCHOUTEN *et al.*, 2005]. This energy transfer causes the light that is emitted at each slit to originate from both lasers rather than one. Moreover, there is a fixed phase relation between the light that is directly transmitted at the first slit, and the light at the second slit that is generated by surface plasmons stemming from the first slit. In other words, the processes of excitation and de-excitation of surface plasmons are phase coherent; hence the plasmon-induced emission at one slit is coherent with the direct emission at the other slit. It is this “plasmon-induced coherence” that causes the field at the two individually illuminated apertures to become fully coherent at both frequencies, i.e.,  $|\mu_{12}(\omega_1)| = |\mu_{12}(\omega_2)| = 1$ . The resulting interference pattern therefore has fringes with a non-zero visibility.

This prediction is experimentally verified by Eliel *et al.* (Leiden University), see [KUZMIN *et al.*, 2005]. Their experimental setup consisted of two separate lasers, a

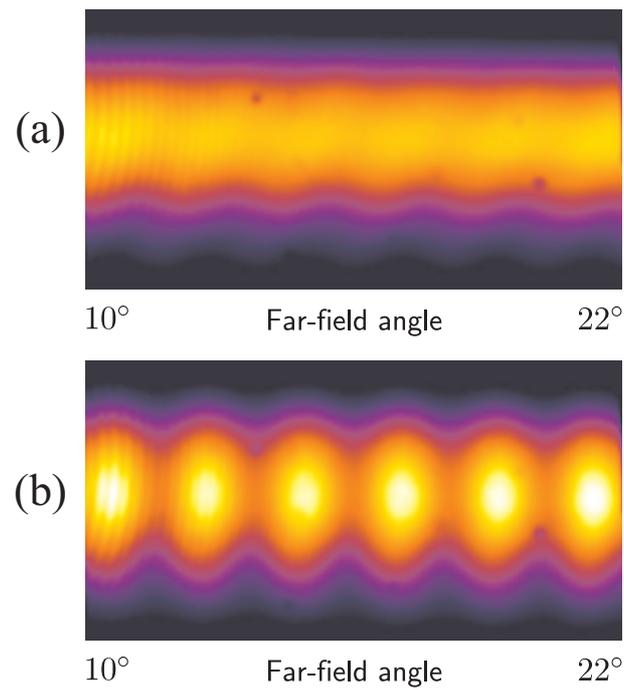


Figure 6.7: Experimental results by Eliel *et al.* (Leiden University): (a) The far-field interference pattern that is measured when the polarization of both laser beams is parallel to the two slits (E-polarization). The Ti:sapphire laser is tuned to 808 nm. The fringe visibility  $\mathcal{V} \approx 0\%$  in this case. (b) The interference pattern that is measured when the polarization of both laser beams is perpendicular to the two slits (H-polarization). The fringe visibility  $\mathcal{V} \approx 28\%$  in this case.

narrowband Ti-Sapphire laser, tunable from 808 to 824 nm, and a semiconductor diode laser operating at 812 nm each illuminate a single slit in a 200 nm thick gold film. Each laser is focused to a spot of approximately  $5 \mu\text{m}$  FWHM. The two parallel slits,  $\sim 25 \mu\text{m}$  apart, are  $50 \mu\text{m}$  long and  $0.2 \mu\text{m}$  wide. The gold film is evaporated on top of a 0.5 mm thick fused quartz substrate with a 10 nm thick titanium adhesion layer between the gold and the quartz. A CCD camera with a 100 ms shutter time is used to record the far-field diffraction pattern. If the polarization of the two beams is parallel to the two slits (E-polarization), the resulting far-field pattern has, as expected, no fringes. An example is shown in Fig. 6.7 (a). However, when the polarization is changed to be perpendicular to the slits (H-polarization), an interference pattern with fringes with good visibility is obtained. This is shown in Fig. 6.7 (b), in which the fringe visibility  $\mathcal{V} = 28\%$ , even though each slit is illuminated by a separate laser.

## 6.7 Conclusions

In summary, we have investigated the coherence properties of the field in Young's interference experiment. In Sec. 6.4 it was shown that for any pair of points within certain planes the light is fully coherent, i.e.  $|\mu| = 1$ , irrespective of the correlation of the field at the two pinholes; in particular, the light could originate from two independent lasers, each illuminating only one of the pinholes.

In Sec. 6.5, we have demonstrated the existence of phase singularities of the spectral degree of coherence of the field at pairs of points in the region of superposition in Young's interference experiment with partially coherent light. The phase of the spectral degree of coherence is shown to make a  $\pm\pi$  jump across such singularities. This phase jump can be observed by means of a second Young's experiment. To our knowledge, this study is the first extension of the field of singular optics to the realm of correlation functions.

Finally, in Sec. 6.6 it was explained how surface plasmons travelling between the two slits can alter the degree of coherence in Young's experiment. In particular, it was predicted that two independent lasers both illuminating only a single slit can still produce interference fringes. This prediction has very recently been verified experimentally [KUZMIN *et al.*, 2005]. Note that this is completely different from the planes of full coherence mentioned before, because here the effect is not restricted to certain pairs of points.



# Bibliography

- G.S. AGARWAL, “Interference in complementary spaces”, *Found. Phys.* 25, pp. 219–228 (1995).
- G.S. AGARWAL, A. DOGARIU, T.D. VISSER, AND E. WOLF, “Generation of complete coherence in Youngs interference experiment with random electromagnetic beams”, *Opt. Lett.* 30, pp. 120–122 (2005).
- L. ALLEN, S.M. BARNETT, AND M.J. PADGET, editors, *Optical Angular Momentum*, Institute of Physics, Bristol (2003).
- G.B. ARFKEN AND H.J. WEBER, *Mathematical Methods for Physicists*, 4th edition, Academic Press, San Diego (1995).
- S. ASTILEAN, PH. LALANNE, AND M. PALAMARU, “Light transmission through metallic channels much smaller than the wavelength”, *Opt. Commun.* 175, pp. 265–273 (2000).
- K.E. ATKINSON, *A Survey of Numerical Methods of Fredholm Equations of the Second Kind*, SIAM, Philadelphia (1976).
- N.H.G. BAKEN, M.B.J. DIEMEER, J.M. VAN SPLUNTER, AND H. BLOK, “Computational modeling of diffused channel waveguides using a domain integral equation”, *J. Lightwave Technol.* 8, pp. 576–586 (1990).
- A. BARBARA, P. QUEÉMERAS, E. BUSTARRET, T. LOPEZ-RIOS, AND T. FOURNIER, “Electromagnetic resonances of sub-wavelength rectangular metallic gratings”, *Eur. Phys. J. D* 23, pp. 143–154 (2003).
- W.L. BARNES, A. DEREUX, AND T.W. EBBESEN, “Surface plasmon subwavelength optics”, *Nature* 424, pp. 824–830 (2003).
- R. BARRETT, M. BERRY, T.F. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. EIJKHOUT, R. POZO, C. ROMINE, AND H. VAN DER VORST, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd edition, SIAM, Philadelphia (1994).

- L. BASANO, P. OTTONELLO, G. ROTTIGNI, AND M. VICARI, “Degree of spectral coherence, space-frequency plots and correlation-induced spectral changes”, *Opt. Commun.* 207, pp. 77–83 (2002).
- M.W. BEIJERSBERGEN, *Phase Singularities in Optical Beams*, Ph.D. thesis, Leiden University, Leiden, (1996).
- M.V. BERRY, “Singularities in waves and rays”, In: R. Balian, M. Kléman, and J.P. Poirier, editors, *Physics of Defects*, volume 35 of *Les Houches*, pp. 453–543, North-Holland, Amsterdam (1981).
- M.V. BERRY, “Wave dislocations reactions in non-paraxial Gaussian beams”, *J. Mod. Opt.* 45, pp. 1845–1858 (1998).
- H.A. BETHE, “Theory of diffraction by small holes”, *Phys. Rev.* 66, pp. 163–182 (1944).
- E. BETZIG, A. HAROOTUNIAN, A. LEWIS, AND M. ISAACSON, “Near-field diffraction by a slit: implications for superresolution microscopy”, *Appl. Opt.* 25, pp. 1890–1900 (1986).
- E. BETZIG AND J.K. TRAUTMAN, “Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit”, *Science* 257, pp. 189–195 (1992).
- J. VAN BLADEL, *Singular Electromagnetic Fields and Sources*, IEEE Press, New York (2000).
- H. BLOK AND P.M. VAN DEN BERG, *Electromagnetic Waves: An Introductory Course*, Delft University Press, Delft (1999).
- C.F. BOHREN, “How can a particle absorb more than the light incident on it?”, *Am. J. Phys.* 51, pp. 323–327 (1983).
- C.F. BOHREN AND D.R. HUFFMAN, *Absorption and Scattering of Light by Small Particles*, Wiley, New York (1983).
- A. BOIVIN, J. DOW, AND E. WOLF, “Energy flow in the neighborhood of the focus of a coherent beam”, *J. Opt. Soc. Am.* 57, pp. 1171–1175 (1967).
- M. BORN AND E. WOLF, *Principles of Optics*, 7th (expanded) edition, Cambridge University Press, Cambridge (1999).
- C.J. BOUWKAMP, “Diffraction theory”, *Rep. Prog. Phys.* 17, pp. 35–100 (1954).

- W. BRAUNBEK AND G. LAUKIEN, “Einzelheiten zur Halbebenen-Beugung”, *Optik* 9, pp. 174–179 (1952).
- Q. CAO AND PH. LALANNE, “Negative role of surface plasmons in the transmission of metallic gratings with very narrow slits”, *Phys. Rev. Lett.* 88, 057403 (2002).
- R.H. CHAN AND M.K. NG, “Conjugate gradient methods for Toeplitz systems”, *SIAM Rev.* 38, pp. 427–482 (1996).
- T.F. CHAN, “An optimal circulant preconditioner for Toeplitz systems”, *SIAM J. Sci. Stat. Comput.* 9, pp. 766–771 (1988).
- W.C. CHEW, “Some observations on the spatial and eigenfunction representation of dyadic Green’s functions”, *IEEE Trans. Antennas Propagat.* 37, pp. 1322–1327 (1989).
- W.C. CHEW, *Waves and Fields in Inhomogeneous Media*, IEEE Press, New York (1995).
- P.J. DAVIS, *Circulant Matrices*, Wiley, New York (1979).
- A. DEREUX, J.-P. VIGNERON, P. LAMBIN, AND A.A. LUCAS, “Theory of near-field optics with applications to SNOM and optical binding”, *Physica B* 175, pp. 65–67 (1991).
- B.T. DRAINE, “The discrete-dipole approximation and its application to interstellar graphite grains”, *Astrophys. J.* 333, pp. 848–872 (1988).
- T.W. EBBESEN, H.J. LEZEC, H.F. GHAEMI, T. THIO, AND P.A. WOLFF, “Extraordinary optical transmission through sub-wavelength hole arrays”, *Nature* 391, pp. 667–669 (1998).
- H. FIZEAU, “Recherches sur plusieurs phenomenes relatifs à la polarisation de la lumière”, *Ann. Chim.* 63, pp. 385–414 (1861).
- I. FREUND, “Saddles, singularities, and extrema in random phase fields”, *Phys. Rev. E* 52, pp. 2348–2360 (1995).
- I. FREUND, “Vortex flowers”, *Opt. Commun.* 196, pp. 63–76 (2001).
- P. DE LA FUENTE, *Singularitäten des zeitgemittelten Energieflusses in elektromagnetischen Feldern*, Ph.D. thesis, Technische Universität München, (1975).
- W. FULTON, *Algebraic Topology, A First Course*, Springer-Verlag, New York (1995).

- F.J. GARCÍA-VIDAL, H.J. LEZEC, T.W. EBBESEN, AND L. MARTÍN-MORENO, “Multiple paths to enhance optical transmission through a single subwavelength slit”, *Phys. Rev. Lett.* 90, 213901 (2003).
- G. GBUR AND T.D. VISSER, “Coherence vortices in partially coherent beams”, *Opt. Commun.* 222, pp. 117–125 (2003).
- H.F. GHAEMI, T. THIO, D.E. GRUPP, T.W. EBBESEN, AND H.J. LEZEC, “Surface plasmons enhance optical transmission through subwavelength holes”, *Phys. Rev. B* 58, pp. 6779–6782 (1998).
- C. GIRARD AND A. DEREUX, “Near-field optics theories”, *Rep. Prog. Phys.* 59, pp. 657–699 (1996).
- G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, 2nd edition, Johns Hopkins University Press, Baltimore (1990).
- F. GORI, M. SANTARSIERO, R. BORGHI, AND S. VICALVI, “Partially coherent sources with helicoidal modes”, *J. Mod. Opt.* 45, pp. 539–554 (1998).
- R.F. HARRINGTON, *Field Computation by Moment Methods*, IEEE Press, New York (1999).
- J.O. HIRSCHFELDER, A.C. CRISTOPH, AND W.E. PALKE, “Quantum mechanical streamlines. I. Square potential barrier”, *J. Chem. Phys.* 61, pp. 5435–5455 (1974a).
- J.O. HIRSCHFELDER, C.J. GOEBEL, AND L.W. BRUCH, “Quantized vortices around wavefunction nodes. II”, *J. Chem. Phys.* 61, pp. 5456–5459 (1974b).
- A.T. DE HOOP, *Handbook of Radiation and Scattering of Waves*, Academic Press, London (1995).
- C.C. HSIUNG, *A First Course in Differential Geometry*, Wiley, New York (1981).
- J.D. JACKSON, *Classical Electrodynamics*, 3th edition, Wiley, New York (1999).
- D.L. JACOBSON, S.A. WERNER, AND H. RAUCH, “Spectral modulation and squeezing at high-order neutron interferences”, *Phys. Rev. A* 49, pp. 3196–3200 (1994).
- D.F.V. JAMES AND E. WOLF, “Spectral changes produced in Young’s interference experiment”, *Opt. Commun.* 81, pp. 150–154 (1991).
- D.F.V. JAMES AND E. WOLF, “Determination of the degree of coherence of light from spectroscopic measurements”, *Opt. Commun.* 145, pp. 1–4 (1998).

- P.B. JOHNSON AND R.W. CHRISTY, “Optical constants of the noble metals”, *Phys. Rev. B* 6, pp. 4370–4379 (1972).
- P.B. JOHNSON AND R.W. CHRISTY, “Optical constants of transition metals: Ti, V, Cr, Mn, Fe, Co, Ni, and Pd”, *Phys. Rev. B* 9, pp. 5056–5070 (1974).
- H.C. KANDPAL, J.S. VAISHYA, M. CHANDER, K. SAXENA, AND K.C. JOSHI, “Field correlations from spectral measurements in Young’s interference experiment”, *Phys. Lett. A* 167, pp. 120–122 (1992a).
- H.C. KANDPAL, J.S. VAISHYA, M. CHANDER, K. SAXENA, D.S. METHA, AND K.C. JOSHI, “Spectral changes due to source correlation in Young’s interference experiment”, *Phys. Lett. A* 167, pp. 114–119 (1992b).
- G.P. KARMAN, M.W. BEIJERSBERGEN, A. VAN DUIJL, D. BOUWMEESTER, AND J.P. WOERDMAN, “Airy pattern reorganization and subwavelength structure in a focus”, *J. Opt. Soc. Am. A* 15, pp. 884–899 (1998).
- G.P. KARMAN, M.W. BEIJERSBERGEN, A. VAN DUIJL, AND J.P. WOERDMAN, “Creation and annihilation of phase singularities in a focal field”, *Opt. Lett.* 22, pp. 1503–1505 (1997).
- E.W. KOLK, N.H.G. BAKEN, AND H. BLOK, “Domain integral equation analysis of integrated optical channel and ridge waveguides in stratified media”, *IEEE Trans. Microwave Theory Techn.* 38, pp. 78–85 (1990).
- V.N. KUMAR AND D.N. RAO, “Two-beam interference experiments in the frequency-domain to measure the complex degree of spectral coherence”, *J. Mod. Opt.* 48, pp. 1455–1465 (2001).
- N. KUZMIN, H.F. SCHOUTEN, G. GBUR, G.W. ’T HOOFT, E.R. ELIEL, AND T.D. VISSER, “Surface-plasmon-induced coherence”, to be published (2005).
- P. LALANNE, C. SAUVAN, J.P. HUGONIN, J.C. RODIER, AND P. CHAVEL, “Perturbative approach for surface plasmon effects on flat interfaces periodically corrugated by subwavelength apertures”, *Phys. Rev. B* 68, 125404 (2003).
- F. LANDSTORFER, H. MEINKE, AND G. NIEDERMAIR, “Ringförmiger Energiewirbel im Nahfeld einer Richtantenne”, *Nachrichtentechn. Z.* 12, pp. 537–576 (1972).
- Y. LEVIATAN, “Study of near-zone fields of a small aperture”, *J. Appl. Phys.* 60, pp. 1577–1583 (1986).

- H.J. LEZEC, A. DEGIRON, E. DEVAUX, R.A. LINKE, L. MARTÍN-MORENO, F.J. GARCÍA-VIDAL, AND T.W. EBBESEN, “Beaming light from a sub-wavelength aperture”, *Science* 297, pp. 820–822 (2002).
- H.J. LEZEC AND T. THIO, “Diffracted evanescent wave model for enhanced and suppressed optical transmission through subwavelength hole arrays”, *Opt. Express* 12, pp. 3629–3651 (2004).
- L. MANDEL AND E. WOLF, *Optical Coherence and Quantum Optics*, Cambridge University Press, Cambridge (1995).
- A.A. MARADUDIN AND D.L. MILLS, “Scattering and absorption of electromagnetic radiation by a semi-infinite medium in the presence of surface roughness”, *Phys. Rev. B* 11, pp. 1392–1415 (1975).
- J.E. MARSDEN AND M.J. HOFFMAN, *Elementary Classical Analysis*, 2nd edition, Freeman, New York (1993).
- O.J.F. MARTIN, N.B. PILLER, H. SCHMID, H. BIEBUYCK, AND B. MICHEL, “Energy flow in light-coupling masks for lensless optical lithography”, *Opt. Express* 3, pp. 280–285 (1998).
- L. MARTÍN-MORENO, F.J. GARCÍA-VIDAL, H.J. LEZEC, A. DEGIRON, AND T.W. EBBESEN, “Theory of highly directional emission from a single subwavelength aperture surrounded by surface corrugations”, *Phys. Rev. Lett.* 90, 167401 (2003).
- L. MARTÍN-MORENO, F.J. GARCÍA-VIDAL, H.J. LEZEC, K.M. PELLERIN, T. THIO, J.B. PENDRY, AND T.W. EBBESEN, “Theory of extraordinary optical transmission through subwavelength hole arrays”, *Phys. Rev. Lett.* 86, pp. 1114–1117 (2001).
- D.L. MILLS AND A.A. MARADUDIN, “Surface roughness and the optical properties of a semi-infinite material; the effect of a dielectric overlayer”, *Phys. Rev. B* 12, pp. 2943–2958 (1975).
- J.F. NYE, “Unfolding of higher-order wave dislocations”, *J. Opt. Soc. Am. A* 15, pp. 1132–1138 (1998).
- J.F. NYE, *Natural Focusing and Fine Structure of Light: Caustics and Wave Dislocations*, Institute of Physics Publishing, Bristol (1999).
- J.F. NYE, “Numerical solution for diffraction of an electromagnetic wave by slits in a perfectly conducting screen”, *Proc. R. Soc. Lond. A* 458, pp. 401–427 (2002).

- J.F. NYE AND M.V. BERRY, “Dislocations in wave trains”, *Proc. R. Soc. Lond. A* 336, pp. 165–190 (1974).
- J.F. NYE, J.V. HAJNAL, AND J.H. HANNAY, “Phase saddles and dislocations in two-dimensional waves such as the tides”, *Proc. R. Soc. Lond. A* 417, pp. 7–20 (1988).
- D.M. PALACIOS, I.D. MALEEV, A.S. MARATHAY, AND G.A. SWARTZLANDER, JR., “Spatial correlation singularity of a vortex field”, *Phys. Rev. Lett.* 92, 143905 (2004).
- M. PAULUS, P. GAY-BALMAZ, AND O.J.F. MARTIN, “Accurate and efficient computation of the Green’s tensor for stratified media”, *Phys. Rev. E* 62, pp. 5797–5807 (2000).
- M. PAULUS, H. SCHMID, B. MICHEL, AND O.J.F. MARTIN, “Contrast mechanisms in high-resolution contact lithography: A comparative study”, *Microelectron. Eng.* 57–58, pp. 109–116 (2001).
- D.W. POHL, W. DENK, AND M. LANZ, “Optical stethoscopy: Image recording with resolution  $\lambda/20$ ”, *Appl. Phys. Lett.* 44, pp. 651–653 (1984).
- S.A. PONOMARENKO, “A class of partially coherent beams carrying optical vortices”, *J. Opt. Soc. Am. A* 18, pp. 150–156 (2001).
- S.A. PONOMARENKO AND E. WOLF, “Coherence properties of light in Young’s interference pattern formed with partially coherent light”, *Opt. Commun.* 170, pp. 1–8 (1999).
- J.A. PORTO, F.J. GARCÍA-VIDAL, AND J.B. PENDRY, “Transmission resonances on metallic gratings with very narrow slits”, *Phys. Rev. Lett.* 83, pp. 2845–2848 (1999).
- E.M. PURCELL AND C.R. PENNYPACKER, “Scattering and absorption of light by nonspherical dielectric grains”, *Astrophys. J.* 186, pp. 705–714 (1973).
- H. RAETHER, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer, Berlin (1988).
- H. RAUCH, “Phase space coupling and interference in EPR experiments”, *Phys. Lett. A* 173, pp. 240–242 (1993).
- LORD RAYLEIGH, “On the passage of waves through apertures in plane screens, and allied problems”, *Phil. Mag.* 43, pp. 259–272 (1897).

- C.E. REED, J. GIERGIEL, J.C. HEMMINGER, AND S. USHIODA, “Dipole radiation in a multilayer geometry”, *Phys. Rev. B* 36, pp. 4990–5000 (1987).
- A. ROBERTS, “Electromagnetic theory of diffraction by a circular aperture in a thick perfectly conducting screen”, *J. Opt. Soc. Am. A* 4, pp. 1970–1983 (1987).
- R. RUPPIN, “Surface polaritons on a left-handed medium”, *Phys. Lett. A* 277, pp. 61–64 (2000).
- M. SANTARSIERO AND F. GORI, “Spectral changes in a Young’s interference pattern”, *Phys. Lett. A* 167, pp. 123–128 (1992).
- H.F. SCHOUTEN, N. KUZMIN, G. DUBOIS, T.D. VISSER, G. GBUR, P.F. ALKEMADE, H. BLOK, G.W. ’T HOOFT, D. LENSTRA, AND E.R. ELIEL, “Plasmon-assisted two slit transmission: Young’s experiment revisited”, *Phys. Rev. Lett.* 94, 053901 (2005).
- H.F. SCHOUTEN, T.D. VISSER, G. GBUR, D. LENSTRA, AND H. BLOK, “Creation and annihilation of phase singularities near a sub-wavelength slit”, *Opt. Express* 11, pp. 371–380 (2003a).
- H.F. SCHOUTEN, T.D. VISSER, G. GBUR, D. LENSTRA, AND H. BLOK, “The diffraction of light by narrow slits in plates of different materials”, *J. Opt. A: Pure Appl. Opt.* 6, pp. S277–S280 (2004).
- H.F. SCHOUTEN, T.D. VISSER, D. LENSTRA, AND H. BLOK, “Light transmission through a subwavelength slit: Waveguiding and optical vortices”, *Phys. Rev. E* 67, 036608 (2003b).
- U. SCHRÖTER AND D. HEITMANN, “Surface-plasmon-enhanced transmission through metallic gratings”, *Phys. Rev. B* 58, pp. 15419–15421 (1998).
- A.W. SNYDER AND J.D. LOVE, *Optical Waveguide Theory*, Chapman and Hall, London (1983).
- M.B. SOBNACK, W.C. TAN, N.P. WANSTALL, T.W. PREIST, AND J.R. SAMBLES, “Stationary surface plasmons on a zero-order metal grating”, *Phys. Rev. Lett.* 80, pp. 5667–5670 (1998).
- M.S. SOSKIN, V.N. GORHKOV, M.V. VASNETSOV, J.T. MALOS, AND N.R. HECKENBERG, “Topological charge and angular momentum of light beams carrying optical vortices”, *Phys. Rev. A* 56, pp. 4064–4075 (1997).

- M.S. SOSKIN AND M.V. VASNETSOV, “Singular optics”, In: E. Wolf, editor, *Progress in Optics*, volume 42, pp. 219–276, North Holland, Amsterdam (2001).
- J.J. STAMNES, *Waves in Focal Regions*, Adam Hilger, Bristol (1986).
- J.J. STAMNES, “Exact two-dimensional scattering of an arbitrary wave by perfectly reflecting elliptical cylinders, strips and slits”, *Pure Appl. Opt.* 4, pp. 841–855 (1995).
- J.R. SUCKLING, A.P. HIBBINS, M.J. LOCKYEAR, T.W. PREIST, J.R. SAMBLES, AND C.R. LAWRENCE, “Finite conductance governs the resonance transmission of thin metal slits at microwave frequencies”, *Phys. Rev. Lett.* 92, 147401 (2004).
- Y. TAKAKURA, “Optical resonance in a narrow slit in a thick metallic screen”, *Phys. Rev. Lett.* 86, pp. 5601–5603 (2001).
- T. THIO, K.M. PELLERIN, R.A. LINKE, H.J. LEZEC, AND T.W. EBBESEN, “Enhanced light transmission through a single subwavelength aperture”, *Opt. Lett.* 26, pp. 1972–1974 (2001).
- M.S. TOMAŠ, “Green function for multilayers: Light scattering in planar cavities”, *Phys. Rev. A* 51, pp. 2545–2559 (1995).
- M. TOTZECK AND H.J. TIZIANI, “Interference microscopy of sub- $\lambda$  structures: A rigorous computation method and measurements”, *Opt. Commun.* 136, pp. 61–74 (1997a).
- M. TOTZECK AND H.J. TIZIANI, “Phase-singularities in 2D diffraction fields and interference microscopy”, *Opt. Commun.* 138, pp. 365–382 (1997b).
- M.M.J. TREACY, “Dynamical diffraction explanation of the anomalous transmission of light through metallic gratings”, *Phys. Rev. B* 66, 195105 (2002).
- L. TSANG, E. NJOKU, AND J.A. KONG, “Microwave thermal emission from a stratified medium with nonuniform temperature distribution”, *J. Appl. Phys.* 46, pp. 5127–5133 (1975).
- T.D. VISSER, H. BLOK, AND D. LENSTRA, “Theory of polarization-dependent amplification in a slab waveguide with anisotropic gain and losses”, *IEEE J. Quantum Electron.* 35, pp. 240–249 (1999).

- 
- H.E. WENT, A.P. HIBBINS, J.R. SAMBLES, C.R. LAWRENCE, AND A.P. CRICK, “Selective transmission through very deep zero-order metallic gratings at microwave frequencies”, *Appl. Phys. Lett.* 77, pp. 2789–2791 (2000).
- A.D. YAGHJIAN, “Electric dyadic Green’s functions in the source region”, *Proc. IEEE* 68, pp. 248–263 (1980).
- F. YANG AND J.R. SAMBLES, “Resonant transmission of microwaves through a metallic slit”, *Phys. Rev. Lett.* 89, 063901 (2002).
- T. YOUNG, “The Bakerian lecture: On the theory of light and colours”, *Phil. Trans. R. Soc.* 92, pp. 12–48 (1802).
- T. YOUNG, *A Course of Lectures on Natural Philosophy and the Mechanical Arts*, volume 1, lecture 39, pp. 364–367, Taylor and Walton, London (1845).
- A.V. ZAYATS AND I.I. SMOLYANINOV, “Near-field photonics: surface plasmon polaritons and localized surface plasmons”, *J. Opt. A: Pure Appl. Opt.* 5, pp. S16–S50 (2003).

# Samenvatting

In dit proefschrift is de lichttransmissie door subgolflengte gaten onderzocht. Het gedrag van licht bij gaten is een onderwerp dat al sinds de negentiende eeuw veel bestudeerd is. Echter, het gaat bijna altijd om gaten met dimensies die veel groter zijn dan de golflengte van het inkomende veld. In de gevallen dat er wel sprake was van subgolflengte gaten, ging het meestal om radio- of microgolven, waarbij, in tegenstelling tot zichtbaar licht, aangenomen mag worden dat het metaal een perfecte geleider is. Dit proefschrift beschrijft een methode om in gevallen van subgolflengte spleten in metalen met een realistische geleidbaarheid de transmissie te berekenen. De resultaten hiervan zijn van groot belang voor toepassingen in bijvoorbeeld het gebied van optische opslag of nabije veld microscopie.

In het tweede hoofdstuk wordt het wiskundig formalisme ontwikkeld om de berekeningen uit te voeren. De basisvergelijkingen van de Elektrodynamica, de Maxwell vergelijkingen, worden daar voor het geval van een lichtverstrooier ingebed in een gelaagd medium omgezet in een integraalvergelijking met als domein het gebied ingenomen door deze verstrooier en met als kern een Greense tensor. Deze Greense tensor wordt afgeleid voor een willekeurig gelaagd medium. Vervolgens wordt de numerieke methode om de integraalvergelijking op te lossen uitgelegd. Bij deze methode wordt de integraalvergelijking omgezet in een stelsel van lineaire vergelijkingen. De laatste sectie van het hoofdstuk beschrijft een methode om dit stelsel efficiënt op te lossen.

Het derde hoofdstuk gaat over de lichttransmissie door een enkele spleet. Hier wordt met name ingegaan op de lichttransmissie als een functie van de spleetbreedte. Ook wordt de rol van de verschillende soorten materiaal waar de plaat uit kan bestaan, de dikte van de plaat en de polarisatie van het inkomende elektrische veld onderzocht. De resultaten blijken goed te verklaren door zowel golfgeleiding als fase singulariteiten te beschouwen. Fase singulariteiten zijn punten in de ruimte waar de amplitude van het veld nul is en de fase van het veld daardoor ongedefinieerd. Deze fase singulariteiten hebben een topologische lading, die behouden is. Het blijkt dat de verhoogde transmissie door een spleet in een dunne metalen plaat gepaard gaat met de annihilatie van zulke fase singulariteiten. Dit fenomeen treedt op als de breedte van de spleet gevarieerd wordt en er een afsnij-

breedte van een geleide mode in de spleet bereikt wordt.

In het vierde hoofdstuk wordt ook naar een enkele spleet in een metalen plaat gekeken. Echter, in plaats van naar de totale lichttransmissie wordt hier naar het stralingspatroon gekeken, d.w.z. naar de verschillende richtingen waarin het licht door de spleet wordt verstrooid. Ook hier spelen fase singulariteiten een rol. Het blijkt namelijk dat de minima in het verre veld stralingspatroon samenhangen met fase singulariteiten in het nabije veld van de spleet. Indien er een systeemp parameter zoals de golflengte van het licht of de dikte van de plaat wordt gevarieerd, dan kan zo'n fase singulariteit van de spleet naar oneindig lopen en daar een singulariteit in het stralingspatroon worden. Deze laatste singulariteit is niet enkel een minimum, maar zelfs een exact nulpunt van het stralingspatroon. Het werd ook aangetoond dat fase singulariteiten op deze manier ook in hun eentje op oneindig kunnen verdwijnen, wat een schending is van topologisch ladingsbehoud. Ook wordt er aangegeven, hoe een verdwijning van een fase singulariteit experimenteel gemeten zou kunnen worden.

De transmissie van licht door twee spleten is het onderwerp van het vijfde hoofdstuk. De transmissie blijkt te worden versterkt of verzwakt als functie van de afstand tussen de twee spleten of als functie de golflengte. Deze modulatie wordt verklaard door zogenaamde oppervlakteplasmonen. Dit zijn elektromagnetische oppervlaktegolven die langs het metaal-luchtraakvlak lopen. Zo'n oppervlakteplasmon wordt aangeslagen bij een spleet, loopt vervolgens naar de andere spleet, waarbij hij kan koppelen aan het doorgelaten veld. De interferentie tussen het licht dat direct door een spleet wordt doorgelaten en licht dat de beschreven "plasmonomweg" neemt, veroorzaakt een interferentiepatroon in de totale transmissie van het licht. Deze voorspelling is experimenteel bevestigd door onderzoekers van de Universiteit Leiden.

Het laatste hoofdstuk verschilt met de rest van het proefschrift in dat het niet aanneemt dat het elektromagnetische veld coherent en monochromatisch is. Het onderwerp is de transmissie van partieel coherent licht door twee gaten. In het eerste deel worden de coherentie eigenschappen van het licht in het interferentie-experiment van Young beschreven. Speciaal wordt gekeken naar paren van punten waar het veld volledig coherent of volledig incoherent is. Het tweede deel voorspelt het effect van oppervlakteplasmonen op de coherentie eigenschappen van Young's experiment. Het blijkt dat er een plasmon-geïnduceerde coherentie tussen de twee spleten kan bestaan, zelfs als ze incoherent belicht worden. Deze laatste voorspelling is inmiddels ook experimenteel bevestigd.

# List of Publications

1. N. Kuzmin, H.F. Schouten, G. Gbur, G.W. 't Hooft, E.R. Eliel and T.D. Visser, “Surface-plasmon-induced Coherence”, submitted (2005).
2. G. Gbur, H.F. Schouten and T.D. Visser, “Achieving Superresolution in Near-field Optical Data Readout Systems using Surface Plasmons”, submitted to Appl. Phys. Lett. (2005).
3. H.F. Schouten and P.S. Carney, “Stokes reciprocity for stratified media”, submitted to Opt. Commun. (2005).
4. H.F. Schouten, N. Kuzmin, G. Dubois, T.D. Visser, G. Gbur, P.F. Alkemade, H. Blok, G.W. 't Hooft, D. Lenstra and E.R. Eliel, “Plasmon-assisted two-slit transmission: Young’s experiment revisited”, Phys. Rev. Lett., vol. 94, 053901 (2005).
5. H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “Connection between Phase Singularities and the Radiation Pattern of a Slit in a Metal Plate”, Phys. Rev. Lett., vol. 93, 173901 (2004).
6. H.F. Schouten, T.D. Visser and D. Lenstra, “Optical Vortices near Sub-wavelength Structures”, J. Opt. B: Quantum Semiclass. Opt., vol. 6, pp. S404–S409 (2004).
7. H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “The diffraction of light by narrow slits in plates of different materials”, J. Opt. A: Pure Appl. Opt., vol. 6, pp. S277–S280 (2004).
8. H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “Phase singularities and enhanced transmission through a subwavelength slit”, Optics & Photonics News, p. 23, in the “What’s New in Optics?” issue, December 2003.

9. H.F. Schouten, T.D. Visser and E. Wolf, “New effects in Young’s interference experiment with partially coherent light”, *Opt. Lett.*, vol. 28, pp. 1182–1184 (2003).
10. H.F. Schouten, G. Gbur, T.D. Visser and E. Wolf, “Phase singularities of the coherence functions in Young’s interference pattern”, *Opt. Lett.*, vol. 28, pp. 968–970 (2003).
11. H.F. Schouten, T.D. Visser, D. Lenstra and H. Blok, “Light transmission through a subwavelength slit: Waveguiding and optical vortices”, *Phys. Rev. E*, vol. 67, 036608 (2003).
12. H.F. Schouten, T.D. Visser, G. Gbur, D. Lenstra and H. Blok, “Creation and annihilation of phase singularities near a sub-wavelength slit”, *Opt. Express*, vol. 11, pp. 371–380 (2003).

# Dankwoord

Als eerste wil ik mijn copromotor Taco Visser bedanken voor zijn geweldige begeleiding. Zijn enthousiasme was zeer aanstekelijk. Het was heerlijk om met hem samen te werken. Ik hoop dan ook dat ik die samenwerking nog voort kan zetten, al zal dat nu een stuk langzamer gaan.

Ik wil verder mijn promotoren Daan Lenstra en Hans Blok bedanken voor het in mij gestelde vertrouwen. Ook wil ik Daan bedanken voor de feedback die hij gaf bij veel van mijn artikelen. Hans wil ik verder bedanken voor zijn uitzonderlijke vermogen om met allerlei zeer relevante (meestal in het Duits geschreven) artikelen uit voor mij onbekende tijdschriften aan te komen.

Verder wil ik Greg Gbur bedanken voor de zeer plezierige en productieve samenwerking. Ook wil ik hem bedanken voor de leuke tijd die we hadden toen hij met mij een kamer op de universiteit deelde. (Greg, in case you wonder, I just said some pretty nasty things about you.)

Ik wil Emil Wolf bedanken voor de mogelijkheid met hem samen te werken op de universiteit van Rochester. De drie weken daar waren wetenschappelijk productief (twee artikelen!) en bovendien erg plezierig. Het was een eer om met hem samen te werken.

Scott Carney wil ik bedanken voor de fantastische tijd aan de universiteit van Illinois. De maand in Urbana-Champaign was een geweldige ervaring. Een aantal andere mensen hebben hier ook flink aan bijgedragen. Met name wil ik Dave Lytle and Bob Schoonover bedanken voor hun gastvrijheid tijdens die weken.

Verder wil ik Eric Eliël en Nikolay Kuzmin bedanken voor het meten van enkele van mijn voorspellingen. Het is geweldig om te zien dat er soms ook nog iets lijkt te kloppen van de dingen die je aan het doen bent.

Ik wil alle leden van de theorie groep bedanken voor de altijd gezellige koffie- en theepauzes. Met name wil ik Fetze Pijlman, Harmen Waringa, Hartmut Erzgräber, Miranda van Iersel, Erik Wessels, Piet Mulders, Ben Bakker en Klaas Allaart bedanken.

Als laatste wil ik mijn ouders, Rob en Toos Schouten, bedanken voor hun steun tijdens al die jaren. Ik wil hen en mijn zusje Ingrid ook bedanken voor de gezellige sfeer thuis.